الاقتصاد القياسي

منشورات جامعة دمشق
كلية الاقتصاد

الاقتصاد القياسي

الدكتور
قاسم النعيمي
أستاذ مساعد في قسم الإحصاء التطبيقي

الدكتور
علي ديبو
مدرس في قسم الاقتصاد

جامعة دمشق
١٤٢٩ - ١٤٣٠ هـ
٢٠٠٨-٢٠٠٩ م

المحتويات

وعند تأليف هذا الكتاب تم الاعتماد على الأبحاث العلمية والمراجع الحديثة، وعلى عددٍ من المنشورات العلمية العربية والأجنبية المتاحة في هذا المجال، بالإضافة إلى أنه تم الاستفادة من المنشورات الموجودة على الانترنت وخاصة المراجع المعتمدة في الجامعات البريطانية والأمريكية . علماً أن جميع المراجع المذكورة في هذا الكتاب متوفرة لدينا ومتاحة لجميع القراء.

وفي الختام نتمنى أن نكون قد وفقنا في إخراج مرجع دراسي لأبنائنا طلاب كلية الاقتصاد والمهتمين بالاقتصاد القياسي. ونتمنى على الزملاء الكرام إبداء النصح وتنبيهنا إلى الأخطاء إن وجدت لتصحيحها . مع فائق الاحترام والتقدير

والله ولي التوفيق

دمشق في – / /2008 م

الدكتور قاسم النعيمي

الفصل الأول

نشأة الاقتصاد القياسي ومفهومه

1-1 - مقدمة:

الاقتصاد القياسي هو أحد فروع علم الاقتصاد التي تهدف إلى الدراسة الظواهر الاقتصادية، بالاعتماد على القياس الفعلي للمتغيرات الداخلة فيها إعادة العلاقات فيما بينها. يعتبر هذا العلم من العلوم الحديثة نسبياً، حيث يرجع هذا العلم إلى أوائل القرن العشرين ففي عام 1930 قدم اثنان من الاقتصاديين هما (راجر فريش)، و(شارل روس) اقتراحاً لتأسيس الجمعية الدولية للاقتصاد القياسيEconometrics International Association وفي عام 1933 استطاعت الجمعية إصدار أول عدد من مجلة الاقتصاد القياسي Econometric للبحث في مجال القياس الاقتصادي. وقد كانت افتتاحية العدد الأول تأكيداً على التجربة التي جمعت بين ثلاث فروع علمية لمواجهة الواقع الاقتصادي الحديث وهذه العلوم هي:

الإحصاء، والنظرية الاقتصادية، والرياضيات، وهذه التجربة ساعدت على فهم التطورات الحديثة والسريعة. ورغم حداثة هذا العلم إلا أنه قد استطاع تحقيق تقدم سريع، حيث أثبت جدواه وأهميته لدراسة العلوم الاقتصادية والتجارية الحديثة.

والاقتصاد القياسي أصبح ضرورة ملحة نظراً إلى اعتماد عمليات التخطيط الاقتصادي على التكنولوجيا الحديثة بما فيه من أساليب البحث الكمي القياسي.

يرى الاقتصادي النمساوي (جوزيف شومتز) أن هذا العلم يرجع في الواقع إلى الاقتصادي البريطاني (وليم بيتي)، حيث كان أول من استخدم تعبير الاقتصاد القياسي في أواخر القرن السابع عشر، واستخدم هذا التعبير كعنوان للبحث اعتماداً على جمع وتركيب البيانات الإحصائية، وذلك كأساس للدراسات السكانية. وهناك من يرجع هذه البدايات إلى الإحصائي الألماني (أرنست) الذي استطاع أن يتوصل في منتصف القرن التاسع عشر إلى أربعة قوانين أساسية تتعلق بالدخل والاستهلاك معتمداً على بحوث نظرية الميزانية لأسرة في بلجيكا.

ويمكننا القول أيضاً: إن البدايات كانت لأعمال (أرفيج فيشر)، عندما نشر كتابه عن القوة الشرائية للنقود، وكذلك أعمال هنري شولتز في كتابه نظرية قياس الطلب، كما كان للإحصائي الاقتصادي النرويجي (وانجر فريش) (Ranger Frisch) أولوية في استخدام مصطلح الاقتصاد القياسي (Econometrics) في عام 1926 والذي جاء على غرار مصطلح (Biometrics) القياس البيولوجي؛ الذي أطلق على علم الأحياء الذي يستخدم الطرق الإحصائية.

كما ظهرت في القرن التاسع عشر بعض الأعمال ذات الطابع القياسي كعمل الإحصائي الألماني (أنكل Enqel أرنست Ernst) (1896 - 1821) الذي وضع قوانينه الخاصة ذاتها بالدخل والاستهلاك في ضوء تحليل بيانات ميزانية الأسرة، وكذلك عمل الاقتصادي الإيطالي (باريتو 1923 – 1848 Vilfredo Parrot) الذي وضع قانونه الخاص بتوزيع الدخول في ضوء البيانات الإحصائية لعدة بلدان، أظهره في عام (1897) في كتابه منهاج الاقتصاد القياسي Cours D'economique Politique.

وفي أثر الحرب العالمية الأولى على وجه الخصوص تطور الاقتصاد القياسي بوصفه أسلوباً لتحليل الظواهر الاقتصادية تحليلاً كمياً؛ إذ تكللت بالنجاح

محاولات رواد الفكر القياسي في تحديد العلاقات الكمية بين المتغيرات الاقتصادية على أساس المعطيات الإحصائية الواقعية، خصوصاً في مجال تحليل السوق (Market Analysis)، وهذه الأساليب عبارة عن سبل للبحث في حل المعضلات الناشئة عن علائق المتغيرات الاقتصادية، التي لا تكفي الطرق الإحصائية التقليدية لمعالجتها.

وتقديراً لأهمية الأعمال القياسية تأسست الجمعية الدولية للقياس الاقتصادي في عام 1930، ومنذ عام 1933 ، وأخذت هذه الجمعية بإصدار مجلة (القياس الاقتصادي) (Econometrics). إن الاقتصاد القياسي قد تطور أولاً في البلدان الرأسمالية وبوجه الخصوص في الولايات المتحدة الأمريكية إذا كان للإنجازات الكبيرة والكبيرة جداً التي قدمها واضعو الفكر القياسي الأوائل في القرن التاسع عشر خلال النصف الأول من القرن العشرين أمثال (مور) (Henry L . Moore) وشولتز (Henry Schultz) وفريش (Rayner Frisch) وستون (Richard Stone) وغيرهم أثر كبير في تطور هذا العلم.

إن نشوء وتطور القياس الاقتصادي في البلدان الرأسمالية في تلك الفترة ذاتها لم يكن مصادفة وإنما كان ضرورية ملحة، لتطورات الاحتكارات الرأسمالية ورأسمالية الدولة. حيث تغير موقع المنتج والدولة في عمليات إعادة الإنتاج الرأسمالية وبذلك تكون قد تغيرت المعضلات الاقتصادية التي يجابهها الأفراد والشركات ومنها:

1) إن المنشآت الفردية لم تعد قادرة على التحكم بالسوق لأنها في ظل ظروف سوق المنافسة الحرة لم تعد المنتجة الوحيدة فهي تشتري موادها الأولية وتبيع منتجاتها النهائية بأثمان خارجة عن إرادتها. وكذلك بالنسبة للشركات الاقتصادية فقد اختلف الأمر كلياً. فالشركات الاحتكارية بدأت تبسط نفوذها على أسواق واسعة تتخطى في معظم الأحيان الحدود القومية. وأصبح لهذه الشركات

قابلية على التأثير في مجرى الحياة الاقتصادية، ليس في نطاق الأمة فحسب بل في ميدان واسع للاقتصاد العالمي.

لذلك لم تعد أسعار المواد الأولية والمنتجات النهائية ناتجة عن أوضاع السوق وإنما أصبحت متغيرات تتأثر قيمها تأثراً مباشراً بسياسة الاحتكارات الإنتاجية والتسويقية.

2) كما أن تنظيم أرباح المنشآت لم يعد يعتمد فقط على أفضل السبل لضغط تكاليف الإنتاج، وإنما أصبح يأخذ ينبغي أن يؤخذ بعين الاعتبار بعض العوامل الخارجية كأسعار المنتجات وكمياتها ومرونات الطلب عليها وعلى بدائلها و (كيفها Quality) و (ديمومتها Durability) وعلى تخصيصات الدعاية ،وغير ذلك؛ لهذا كله أصبح من الضروري لحل مشكلات المنشآت الاقتصادية إدخال الأسلوب الرياضي والإحصائي المتطور.

3) تطور لقياس الاقتصادي وأصبح إحدى الأدوات التي تستخدم في تحليل المعضلات الاقتصادية الجديدة؛ التي تمخض عن تطور الاحتكارات ورأسمالية الدولة.

4) إن العملية التخطيطية أصبحت معقدة وغدت بحاجة إلى تقنية حديثة لمعرفة الأثر المتبادل بين النشاطات الاقتصادية المتداخلة، وتقدير الطلب على السلع الاستهلاكية وتقدير تأثيره بدخل الأفراد وبأسعار هذه السلع، وكذلك تقدير الطلب الوسيط على السلع الاستراتيجية كالطاقة، والمعادن الأساسية ، والمواد الإنشائية الرئيسية.

5) لقد تم استخدام أساليب القياس الاقتصادي في إيجاد بعض المترابطات الهامة بين بعض المتغيرات الاقتصادية الأساسية؛ كالعلاقة بين الادخار وإنتاجية العمل في تحليل نمو الناتج المحلي، كما استخدم في تقييم التطور والنمو، وفي تقييم السياسات الاقتصادية خلال الخطط السابقة. ومن الجدير بالذكر فإن

الاقتصاد الاشتراكي ساعد في إمكانية بناء النماذج القياسية، نظراً لاعتماد هذا النوع من الاقتصاد على الملكية الجماعية لوسائل الإنتاج، حيث إن إنتاج المؤسسات الإنتاجية الكبيرة المتشابكة بصورة مركزية، كذلك ساعد على توفر الإحصاءات بصورة شاملة ودقيقة في الاقتصاد الاشتراكي أتاح المجال إلى حد كبير لتطبيق أساليب القياس الاقتصادي، بينما تقف طبيعة الإدلاء الإحصائي في الاقتصاد الرأسمالي حائلاً دون توفير إحصاءات سليمة. لأنه يقوم على إخفاء الأرقام الحقيقية للأرباح والمبيعات بهدف التهرب من الضرائب، وغير ذلك من اتباع الأساليب الملتوية في الالتفاف على القوانين والأنظمة النافذة بغرض تحقيق الربح.

ولا ننسى أن توفر الإحصاءات السليمة يعتبر من أهم العوامل اللازمة لعمليات التقدير الكمي للظواهر الاقتصادية، فبقدر صحة الإحصاءات تكون النتائج صحيحة. كما أن توفر الإحصاءات حول مختلف المتغيرات الاقتصادية يتح المجال أمام الباحثين الأخذ بعين الاعتبار قدراً أكبر من المتغيرات؛ مما يزيد من دقة النتائج، ويقلل من احتمالات الخطأ.

ولابد من الإشارة إلى أن أساليب الاقتصاد القياسي تستخدم كسلاح ماض بيد الاحتكارات ضد مصلحة المستهلكين، فهي تستخدم لتعظيم الأرباح على حساب ميزانية المستهلك. كما تستخدم في كثير من الأحيان لأغراض التضليل الفكري، وذلك عن طريق إيجاد ترابطات إحصائية بين متغيرات غير مرتبطة سببياً بالأساس، أو غير ذلك من الأساليب. وقد شوهت هذه الحقيقة مضمون الاقتصاد القياسي كعلم؛ مما جعل بعض الاقتصاديين في البلدان الاشتراكية يعتقدون بأن القياس الاقتصادي أحد فروع الاقتصاد الرأسمالي، إلا أن استخدام الأساليب القياسية بنجاح وبشكل واسع في تخطيط الاقتصاد الاشتراكي منذ

تساعد هذه النظرية على حل الصعوبات في اتخاذ القرارات الاقتصادية؛ عندما تتداخل جملة من المتغيرات فيما بينها، حيث ينجم عن اتخاذ قرار ما جملة من القرارات لتساعد في تنفيذه.

مثال ذلك (اتخاذ قرار بزيادة طاقة الإنتاجية للتشييد بنسبة معينة؛ فإن ذلك يتوقف على مدى توفر مستلزمات الأولية لهذه العملية الإنتاجية من إسمنت، ورمل، و بحص، وحديد التسليح، و نقل وحجم عمالة مدربة في البناء، وهذا كله ناجم عن اتخاذ قرار واحد وبسيط، فما بالك في قرارات تؤخذ كل يوم وباستمرار وعلى مستويات مختلفة.

إن أي تقصير في متغير من المتغيرات كالحديد أو النقل أو الأسمنت أو العمالة ... إلخ سوف ينتج عنه نقطة اختناق Bottle-neck حتى لو توفرت جميع المتغيرات الأخرى بالكميات المطلوبة، وهكذا نجد أن نظرية البرمجة ضرورية لإجراء عمليات التنسيق والربط بين جميع المتغيرات الاقتصادية (المتشابكة Interdependent) وكان الاقتصادي السوفياتي البروفسور (كانتاروفتش L . W. Kantarowitsch) أول من عبر عن المفهوم الأساسي لنظرية البرمجة الحديثة بكتابة (الطرق الرياضية لتنظيم وتخطيط الإنتاج Mathematical Methods of Organization and Planning Production) الذي نشره في عام 1920. وكذلك تطورت في الولايات المتحدة الأمريكية خلال الحرب العالمية الثانية، وذلك لارتباطها الوثيق بمشكلات الاقتصاد الحربي (War Economy) أو ما يسمى اليوم بالدعم اللوجيستي على وجه الخصوص ضرورات تزويد القوات المسلحة باحتياجاتها من طائرات ودبابات وعتاد ...إلخ دون خلق نقاط اختناق.

كذلك يتناول القياس الاقتصادي قضايا اقتصادية أخرى نخص منها بالذكر الإنتاج وتكاليفه، ولهذه الدراسات أهمية كبيرة في تخطيط الإنتاج في الوحدات

والقطاعات الإنتاجية المختلفة. كذلك للدراسات الخاصة بدوال الإنتاج على الصعيد الكلي دور في تحليل النمو والوقوف بدقة على أثر عوامل الإنتاج (العمل ورأس المال والمتطور التقني).

د- **رسم السياسات الاقتصادية للدول** : يطبق الاقتصاد القياسي من أجل الحصول على تقديرات رقمية لمعاملات العلاقات الاقتصادية بهدف صياغة السياسات العامة للحكومة. وبتحديد هذه القيم يمكننا من تقدير معاملات النظرية الاقتصادية مثل المرونات ، المضاعفات، معاملات الإنتاج الفنية ، التكاليف الحدية، العوائد الحدية و إلخ... . ويعتبر تحديد القيم العددية لهذه المتغيرات ضرورة لاتخاذ القرارات في المؤسسات والإدارات الحكومية لصياغة السياسة الاقتصادية والحكومية لدراسة القرارات البديلة في السياسة الاقتصادية. على سبيل المثال إذا رغبت الحكومة في تخفيض قيمة العملة لإزالة العجز من ميزان المدفوعات ؛ نلاحظ أن أهم معاملين يمكن أن تعتمد عليهما الحكومة هما الميل الحدي للاستيراد ومرونة أسعار الصادرات والواردات، أي: ينظر إلى مرونة أسعار الصادرات والواردات أقل من واحد كان تخفيض العملة دون فائدة.

أما على صعيد المنشآت الاقتصادية الخاصة نلاحظ على سبيل المثال إذا كان الطلب بالنسبة للسعر لسلعة ما أقل من الواحد (أي الطلب غير مرن) فإن المنتج لن يستفيد من قيمة سعر هذه السلعة لأنه يخفض من إيرادات المنشأة . والعكس صحيح إذا كانت مرونة الطلب أكبر من الواحد يتوجب على الحكومة في هذه الحالة الامتناع عن زيادة الضريبة المفروضة على هذه السلعة؛ لان زيادة الضرائب سوف يؤدي إلى تخفيض إيرادات الحكومة الضريبية بسبب انخفاض الطلب على هذه السلعة بارتفاع سعرها، وهنا تتجلى أهمية الاقتصاد القياسي في تقديم تقديرات قيمية لهذه المعاملات الاقتصادية؛ ليصار إلى اتخاذ القرار المناسب.

هـ- **التنبؤ بالقيم المستقبلية للمتغيرات الاقتصادية :** كلما كانت لدينا القدرة على التنبؤ بالقيم الاقتصادية للمتغيرات بشكل دقيق؛ كلما كانت لدينا القدرة على اتخاذ القرارات السياسة الاقتصادية بشكل سليم . على سبيل المثال الحكومة تريد اتخاذ قرار بخوص سياستها التوظيفية ، فإنه يتطلب منها تحديد الوضع الحالي الوظيفي، وحال هذا الوضع لمدة خمس سنوات مثلاً في حل عدم تدخل الحكومة. هذا الإجراء تأمنه لنا طرق الاقتصاد القياسي ، إذا كان مستوى التوظيف دون الحد الأدنى؛ فإن على الحكومة أن تعمل على سياسة التوظيف، أما في حال العكس فإنه يتطلب من الحكومة الحد من التوظيف، من أجل عدم حصول تضخم مقنع في القوة العاملة المتوقعة.

إن الدول النامية في أشد الحاجة وإلى وسائل ضبط مخططات التنمية بما في ذلك القياس الاقتصادي، وذلك بسبب تناقضها الشديد بين التخلف الواسع، وضرورات التسريع، وتأثير التنمية من جهة أخرى، وكذلك على الدول النامية مواجهة كفاح شديد داخل السوق الرأسمالية العالمية، وخاصة مصالح الاحتكارات الإمبريالية المتمثلة بالاستعمار الحديث؛ لاستغلال ثروات هذه البلدان.

لذلك بدأت بعض هذه البلدان تأخذ بالأساليب المتطورة للقياس الاقتصادي، ومنها الأقطار العربية، وذلك على أساس برامج دقيقة التخطيط نسبياً. وتعتبر بلدان شرق آسيا والهند في مقدمة الدول النامية؛ التي تطور لديها استخدام الاقتصاد القياسي.

1-4- منهجية الاقتصاد القياسي

لا يمكن أن تكون بحاراً ناجحاً مكتفياً في قراءة لكتب وأنت جالس في البيت، وكذلك لا يمكن للطالب أن يصبح اختصاصياً بالاقتصاد القياسي من خلال قراءة هذا الكتاب. الحياة العملية والواقعية أعمق من أن تصفها الكتب

والنماذج الموضوعة، فعندما تمارس ذلك سوف تلاقي الكثير من المشاكل التي لا تجدها في الكتب الموضوعة والأبحاث المنشورة. يمكن أن تصبح اختصاصياً إذا تمتعت بالنظرة التحليلية العلمية العميقة، و وجدت الأستاذ الاختصاصي الذي يقدم لك خبرته في الحياة العملية؛ لذلك لا يمكن لدارس النظرية أن يكون واقعياً، ولا يمكن لممارس العمل وحده دون النظرية أن يكون موضوعياً. لا تفكر أبداً أن الاقتصاد القياسي هو عبارة دراسة أهم المتغيرات الاقتصادية وإجراءات بناء النموذج. الاقتصاد القياسي هو الاقتصاد التطبيقي الذي يوازن بين النظرية الاقتصادية والبيانات الاقتصادية المتوفرة، مع تلاقح الأفكار، وفهم معمق لنظرية الاقتصاد القياسي.

في هذا الفصل سوف نستعرض العلاقة الحميمة بين الاقتصادي القياسي والفيزيائي والاقتصادي القياسي والرياضي الاقتصادي والإحصائي الرياضي ، أي: ندرس العلاقة والفجوة بين النظرية والتطبيق وبين الأساليب والطرق من الأعلى إلى الأسفل ومن الأسفل إلى الأعلى، والحلقة الضعيفة تتجلى في إمكانية الاستفادة من تجارب ودراسات الآخرين . هذه التجربة تجمع جميع عناصر الاقتصاد القياسي.

ما هو عمل الاقتصادي القياسي على وجه الخصوص ؟

تصور أن هناك باخرة كبيرة تزن حوالي 60000 طن. الباخرة صنعت منذ 25 سنة، خلت فيها قاعة الاستراحة، وفيها مطاعم و سينما ومقهى وبار ومكتبة ونايت كلوب ومركز تجاري وغيرها من الخدمات. والسؤال الذي يطرح نفسه: ما هو عمر الرُبَان لهذه السفينة؟

للبرهة الأولى، لا نستطيع مباشرة الإجابة عن هذا السؤال؛ لأن المعلومات التي أعطيت عن الباخرة لا شيء منها يحدد أو يشير إلى تحديد عمر الربان. ولكن هذا ليس كذلك. الباخرة تزن 60 ألف طن هذا يحدد باخرة كبيرة جداً ،

أي أن المسؤولية التي تقع على عاتق الربان كبيرة، وفقط شخص ذو خبرة كبيرة يمكن له قيادتها. وفي الوقت نفسه مطلوب أن يكون الربان متمتعاً بمهارة فيزيائية، وكذلك حذاقة خلاّبة للمفاجئات. إذاً العمر المتوقع للربان يجب أن يكون بين الـ40 و 60 سنة. الاقتصادي القياسي حدد العمر المتوقع للكابتن بـ 50 سنة بانحراف معياري مقداره 5 سنوات. هذا مثال جيد حيث استطاع الاقتصادي القياسي أن يحدد عمر الربان بدون أن تتوفر لديه بيانات كافية، وذلك بالاعتماد على الخبرة، أي: بالاعتماد على نظرية الاحتمالات دون البيانات الإحصائية الأساسية.

الأمثلة الاقتصادية الواقعية عادة تكون صعبة أكثر مما نتوقع. الكثيرون من القياسيين يعتقدون أن الهدف الأساسي من الاقتصاد القياسي التطبيقي هو معرفة النظرية الاقتصادية من جهة، ومن جهة ثانية المشاهدات الإحصائية. هذا يحمل في نفسه اختبار الفرضية، مثال نظرية الترشيد في الاستهلاك. وتنحصر وظيفة الاقتصادي القياسي في تأكيد النظرية الاقتصادية و تطابقها مع البيانات الإحصائية المتوفرة. لا أحد يقول أن هذه مسألة سهلة، ولكن هذا ممكن؟ في عام 1995 ناقش كازينكامب Keuzenkamp و مجانيوس Magnus في مقالة موجهة إلى القراء بعنوان: هل تغيرت آراء القراء في بعض القضايا الاقتصادية. إذا وجدت إجابة على مثل هذا العمل فإنه يعتبر اختباراً ناجحاً لاختبار النظرية، ولكن لم يجب أحد.

الاقتصاد القياسي والفيزياء : من الشيء الطبيعي لا يمكن صياغة قانون عام شامل يصف الظواهر الاقتصادية كافة، أو يكون على مستوى التعميم في مختلف البلدان. من محددات هذا القانون وعدم إمكانيته في التطبيق أو الحصول على النتائج التي نحصل عليها منه، تعالوا نناقش قانوناً شاملاً بشكل مختصر. لنأخذ قانون متغير من الفيزياء حيث كان قانوناً صحيحاً يمكن استخدامه للكشف

عن قانون أخر. لنراقب أقمار جوبتر يمكن ملاحظة أنهم أحياناً يسبقوا ثماني دقائق، وأحياناً أخرى يتأخروا بمقدار 8 دقائق من وقت لآخر. ووجدوا أن أقمار جوبتر تتسارع عندما تكون قريبة من الأرض، وتتباطأ عندما تكون بعيدة عن الأرض، وهذا عكس قانون نيوتن . وكان أولاف ريمر في عام 1675 قد قاس سرعة الضوء وقدرها بـــ 214000 كم/ثانية، وهي قيمة قريبة من الفعلية تنقص عنها بحوالي 30%.

لكن هذه الأفعال لا يمكن أن تحصل في القياس الاقتصادي. يكون تفاعل الناس والشركات والمنظمات مع بعضها بعض في مستويات مختلفة معقداً؛ لدرجة يكون من الصعب صياغة وضع نموذج رياضي، أو التنبؤ بأفعالها.

النظرية والتطبيق: يظهر للوهلة الأولى: أن الفجوة بين النظرية والتطبيق في الاقتصاد القياسي كبيرة مقارنة مع العلوم الأخرى، مثل الفيزيـــاء والطـــب وغيرها. ومن السائد في أوساط الاقتصاديين إما أن يكون الاقتصادي متمســـكاً بالنظرية الاقتصادية ويدافع عنها وإما أن يكون متمسكاً بالتطبيق ويدافع عنـــه. مما يعقد الأمور التفاهم بين النظرية والتطبيق؛ لذلك عندما نجد شخصاً ما يجمع النظرية مع التطبيق فإنه يكون قادراً على وضـــع حلـــول لـــبعض المشـــاكل الاقتصادية التي تعترضه، وهذا يعني النظرية بدون التطبيق لا تؤدي الغـــرض المطلوب، وكذلك التطبيق دون النظرية لا يؤدي الغرض.

يحضرني المثل الذي يضرب للمقارنة بين النظرية والتطبيق، التقى شخص صديقاً له في الليل يبحث عن مفتاح قد فقده، فسأل الشخص عن ماذا تبحث في هذا الليل الدامس تحت مصباح عامود الكهرباء؟ فأجابه صديقه: لقد فقدت مفتاحي. فسأله: أين فقدته؟ فقال: هناك مشيراً إلى بعد خمسين متراً. إذا لماذا لا تبحث عنه هناك حيث فقدته؟! وهنا يظهر الفرق ما بين النظرية

والتطبيق، النظرية تبحث عن الحلول حيث حصلت المشكلة في الوقت الذي فيه التطبيق يبحث عن الحل في الضوء.

1-5- أساليب الاقتصاد القياسي:

هل توجد أساليب للاقتصاد القياسي؟ نعم ، هذه الأساليب موجودة، ولكن هذه الأساليب ليست مصاغة على شكل علاقات جاهزة تقدم الحلول لكل مشكلة بشكل فوري وجذري. ولو كان ذلك محققاً لفقد الاقتصاديون القياسيون في العالم لقمة عيشهم كما فقد جزء منها الإحصائيون. هل هذا يعني وجود مدخل بحثي إلى هذا العلم ؟ إذا نظرنا في المقالات المنشورة في المجلات المحكمة فإننا نجد مثل هذا المدخل؛ لأن أي بحث منشور سنلاحظ فيه ما يلي: المقدمة، الدراسات السابقة، نموذج معالجة البيانات، بعض المشاكل الاقتصادية القياسية، وكيف استطاع الباحث معالجتها بالأساليب التطبيقية، النتائج العملية، ومن ثم الخاتمة. أي: يتوفر المدخل إلى كيفية معالجة النتائج التطبيقية للعمل، ولكن يملك علاقة ضعيفة في ترجمة خصوصية الدراسة. ربما المقالة لا تحتوي قسم تجارب المجالات التي تبين كيف عمل البحث بحثه، وما هي الأخطاء التي وقع فيها وغيرها.

الاقتصاد القياسي التقليدي: يبين أن الباحث عمل على بناء نموذج، جمع البيانات، واختار الأسلوب الأكثر مواءمة ومن ثم اختبر النموذج. بعد ذلك كيف اختار النموذج الموافق، وهذه الخطوة يمكن أن نبينها بأشكال مختلفة: قدر معلمات النموذج (مثال: المرونات) ، اختبر الفرضيات، بين الأهمية، عمل تنبؤ، أو عمل توصيات ومقترحات في السياسة الاقتصادية. هذا أسلوب جيد، ولكنه ليس عملياً، لأنه متفائل كثيراً .

وهذا ما يؤكده الاختلاف بين الفيزياء والاقتصاد، بأنه لا يوجد نموذج قياسي يمكن استخدامه في كل الحالات. والشيء الذي يمكن أن نطبقه، هو إمكانية تطبيق النموذج على شيء محدد؛ وهذا الأمر يتعلق بالسؤال الذي يجب أن يجيب عنه الباحث، ونسميه أدوات الباحث.

وبرأيي إن اختيار أدوات الدراسة يعتبر أهم خطوة في الطريق الصحيح للبداية. فكل الخطوات اللاحقة من اختيار للنموذج والبيانات الضرورية، وأسلوب التقدير تتعلق به. وكل ما قيل سابقاً يمكن أن يساعد الاقتصاديين القياسيين.

بالطبع فإن الاقتصاديين القياسيين عندهم تصورات مختلفة لإمكانية إجراء الأبحاث التطبيقية. المؤتمر الدولي المنعقد في كمبردج عام 1985 ومؤتمر الأسترالي- الآسيوي المنعقد في كانبير عام 1988 تضمنت برامجها مناقشاتها الأساليب القياسية. في مجمل المناقشات كان القياسيون قد اتفقوا على نقطة واحدة وهي الدخول من الأعلى إلى الأسفل في البحث. و ينص هذا الأسلوب (من الأعلى إلى الأسفل) على البدء في النماذج الكبيرة المتضمنة كمية كبيرة من المتغيرات واختبار معنوياتها. إذا كانت نتائج الاختبارات الإحصائية للمتغيرات غير معنوية يمكن حذف هذه المتغيرات من النموذج. بعد جراء عدة خطوات من هذا النوع نصل إلى النموذج الذي يعكس جوهر المتغيرات للدراسة. مبدئياً هذه الأساليب تعتبر رائعة في الوصول إلى القوة النظرية في التحديد. ولكن يوجد الكثير من الصعوبات، نذكر منها: يبدو هذا الأسلوب ليس عملياً، وخاصة إذا بدأت العمل بنموذج يحتوي كل المتغيرات التي تعكس الظاهرة المدروسة، وتبدأ بحذف المتغيرات حسب المعنوية الإحصائية ، فتصل في النهاية إلى نموذج لا معنى له ولا يعبر عن الدراسة المطلوبة. تبين من خلال هذا الأسلوب أن لا أحد من الاقتصاديين القياسيين يسلك هذا الأسلوب. وعادة هم يتبعون أسلوب من

الأسفل إلى الأعلى. من خلال هذا الأسلوب يبدأ الباحث في النموذج المبسط ويركبه. وهذا الأسلوب هو المتبع في كثير من مجالات البحث الأخرى.

وفي النتيجة يوجد عدد من النظريات تصف التحليل من الأعلى إلى الأسفل، ولكن لا يوجد نظرية لوصف الحالة من الأسفل إلى الأعلى التي تستخدم في التطبيق. لا يوجد نظرية تحدد للاقتصاديين القياسيين كيفية الانتقال من الأقل إلى الأكثر، على سبيل المثال كيفية اختيار المتغيرات الجديدة إذا كان النموذج حساساً ، حيث إن عدد المتغيرات غير كافٍ. وهذه ليست مسألة سهلة للنظرية القياسية، ولكنها ممكنة التطبيق.

1--6- خطوات البحث القياسي الاقتصادي:

عند إجراء البحث القياسي الاقتصادي دائماً يجب مراعاة الخطوات التالية:

1) بناء النموذج الاقتصادي الرياضي وصياغة علاقاته الرياضية.

2) تحويل النموذج الرياضي إلى نموذج إحصائي احتمالي.

3) جمع وتبويب البيانات اللازمة لتقدير معلمات النموذج.

4) تميز العلاقات المكونة للنموذج.

5) التأكد من إمكانية قياس كل متغير تفسيري على حدة.

6) الإلمام بالآثار المرتقبة لأخطاء التجميع على تقدير معلمات النموذج.

7) اختبار أسلوب القياس المناسب لتقدير معلمات النموذج.

8) اختبار معنوية التقديرات لمعلمات النموذج.

9) تحديد القدرة التنبؤية للنموذج.

10) المعالجة على الحاسوب.

1) بناء النموذج الرياضي للظاهرة موضع البحث:

يتعامل الاقتصاد الكلي مع المتغيرات الاقتصادية كمجموع مثل الدخل، الاستثمار، الصادرات، الواردات، والناتج، لذا توصف العمليات الاقتصادية في شكل علاقات بين المتغيرات الاقتصادية الكلية. ويتم تحديد الظاهرة عادة في ضوء المعلومات التي يمكن أن تتوفر فيها المصادر التالية:

أ- النظرية التي يقدمها علم الاقتصاد بخصوص تفسير الظاهرة محل البحث؛ بحيث تساعد في تحديد أهم المتغيرات المحددة للظاهرة والشكل الرياضي للعلاقة بين المتغيرات، مع الأخذ بعين الاعتبار إشارة بعض المعلمات، أو مدى القيم التي تأخذها.

ب- كذلك يجب الأخذ بعين الاعتبار المعلومات الفنية أو التكنولوجية للظاهرة موضع البحث، مثال: استجابة لهكتار من محصول معين لنوع معين من الأسمدة، أو تحديد طول الفترة لظهور منتج في السوق والقرار (فترة الإبطاء) أو تحديد بعض المعالم التي تؤثر على استجابة تغيرات الأسعار، كوجود دعم لسلعة معينة، أو وجود تعاقدات مسبقة تحدد جودة الإنتاج.

جـ- الدراسات التي سبق لها أن تعرضت للظاهرة موضع البحث، سواء كان ذلك من الناحية التحليلية أو النظرية التطبيقية، والتي لم يتم دمجها بعد في النظريات الاقتصادية.

مما سبق نستنتج الخطوات الأساسية لوضع النموذج بشكل عام. و نجمل هذه الخطوات على النحو الآتي:

1) دراسة الجملة المختارة وجمع المعطيات وتتم على مرحلتين أو ثلاثة مراحل:

ينبغي أن تقيس الأثر على الإنتاج؛ نتيجة تغير عامل من عوامل الإنتاج مع ثابت الكمية المستخدمة من العامل الآخر.

ولكن هل يمكن قياس أثر كل عامل على حدة علماً أن البيانات المستخدمة في الغالب يتغير فيها العاملان معاً وفي الوقت نفسه؟

يكون ذلك ممكناً إذا كان الارتباط غير قوي بين المتغيرات التفسيرية. أما إذا كان الارتباط قوياً فإن المعلمات التي نقدرها لن تكون دقيقة، أي أنها لن تعكس كل متغير تفسيري على حدة. وكذلك في الحالة التي يكون فيها المتغيران التفسيريان الارتباط بينهما تام، فإنه لا يمكن قياس أثر كل منهما. وسوف نستعرض ذلك بالتفصيل لدى دراسة الافتراضات الخاصة عند نموذج الانحدار الخطي المتعدد.

6) الإلمام بالآثار المرتقبة لأخطاء التجميع على تقدير معلمات النموذج:

إن أغلب البيانات التي يستخدمها الاقتصاديون تنطوي على درجة ما من التجميع الذي يسبب وجوده وجود بعض الأخطاء، والتي تسبب بعض التحيز في تقدير معلمات النموذج.

حيث يعتبر الدخل القومي مفهوماً تجمعياً ؛ لأنه يمثل مجموع دخول الأفراد، وكذلك الإنتاج القومي وإنتاج مزرعة معينة في سنة محددة يعتبر مفهوماً تجمعياً؛ لأن إنتاجها عبارة عن جمع المنتجات المتحققة خلال المواسم الزراعية خلال السنة المحددة نفسها. كذلك فإن دوال الإنتاج يختلف بعضها عن بعض، وذلك حسب نوع المنتج، وهذا يؤدي إلى تميز في قياس معلمات دالة الإنتاج الكلي.

ومن المعروف أنه إذا أردنا توخي الدقة، فمن المفروض الحصول على دوال الإنتاج الجزئية. ولكن عادة لا تكون دوال الإنتاج المتاحة إلا على

المستوى الكلي. والسؤال الذي نقف عنده الآن: هل يمكن تفسير العلاقة بين هذه المتغيرات التجميعية لمجموع المنشآت كدالة الإنتاج بالمعنى بنفسه الذي تأخذه دالة الإنتاج على المستوى الجزئي، وبالطبع فإن الجواب يتوقف على ما إذا كان للدالة التجميعية الخصائص نفسها و الشكل ذاته الذي تأخذ الدوال الجزئية، بحيث يمكن مساواة الإنتاجية الحدية للعمل المشتقة من الدالة الكلية بمعدل الأجر، كما تطبق في حالة الدالة الجزئية.

وهذا أمر لا يتوقع حدوثه عادة بمجرد ربط مجموع إنتاج المنشآت بمجموع كميات عناصر الإنتاج المختلفة، وإنما يتطلب الأمر قواعد معينة لإتمام عملية التجميع، واستخدام أشكال جبرية للدالة الإنتاجية يسهل معها تحويل الدوال الجزئية إلى دالة تجميعية، وعموماً يمكن التوصل إلى دالة الإنتاج التجميعية إذا كانت لدوال الإنتاج الجزئية الشكل الجبري التالي:

$$V = \alpha L^{b} + \beta K^{c}$$

أي أنه يمكن تقسيم الإنتاج إلى مكونتين مضافتين، مكونة ترجع لعنصر رأس المال، ومكونة ترجع لعنصر العمل.

أي أنه يفترض عدم وجود تفاعل بين العمل ورأس المال مثلما يحدث إذا كانت العلاقة تحوي على حاصل ضربهما. وبالتالي هذا الشرط لا ينطبق على دالة كوب دوغلاس Cobb-Douglas وبالتالي يمكن القول أنه لا يمكن الحصول على دوال الإنتاج الجزئية إذا كانت الدالة التجميعية من هذا النوع.

7) اختيار الأسلوب القياسي المناسب لتقدير معلمات النموذج:

نستطيع تقسيم أساليب القياس المناسبة والمتاحة إلى قسمين رئيسيين هما:

1- الأساليب الخاصة لتقدير كل معادلة على حدة (أسلوب المعادلة البسيطة).

2- الأساليب الخاصة لتقدير معادلات نموذج متعدد مرة واحدة.

يمكن إدخالها في النموذج، أو محاولة تقديرها من بيانات أدق حيث غالباً ما يعزى سوء التقدير إلى أخطاء في البيانات.

2- اختبار معنوية التقديرات للمعلمات وفقاً للمعايير الإحصائية:

الاختبارات الإحصائية يقرر الباحث من خلالها أمرين هامين:

أ- مقدرة النموذج على تفسير الظاهرة محل البحث.

ب- مدى الثقة في تقديرات معلمات النموذج.

للتوضيح نستعرض المثال التالي:

في دراسة إنتاج الألبان جمعت بيانات من (30) مزرعة لقياس علاقة تفسير المتغيرات في نسبة حصيلة الألبان إلى الإيرادات الكلية للمزرعة.

$$Y_i = a + b X_{1i} + cX_{2i} + u_i$$

حيث: X_{1i} عدد الأبقار الموجودة في المزرعة.

X_{2i} المساحة الكلية في المزرعة بالهكتارات

وبتطبيق طريقة المربعات الصغرى العادية حصلنا على العلاقة التالية:

$$\hat{Y}_i = 4.5 + 0.42x_{1i} + 0.11x_{2i}$$

وعند دراسة مقدرة العلاقة المقدرة على تفسير المتغيرات في نسبة مبيعات الألبان إلى الإيرادات الكلية في المزرعة (Y_i) فإن المعيار الإحصائي المناسب هو معامل التحديد (r^2) والذي يعطي نسبة التغير في (Y). أي النسبة من تباين القيمة (Y) التي يمكن تفسيرها باستخدام المعادلة الثابتة، وعموماً كلما ارتفعت قيمة (r^2) كلما كان ذلك دليلاً على قوة المعادلة المستخدمة في تفسير المتغير التابع. في هذا المثال وجدنا أن ($R^2 = 0.41$) أي أن العلاقة المقدرة لا تفسر إلا 0.41 من التغير الكلي من نسبة حصيلة الألبان إلى الإيرادات الكلية في المزرعة. وهذه القيمة تدل على ضعف العلاقة المقدرة من زاوية الظاهرة محل البحث.

لابد من الانتباه إلى أنه هناك اعتبارات كثيرة تؤخذ بعين الاعتبار لمسؤوليتها عن انخفاض قيمة معامل التحديد.

أحد هذه الاعتبارات: أن العلاقة الحقيقية هي علاقة غير خطية، في حين أن R^2 مقياس لمدى الارتباط الخطي. ومن ناحية أخرى قد تشير R^2 إلى أن هناك متغيرات تفسيرية هامة لم تؤخذ بعين الاعتبار؛ لذلك ينبغي إدراج هذه المتغيرات فيها.

وعموماً يمكن زيادة معامل التحديد (R^2) بإضافة متغيرات تفسيرية جديدة، ولكن هذا يتعارض مع أحد معايير النموذج، وهي أن تكون له القدرة على تفسير الظاهرة بعدد محدود من المتغيرات التفسيرية. وهذا ما تعرضنا له سابقاً في فقرة اختيار الأسلوب القياسي المناسب لتقدير معلمات النموذج.

كذلك من الممكن الحصول على معامل تحديد R^2 مرتفع من غير أن يعني ذلك ارتفاعاً حقيقياً في القوة التفسيرية للعلاقة المقدرة؛ نتيجة وجود اتجاهات عامة قوية في متغيرات العلاقة.

لذلك ينبغي مراعاة الحصر الشديد في تفسير معامل التحديد، وعدم التسرع في إعطاء معنى اقتصادي للعلاقة المقدرة في ضوء قيمة معامل التحديد وحده.

ويجب معرفة أن قدرة النموذج تكون ضعيفة عندما تكون معلمات (أو نسبة كبيرة منها غير معنوية) حتى ولو كانت R^2 عالية جداً، وهذا أمر متكرر الحدوث في التطبيقات العملية.

مدى الثقة في تقديرات معلمات النموذج من الناحية الإحصائية: يقصد بذلك تحديد إمكانية الاعتماد عليها كتقديرات المعلمات الإحصائية في المجتمع الذي أخذت منه العينة المدروسة، أي: نريد أن نتحقق التقديرات التي حصلنا عليها ليست ناتجة عن الصدفة المصاحبة لاختيار العينة، وإنها انعكاس أصيل

للعلاقة السارية للمجتمع المدروس، ويمكن التحقق من ذلك بإجراء الاختبارات المعنوية المعروفة.

مثل اختبار (فشير F وستودينت t) لاختبار معنوية المعلمات، وعلى بالرغم من أن هذه الاختبارات تدور حول إذا ما كانت قيمة المعلمات المقدرة تختلف عن الصفر؛ إلا إن الاختبارات هذه ضعيفة، وفائدتها قليلة.

فقد يكون الأمر الهام في تطبيقات عديدة هو التأكد من قياس مدى اختلاف القيمة المقدرة لمعامل معين عن قيمة محددة.

مثال ذلك اختبار أن $\sigma = 6$ في المثال السابق، وذلك باستخدام اختبار ستودنت (t).

ولكن الذي يجب معرفته عدم كفاية هذه الاختبارات المعنوية لتحديد قيمة النموذج، وإنما يجب أن تترافق مع معاملات عالية نسبياً، وأن تكون المعادلة ذات مغزى اقتصادي.

تأتي الاختبارات الإحصائية في المرتبة الثانية بعد الاختبارات الاقتصادية التي يجب أن تتحقق، وهذا ما ذكرناه سابقاً، وذلك لأجل تسهيل الاختبارات الإحصائية؛ لتكون على أساس سليم في رفض أو قبول التقديرات الناتجة.

3- اختبار معنوية التقديرات طبقاً لمعنوية الاقتصاد القياسي:

لاختبار معنوية التقديرات وفقاً لمعايير الاختبار الاقتصادي هو التأكد من صحة الاختبارات؛ التي يشترط فيها توفر تطبيق أسلوب الاقتصاد القياسي المستخدم في تقدير هذه المعلمات؛ ذلك لأن هذه الافتراضات لها أهمية من الناحيتين التاليتين:

أ- إن الاختبارات الإحصائية التي سبق ذكرها لا تصح إلا إذا تحققت هذه الافتراضات فمثلاً يفترض في طريقة المربعات الصغرى العادية أن قيم المتغير العشوائي الموجود في العلاقة غير مرتبطة مع بعضها بعضٍ. وإذا لم يتحقق

ذلك فإن الأخطاء المعيارية للمعلمات المقدرة تصبح غير دقيقة في قياس انتشار تقديرات كل معلمة حول القيمة الصحيحة لهذه المعلمة، وحيث إن الخطأ المعياري عنصر أساسي في اختبار المعنوية فإنه يترتب على عدم الثقة في قياسه عدم سلامة اختبار المعنوية ذاتها. ولذلك يصبح من الضروري التحقق من سلامة الغرض بالنسبة للبيانات المستخدمة.

ب- إن تقديرات المعلمات التي نحصل عليها باستخدام أسلوب القياس الاقتصادي، يمكن أن يتمتع ببعض الخصائص المرغوب فيها. وإن مدى تحقق هذه الخصائص يتوقف على مدى الافتراضات الخاصة لنموذج الانحدار، وأهمها الافتراضات الخاصة بتوزيع عنصر الخطأ وعلاقته بالمتغيرات داخل النموذج؛ ولذلك أصبح من الضروري التحقق من تلك الافتراضات حتى يمكن القول فيما إذا كانت الخصائص المرغوب فيها من التقديرات متحققة في الحالة محل البحث أم لا.

4) خصائص جودة المعلمات المقدرة:

يوجد بعض الخصائص، حيث يدل توفرها في تقديرات المعلمات لعلاقة اقتصادية عن جودة هذه التقديرات وسلامة التقدير الإحصائي، وهي:

أ- عدم التحيز unbiased

التحيز هو عبارة عن الفرق بين الأمل الرياضي أو الوسط الحسابي لتقديرات المعلمة $(\hat{a})$ والقيمة الحقيقية للمعلمة (a).

أي : $\mathbf{Bias = [E(\hat{a}) - a]}$

وبناء على هذه العلاقة نقول إن المعلمة $(\hat{a})$ هي تقدير غير متحيز للمعلمة (a) ، إذا كانت قيمة التحيز صفر، أي إذا كان $[E(\hat{a}) = a]$.

ب- أصغر تباين Minimum Variance:

نفترض أننا أجرينا قياس للمعلمة (a) باتباع أحد أساليب القياس المعتمدة باستخدام عدد كبير من العينات ذات الحجم الثابت (n) وإننا حسبنا التباين لهـذه التقديرات المختلفة، ويساوي إلى: $\mathrm{Var}(\hat{a}) = \mathrm{E}(\hat{a} - \mathrm{E}(\hat{a}))^2$

فإذا كان التباين $(\hat{a})$ أصغر من تباين كل التقديرات التي يمكن الحصول عليها بتطبيق أساليب قياس أخرى، فإننا نقول: التقدير $(\hat{a})$ يتمتع بخاصية أصغر تباين . فإذا كانت $(\hat{a})$ هي تقدير آخر للمعلمة a فإن خاصية أصغر تباين فـي التقدير $(\hat{a})$ تتوفر عندما يكون: $\mathrm{Var}(\hat{a}) < \mathrm{Var}(\hat{\hat{a}})$

أي: $\mathrm{E}(\hat{a} - \mathrm{E}(\hat{a}))^2 < \mathrm{E}(\hat{\hat{a}} - \mathrm{E}(\hat{\hat{a}}))^2$

ويدعى التقدير $(\hat{a})$ في هذه الحالة أفضل تقدير للمعلمة a

ج- الكفاءة Efficiency

نقول عن تقدير معين $(\hat{a})$ للمعلمة a تقدير كفء عنـدما يتـوفر فيـه الخاصيتان السابقتان الوقت نفسه، أي: عندما يكون التقدير غير متحيـز، ولـه أصغر تباين بالمقارنة مع كل التقديرات غير المتحيزة الأخرى للمعلمــة، ولــذا يوصف التقدير $(\hat{a})$ في هذه الحالة بأنه أفضل تقدير غير متحيز للمعلمة a.

Best unbiased Estimate عندما $E(\hat{a}) = a$

$$Var(\hat{a}) < \mathrm{Var}(\hat{\hat{a}})$$

نقول: إن $(\hat{a})$ أفضل تقدير غير متحيز للمعلمة (a)

4- أفضل تقدير خطي غير متحيز (Best Linear unbiased Estimator (B.L.U.E.)):

نعتبر التقدير $(\hat{a})$ تقديراً خطياً في دالة ما عندما تكون $(\hat{a})$ دالة خطية في قيم متغير التابع لنفترض أننا نقيس العلاقة:

$$Y_i = a + bX_i + u_i$$

وبطريقة القياس حصلنا على التقديرين $(\hat{a})$ و $(\hat{b})$ في هذه الحالــة نقــول: إن $(\hat{a})$ و $(\hat{b})$ يعتبران أفضل تقديرين خطيين غير متحيزين إذا كان:

$$(\hat{a}) = vY_1 + vY_2 + \ldots\ldots + vY_n$$
$$(\hat{b}) = wY_1 + wY_2 + \ldots\ldots + wY_n$$

حيث $v_1 , v_2 , .. v_n$ و $w_1 , w_2 , .. w_n$ ثوابت

فإن الوسط الحسابي $(\hat{w})$ للعينة $(w_1 \ldots w_n)$ يعتبر تقديراً خطياً، حيث إنه يحسب كدالة خطية لقيم العينة، و يتضح ذلك بإعادة كتابة تعريف الوسط الحسابي على الشكل:

$\hat{Y} = \frac{1}{n}\sum_{i=1}^{n} X_i$ والتي يمكن كتابتها بالشكل التالي:

$$\overline{Y} = \frac{1}{n}Y_1 + \frac{1}{n}Y_2 + \ldots\ldots + \frac{1}{n}Y_n$$

أما الوسيط والمنوال فهما ليسا تقديران خطيان لقياس النزعة المركزية للتوزيع، ويرد برهان هذه الحقيقة تحت عنوان (نظرية غاوس – ماركوف) Gauss – Markov

ومما سبق يمكننا أن نعتبر أي تقدير معين لـ $(\hat{a})$ هو أفضل تقدير خطي غير متحيز للمعلمة a إذا توفرت فيه الشروط التالية:

أ– $(\hat{a})$ تقدير غير متحيز (a) أي: $(\hat{a})\ a = E$

ب– $(\hat{a})$ تقدير له أصغر تباين بالمقارنة بكل التقديرات الأخرى:

$$(\hat{a}) \le \mathrm{Var}\ (\hat{\hat{a}})\ \mathrm{Var}$$

جـ– $(\hat{a})$ عبارة عن تشكيلة خطية من قيم مشاهدات معينة.

5– أصغر متوسط مربعات الأخطاء (M. S. E) : تعتبر هذه الخاصية مزيجاً من خاصية عدم التحيز وخاصية أصغر تباين، والمقصود بالمتوسط مربعات الأخطاء للتقدير $(\hat{a})$ هو التوقع الرياضي لمربعات الفروق بين تقديرات المعلمة $(\hat{a})$ والقيمة الحقيقية للمعلمة a

$$E\ (\hat{a} - a)^2 = MSE(\hat{a})$$

بإضافة توقع (â) ثم طرحه داخل القوس بالطرف الأيمن من المعادلة نجد:

$$\begin{aligned}\mathbf{MSE} &= \mathbf{E}\left(\hat{a} - \mathbf{E}(\hat{a}) + \mathbf{E}(\hat{a}) - a\right)^2 \\ &= \mathbf{E}\left[\left(\hat{a} - \mathbf{E}(\hat{a})\right) + \left(\mathbf{E}(\hat{a}) - a\right)\ \right]^2 \\ &= \mathbf{E}\left(\left(\hat{a} - \mathbf{E}(\hat{a})\right)^2 + \left(\mathbf{E}(\hat{a}) - a\right)^2\right) + 2\mathbf{E}\left[\left(\hat{a} - \mathbf{E}(\hat{a})\right)\left(\mathbf{E}(\hat{a}) - a\right)\right]\end{aligned}$$

نلاحظ أن المقدار الأول من الطرف الأيمن في المعادلة الأخيرة عبارة عن تباين (â) ((â) Var).

وإن المقدار الثاني هو مربع تحيزه (â) $(\hat{a})\ (\mathrm{Bias})^2$

أما المقدار الأخير فهو يساوي الصفر لأنه عبارة مجموع فروفات القيم عن أوساطها الحاسبية .

وتصبح العلاقة : $\mathbf{MSE} = \mathbf{Var}(\hat{a}) + \left(\mathbf{Bias}(\hat{a})\right)^2 + 0$

6- الكفاية Sufficiency: نقول: إن (â) تقديراً كفاياً إذا كان مقدراً بطريقة تستخدم كل المعلومات المتوفرة في العينة عن المعلمة الحقيقية، أي: إذا كانت تستخدم في تقدير a كل مشاهدات العينة.

فالوسط الحسابي $\overline{\mathbf{Y}}$ للعينة (Y_1 .... Y_n) يعتبر تقديراً كافياً ، لأنه يستخدم كل معلومات العينة في حين أن الوسيط والمنوال ليسا من التقديرات الكفاية ، لأننا نستخدم في تقدير الوسيط أو المنوال إلا بعض المشاهدات العينة.

وفي النهاية نخلص إلى القول: إن أفضل طريقة لقياس معلمات نموذج معين هي تلك التي تعطي تقديرات تتمتع بعدد من الخصائص المرغوب فيها؛ أكبر من عدد الخصائص التي تتمتع بها أي تقديرات يمكن الحصول عليها بأية طريقة أخرى للقياس، ويمكن القول بأنه يفضل عادة ذلك التقدير الذي يجمع بين خاصتين: عدم التحيز، وأصغر تباين في الوقت بنفسه.

9) اختبار القدرة التنبؤية للنموذج:

بما أن أحد الأهداف الهامة للقياس الاقتصادي هو استخدام النماذج التي يتم قياسها في التنبؤ، وبالتالي التعرف على مسار الظاهرة محل البحث في المستقبل، ولذا فإنه من الأهمية بمكان أن يتم اختبار قدرة النموذج على التنبؤ، فقد تكون تقديرات النموذج فرضية تماماً وفقاً للمعايير السابقة، مع ذلك يكون النموذج قاصراً عن إعطاء تقديرات أو تنبؤات يمكن الاعتماد عليها ، يمكن إجراء اختبارات القدرة النموذجية للتنبؤ بالشكل التالي:

أ- وذلك عن طريق مدى استقرار التقديرات، أي: مدى حساسيتها للتغير مع تغير حجم العينة: ويمكن معرفة ذلك بإضافة متغيرات جديدة للعينة السابق استخدامها في التقدير ومن ثم إعادة التقدير باستخدام العينة الموسعة ومن المتوقع أن تختلف التقديرات الجديدة للمعلمات عن تقديرات المعلمة الأصلية ومن المهم أن لا تكون الفوارق من التقديرات كبيرة جداً من الناحية الإحصائية أي أن لا تكون هذه الفوارق معنوية إحصائية ويمكن ذلك باستخدام المقاييس الإحصائية المعروفة.

ب- استخدام النموذج المقدر في التنبؤ في مسار الظاهرة، وذلك خلال فترة ماضية لم تدخل بياناتها في العينة الأصلية؛ التي تم استخدامها في تقدير معلمات النموذج.

ومن المتوقع أيضاً في هذه الحالة أن تختلف القيم التنبؤية للمتغيرات التابعة عن قيمها الحقيقية وإذا ثبت من الاختبارات المعنوية أيضاً أن الفوارق بين القيم التنبؤية والفعلية لفترة معلومة هي فوارق معنوية إحصائياً؛ فإنه يلزم

في هذه الحالة إعادة النظر في تقديرات النموذج، سواء في استخدام بيانات أحدث أو بيانات أدق، أو الاستعانة بمعلومات إضافية عن الظاهرة المدروسة.

10) المعالجة على الحاسوب : بعد الانتهاء من استكمال خطوات البحث الأساسية كافة، نقوم بوضع الخطوات الأساسية للحل (أي خوارزمية الحل)، ورسم المخطط التدفقي لسير عمليات والخطوات اللازمة لخوارزمية الحل. ومن ثم ترجمة هذه الخوارزمية إلى برنامج مكتوب بإحدى لغات البرمجة المعروفة. لكن يجب ألا يغيب عن أذهاننا أنه يوجد الكثير من البرامج الإحصائية الجاهزة للقيام بمثل هذا النوع من التحليل. من هذه البرامج على سبيل المثال لا الحصر:
(BMDP , SAS , SPSS ,Statistica, Stp, Statgraphics)

1-7- أنواع البيانات الإحصائية:

نظراً لأهمية البيانات الإحصائية في الدراسات القياسية وجدنا من الضرورة بمكان التعريف بالبيانات الإحصائية بأشكالها المختلفة .

تعتبر البيانات الإحصائية الركيزة الأساسية لعلم الإحصاء، وهذه البيانات تأتي عادة من الملاحظات المسجلة عن بعض أفراد المجتمع الإحصائي، وقد يعبر عنها بأرقام مختلفة ترمز إلى وحدات قياسية، أو هي أرقام مطلقة، إذ قد تدل على الوزن والطول والحجم والزمن، والمستوى الثقافي أو نتائج اختبارات أو تجارب معينة. نطلق عادة على البيانات تسمية المتغيرات. نعرف **المتغير**: وهو عبارة عن قيم النتائج التي تخص ظاهرة ما، أو هي ما ندعوه بمجموعة المشاهدات للظاهرة المدروسة. وهذه المشاهدات غالباً ما تأخذ قيماً كمية أو نوعية، وهذه القيم متغيرة (أي: متحولة وليست ثابتة). وقد تتصف المشاهدات الخاصة بالمتغير بالقياسات الكمية وهي عبارة عن استقصاءات متكررة تتعلق بظاهرة ما لفترات زمنية مختلفة وتأخذ قيماً قياسية. وعادة تسجل سلوك المجموعة نفسها من الوحدات الاقتصادية الجزئية خلال فترة زمنية محددة، أو قد تكون صفة المتغير منفصلة كالنوع، والجنس، والجنسية، واللون، والدين، وغيرها.

وأيضاً يمكن القول أن المتغيرات قد تعبر عن سلاسل زمنية لظاهرة ما. مثال ذلك البيانات المتعلقة بإجمالي الصادرات والواردات في سورية خلال الفترة الزمنية 1970 ولغاية 2004. أو أن تكون قيم المتغيرات عبارة عن بيانات عرضية (مقطعية) وذلك عندما نعبر بها عن ظاهرة ما عند نقطة زمنية معينة، أو خلال فترة زمنية محددة، كأن ندرس ميزانية الأسرة في عام 2003 في جمهورية اليمن. وحسب معيار دلالة هذه البيانات، أو المعطيات، أو المتغيرات التي يمكننا تصنيفه إلى الأنواع التالية:

المتغيرات التي تصادفنا في الدراسات الاقتصادية والاجتماعية المحددة لأي ظاهرة كانت من الظواهر المدروسة:

أ - **المتغيرات المستقلة (Independent Variable):** وهي تلك المتغيرات التي تؤثر في المتغير التابع، وهي أيضاً عبارة عن متغيرات السببية، و تنقسم إلى قسمين:

1- المتغيرات الداخلية: وهي المتغيرات التي تؤثر بشكل مباشر على المتغير التابع.

2- المتغيرات الخارجية: وهي المتغيرات التي تؤثر بشكل غير مباشر على المتغير التابع.

ب - **المتغير التابع (Dependent Variable):** وهو ذلك المتغير الذي يتأثر بالمتغيرات المستقلة، أي: هو متغير النتيجة.

مثال: فإذا كانت درجة المؤهل العلمي هو المتغير المستقلY فإن الدخل التصرفي X يكون متغير تابع، وذلك لأنه كلما ارتفعت درجة المؤهل العلمي كلما ارتفع دخل الفرد: $X \xrightarrow{\;/\;} Y$

وتصادفنا في دراسة الظواهر الاقتصادية والاجتماعية بعض المتغيرات ؛ التي ترتبط فيما بينها بعلاقة متبادلة، أي: كل متغير يؤثر في المتغير الآخر ويتأثر به. وتدعى المتغيرات المتبادلة، ومثال على ذلك: من أجل تحصيل التعليم العالي لابد من دفع المصاريف والتي بدورها تحتاج إلى دخل مرتفع لتسديدها. وعندما نحصل على التعليم العالي نحصل على دخولاً مرتفعة:

المتغيرات المتبادلة $X \rightleftarrows Y$

جـ - المتغير المتداخل (Intervening Variable): هو المتغير الذي يكون نتيجة من نتائج المتغير المستقل، ومحدوداً، أو شرطاً لحدوث المتغير التابع. تظهر هذه الحالة أثناء تحليل البيانات، وتحديد المتغيرات المستقلة والمتغير التابع. عندما يحاول الباحث إبعاد أثر عدد من المتغيرات فإن العلاقة تضعف نتيجة إبعاد متغير متداخل، و تصبح قوية في حال وجوده (ونسميه في بعض الأحيان بالمتغير الوسيط):

المتغير Z متغير متداخل (وسيط)

X ⟶ Z ⟶ Y

مثال: عند دراسة جنوح الأطفال نجد أن الأسر التي حدث فيها طلاق تساهم بنسبة أعلى في ظاهرة الجنوح عند الأحداث؛ من الأسر التي يعيش فيها الأبوان تحت سقف واحد معاً. وعند تحليل الأسباب، نجد أن من أهم الأسباب التي يمكن أن يفكر بها الباحث هو دور الأب في الإشراف على تربية الطفل. وهذا يدفعنا إلى القول: إن الأطفال الذين يعيشون في منزل فيه أم فقط لا يحصلون على العناية الضرورية لتشرب القيم الأساسية، ولتعلم احترام النظم والقوانين الرئيسية.

المتغير متداخل: تفكك أسري ⟵ غياب الأب ⟵ جنوح الأطفال

نلاحظ أن هذه العلاقة تزداد شدتها في غياب دور الأب بعين الاعتبار وتضعف، بحضوره ولكنها لا تزول نهائياً في حالة استبعاد دور الأب.

د - المتغير الدخيل (Extraneous Variable): هو عبارة عن المتغير الذي يكون السبب في ظهور علاقة بين متغيرين (المتغير المستقل والمتغير التابع)، فقد يكون السبب الحقيقي للعلاقة التي ظهرت أثناء تحليل البيانات بين المتغيرين، وإن كليهما كان نتيجة السبب المشترك في ظهور مثل هذه العلاقة.

وفي واقع الحال لا توجد العلاقة بين المتغيرين التي افترضها الباحث، و إن هناك متغيراً تابعاً ومتغيراً مستقلاً:

$$\begin{tikzcd} & Y \\ Z \arrow[ur, dashed] \arrow[dr, dashed] & \\ & X \arrow[uu, dashed] \end{tikzcd}$$

مثال : بفرض أن عدد المسافرين بالطائرات يزداد كلما ارتفع عدد الأسر الحاصلين على خدمة الهاتف في المجتمع، أي يعتقد الباحث أنه بسبب زيادة خدمة الهاتف أدت إلى سهولة تأمين الحجز بالطائرات والاستفسارات عن السفر. بينما واقع الحال أن كلا الظاهرتين هما نتيجة التقدم العلمي والتقني الذي دخل المجتمع الحديث، أي إنها نتيجة مباشرة لظاهرة واحدة.

هـ - المتغير السابق (Antecedent Variable): هو عبارة عن المتغير الذي يحدث في زمن سابق للمتغير المستقل ويؤثر فيه، وبذلك يساهم في حدوث الظاهرة أو التأثير في المتغير التابع بشكل غير مباشر :

$$Z \rightarrow X \rightarrow Y$$

مثال: وجود علاقة بين ظاهرة التدين في المجتمع وظاهرة تقديم الخدمات الاجتماعية للغير دون مقابل. نجد أن أهمها هو ظاهرة التدين في المجتمع، لأنه كلما كان المجتمع ملتزماً دينياً كلما ارتفعت درجة احتمال أن يكون أبناء هذا المجتمع أكثر تكافلاً اجتماعياً. ومنه نستنتج أن المتغير الذي يعكس تدين المجتمع هو متغير سابق لإظهار العلاقة بين التدين عند أفراد المجتمع، وتبادل تقديم الأعمال الخيرية:

تدين المجتمع ← أفراد متدينون ← تبادل تقديم الأعمال الخيرية

وأيضاً يمكننا تقسيم المتغيرات إلى متغيرات كمية ومتغيرات نوعية:

المتغيرات الكمية: وهي بدورها تنقسم إلى متغيرات عشوائية مستمرة، ومتغيرات عشوائية منفصلة. والمتغيرات النوعية تنقسم إلى متغيرات ثنائية ومتغيرات متعددة.

أنواع البيانات

البيانات الطولية (بيانات السلاسل الزمنية Time Series Data): وهي عبارة عن البيانات التي تعرض قراءة لمتغير ما خلال فترة زمنية محددة	**بيانات العرضية (بيانات عرضية المقطع Cross Section Data)**: وهي عبارة عن البيانات التي تعكس متغيراً ما لظاهرة ما خلال فترة زمنية معينة	**بيانات طولية عرضية (سلسلة زمنية مقطعية):** وهي عبارة عن بيانات تجمع ما بين صفات السلسلة الزمنية والصفة المقطعية (العرضية)
البيانات الرتبية: عبارة عن المتغيرات المرتبة ترتيباً تصاعدياً أو تنازلياً، أو حسب معيار معين دون اعتبار مدى الفرق بين وحدات هذه القياسات	**البيانات الموقعية:** وهي عبارة عن البيانات التي لا يمكن تصنيفها فيما بينها في وضع نظام من الظواهر	**البيانات الاسمية:** هي عبارة عن البيانات التي تصنف المشاهدات في المرتبة الموائمة لها، ثم تسجيل أعداد أفراد المرتبة في الخانة المقابلة لها، بحيث إن كل فرد يدخل في صنف واحد أو طبقة واحدة دون الأخرى، أي أن كل فئة تقتصر على أفرادها دون غيرهم

أنواع المتغيرات

المتغيرات المستقلة (Independent Variable): وهي تلك المتغيرات التي تؤثر في المتغير التابع
المتغير التابع (Dependent Variable): وهو ذلك المتغير الذي يتأثر بالمتغيرات المستقلة $Y \leftarrow X$
المتغير المتداخل (Intervening Variable): هو المتغير الذي يكون نتيجة من نتائج المتغير المستقل، ومحدوداً، أو شرطاً لحدوث المتغير التابع $Y \leftarrow Z \leftarrow X$
المتغير السابق (Antecedent Variable): هو عبارة عن المتغير الذي يحدث في زمن سابق للمتغير المستقل ويؤثر فيه، وبذلك يساهم في حدوث الظاهرة أو التأثير في المتغير التابع بشكل غير مباشر: $Y \leftarrow X \leftarrow Z$
المتغير الدخيل (Extraneous Variable): هو عبارة عن المتغير الذي يكون السبب في ظهور علاقة بين متغيرين (المتغير المستقل والمتغير التابع)، فقد يكون السبب الحقيقي للعلاقة التي ظهرت أثناء تحليل البيانات بين المتغيرين، وإن كليهما كان نتيجة السبب المشترك في ظهور مثل هذه العلاقة 

متغيرات منفصلة: عندما يأخذ المتغير قيماً صحيحة محدودة، ولا يوجد فيها أرقام كسرية		**متغيرات مستمرة:** عندما يأخذ المتغير أية قيمة عددية قد تكون كسرية أو صحيحة ضمن مجال معين

تمارين غير محلولة

1- ما المقصود بالمتغير العشوائي . وما مبررات استخدامه في النموذج الاقتصادي؟

2- ماذا تعني العبارة التالية: "المتغير (X) له توزيع طبيعي" ؟

3- ما الفرق بين العلاقة الرياضية والعلاقة الإحصائية والعلاقة القياسية استشهد ببعض الأمثلة من الواقع .

4- ما أوجه الاختلاف والشبه بين النظرية الاقتصادية وكل من الاقتصاد الرياضي، الاقتصاد القياسي ، الإحصاء الاقتصادي؟

5- هل يعد موضوع الاقتصاد القياسي مرادفاً لموضوع تحليل الانحدار؟ وكيف ذلك؟

6- تنص النظرية الاقتصادية "على وجود علاقة طردية بين ارتفاع الدخل القومي وزيادة الإنفاق على الاستهلاك " كيف يمكن أن يعبر عن هذه علاقة رياضياً وقياسياً؟ اذكر خمس علاقات اقتصادية رياضية وقياسية متشابهة ؟

7- كيف يختلف ويتطابق الأسلوب الوصفي مع الأسلوب الكمي في معالجة الظواهر الاقتصادية ؟ أيهما تفضل؟ ولماذا؟

8- حدد أهم وظائف الاقتصاد القياسي؟ وبين كيف يختلف عن العلوم النظرية؟

9- بأية طريقة، ولأي غرض يجتمع كل من النظرية الاقتصادية والرياضيات والتحليل الإحصائي لتكوين علم الاقتصاد القياسي؟

10- ما المقصود بالعبارة الأتية: "ليس من الضروري أن تتضمن علاقة الارتباط بين متغيرين علاقة سببية"؟. أعط خمسة أمثلة على حالات الارتباط اللامعقول بين المتغيرات .

11- اذكر مراحل خطوات المعالجة القياسية للظواهر الاقتصادية قيد الدرس، وناقشها بتركيز، مع إعطاء أمثلة من النظرية الاقتصادية.

12- ضع تخطيطاً توضحيًا تبين فيه مراحل المعالجة القياسية للنظرية الاقتصادية .

13- تحدث عن أهداف الاقتصاد القياسي.

14- ما رأيك في منهج "الاقتصاد القياسي دون النظرية"؟ أعط بعض الأمثلة لتوضيح مدى العون الذي يمكن أن يستمده الباحث من النظرية الاقتصادية، وهو بصدد تقدير بعض العلاقات الاقتصادية".

15- ناقش العبارة التالية: "يعد الاقتصاد القياسي هو المقيم للنظرية الاقتصادية ".

16- ناقش العبارة الآتية:"تعد مرحلتا التطبيق والتنبؤ أهم خطوتين من خطوات المعالجة القياسية" .

17- تحدث عن أنواع المتغيرات.

18- عرف ما يلي: البيانات الفعلية – المتغير المستمر – المتغير المنقطع – المتغيرات الرتبية – البيانات الموقعية – البيانات الاسمية –البيانات الطولية العرضية – المتغيرات المستقلة – المتغير التابع – المتغير المتداخل – المتغير الدخيل – المتغير السابق .

19- تحدث عن أنواع المتغيرات.

20- كيف نميز بين المتغيرات ؟

الفصل الثاني

المتوسطات والتباين المشترك ومعامل الارتباط

لقد درسنا في السنوات السابقة مقاييس النزعة المركزية والتشتت وسنحاول في هذا الفصل التذكير بأهمها مقاييس النزعة المركزية: الوسط الحسابي – الوسيط – الربيع الأعلى والربيع الأدنى – المنوال – الوسط الهندسي – الوسط التوافقي – الوسط التربيعي. أما مقاييس التشتت: المدى المطلق، الانحراف المتوسط، والانحراف الربيعي ثم الانحراف المعياري ومعامل الاختلاف.

2-1- الأوساط الحسابية:

2-1-1- الوسط الحسابي (Arithmetic Mean) : الوسط الحسابي لمجموعة من البيانات هو عبارة عن مجموع هذه البيانات مقسومة على عددها. أي هو القيمة التي إذا أعطيت لكل المفردات لما تغير المجموع الكلي للمفردات قبل تبديل وبعده .

1- حساب الوسط الحسابي لبيانات غير مبوبة:

$$\overline{X}=\frac{X_1+X_2+...+X_n}{n}=\frac{1}{n}\sum_{i=1}^{n}X_i$$

حيث إن : n– حجم العينة.

4- لا يمكن حسابه بالنسبة للبيانات النوعية (الوصفية) بشكل مباشر.

5 - يتأثر بالقيم المتطرفة والشاذة.

6- يعتبر الوسط الحسابي ضابطاً تقارن به كل المفردات الأخرى للتوزيع، عندها يكون قيمة ممثلة لهذا التوزيع.

7- يعتبر من أكثر المتوسطات الحسابية سهولة ووضوحاً في التفسير.

2-1-2- الوسيط (Median): الوسيط لمجموعة من المشاهدات هو عبارة عن القيمة الوسطى يسبقها ويليها تنازلياً أو تصاعدياً 50%. أي أن الوسيط يعتبر مقياس موقع لأنه يتحدد من خلال موقعه بين المشاهدات وليس مقياس موضوعياً كما هو الوسط الحسابي الذي يأخذ بكل المفردات بعين الاعتبار والوسيط لا يهتم إلا بالقيمة التي تقسم المفردات إلى قسمين لكونه مقياس موقع.

حساب الوسيط في البيانات غير المبوبة :

لحساب الوسيط في البيانات غير المبوبة نتبع الخطوات التالية:

1- نرتب البيانات ترتيباً تصاعدياً أو تنازلياً.

2- نحدد موقع الوسيط ومن خلال العلاقة: $Loc(Med)=\frac{N+1}{2}$.

3- نبحث في البيانات المرتبة تصاعدياً عن قيمة الموقع ($Loc(Med)$)، ونحدد عندئذ قيمة الوسيط ونرمز له (Med).

حساب الوسيط في البيانات المبوبة :

يمكننا حساب الوسيط في البيانات المبوبة وذلك بإتباع الخطوات التالية:

1- نحسب التكرار التجميعي الصاعد ،ونرمز له ($\sum F_i\uparrow$) والذي عرفناه سابقاً على أنه تكرار الفئة مضافاً إلى مجموع تكرار الفئات التي تسبقها.

يسبقها 1/100 من البيانات ويليها 99/100 من البيانات وذلك بعد ترتيب البيانات بشكل تصاعدي.

موقع المئين الأول: $Loc\,(P_1) = \dfrac{\sum F_i}{100}$

يحسب من العلاقة: $P_1 = L + \dfrac{C}{F_m}\left(0.01\,N - F_b\right)$

و على نفس المنوال نحسب أي قيمة موقعية:

و تعميم العلاقة الموقعية كمقياس للنزعة مركزية نقترح العلاقة التالية:

$$K_1 = L + \frac{C}{F_m}\left(Loc\,(K_1) - F_b\right)$$

حيث أن: K_1 – القيمة الموقعية المراد تحديدها.

أي أن: $(K_1 = Med\,,\, Q_1\,,\, Q_2\,,\, Q_3\,,\, D_1 - D_9\,,\, P_1 - P_{100})$

L – الحد الأدنى للفئة الموقعية المراد حسابها.

C – طول الفئة الموقعية.

F_m– تكرار الفئة الموقعية المطلوبة.

F_b– التكرار الذي يسبق تكرار الفئة الموقعية في التكرار التجميعي الصاعد .

$Loc\,(K_1)$ – القيمة الموقعية المطلوبة (للموقع المطلوب).

يمكننا أن نستنتج التالي : $Med = Q_2 = D_5 = P_{50}$

أي أن الوسيط = الربيع الثاني = العشير الخامس = المئين الـ 50.

وبالطبع لهذه المقاييس أهمية في الدراسات الإحصائية النوعية، كما هو الحال في الدراسات الاجتماعية والتربوية وغيرها. فهي تستخدم لوصف المجموعات أو المشاهدات، وتستخدم هذه المقاييس للدراسات الأولية.

2-1-6- المنوال (Mode):

I – حساب المنوال لبيانات غير مبوبة : المنوال لمجموعة من المشاهدات هو القيمة الأكثر تكراراً وشيوعاً بين مفردات المشاهدات ويكفي لحساب المنوال للبيانات غير المبوبة بان نقوم بعد تكرار كل مفردة من المشاهدات والمفردة الأكثر تكراراً تعتبر قيمتها القيمة التمثيلية للمنوال. ونرمز للمنوال بـ Mod .

II – حساب المنوال في البيانات المبوبة : لتحديد المنوال في البيانات المبوبة نتبع الخطوات التالية:

1- نحدد الفئة المنوالية وهي الفئة التي يقابلها أكبر تكرار.

2- نحسب Δ_1 الفرق بين تكرار الفئة المنوالية وتكرار الفئة التي تسبقها.

3- نحسب Δ_2 الفرق بين تكرار الفئة المنوالية وتكرار الفئة التي تليها.

4- نطبق العلاقة: $$Mod = L + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right) * C$$

حيث أن: L- الحد الأدنى للفئة المنوالية.

C- طول الفئة المنوالية.

علاقة بيرسون في التماثل :

يرتبط الوسيط والوسط الحسابي والمنوال في علاقة موضحة أدناه للحكم على سوية التوزيع الإحصائي للبيانات المدروسة ذات المنوال الوحيد وفيه شيء من الالتواء فإن هناك علاقة تدعى علاقة بيرسون تحكم هذه المقاييس.

$$\bar{X} - Mod = 3\left(\bar{X} - Med\right)$$

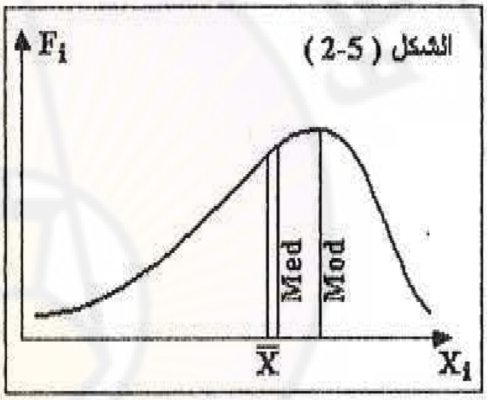
أما إذا كان التوزيع التكراري منتظماً أي أن كل فئتين تبتعدان نفس البعد عن فئة المركز لهما نفس التكرار عندئذ جميع المقاييس الإحصائية آنفة الذكر تتطابق أي أن: وسط الفئة ذات التكرار الطبيعي = $\bar{X} = Med = Mod$.

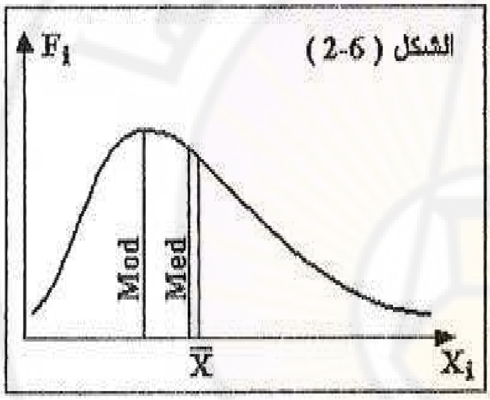
أما إذا كان التوزيع التكراري غير متماثل (غير منتظم) فإننا نميز الحالات التالية :

أ- التوزيع متلويا نحو اليمين: $\bar{X} - Mod > 0 \Rightarrow \bar{X} > Mod$

ب- التوزيع متلويا نحو اليسار: $\bar{X} - Mod < 0 \Rightarrow \bar{X} < Mod$

أنظر الرسم البياني الشكل (1 - 2) والشكل (2 - 2) .





إن الاستدلال الإحصائي عن المجتمعات المدروسة يكون أكثر سهولة عند التعامل مع الوسط الحسابي كمقياس للنزعة المركزية. وهذا ليس جزم وإنما إشارة لأنه في النهاية لكل ظاهرة إحصائية الأسلوب الأنجع في اختيار مقياس "نزعة المركزية المناسب.

2-1-7- الوسط الهندسي (Geometric Mean)

يعتبر الوسط الهندسي مقياس لحساب النزعة المركزية في البيانات الإحصائية المعطاة على شكل نسب مئوية أو أرقام قياسية. حيث يعتبر استخدام الوسط الهندسي في الأرقام القياسية أحد أهم استخداماته على الإطلاق. ويستخدم أيضاً لإيجاد متوسط نسب الزيادة في قيم المتغيرات في حالات إيجاد معدلات البيع أو التزايد السكاني.

الوسط الهندسي بالتعريف لمجموعة من القيم $x_1 , x_2 , ... , x_n$ والتي عددها N هو حاصل ضربها مجذوراً إلى عددها الجبري. ونرمز له بـ $\overline{G}$ ؛

أي أن :

$$\overline{G} = \sqrt[n]{X_1 \cdot X_2 \cdot ... \cdot X_n}$$

وإن حسابه من السهولة بمكان في ظل وجود الحاسب الآلي والحاسبات الإلكترونية المتوفرة بكثرة وبمتناول الجميع.

وبما أن الوسط الهندسي أسلوب رياضي فإنه يمكننا أيضاً أن نطبق عليه الخواص الرياضية مثلاً باستخدام خواص القوى نكتبه على الشكل التالي:

$$\overline{G} = (X_1 \cdot X_2 \cdot ... \cdot X_n)^{\frac{1}{n}}$$

أو أن نطبق عليه خواص اللوغاريتمات ويصبح على الشكل التالي:

$$Log(\overline{G}) = \frac{1}{n} Log(X_1 \cdot X_2 \cdot ... \cdot X_n) =$$

$$= \frac{1}{n}(Log(X_1) + Log(X_2) + ... + Log(X_n))$$

حساب الوسط الهندسي في البيانات المبوبة:

لحساب الوسط الهندسي في البيانات المبوبة نتبع الخطوات التالية:

1- نحسب أوساط الفئات (X_i)

2- نحسب لوغاريتم أوساط الفئات $(Log\ (X_i))$

3- نحسب جداء لوغاريتم وسط كل فئة بتكرارها ونجمع الناتج:

$$\left(\sum F_i * Log\ (X_i)\right)$$

4- نقسم $\left(\sum F_i * Log\ (X_i)\right)$ على مجموع التكرارات $\left(\sum F_i\right)$ فنحصل على لوغاريتم الوسط الهندسي للبيانات $Log\ (G)$

5- نوجد مقابل $Log\ (G)$ ، فنحصل على قيمة الوسط الهندسي .

$$\overline{G} = \sqrt[n]{\sum F_i\ X_i}$$

$$Log\ \overline{G} = \frac{1}{n} \sum F_i\ Log\ (X_i)$$

ونحن نعلم أن : $N = \sum F_i$

2-1-8- الوسط التوافقي (The Harmonic Mean):

I - حساب الوسط التوافقي في البيانات غير المبوبة:

الوسط التوافقي لمجموعة من المشاهدات هو عبارة عن مقلوب الوسط الحسابي لمقلوبات هذه القيم فإذا كان لدينا N مشاهدة ولتكن $X_1 , X_2 , ... ,$ X_n . فإن الوسط التوافقي لها يحسب من العلاقة التالية:

$$\overline{H} = \frac{1}{\frac{\sum_{i=1}^{n} \frac{1}{X_i}}{N}} = \frac{N}{\sum_{i=1}^{n} \frac{1}{X_i}}$$

قد تكون استعمالات هذا المقياس قليلة. ولكن له استعمالات خاصة ومنها على سبيل المثال، حساب متوسط السرعة لقطع مسافة معينة.

حساب الوسط التوافقي في البيانات المبوبة:

يمكننا حساب الوسط التوافقي في البيانات المبوبة وذلك بإتباع الخطوات التالية

1- نحسب أوساط الفئات (X_i).

2- نحسب مجموع نسب التكرارات على أوساط الفئات $\left(\sum_{i=1}^{n} \frac{F_i}{X_i}\right)$.

3- نطبق العلاقة التالية : $\overline{H} = \frac{\sum F_i}{\sum_{i=1}^{n} \frac{F_i}{X_i}}$

ملاحظة: إن قيمة الوسط الهندسي أكبر من الوسط التوافقي و أقل من الوسط الحسابي، أي: $\overline{H} < \overline{G} < \overline{X}$

2-1-9- الوسط التربيعي أو جذر متوسط المربعات (The Quadratic Mean)

ويرمز له بـــ $\overline{Qm}$ وهو عبارة عن الجذر التربيعي للوسط الحسابي لمربعات قيم المشاهدات أو المفردات.

ويعطى بالعلاقة: $\overline{Qm} = \sqrt{\frac{\sum X_i^2}{n}}$

حساب الوسط التربيعي في البيانات المبوبة:

نتبع الخطوات التالية :

1- نحسب أوساط الفئات (X_i).

2- نربع أوساط الفئات ونجمعها أي نحسب: $\sum X_i^2$.

3- نحسب حاصل ضرب X_i^2 في F_i ونجمع كافة النتائج، أي نحسب $\sum F_i X^2$.

4- نطبق العلاقة:

$$\overline{Qm} = \sqrt{\frac{\sum_{i=1}^{n} F_i X_i^2}{\sum F_i}}$$

2-2- مقاييس التشتت:

2-2-1- المدى المطلق (Absolute Range)

يعرف المدى المطلق: على أنه القيمة الناتجة من حساب الفرق بين أكبر مفردة في الترتيب وأصغر مفرده فيه . ونرمز له : $AR = X_{max} - X_{min}$ ويمكن أن ننسب المدى المطلق إلى الوسط الحسابي عند المقارنة لتميز الحالات المتشابهة في طول المدى والمختلفة في الوسط الحسابي بعضها عن بعض. ونحصل عندئذ على ما يسمى بالمدى النسبي : $PAR = \frac{AR}{\bar{X}} * 100$.

2-2-2- الانحراف الربيعي (Quadratic Dispersion)

الانحراف الربيعي هو عبارة عن الفرق بين قيمة الربيع الثالث والربيع الأول مقسوماً على 2 أي هو عبارة عن نصف المدى بين الربيع الثالث والأول ويعطى على الشكل التالي: $RQ = \frac{Q_3 - Q_1}{2}$

يعتبر كلاً من الربيع الأول والثالث من مقاييس النزعة المركزية، حيث بين كيفية حسابهما في الفصل السابق.

حساب الانحراف الربيعي النسبي لبيانات مبوبة:

يمكننا أن نسب الانحراف الربيعي إلى الوسيط أي الربيع الثاني:

$$PRQ = \frac{Q_3 - Q_1}{2Q_2} * 100 = \frac{Q_3 - Q_1}{2\,Med} * 100$$

أو يمكننا أن نأخذ العلاقة التالية لنعبر عن الانحراف الربيعي النسبي:

$$PRQ = \frac{Q_3 - Q_1}{Q_3 + Q_1} * 100$$

2-2-3- الانحراف المتوسط (Mean Deviation): وهو عبارة عن الوسط الحسابي لانحراف قيم المشاهدات عن أحد مقاييس النزعة المركزية بالقيمة المطلقة. وعادة يأخذ إما عن الوسط الحسابي أو الوسيط.

حساب الانحراف المتوسط في البيانات غير المبوبة: نتبع الخطوات التالية:

1- نحسب الوسط الحسابي أو الوسيط.

2- نحسب انحراف كل قيمة (كل مشاهدة) عن مقياس النزعة المركزية المختار (الوسط الحسابي أو الوسيط). ونرمز له d. ونهمل الإشارة الجبرية.

3- نجمع الانحرافات ونقسمها على عددها $\sum d_i/n$.

4- نطبق العلاقة: $MD = \frac{\sum |d_i|}{n}$.

حساب متوسط الانحرافات في البيانات المبوبة: نتبع الخطوات التالية لحسابه:

1- نحسب الوسيط (Med) أو الوسط الحسابي $(\bar{X})$.

2- نحسب أوساط الفئات (X_i).

3- نحسب انحراف وسط كل فئة عن الوسيط ونأخذ القيمة المطلقة (d_i).

4- نحسب جداء تكرار كل فئة بانحراف وسطها المطلق عن الوسيط .

5- نطبق العلاقة التالية :

$$MD = \frac{\sum_{i=1}^{n} F_i |d_i|}{\sum_{i=1}^{n} F_i}$$

2-2-4- الانحراف المعياري (Standard Deviation)

لا تقتصر أهميته كونه مقياس تشتت، و إنما تنبع أهميته أيضاً كمؤشر أساسي من مؤشرات التوزيعات الاحتمالية، حيث يعتبر الانحراف المعياري من أهم مقاييس التشتت وأكثرها أهمية ودلالة. فكل البيانات الإحصائية سواء كانت مبوبة أو غير مبوبة يمكن عرضها على شكل رسوم بيانية أو على شكل منحنى احتمالي للتوزيع والذي يختلف شكله من ظاهرة إلى أخرى. ولكن هذه المنحنيات الاحتمالية (المنحنيات التكرارية) يمكن التميز فيما بينها من خلال مؤشرين أساسيين وهما:

1 – **الوسط الحسابي**: حيث يحدد لنا تمركز البيانات أي يبين لنا موقع قمة هذا المنحنى.

2– **الانحراف المعياري**: يبين لنا كيف توزعت المشاهدات للظاهرة المدروسة حول نقطة الوسط الحسابي. وهذا يسمح لنا بالتعرف على شكل المنحنى واتساعه. كما إن الانحراف المعياري يسمح لنا بتحديد شدة الالتواء من خلال العلاقة: $\frac{\overline{X} - Mod}{S}$

ومن هنا تنبع أهمية الانحراف المعياري كضابط لكل التوزيعات التكرارية (أي الاحتمالية) التي تخضع لها البيانات الإحصائية.

حساب الانحراف المعياري لبيانات غير مبوبة :

الانحراف المعياري لمجموعة من المشاهدات هو عبارة عن الجذر التربيعي لمتوسط مربع انحرافات المشاهدات عن وسطها الحسابي، فإذا كانت

لدينا N مشاهدة قيمها: $X_1 , X_2 , \ldots , X_n$ ، فإن الانحراف المعياري لهـذه القيم:

$$\sigma = \sqrt{\frac{\sum (X_i - \bar{X})^2}{N}}$$

حساب الانحراف المعياري لبيانات غير مبوبة بإسلوب الانحرافات :

بفرض أن الوسط الحسابي الفرضي هو $\bar{A}$ إذاً يمكننا أن نكتب $d_i = X_i - \bar{A}$

حيث d_i - انحرافات القيم عن الوسط الحسابي الفرضي:

$$\therefore \sum X_i = N\bar{A} + \sum d_i \Rightarrow \bar{X} = \bar{A} + \bar{d}$$

وحيث أن: $\bar{d} = \frac{\sum d_i}{N}$

نعوض في العلاقة: $\sigma = \sqrt{\frac{\sum [(X_i - \bar{A}) - \bar{d}]^2}{N}}$

فنجد أن: $\sigma = \sqrt{\frac{\sum (X_i - (\bar{A} + d))^2}{N}} = \sqrt{\frac{\sum [(X_i - \bar{A}) - \bar{d}]^2}{N}}$

$$\sigma = \sqrt{\frac{\sum (d_i - \bar{d})^2}{N}} = \sqrt{\frac{\sum d_i^2}{N} - \left(\frac{\sum d_i}{N}\right)^2}$$

حساب الانحراف المعياري في البيانات المبوبة:

أ- **الأسلوب المباشر:** ولحسابه نتبع الخطوات التالية:

1- نحسب أوساط الفئات (X_i).

2- نحسب مربع أوساط الفئات (X_i^2).

3- نحسب $(F_i X_i)$.

4- نضرب تكرار كل فئة بـــ (X_i^2) ونحصل على $(F_i X_i^2)$.

5- نطبق العلاقة التالية : $\sigma = \sqrt{\frac{\sum F_i X_i^2}{\sum F_i} - \left(\frac{\sum F_i X_i}{\sum F_i}\right)^2}$

ب- **بأسلوب الانحرافات** : ولحسابه نتبع الخطوات التالية :

1- نحسب أوساط الفئات (X_i).

2- نحسب انحرافات X_i عن الوسط الحسابي المفترض $(\overline{A})$ أي:

$(d_i = X_i - \overline{A})$.

3- نحسب $\sum F_i d_i$.

4- نحسب d_i^2 .

5- نحسب $(F_i d_i^2)$.

6- نطبق العلاقة التالية: $\sigma = \sqrt{\frac{\sum F_i d_i^2}{\sum F_i} - \left(\frac{\sum F_i d_i}{\sum F_i}\right)^2}$

ج- **أسلوب الانحرافات المختصرة** : ويستخدم هذه الأسلوب إذا كانت الفئات متساوية. ولحسابه نتبع الخطوات التالية :

1- نحسب أوساط الفئات (X_i).

2- نحسب انحرافات القيم عن الوسط الحسابي المفترض A أي:

$(d_i = X_i - \overline{A})$.

3- نقسم d_i على طول الفئة C فنحصل على: $U_i = \frac{d_i}{C}$.

4- نحسب $\sum F_i U_i$.

5- نحسب U_i^2 .

6- نحسب $(\sum F_i U_i^2)$.

7- نطبق العلاقة التالية: $\sigma = \left(\sqrt{\frac{\sum F_i U_i^2}{\sum F_i} - \left(\frac{\sum F_i U_i}{\sum F_i}\right)^2}\right) * C$

2-2-5- معامل الاختلاف (Coefficient of Variation)

عند مقارنة التشتت بالنسبة للمشاهدات بالنسبة للظواهر المتماثلة نجـد أن هناك صعوبة في المقارنة وإذا ما قارنا تكون العملية خاطئة لذلك نلجأ إلى مـا يسمى بمعامل الاختلاف (CV) وهو عبارة عن المقياس النسبي الذي يمكننا من التخلص من تأثيرات وحدة القياس التي لم نستطيع تفسيرها في مقاييس التشتت. ومعامل الاختلاف هو عبارة عن نسبة الانحراف المعياري إلى الوسط الحسابي والضرب بمائة، أي :

$$CV = \frac{\sigma}{\bar{X}} * 100$$

2-3- التباين (Variance):

التباين هو مربع الانحراف المعياري أي أن التباين هو متوسط مربع الفروق لقيم المشاهدات عن وسطها الحسابي. ولحساب التباين نطبق كافة العلاقة التي نحسب من خلالها الانحراف المعياري بعد تربيعها.

خواص الانحراف المعياري والتباين:

1- عادة نرمز للانحراف المعياري لبيانات العينة بالرمز S عندئذ التباين S^2 للعينة أما للمجتمع الإحصائي نرمز له بـــ σ_x ولتباين المجتمع بـــ σ_x^2 .

2- إذا أضفنا (أو طرحنا) عدد ثابت (a) إلى كل قيمة (أو من كل قيمة) X_i من قيم المشاهدات فإن الانحراف المعياري والتباين لا يتغير.

3- إذا حسبنا الانحراف المعياري أو التباين عن أية قيمة أخرى B غير الوسط نجد أن قيمة الانحراف المعياري عن الوسط الحسابي هي أقل ما يمكــن عن أية قيمة أخرى غير الوسط الحسابي:

$$\sigma_B = \sqrt{\frac{1}{N}\left(\sum_{i=1}^{n} X_i - B\right)^2} > \sigma_x$$

4- إذا كان توزيع البيانات يخضع للتوزيع الطبيعي أي شكل المنحنى كشكل الجرس المقلوب المتماثل حول الوسط الحسابي فإن 68.28 % من مساحة المنحنى تقع ضمن المجال $(\overline{X} \mp \sigma_x)$ أي يبعد انحراف معياري واحد عن الوسط الحسابي للتوزيع على الجانبين. و 95.45 % من المساحة تحت المنحنى أي من مشاهدات التوزيع تتوزع ضمن مجال $(\overline{X} \mp 2\sigma_x)$ أي يبعد انحرافتين معياريين على جانبي الوسط الحسابي. و 99.73 % من المساحة تحت المنحنى أي من قيم المشاهدات تتوزع ضمن المجال $(\overline{X} \mp 3\sigma_x)$ أي يبعد ثلاثة انحرافات على جانبي الوسط الحسابي. أنظر الشكل (1-2).

2-4-العلاقة بين مقاييس التشتت :

حالة خاصة : إذا كانت التوزيعات التكرارية ملتوية قليلاً نجد أن:

$$MD = 4/5 * \sigma_x$$

ومن أجل التوزيع الطبيعي $MD = 0.7979$

وإذا عرفنا متغيراً جديداً على إنه Z فإن: $Z = \dfrac{X_i - \overline{X}}{\sigma_x}$

فإننا نطلق عليه التوزيع الطبيعي المعياري. فإن الوسط الحسابي لهذا التوزيع الجديد Z يساوي الصفر والانحراف المعياري والتباين يساوي الواحد. ونرمز له بـ $N \sim (0.1)$.

2-5- قواعد تباين العينة:

تعريف تباين العينة Definition Of Sample Variance : على فرض أنه لدينا عدداً من المشاهدات العشوائية للمتغير Y فإن تباين المتغير Y هو متوسط مربع انحرافات القيم عن وسطها الحسابي. إن تباين العينة يعامل

كحالة خاصة من التباين المشترك للعينة. انظر إلى هذه العلاقة خارج معادلة التربيع .

$$\begin{aligned}\mathbf{Var}(X) &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})(X_i - \overline{X}) \\ &= \mathbf{Cov}(X, X)\end{aligned}$$

أي أننا حصلنا على علاقة تباين مشترك لعينة المتغير X مع نفسه، أي يمكننا اشتقاق بعض قواعد التباين من قواعد التباين المشترك. في البداية يمكننا أخذ تباين مجموع متغيرين عشوائيين.

القاعدة الأولى للتباين : إذا كان لدينا العلاقة التالية:

If $Y = V + W$,

فإن تباين Y يساوي:

$$\text{Var}(Y) = \text{Var}(V) + \text{Var}(W) + 2\text{Cov}(V, W)$$

البرهان:

كتابة تباين المتغير Y مع نفسه كتباين مشترك، ومن ثم نبدل واحدة من Y.

$$\text{Var}(Y) = \text{Cov}(Y, Y) = \text{Cov}(Y, [V + W])$$

و بتطبيق القاعدة الأولى للتباين المشترك:

$$\text{Var}(Y) = \text{Cov}(Y, V) + \text{Cov}(Y, W)$$

لنبدل قيمة Y في كل حالة من الحالات:

$$\text{Var}(Y) = \text{Cov}([V + W], V) + \text{Cov}([V + W], W)$$

نستخدم القاعدة الأولى للتباين المشترك مرتين فنحصل على:

$$\begin{aligned}\text{Var}(Y) = \text{Cov}(V, V) + \text{Cov}(W, V) \\ + \text{Cov}(V, W) + \text{Cov}(W, W)\end{aligned}$$

مع الأخذ بعين الاعتبار أن التبيان المشترك $\mathrm{Cov}(V, V)$ تماماً هو تباين: $\mathrm{Var}(V)$ و التبيان المشترك $\mathrm{Cov}(W, W)$ تماما هو تباين $\mathrm{Var}(W)$ ، ومع ملاحظة أن ترتيب المتغيرات بالنسبة لعلاقة التباين المشترك لا تؤثر عليه. أي $\mathrm{Cov}(W, V)$ هي العلاقة نفسها $\mathrm{Cov}(V, W)$:

$$= \mathrm{Var}(V) + \mathrm{Var}(W) + 2\mathrm{Cov}(V, W)$$

القاعدة الثانية للتباين:إذا كان لدينا العلاقة : $\boldsymbol{Y = bZ}$

حيث b قيمة ثابتة

فإن : $\mathrm{Var}(Y) = b^2\mathrm{Var}(Z)$

البرهان : نكتب التباين كتباين مشترك للمتغير Y مع نفسه، و نبدل إحدى حالات المتغير Y و نستخدم القاعدة الثانية من التباين المشترك، فنحصل على الشكل التالي:

$$\mathrm{Var}(Y) = \mathrm{Cov}(Y, Y) = \mathrm{Cov}(Y, bZ)$$
$$= b\mathrm{Cov}(Y, Z)$$

نستبدل حالة Y الثانية فنجد :

$$= b\mathrm{Cov}(bZ, Z)$$
$$= b^2\mathrm{Cov}(Z, Z)$$

في نهاية المطاف نستخدم مرة أخرى قاعدة التباين المشترك الثانية، فنحصل على المطلوب:

$$\mathrm{Var}(Y) = b^2\mathrm{Var}(Z)$$

القاعدة الثالثة للتباين: معلوم أن التباين المشترك للثابت يساوي الصفر، فإذا كان لدينا العلاقة : $\boldsymbol{Y = b}$ حيث b قيمة ثابتة

فإن: $\mathrm{Var}(Y) = 0$

البرهان : من أجل برهان ذلك سوف نستخدم الحالة الثالثة من قاعدة التباين المشترك.

$$\mathrm{Var}(Y) = \mathrm{Cov}(Y, Y)$$

$$= \mathrm{Cov}(b, b)$$
$$= 0$$

القاعدة الرابعة للتباين: إذا كان لدينا : $Y = V + b,$

حيث b قيمة ثابتة

فإن: $\mathrm{Var}(Y) = \mathrm{Var}(V)$

البرهان : من أجل برهان ذلك نبدأ بتطبيق القاعدة الأولى للتباين:

$$\mathrm{Var}(Y) = \mathrm{Var}(V + b)$$
$$= \mathrm{Var}(V) + \mathrm{Var}(b) + 2\mathrm{Cov}(V, b)$$

وباستخدام القاعدة الثالثة للتباين و قاعدة التباين المشترك نلاحظ أن الحدين الأخيرين مساويين للصفر، ونحصل على العلاقة:

$$\mathrm{Var}(Y) = \mathrm{Var}(V)$$

لماذا إضافة الثابت إلى المتغير غير مؤثر بالتباين؟

إن إضافة قيمة ثابتة للمتغير Y سوف يؤدي إلى إزاحة النقاط نحو اليمين بمقدار قيمة الثابت، وبالتالي المتوسط أيضاً ينزاح بالقيمة بنفسها.

إذاً لا يحصل انحراف جديد في العينة، وبالتالي فإن تباين العينة أيضاً لا يتأثر ؛ بذلك لأنه يعتمد فقط على مربع الانحرافات القيم المشاهدات عن المتوسط.

وفي هذا السياق نجد أن التباين المشترك للعينة يكتب على النحو الآتي :

$$\mathrm{Cov}(X,Y)=\frac{1}{n}\left[\sum_{i=1}^{n}X_iY_i\right]-\overline{X}\overline{Y}$$

وبالتالي يمكن اشتقاق علاقة تباين العينة كحالة خاصة من التباين المشترك، أي:

$$\mathrm{Var}(X)=\mathrm{Cov}(X,X)=\frac{1}{n}\left[\sum_{i=1}^{n}X_iX_i\right]-\overline{X}\overline{X}$$

$$=\frac{1}{n}\left[\sum_{i=1}^{n}X_i^2\right]-\overline{X}^2$$

وهذا يعني أن تباين العينة يعتمد على تقدير تباين المجتمع. ويمكننا أن نستنتج ذلك بسهولة بالاعتماد على معامل التصحيح $(n-1)/n$.

$$E[\mathrm{Var}(X)]=\frac{n-1}{n}\sigma_X^2$$

ومعلوم أنه يمكننا تحديد التقدير غير المتحيز لتباين المجتمع تماماً وذلك بضرب تباين العينة بالعلاقة $n/(n-1)$

$$s_X^2=\frac{n}{n-1}\mathrm{Var}(X)$$

إن المضاريب تختصر مع بعضها بعض ،وهنا يظهر تباين العينة S^2 كتقدير غير متحيز لتباين المجتمع :

$$E\left[s_X^2\right]=E\left[\frac{n}{n-1}\mathrm{Var}(X)\right]$$

$$=\frac{n}{n-1}E[\mathrm{Var}(X)]$$

$$=\frac{n}{n-1}\times\frac{n-1}{n}\sigma_X^2=\sigma_X^2$$

2-6- تحديد التباين المشترك للعينة:

Definition of Sample Covariance

من جملة الأسئلة التي نهتم في الإجابة عنها عند دراسة متغيرين : هل يرتبط هذان المتغيران بعلاقة طردية أم عكسية ؟ أي: هل القيم الكبرى لواحد منهما عادة ما تقترن بالقيم الكبرى للآخر علاقة طردية؟ أم أن القيم الكبرى للمتغير الأول عادة ما تكون مصاحبة للقيم الصغرى للثاني (علاقة عكسية)؟ وتسمى المعلمة التي تحسب هذه العلاقة سواء كانت عكسية أم طردية بالتباين المشترك (التغاير) ، وعلى فرض أنه لدينا عينة مؤلفة من n مشاهدة للمتغيرين X,Y فإن التباين المشترك للعينة هو متوسط ناتج جداء كل متغير عن وسطه الحسابي ونعبر عن ذلك رياضياً:

$$\mathbf{Cov}(X,Y)=\frac{1}{n}\left[(X_1-\bar{X})(Y_1-\bar{Y})+\ldots+(X_n-\bar{X})(Y_n-\bar{Y})\right]$$
$$=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})$$

مثال(2-1): لنفترض انه لدينا الجدول التالي الذي يبين عدد سنوات التدريب S و Y ما يكسبه في الساعة بالدولار لعشرين فرد في أمريكا، من خلال المسح الوطني للشباب عام 1992:

الجدول رقم (1-2)

N	Y	S	$Y-\bar{Y}$	$X-\bar{X}$	$(Y-\bar{Y})(X-\bar{X})$
1	18.38	11	6.4	-0.8	-5.104
2	10.08	12	-1.9	0.2	-0.384
3	17	12	5.0	0.2	1
4	17.92	15	5.9	3.2	18.944
5	12	13	0.0	1.2	0
6	12	9	0.0	-2.8	0
7	18	13	6.0	1.2	7.2
8	6.83	12	-5.2	0.2	-1.034
9	9	12	-3.0	0.2	-0.6
10	12	11	0.0	-0.8	0
11	12	12	0.0	0.2	0
12	13	16	1.0	4.2	4.2
13	16	13	4.0	1.2	4.8
14	9.38	12	-2.6	0.2	-0.524
15	9.61	11	-2.4	-0.8	1.912
16	6.5	9	-5.5	-2.8	15.4
17	5.5	14	-6.5	2.2	-14.3
18	14	8	2.0	-3.8	-7.6
19	8.65	7	-3.4	-4.8	16.08
20	12	13	0.0	1.2	0
المجموع	239.9	235	-0.1	-1	39.99
المتوسط	12.0	11.8			1.9995

والمطلوب: حساب التباين المشترك للبيانات أعلاه.

الحل: في البداية نقوم برسم شكل الانتشار للعينتين:

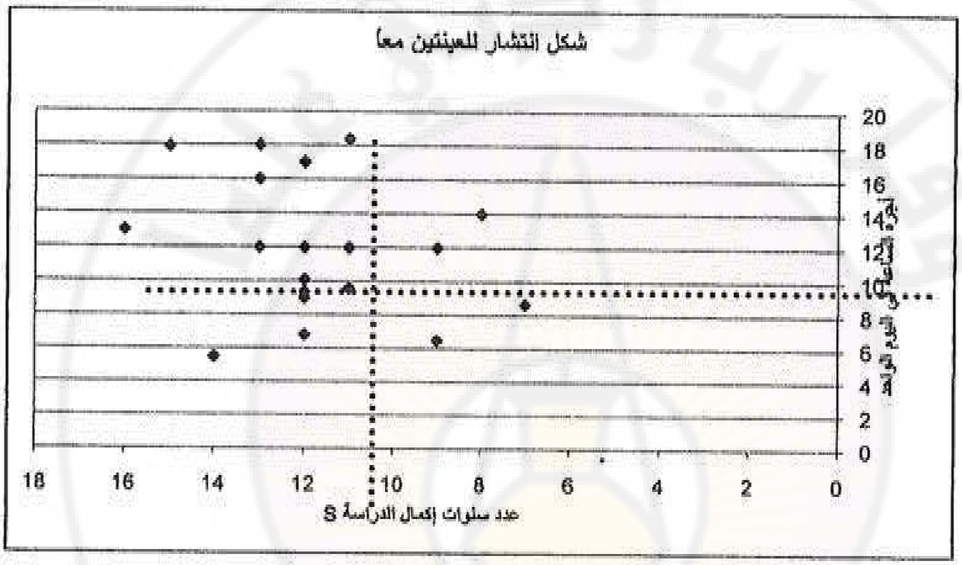
1- نبدأ بحساب الأوساط الحسابية للمتغيرين S و Y انظر الجدول لمشاهدة تفاصيل الحسابات انظر الشكل (3-2).

2- نحسب فروق المشاهدات عن الأوساط الحسابية لكل متغير على حدة.

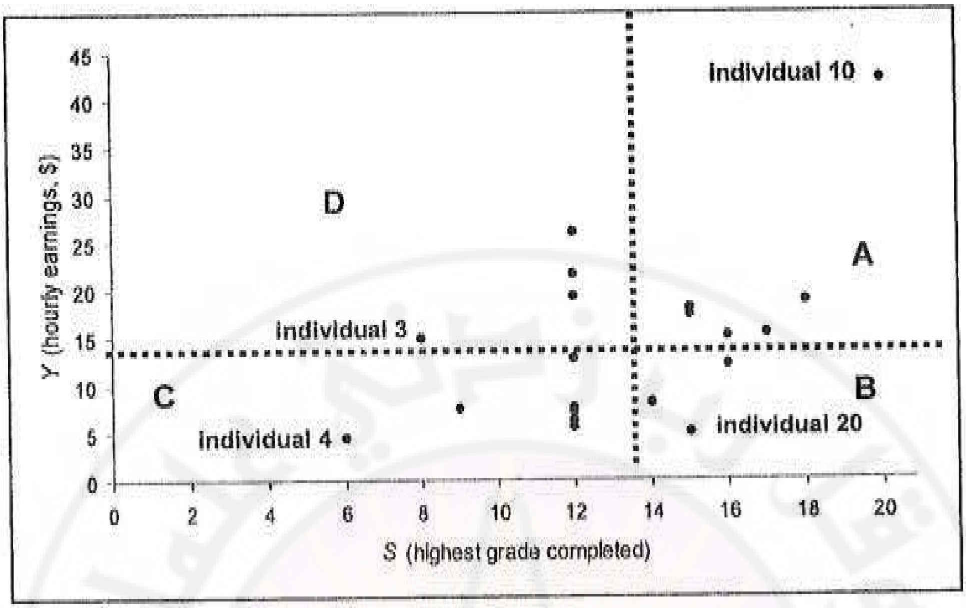
3- نطبق العلاقة فنحصل :

$$\mathrm{Cov}(X,Y)=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})=\frac{40}{20}=2.0$$

الإشارة الموجبة للتباين المشترك تدل على أن الارتباط موجب بين المتغيرين S و Y وهذا بالطبع ما يجب توقعه.



الشكل رقم (3-2)



الشكل رقم (4-2)

من الشكل البياني رقم (4-3) نلاحظ أن النقاط الواقعة في المربع **A** لها انحراف موجب بين المتغيرين وهي تملك مساهمة موجبة في التباين المشترك. ونلاحظ أن الحالة 4 من هذه المشاهدات تقع في هذا القسم. بينما المشاهدة 9 تقع في القسم **B** ، أي وهو الذي أنهى 12 سنة من الدراسة ، ولكن لا يستطيع أن يكسب إلا 9.00 دولارات في ساعة العمل .

أما المشاهدة 18 التي لمن أنهى 9 سنوات في الدراسة، ولكن لا يستطيع ان يحصل إلا على عمل يدوي بالمزرعة بأجر قليل جداً 6.5 قدره دولاراً، ويقع في القسم D ، وتكون مساهمته في التباين موجبة.

في آخر المطاف نلاحظ أن المشاهدات في القسم C هي لمن يملك عدد قليل من سنوات الدراسة، ولكن يعمل في البناء، هو يملك متوسط دخل عالٍ في الساعة كما في الحالة 18 ، وهذا يساهم بقيم سالبة في التباين.

من الشكل البياني (4-2) نلاحظ أن اتجاه المشاهدات في المربعين A و C موجبة ؛ لذلك فهي تساهم بقيم موجبة تفوق القيم السالبة في القسمين B, D.

التتابع في حساب التباين المشترك للعينة في القيم الموجبة بين سنوات الدراسة والكسب المتحقق لساعة العمل الواحدة، بالطبع يجب توقعه أن يكون موجب لان هناك ارتباطاً بين المتغيرين بحكم علاقة السبب و النتيجة.

2-7- قواعد التباين المشترك (Covariance Rules)

سوف نستعرض ثلاثة من قواعد تحليل التباين المشترك، ونحاول إثباتها تبعاً .

- القاعدة الأولى: إذا كان لدينا : $Y = V + W,$

فإن: $\mathrm{Cov}(X, Y) = \mathrm{Cov}(X, V) + \mathrm{Cov}(X, W)$

$$\mathbf{Cov}(X,Y) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$$

$$= \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})([V_i + W_i] - [\bar{V} + \bar{W}])$$

نستعيض عن متوسط القيمة Y بمتوسط القيمتين V و W.

$$\mathbf{Cov}(X,Y) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$$

$$= \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})([V_i + W_i] - [\bar{V} + \bar{W}])$$

$$= \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})([V_i - \bar{V}] + [W_i - \bar{W}])$$

و الآن نستطيع كتابة العلاقة بعد فك الأقواس كما هو مبين أدناه:

$$\mathrm{Cov}(X,Y) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(V_i - \bar{V}) + \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(W_i - \bar{W})$$

هذا يوصلنا إلى النتيجة التي نريدها.

$$\text{Cov}(X,Y) = \text{Cov}(X,V) + \text{Cov}(X,W)$$

- **القاعدة الثانية:**

الحالة التي يوجد في العلاقة قيمة ثابتة، أي شكل العلاقة: $\boldsymbol{Y = bZ}$

فإن: $\text{Cov}(X, Y) = \text{Cov}(X, bZ) = b\text{Cov}(X, Z)$

نتابع تطبيق القاعدة على مشاهدات المتغير ومضاعفتها بالثابت. بعد أن نبدل قيم المتغير Y بما يقابلها من قيم Z ثم نخرج قيم الثابت العامل المشترك فنحصل بالنتيجة على الشكل التالي:

$$\begin{aligned}
\textbf{Cov}(X,Y) &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) \\
&= \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(bZ_i - b\bar{Z}) \\
&= b\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(Z_i - \bar{Z}) \\
&= b\textbf{Cov}(X,Z)
\end{aligned}$$

مثال (2–2): ليكن لدينا العلاقة التالية: $\text{Cov}(X, 3Z) = 3\text{Cov}(X, Z)$

القاعدة الثالثة: قيمة المتغير التي تساوي قيمة ثابتة: $\boldsymbol{Y = b}$

فإن: $\text{Cov}(X, Y) = \text{Cov}(X, b) = 0$

البرهان : نبدل قيم Y المتغير بما يقابلها من b ، مع الأخذ بعين الاعتبار أن متوسط القيمة الثابتة يساويها، وهذه الحالة تساوي الصفر؛ لأن العلاقة مضروبة بصفر، كما هو واضح من الشكل الرياضي التالي:

$$\begin{aligned}\text{Cov}(X,Y)&=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})\\&=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})(b-\bar{b})\\&=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})(b-b)\\&=0\end{aligned}$$

مثال(2-3) : $\text{Cov}(X, 10)=0$

مثال(2-4): على فرض لدينا علاقة من الشكل التالي:

$$Y=b_1+b_2Z$$

فإن الحل يكون : $\text{Cov}(X, Y)=\text{Cov}(X, [b_1+b_2Z])$

و هنا نلاحظ الاستخدام النموذجي للقواعد. حيث نفترض أن Y يرتبط بعلاقة خطية مع المتغير Z ونرغب في تحليل التباين المشترك بين المتغيرين .

أولاً: نستخدم القاعدة الأولى للتباين فنجد:

$$\text{Cov}(X, Y)=\text{Cov}(X, b_1)+\text{Cov}(X, b_2Z)$$

ثم نقوم بتطبيق القاعدة الثالثة نحصل على :

$$=0+\text{Cov}(X, b_2Z)$$

و من ثم نقوم بتطبيق القاعدة الثانية فتصبح العلاقة:

$$\text{Cov}(X, Y)=b_2\text{Cov}(X, Z)$$

العلاقة البديلة للتباين المشترك : يمكن البرهان على أن التباين المشترك للمتحولين (X,Y) يساوي:

$$\text{Cov}(X,Y)=\frac{1}{n}\left[\sum_{i=1}^{n}X_iY_i\right]-\overline{X}\overline{Y}$$

الخطوة الأولى في البرهان نفك ثنائي الحدين في العلاقة الأساسية فنجد أن:

$$\mathbf{Cov}(X,Y)=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})$$
$$=\frac{1}{n}\sum_{i=1}^{n}(X_iY_i-X_i\bar{Y}-\bar{X}Y_i+\bar{X}\bar{Y})$$

والآن ندخل إشارة المجموع فنحصل على :

$$=\frac{1}{n}\left[\sum_{i=1}^{n}X_iY_i-\sum_{i=1}^{n}X_i\bar{Y}-\sum_{i=1}^{n}\bar{X}Y_i+n\bar{X}\bar{Y}\right]$$

سوف نقسم هذه العلاقة إلى أربع حدود، و نخرج قيمة المتوسط $\bar{Y}$ من الحد الثاني كعامل مشترك للعلاقة، و بالمثل للحد الثالث نخرج قيمة المتوسط $\bar{X}$ كعامل مشترك من الحد الثالث. نلاحظ أن ما داخل الأقواس في الحد الثاني هو متوسط قيمة X . وبالمثل للحد الثالث نجد أن قيمة القوس هو متوسط Y. نلاحظ أن الحدين الوسطيين متساويين ومتعاكسين في الإشارة نحذفهما، ونصل إلى العلاقة البديلة، مع الأخذ بعين الاعتبار($1/n$) فقط للحد الأول:

$$\mathrm{Cov}(X,Y)=\frac{1}{n}\left[\sum_{i=1}^{n}X_iY_i\right]-\frac{1}{n}\left[\sum_{i=1}^{n}X_i\bar{Y}\right]-\frac{1}{n}\left[\sum_{i=1}^{n}\bar{X}Y_i\right]+\bar{X}\bar{Y}$$
$$=\frac{1}{n}\left[\sum_{i=1}^{n}X_iY_i\right]-\bar{Y}\left[\frac{1}{n}\sum_{i=1}^{n}X_i\right]-\bar{X}\left[\frac{1}{n}\sum_{i=1}^{n}Y_i\right]+\bar{X}\bar{Y}$$
$$=\frac{1}{n}\left[\sum_{i=1}^{n}X_iY_i\right]-\bar{Y}\bar{X}-\bar{X}\bar{Y}+\bar{X}\bar{Y}$$
$$\mathrm{Cov}(X,Y)==\frac{1}{n}\left[\sum_{i=1}^{n}X_iY_i\right]-\bar{X}\bar{Y}$$

2-8- التباين المشترك للمجتمع POPULATION COVARIANCE

التباين المشترك للمجتمع مكون من متغيرين عشوائيين، هو عبارة عن مضاريب القيمة المتوقعة لانحرافات المتغيرين عن الوسط الحسابي للمجتمع.

$$\textbf{population covariance of } X \textbf{ and } Y = \sigma_{XY} = E[(X-\mu_X)(Y-\mu_Y)]$$

التباين المشترك للعينة يكون عبارة عن التقدير غير المتحيز للتباين المشترك للمجتمع الإحصائي، وهذا يعتمد على تعديل التباين المشترك للمجتمع بضربه بالعلاقة التالية: $(n - 1)/n$

$$E[\text{Cov}(X,Y)] = \frac{n-1}{n}\sigma_{XY}$$

إذاً يمكن الحصول مجازياً على التباين المشترك للعينة من التقدير غير المتحيز للتباين المشترك للمجتمع:

$$\therefore E\left[\frac{n}{n-1}\text{Cov}(X,Y)\right] = \sigma_{XY}$$

إذا كان المتغيران مستقلان فإن التباين المشترك للمجتمع يساوي الصفر. هنا يظهر التباين المشترك للمجتمع مبدئياً كقيمة متوقعة لمضاريب هذه العوامل σ_{XY}. أي :

If X and Y are independent, $\sigma_{XY} = 0$

$$\begin{aligned} E[(X-\mu_X)(Y-\mu_Y)] &= [E(X-\mu_X)][E(Y-\mu_Y)] \\ &= [E(X)-E(\mu_X)][E(Y)-E(\mu_Y)] \\ &= [\mu_X-\mu_X][\mu_Y-\mu_Y] = 0 \times 0 = 0 \end{aligned}$$

حيث: $\mu_X ; \mu_Y$ تمثلان متوسط المجتمع للمتغيرين Y و Xعلى التوالي.

نستطيع عمل ذلك فقط إذا كان المتغيران X و Y مستقلين (انظر السلاسل المستقلة). أما القيم المتوقعة لكلا العوامل تساوي الصفر لأن:

$$E(\mu_X) = \mu_X \text{ و } E(Y) = \mu_Y \text{ و } E(X) = \mu_X$$

$E(\mu_Y) = \mu_Y$ لأنها قيم ثابتة.

2-9- تقدير تباين المجتمع من متوسط العينة
POPULATION VARIANCE OF THE SAMPLE MEAN

يمكننا باستخدام سلسلة الاشتقاقات المتعاقبة أن نحدد تباين المجتمع بالاعتماد على متوسط العينة، ويمكننا صياغة أول قاعدة تباين بهذا الخصوص إذا كان المتغيران X و Y مستقلين فنجد أن:

$$\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y + 2\sigma_{XY} = \sigma^2_X + \sigma^2_Y$$

لأنه إذا كان Y وX متغيران مستقلين فإن التباين المشترك للمجتمع يساوي الصفر كما ذكرنا أعلاه، وكذلك فإن تباين المجتمع للمتغيرين $(X + Y)$ هو عبارة عن مجموع التباينات للمتغيرين Y و X. سوف نطبق هذه النتيجة على متوسط المجتمع، فنحصل على:

$$\sigma^2_{\bar{X}} = \left[\sigma^2_{\frac{1}{n}(X_1+\ldots+X_n)}\right] = \frac{1}{n^2}\left[\sigma^2_{X_1+\ldots+X_n}\right]$$

بتطبيق القاعدة الثانية من التباين:

$$\sigma^2_{\bar{X}} = \left[\sigma^2_{\frac{1}{n}(X_1+\ldots+X_n)}\right] = \frac{1}{n^2}\left[\sigma^2_{X_1+\ldots+X_n}\right]$$
$$= \frac{1}{n^2}\left[\sigma^2_{X_1}\ldots+\sigma^2_{X_n}\right]$$

والآن القاعدة الأولى: سوف نفترض أن المشاهدات مستقلة، و كذلك التباين المشترك للمجتمع يساوي الصفر .

$$\sigma_{\bar{X}}^2 = \left[\sigma^2_{\frac{1}{n}(X_1+\ldots+X_n)}\right] = \frac{1}{n^2}\left[\sigma^2_{X_1+\ldots+X_n}\right]$$

$$= \frac{1}{n^2}\left[\sigma^2_{X_1}\ldots+\sigma^2_{X_n}\right]$$

$$= \frac{1}{n^2}\left[\sigma^2_X\ldots+\sigma^2_X\right] = \frac{1}{n^2}\left[n\sigma^2_X\right] = \frac{\sigma^2_X}{n}$$

$$\sigma_{\bar{X}}^2 = Var(\bar{X}) = Var(\frac{1}{n}\sum X_i)$$

$$= \frac{1}{n^2}Var(\sum X_i) = \frac{1}{n^2}\left[n\sigma^2_X\right] = \frac{\sigma^2_X}{n}$$

2-9- معامل الارتباط CORRELATION COEFFICIENT:

2-9-1- الارتباط البسيط للمتغيرات الكمية:

بعد أن يتأكد الباحث من وجود علاقة سببية بين المتغيرات قيد الدراسة يحدد المتغير المستقل والمتغير التابع، ويقوم بجمع البيانات عن الظاهرتين، حيث من خلال هذه البيانات يستطيع أن يحدد متانة هذه العلاقة وطبيعتها واتجاهها بالطرائق الآتية:

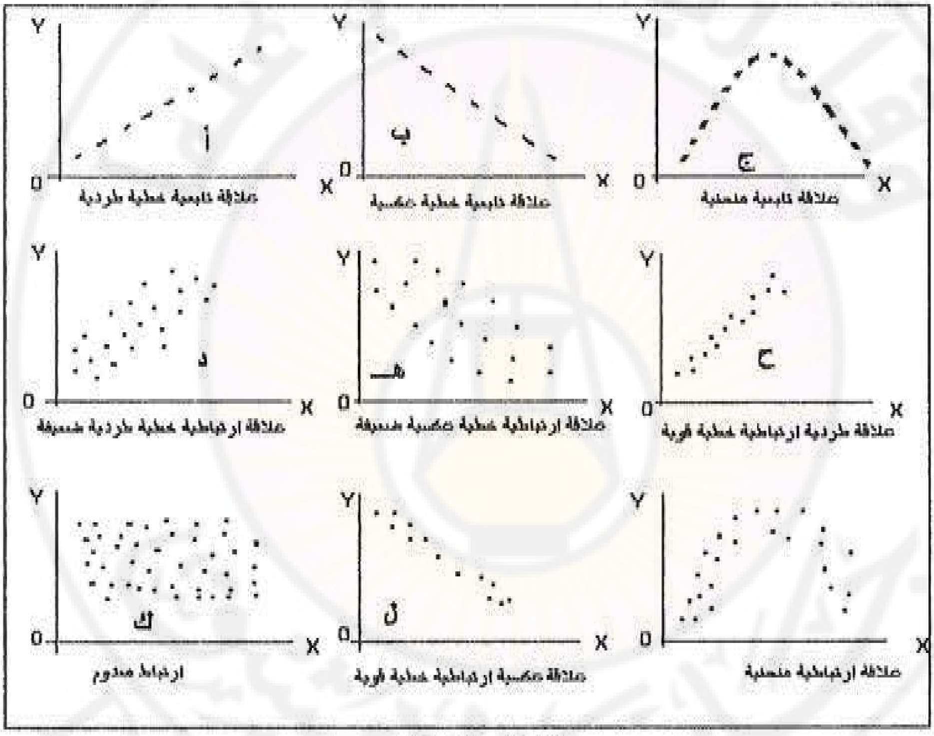
2-9-1-1- التحليل المباشر :

إذا تبين من خلال قراءة بيانات المتغيرين أن القيم الدنيا للمتغير المستقل تتقابل مع القيم الدنيا للمتغير التابع، والعليا مع العليا، فالعلاقة بين المتغيرين هي علاقة طردية، أما إذا كانت القيم الدنيا للمتغير المستقل تتقابل مع القيم العليا للمتغير التابع وبالعكس، فالعلاقة عكسية، أي أن الزيادة في المتحول المستقل يقابلها تناقص بالمتغير التابع، ولمعرفة نوع العلاقة فيما إذا كانت خطية أو منحنية، نقارن الفروق بين كل قيمة للمتغير المستقل والقيمة السابقة لها مع

مثيلاتها للمتغير التابع، فإذا كانت متقاربة فالعلاقة خطية، أما إذا كانت متباعدة فالعلاقة غير خطية .

2-9-1-2- شكل الانتشار (Scatter Diagram)

وهو عبارة عن تمثيل البيانات بشكل أزواج من القيم على المحاور الإحداثية (XOY) بهدف التعرف على وجود علاقة أو عدم وجودها، وإذا وجدت يجب تحديد ما نوعها؟ خطية أو غير خطية.



الشكل (2-5)

إن الانتشار في أحد الأشكال الآتية الواردة في الشكل رقم (2-5)، ومن خلال قراءة هذه الأشكال يمكننا تحديد طبيعة ونوع ومتانة العلاقة بين المتغيرات، فإذا كانت نقاط الانتشار تقع على خط مستقيم واحد تكون العلاقة

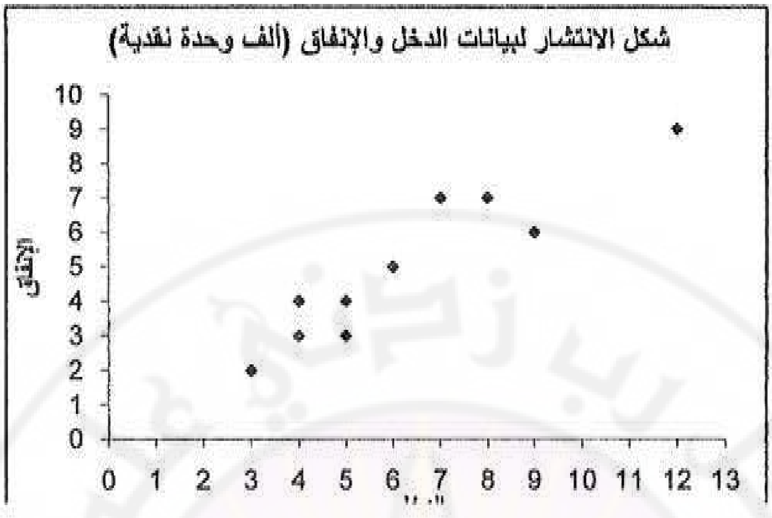
قوية جداً، وبالتالي يكون الارتباط تاماً. ويمكن أن تكـون العلاقـة عكسـية أو طردية حسب موقع المستقيم الأشكال، انظر الشكلين : أ، ب في الشكل (5-2)، وإذا كانت تقع على منحن واحد كانت العلاقة تابعية منحنية الشـكل ج-(5-2). وعندما تكون نقاط الانتشار على شكل حزمة من النقاط ولها اتجاه واحد تكـون العلاقة طردية أو عكسية ، وتكون العلاقة أقوى كلما كـان عـرض الحزمـة صغيراً، لاحظ الأشكال د،هـ،ح ،ل، م في الشكل رقم(5-2) وعنـدما تكـون النقاط مبعثرة دون أن يكون لها اتجاه معين يكـون الارتبـاط معـدوماً بـين المتغيرين، الشكل) ك- (5-2) .

مثال **(5-2)** : الجدول الآتي يمثل بيانات الدخل الشهري، والإنفاق الشهري لعينة مؤلفة من 10 أسر .

جدول رقم (2-2)

الأسرة	1	2	3	4	5	6	7	8	9	10
الدخل (بآلاف الوحدات)	4	5	3	6	7	8	9	4	5	12
الإنفاق (بآلاف الوحدات)	3	4	2	5	7	7	6	4	3	9

إن رسم شكل الانتشار للبيانات السابقة يعطي الشكل الآتي رقم (6-2).



الشكل رقم (6-2)

يمكننا من خلال هذا الشكل ملاحظة طبيعة العلاقة، حيث القيم الدنيا للدخل تترافق مع القيم الدنيا للإنفاق، والعليا مع العليا ، مما يبين أن العلاقة طردية. كما يمكن من الشكل معرفة متانة العلاقة، حيث تكون العلاقة تابعية عندما تقع النقاط على منحن واحد، وتكون ارتباطية وقوية كلما اقتربت النقاط من المنحني، ففي الشكل رقم (6-2) تنتشر النقاط حول مستقيم يمكن تمريره من بين النقاط؛ مما يسمح لنا بالقول: إن العلاقة خطية وقوية.

2-9-3- معامل ارتباط بيرسون (Pearson)

يساعد معامل ارتباط بيرسون (Pearson) على تحديد متانة وطبيعة العلاقة بين المتغيرين X و Y وتعطى قيمة هذا المعامل بالعلاقة :

$$r_{xy} = \frac{\Sigma (X - \bar{X})\ (Y - \bar{Y})}{n\ S_x\ S_y} \qquad (1)$$

حيث أن : $\overline{X}$و $\overline{Y}$ الوسطان الحسابيان للمتغير المستقل X والمتغير التابع Yعلى التوالي . S_x و S_y الانحرافات المعيارية للمتغير المستقل والمتغير التابع. و n عدد المشاهدات .

ويمكن إصلاح العلاقة السابقة وكتابتها بأشكال أخرى أكثر سهولة لحساب معامل الارتباط :

وإذا فرضنا أن: $y = \left(Y_i - \bar{Y}\right) \quad , x = \left(X_i - \bar{X}\right)$

فإن علاقة بيرسون تصبح على الشكل التالي:

$$(2)\, r = \frac{\sum x\, y}{n S_X S_Y}$$

ويمكننا ان نكتب العلاقة بشكلها المفصل على النحو الآتي:

$$r_{xy} = \frac{n(\Sigma XY) - (\Sigma X)(\Sigma Y)}{\sqrt{\left[n\,\Sigma X^2 - (\Sigma X)^2\right]\left[n\,\Sigma Y^2 - (\Sigma Y)^2\right]}} \qquad (3)$$

وبتقسيم البسط والمقام على n نحصل على العلاقة التالية:

$$r_{xy} = \frac{\Sigma XY - n\,\overline{X}\,\overline{Y}}{\sqrt{\left[\Sigma X^2 - n\,\overline{X}^2\right]\left[\Sigma Y^2 - n\,\overline{Y}^2\right]}} \qquad (4)$$

مثال(6–2): سوف نوضح فيما يأتي كيفية حساب معامل الارتباط بالعلاقات الثلاث السابقة من خلال بيانات إلى المثال رقم(5–2) السابق .

جدول رقم (3-2)

الأسرة	الدخل (الف) X	الإنفاق (الف) Y	XY	X^2	Y^2	x	y	Σxy
1	4	3	12	16	9	- 2.3	- 2	4.6
2	5	4	20	25	16	- 1.3	- 1	1.3
3	3	2	6	9	4	- 3.3	- 3	9.9
4	6	5	30	36	25	- 0.3	0	0
5	7	7	49	49	49	0.7	2	1.4
6	8	7	56	64	49	1.7	2	3.4
7	9	6	54	81	36	2.7	1	2.7
8	4	4	16	16	16	- 2.3	- 1	2.3
9	5	3	15	25	9	- 1.3	- 2	2.6
10	12	9	108	144	81	5.7	4	22.8
المجموع	63	50	366	465	294	0	0	51

الحل حسب الصيغة رقم (1) أو (2).

-نحسب المتوسطين $\overline{X}$و $\overline{Y}$ فنجد ان: $\overline{X} = 6.3$و $\overline{Y} = 5$

- نحسب الانحراف المعياري S_xلـــ X وكذلك S_y لـــ Y من العلاقتين:

$$S_x = \sqrt{\frac{\Sigma X^2}{n} - \overline{X}^2}$$

$$S_x = \sqrt{\frac{465}{10} - (6.3)^2} = 2.6096$$

$$S_y = \sqrt{\frac{\Sigma Y^2}{n} - \overline{Y}^2}$$

$$S_y = \sqrt{\frac{294}{10} - (5)^2} = 2.0976$$

ومنــه :

$$r_{xy} = \frac{\Sigma (X - \overline{X})\ (Y - \overline{Y})}{n\ S_x\ S_y} = \frac{\sum (x * y)}{n\ S_x\ S_y}$$

$$r_{xy} = \frac{51}{(10)\,(2.6096)\,(2.0976)} = +\ 0.9317$$

الحل حسب الصيغة رقم (3) :

$$r_{xy} = \frac{n\,\Sigma\,XY - \Sigma X\,\Sigma Y}{\sqrt{\left[n\,\Sigma X^2 - (\Sigma X)^2\right]\left[n\,\Sigma Y^2 - (\Sigma Y)^2\right]}}$$

$$= \frac{10\,(366) - (63)\,(50)}{\sqrt{\left[10\,(465) - (63)^2\right]\left[10\,(294) - (50)^2\right]}} = +\ 0.9317$$

الحل حسب الصيغة رقم (4) :

$$r_{xy} = \frac{\Sigma\,XY - n\,\overline{X}\overline{Y}}{\sqrt{\left[\Sigma X^2 - n\,\overline{X}^2\right]\left[\Sigma Y^2 - n\,\overline{Y}^2\right]}}$$

$$= \frac{366 - 10\ (6.3)\,(5)}{\sqrt{\left[465 - 10\,(6.3)^2\right]\left[294 - 10\,(5)^2\right]}} = +\ 0.9317$$

مثال(7-2): تبين لأحد الباحثين أن هناك علاقة بين حجم الأسرة X وحجم إنفاقها Y . وكانت البيانات على الشكل التالي:

جدول (4-2)

14	11	9	8	5	4	3	2	X_i
9	8	7	5	4	3	3	1	Y_i

والمطلوب: حساب شدة الارتباط وبيان نوعه بين المتغيرين X و Y .

الحل :

أولاً ، نحسب الوسطين الحسابين فنجد أنها يساويان:

$$\bar{X} = \frac{56}{8} = 7 \qquad \text{أما} \qquad \bar{Y} = \frac{40}{8} = 5$$

ثانياً: نكون الجدول التالي:

الجدول (5-2)

Y	X	$y=(Y_i-\bar{Y})$	$x=(X_i-\bar{X})$	$\sum y^2$	$\sum x^2$	$\sum xy$
1	2	-4	-5	16	25	20
3	3	-2	-4	4	16	8
3	4	-2	-3	4	9	6
4	5	-1	-2	1	4	2
5	8	0	1	0	1	0
7	9	2	2	4	4	4
8	11	3	4	9	16	12
9	14	4	7	16	49	28
40	56	0	0	54	124	80

ثالثاً : نحسب الانحراف المعياري للمتغيرين X , Y فنجد أن:

$$S_X = \sqrt{\frac{\sum x^2}{n}} = \sqrt{\frac{124}{8}} = \sqrt{15.5} = 3.937$$

$$S_Y = \sqrt{\frac{\sum y^2}{n}} = \sqrt{\frac{54}{8}} = \sqrt{6.75} = 2.598$$

رابعاً : نطبق علاقة بيرسون :

$$r = \frac{\sum x\,y}{N * S_X S_Y} = \frac{80}{8*(3.937)*(2.598)} = 0.9777$$

التفسير : إن الارتباط بين عدد الأفراد في الأسرة وحجم الإنفاق قوي جداً، حيث إن قيمة المعامل قريبة من الواحد الصحيح، وإن العلاقة بينهما طردية لأن إشارة معامل الارتباط موجبة.

حساب معامل الارتباط لبيانات غير مبوبة بأسلوب الانحرافات حسب علاقة بيرسون:

في كثير من الأحيان يكون من الأفضل اختيار وسط حسابي فرضي ذي قيمة صحيحة بدلاً من الوسط الحسابي؛ الذي يحتوي على كسور عشرية مما يعقد العمليات الحسابية أثناء استخدام الطرق اليدوية. وبالطريقة بنفسها التي اتبعت في حسابات الوسط الحسابي بطريقة الانحرافات، وأيضاً الانحراف المعياري بطريقة الانحرافات لبيانات غير مبوبة، يمكننا كتابة علاقة بيرسون بأسلوب الانحرافات على الشكل التالي:

$$r = \frac{\sum dx\, dy - \frac{\sum dx \sum dy}{n}}{n\, S_x\, S_y}$$

حيث إن: $dx = (X_i - A)$ و A - هي الوسط الحسابي الفرضي للمتغير X

$dy = (Y_i - B)$ و B - هي الوسط الحسابي الفرضي للمتغير Y

أما الانحراف المعياري بطريقة الانحرافات وقد حسبناه في مقاييس التشتت:

$$S_x = \sqrt{\frac{\sum d_x^2}{n} - \left(\frac{\sum dx}{n}\right)^2} \quad ; \quad S_y = \sqrt{\frac{\sum d_Y^2}{n} - \left(\frac{\sum dy}{n}\right)^2}$$

مثال (8-2): إذا عالجنا المثال باستخدام طريقة الانحرافات عن وسط حسابي فرضي، على افترض أن: الوسط الفرضي لـ X هو $A = 8$ أما الوسط الحسابي الفرضي لـ Y فهو $B = 4$.

نحسب الانحراف المعياري لـ X و Y:

$$S_y = \sqrt{\frac{62}{8} - \left(\frac{8}{8}\right)^2} = 2.598 \quad ; \quad S_x = \sqrt{\frac{132}{8} - \left(\frac{-8}{8}\right)^2} = 3.937$$

جدول (2-6)

n	Y	X	dy	dx	dy^2	dx^2	dx*dy
1	1	2	-3	-6	9	36	18
2	3	3	-1	-5	1	25	5
3	3	4	-1	-4	1	16	4
4	4	5	0	-3	0	9	0
5	5	8	1	0	1	0	0
6	7	9	3	1	9	1	3
7	8	11	4	3	16	9	12
8	9	14	5	6	25	36	30
المجموع	40	56	+8	-8	62	132	72

وبالتعويض في علاقة بيرسون بأسلوب الانحرافات نجد :

$$r = \frac{72 - \dfrac{(8)*(-8)}{8}}{8*(3.937)*(2.598)} = 0.9774$$

حساب معامل الارتباط حسب علاقة بيرسون للبيانات غير المبوبة بالأسلوب المباشرة ، وذلك باستخدام العلاقة التالية:

$$r = \frac{\sum XY - n\,\overline{XY}}{n\,S_x\,S_y}$$

وفي الفصل السابق بين كيف حساب الانحراف المعياري بالأسلوب المباشر من خلال العلاقة:

$$S_x = \sqrt{\frac{\sum X^2}{n} - \left(\frac{\sum X}{n}\right)^2} = \sqrt{\frac{516}{8} - \left(\frac{56}{8}\right)^2} = 3.937$$

$$S_y = \sqrt{\frac{\sum Y^2}{n} - \left(\frac{\sum Y}{n}\right)^2} = \sqrt{\frac{254}{8} - \left(\frac{40}{8}\right)^2} = 2.598$$

نكون جدول لحساب المجاميع : $\sum X^2$; $\sum Y^2$; $\sum XY$

نحسب الأوساط الحسابية: $\bar{Y} = \frac{40}{8} = 5 \quad ; \quad \bar{X} = \frac{56}{8} = 7$

جدول (7-2)

المتغيرات	1	2	3	4	5	6	7	8	المجاميع
X_i	2	3	4	5	8	9	11	14	56
Y_i	1	3	3	4	5	7	8	9	40
XY	2	9	12	20	40	63	88	126	360
X^2	4	9	16	25	64	81	121	196	516
Y^2	1	9	9	16	25	49	64	81	254

نعوض في علاقة بيرسون حسب الأسلوب المباشر فنجد:

$$r = \frac{360 - 8 * 7 * 5}{8 * (3.937) * (2.598)} = 0.9774$$

نلاحظ أنه قد حصلنا على النتيجة نفسها التي حصلنا عليها سابقاً؛ لذلك يتوجب علينا اختيار الطريقة الأنسب بالنسبة للبيانات المدروسة.

خصائص معامل ارتباط بيرسون :

1- يأخذ معامل ارتباط r قيمة تتراوح بين -1 إلى +1 ويمكن إثبات ذلك رياضياً.

2- تكون قيمة r تساوي الواحد عندما تكون العلاقة تامة، وتبلغ قيمته الصفر عندما يكون الارتباط معدوماً، وتقترب من الواحد عندما تكون العلاقة قوية جداً.

3- يأخذ معامل الارتباط قيماً موجبة عندما تكون العلاقة طردية، وسالبة عندما تكون عكسية.

4- قيمة معامل الارتباط مجردة من وحدات القياس، وهذه من أهم مزاياه حيث يتم التخلص من وحدات القياس عند التقسيم على الانحرافات المعيارية (لاحظ الصيغة 1).

في حين أن أهم مساوئ معامل ارتباط بيرسون أنه لا يعبر عن متانة العلاقة بشكل صحيح إلا إذا كانت العلاقة خطية، أما إذا كانت العلاقة منحنية، فإن قيمته لا تعبر عن متانة العلاقة، وفي هذه الحالة يجب استخدام مقاييس أخرى، كما أنه لا يمكن استخدام معامل بيرسون في حالة المتغيرات النوعية.

2-9-4- حساب معامل الارتباط بالاعتماد على التباين المشترك:

يمكننا حساب معامل الارتباط بالاعتماد على التباين المشترك ، نرمز لمعامل ارتباط المجتمع ρ_{XY}، من أجل متغيرين X و Y هو يعطى بالشكل التالي:

$$\rho_{XY} = \frac{\sigma_{XY}}{\sqrt{\sigma_X^2\ \sigma_Y^2}}$$

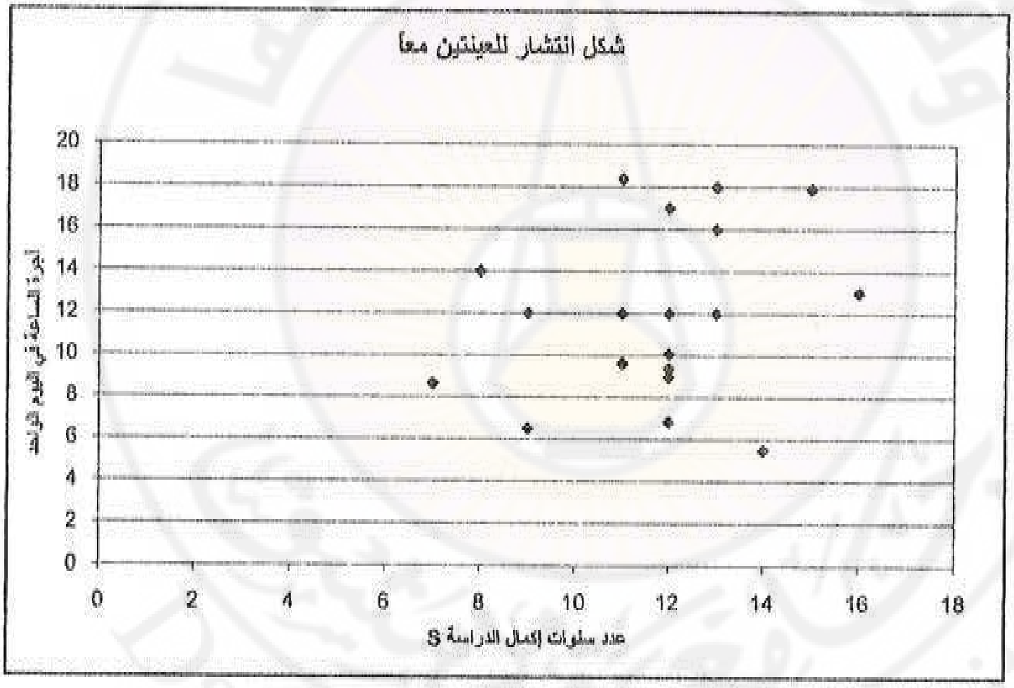
يكون معامل الارتباط مساوياً الواحد إذا كانت العلاقة تامة طردية، ويساوي الناقص واحد إذا كانت العلاقة تامة وعكسية . ويساوي الصفر في حال عدم وجود علاقة بين المتغيرين.

ولنفرض أن مشاهدات العينة تستخدم لتقدير معامل الارتباط ρ_{XY} من خلال التباين المشترك للمجتمع، والتباين غير المتحيز وذلك من خلال العلاقة الآتية:

$$r_{XY} = \frac{\frac{n}{n-1}\mathbf{Cov}(X,Y)}{\sqrt{\frac{n}{n-1}\mathbf{Var}(X)\ \frac{n}{n-1}\mathbf{Var}(Y)}} = \frac{\mathbf{Cov}(X,Y)}{\sqrt{\mathbf{Var}(X)\ \mathbf{Var}(Y)}}$$

نلاحظ أن القيمة $\frac{n}{n-1}$ حذفت من العلاقة لأنها اختصرت مع مربع القيمة تحت الجذر. كما يلاحظ أنه حتى إذا كان لا يوجد علاقة حقيقية بين المتغيرين Y و X فان r_{XY} بشكل عام لن تساوي الصفر. هذه العلاقة العشوائية سيكون لها توزيع قريب من الصفر.

لنعدْ إلى مثالنا السابق لحساب معامل الارتباط بين عدد ساعات التدريب وأجور الساعة، وباستخدام حسابات التباين المشترك للعينة؛ نجد أن شكل الانتشار له التوزيع التالي:



الشكل رقم (7-2)

لنعد جدولاً بأهم الحسابات المطلوبة لتطبيق العلاقة السابقة لحساب معامل الارتباط:

جدول (8–2)

$(Y-\overline{Y})(X-\overline{X})$	$(X-\overline{X})^2$	$(Y-\overline{Y})^2$	$X-\overline{X}$	$Y-\overline{Y}$	X	Y	N
-5.104	0.6	40.7	-0.8	6.4	11	18.38	1
-0.384	0.0	3.7	0.2	-1.9	12	10.08	2
1	0.0	25.0	0.2	5.0	12	17	3
18.944	10.2	35.0	3.2	5.9	15	17.92	4
0	1.4	0.0	1.2	0.0	13	12	5
0	7.8	0.0	-2.8	0.0	9	12	6
7.2	1.4	36.0	1.2	6.0	13	18	7
-1.034	0.0	26.7	0.2	-5.2	12	6.83	8
-0.6	0.0	9.0	0.2	-3.0	12	9	9
0	0.6	0.0	-0.8	0.0	11	12	10
0	0.0	0.0	0.2	0.0	12	12	11
4.2	17.6	1.0	4.2	1.0	16	13	12
4.8	1.4	16.0	1.2	4.0	13	16	13
-0.524	0.0	6.9	0.2	-2.6	12	9.38	14
1.912	0.6	5.7	-0.8	-2.4	11	9.61	15
15.4	7.8	30.3	-2.8	-5.5	9	6.5	16
-14.3	4.8	42.3	2.2	-6.5	14	5.5	17
[illegible]	14.4	4.0	-3.8	2.0	8	14	18
16.08	23.0	11.2	-4.8	-3.4	7	8.65	19
0	1.4	0.0	1.2	0.0	13	12	20
39.99	93.8	293.5	-1	-0.1	235	239.9	المجموع
1.9995					11.8	12.0	المتوسط

وبالتطبيق نجد:

$$r_{SY} = \frac{\mathrm{Cov}(S,Y)}{\sqrt{\mathrm{Var}(S)\mathrm{Var}(Y)}} = \frac{2}{\sqrt{293.5 \times 93.8}} = \frac{2}{165.9} = 0.012$$

مثال(9–2): على فرض أن المتغير Y يتحدد بدلالة X بالعلاقة الخداية التالية:

$$Y = b_1 + b_2X$$

وعلى فرض أن عينة المشاهدات للمتغيرين X,Y و تتضمن متغيراً ثالثاً Z. والبرهان على أن معامل الارتباط للعينة بين المتغيرين Y,Z يساوي المعامل بين X,Z ، إذا كانت قيمة b_2 موجبة.

والمطلوب: أثبت أن: $r_{Y,Z} = r_{X,Z}$

معامل الارتباط بين Z و Y محدد بالعلاقة التالية:

$$r_{Y,Z} = \frac{\mathbf{Cov}(Y,Z)}{\sqrt{\mathbf{Var}(Y)\ \mathbf{Var}(Z)}}$$

أولاً: نبدل قيمة Y بما تساويه في حد التباين المشترك

$$\mathbf{Cov}(Y,Z) = \mathbf{Cov}([b_1 + b_2X], Z)$$

نستخدم القاعدة (1) من التباين المشترك، نستطيع تحليل مركبات التباين المشترك إلى حدين.

نستخدم القاعدة (3) من التباين المشترك، فنجد أن أول حد يساوي الصفر لأن b_1 قيمة ثابتة. نستخدم القاعدة (2) من التباين المشترك ونخرج قيمة b_2خارج القوس؛ لأنها عامل مشترك ثابت، فنحصل على:

$$\begin{aligned}\mathbf{Cov}(Y,Z) &= \mathbf{Cov}([b_1 + b_2X], Z)\\ &= \mathbf{Cov}(b_1, Z) + \mathbf{Cov}(b_2X, Z)\\ &= 0 + b_2\mathbf{Cov}(X,Z)\end{aligned}$$

نبدل قيمة Y بما يساويها في علاقة التباين $\mathrm{Var}(Y)$فنجد:

$$\mathrm{Var}(Y) = \mathrm{Var}(b_1 + b_2X) = \mathrm{Var}(b_2X)) = b_2^2 Var(X)$$

لأنه عندما نستخدم القاعدة (4) للتباين ، نحذف b_1 لأنها قيمة ثابتة، و نستخدم القاعدة (2) للتباين ، نخرج b_2 خارج القوس التباين ، ولكنها قيمة مربعة . نعلم ان $\mathrm{Var}(b_2X)$ نفس التباين المشترك $\mathrm{Cov}(b_2X, b_2X)$، وكذلك b_2 عامل مشترك ثابت لذلك نضاعفه، فنحصل:

$$\mathbf{Var}(Y) = \mathbf{Var}(b_1 + b_2X) = \mathbf{Var}(b_2X) = b_2^2\mathbf{Var}(X)$$

الآن نقوم بتبديل $\text{Cov}(Y, Z)$ و $\text{Var}(Y)$ في علاقة معامل الارتباط فنجد:

$$r_{Y,Z} = \frac{\mathbf{Cov}(Y,Z)}{\sqrt{\mathbf{Var}(Y)\,\mathbf{Var}(Z)}} = \frac{b_2\mathbf{Cov}(X,Z)}{\sqrt{b_2^2\mathbf{Var}(X)\,\mathbf{Var}(Z)}} = r_{X,Z}$$

نختصر b_2 من البسط مع القيمة نفسها b_2 من المقام . وهكذا بكل وضوح يظهر أن معامل الارتباط بين Y و Z هو نفسه بين X و Z.

مثال(10--2): على سبيل المثال، نفرض أن C متغير درجة الحرارة في مكان العمل مقاسة بالدرجة المئوية، F عبارة عن درجة الحرارة مقاسة بالفرنهايت ، I متغير مستوى النشاط الثقافي فنجد:

$$r_{I,C} = \frac{\mathbf{Cov}(I,C)}{\sqrt{\mathbf{Var}(I)\,\mathbf{Var}(C)}} = \frac{\frac{5}{9}\mathbf{Cov}(I,F)}{\sqrt{\mathbf{Var}(I)\left(\frac{5}{9}\right)^2\mathbf{Var}(F)}} = r_{I,F}$$

حيث إن : $C = \frac{5}{9}(F-32)$

I- المستوى الثقافي و C- درجة الحرارة مقاسة بالدرجة المئوية.

ولحساب معامل الارتباط بين I و C و بين I و F، نبدأ بتبديل قيمة C في التباين المشترك للعلاقة :

$$\begin{aligned}\mathbf{Cov}(I,C) &= \mathbf{Cov}\left(I, \frac{5}{9}[F-32]\right) \\ &= \mathbf{Cov}\left(I, \frac{5}{9}F\right) + \mathbf{Cov}\left(I, -\frac{5}{9}\times 32\right) = \frac{5}{9}\mathbf{Cov}(I,F)\end{aligned}$$

جدول (2-11)

	حصن ضد الشلل	لم يحصن ضد الشلل	المجموع
أصيب بالشلل	39	65	104
لم يصب	159	26	185
المجموع	198	91	289

والمطلوب : احسب معامل الاقتران، وفسر النتيجة .

الحل: نطبق علاقة معامل الاقتران فنجد:

$$r_c = \frac{(39 * 26) - (65 * 159)}{(39 * 26) + (65 * 159)} = \frac{-9321}{11349} = -0.821$$

التفسير: يوجد علاقة ما بين الإصابة بالمرض ومن حصن ضد هذا المرض، وهي علاقة قوية، وهذا يشير إلى ضرورة تحصين الأطفال ضد الشلل.

2-9-7- معامل التوافق:

رأينا أن معامل الاقتران يستخدم لقياس متانة العلاقة بين متغيرين نوعيين، لكل منهما صفتان (أي الجدول يحوي أربع خلايا)، في حين لا يمكن استخدامه إذا كان لأحد المتغيرين على الأقل أكثر من صفتين، حيث يصبح الجدول في هذه الحالة مركباً في أكثر من أربع خلايا، والمعامل الذي يمكن استخدامه هو معامل التوافق، وصيغته كما يأتي:

$$r_A = \sqrt{\frac{G-1}{G}}$$

حيث :

$$G = \sum \left(\frac{F_{ij}^2}{F_i * F_j} \right)$$

F_{ij} : عدد التكرارات في الخلية الواقعة في السطر i والعمود j .

F_i : مجموع السطر i .

F_j : مجموع العمود j .

مثال (13-2) : بهدف دراسة تأثير الحالة التعليمية للأم على مستوى الابن في المدرسة الابتدائية تم جمع البيانات الآتية عن 200 تلميذ في إحدى المدارس.

جدول رقم **(12-2)**

الحالة التعليمية للأم / مستوى الابن في المدرسة	امية وابتدائية	ثانوية وإعدادية	فوق الثانوية	المجموع
متفوق	5	15	30	50
وسط	5	40	15	60
ضعيف	70	15	5	90
المجموع	80	70	50	100

$$G = \frac{5^2}{80 \times 50} + \frac{5^2}{80 \times 60} + \frac{70^2}{80 \times 90} + \frac{15^2}{70 \times 50} + \frac{40^2}{70 \times 60} +$$

$$+ \frac{15^2}{70 \times 90} + \frac{30^2}{50 \times 50} + \frac{15^2}{60 \times 50} + \frac{5^2}{50 \times 90} = 1.613$$

$$r_A = \sqrt{\frac{G-1}{G}} = \sqrt{\frac{1.613-1}{1.613}} = 0.616$$

وهذا يعني أن هناك علاقة قوية بين حالة الأم التعليمية ومستوى الابن في الصف. وبشكل عام نقول: إن هناك علاقة قوية أو جيدة بين المتغيرين، كلما كانت قيمة معامل التوافق أكبر من 0.5 .

مثال (2-14): احسب علاقة الارتباط بين تقدير الطالب في مقرر الإحصاء وفي مقرر اللغة العربية، وذلك وفق تكرارات الجدول التالي:

جدول **(13-2)**

درجات الطلاب في مقرر اللغة العربية	درجات الطلاب في مقرر الإحصاء				
	100-90	89-80	79-65	64-50	المجموع
50 وأقل من 65	21	35	27	15	98
65 وأقل من80	75	123	155	68	421
80 وأقل من 90	32	45	75	41	193
المجموع	128	203	257	124	712

الحل :

$$G = \frac{(15)^2}{124 * 98} + \frac{(68)^2}{124 * 421} + \frac{(41)^2}{124 * 193} + \frac{(27)^2}{257 * 98} + \frac{(155)^2}{257 * 421} +$$

$$+ \frac{(75)^2}{257 * 193} + \frac{(35)^2}{203 * 98} + \frac{(123)^2}{203 * 421} + \frac{(45)^2}{203 * 193} + \frac{(21)^2}{128 * 98} +$$

$$+ \frac{(75)^2}{128 * 421} + \frac{(32)^2}{128 * 193} = 1.013$$

$$r_p = \sqrt{\frac{G - 1}{G}} = \sqrt{\frac{1.013 - 1}{1.013}} = \sqrt{0.01284} = 0.1133$$

التفسير : إن علاقة الارتباط بين مقرري الإحصاء و اللغة العربية ضعيفة جداً في تقديرات الطلاب.

2-9-8- معامل ارتباط الرتب (Coefficient of Rank Correlation)

يستخدم معامل ارتباط الرتب أو معامل ارتباط سبيرمان (Sperman) نسبة إلى واضعه، لقياس متانة العلاقة بين متغيرين نوعيين، لكل منهما عدد كبير من الصفات، أو إذا كان أحد المتغيرين نوعياً والآخر كمياً. وكذلك الأمر يمكن استخدامه في حالة المتغيرات الكمية أحياناً لاختصار الحسابات .

يتم حساب ارتباط الرتب باستبدال الصفات النوعية المرتبة بقيم عددية تعبر عنها وتسمى الرتب Rank، حيث يتم وضعها عن طريق إيجاد العلاقة

الداخلية للصفات ولكل متغير على حدة، فنبدأ بأضعف الصفات مثلاً ونعطيها الرتبة 1 والأقوى منها رقم 2 وهكذا، بحيث يتولد لدينا متوالية عددية لكل متغير يكون مجموعها متساوياً، لأن عدد الصفات لكل متغير متساوية، وبعد ذلك نحسب معامل ارتباط الرتب بالعلاقة:

$$r_s = 1 - \frac{6 \Sigma D^2}{n(n^2 - 1)}$$

حيث D : فرق رتبة X من رتبة Y لكل ثنائية .

n : عدد المشاهدات .

مثال (15-2) : ليكن لدينا الجدول الآتي الذي يمثل البيانات الخاصة بشكل علبة الحلوى والطلب على هذا النوع من الحلوى .

جدول رقم (14-2)

شكل علبة الحلوى (X)	*نوع الطلب* (*Y*)	رتب X	رتب Y	D	D^2
علبة بسيطة دون رسوم	ضعيف	1	1	0	0
علبة بسيطة برسوم	وسط	2	2	0	0
علبة بسيطة برسوم وغلاف	جيد جداً	3	4	1	1
علبة تقليدية	جيد	4	3	1	1
علبة مزخرفة	عالٍ جداً	5	5	0	0
المجموع					2

$$r_s = 1 - \frac{6 \Sigma D^2}{n(n^2 - 1)}$$

$$= 1 - \frac{6 \times 2}{5(5^2 - 1)} = 1 - \frac{12}{120} = 0.9$$

أي أن هناك علاقة قوية جداً بين شكل العلبة ونوع الطلب عليها. وتجب الإشارة هنا إلى أن معامل ارتباط الرتب يتمتع بالمزايا نفسها التي يتمتع بها معامل ارتباط بيرسون، باعتبار أنه مستنتج من معامل ارتباط بيرسون بعد تعويض$\sum X$و$\sum Y$ و $\sum X^2$ و $\sum Y^2$ بما يقابلها باعتبارها متواليات عددية بعد إعطاء الرتب للمتغيرين .

2-9-9- معامل ارتباط الرتب في حالة تكرار عدد من الصفات:

مثال (2-16) : تبين البيانات الآتية غزارة الأمطار ولون الغيم في إحدى المدن في فصل الشتاء.

جدول رقم (2-15)

لون الغيم (X)	*كمية المطر مليمتر/الدقيقة (Y)*	رتب X	رتب Y	D	D^2
ابيض متفرق	0	1	1	0	0
ابيض كثيف	2	2	2	0	0
رمادي	3	3	3.5	0.5 -	0.25
رمادي غامق	3	4	3.5	0.5	0.25
أسود	4	5	5	0	0
المجموع		15	15		0.5

نلاحظ من هذه البيانات أن أحد المتغيرين نوعي، وهو لون الغيوم والآخر كمي.

ولقد تم وضع رتب X مبتدئين بالأضعف حسب معرفتنا بحالات هطول المطر، والأضعف في هذه الحالة إذا كان لون الغيم أبيض متفرقاً، وأعطيناه الرتبة 1، والأفضل هو أبيض كثيف 2 وهكذا... أما المتغير الأخر فهو كمية المطر وبياناته كمية، وهنا لن تكون هناك صعوبة في وضع الرتب، حيث أقل

هذه البيانات سوف يأخذ الرتبة 1 وهكذا، ولكن هناك مشكلة تكرار إحدى صفات الرقم 3، في هذه الحالة تعطى الرتب كما يأتي : حسب الترتيب أول صفة يجب أن تعطى الرتبة 3 في البيانات، والصفة المكررة الأخرى يجب أن تعطى الرتبة 4 ، ولكن بما أن الصفتين متساويتان فيجب أن تساوي بين الرتبتين، ويتم ذلك من خلال إيجاد الوسط الحسابي للرتبتين 3 و 4 وهو 3.5، فيعطى كل منها الرتبة 3.5. بهذه الطريقة نكون قد حافظنا على مجموع الرتب لكل من المتغيرين بحيث يكون متساوياً .

مثال (17-2): لدينا درجات ستة طلاب في مقرري الإحصاء والرياضيات، والمطلوب: احسب معامل ارتباط الرتب بين درجات المقررين.

جدول (16-2)

مقرر الإحصاء X	50	70	90	100	80	60
مقرر الرياضيات Y	15	45	30	90	75	60

الحل: لنستخدم الترتيب التصاعدي حيث نأخذ أصغر درجة ونعطيها الرتبة (1) ثم الدرجة التي تليها الرتبة (2) وهكذا لكلا المتغيرين، ومن ثم نحسب الفروق ما بين الرتب كما هو واضح في الجدول (17-2).

جدول (17-2)

التسلسل	X_I	Y_I	d_x	d_y	$D=d_x-d_y$	D^2
1	50	15	1	1	0	0
2	70	45	3	3	0	0
3	90	30	5	2	3	9
4	100	90	6	6	0	0
5	80	75	4	5	-1	1
6	60	60	2	4	-2	4
	450	315			0	14

نطبق علاقة سبيرمان: $R = 1 - \frac{6 * 14}{6 * (36 - 1)} = 1 - 0.4 = 0.6$

التفسير: إن شدة العلاقة بين درجات الطلاب في مقرر الإحصاء والرياضيات قوية نسبياً حيث بلغت 0.60 وهي علاقة موجبة (أي طردية).

يمتاز معامل ارتباط الرتب بالخصائص التالية :

1- يعطي قيمة تقريبية مختلفة عن معامل بيرسون؛ وذلك لأنه لا يأخذ قيم المشاهدات بعين الاعتبار، وإنما يعتمد على رتبها.

2- يمكننا معامل ارتباط الرتب من دراسة العلاقة ما بين بيانات نوعية سواء كان ذلك بين الظاهرتين أو أحدهما نوعية والأخرى كمية، وهذا غير ممكن باستخدام علاقة بيرسون.

مثال (18-2): احسب شدة الارتباط بين درجات الطلبة في مقرر الرياضيات ودرجة ذكائهم؛ علماً بأن العينة مؤلفة من 8 طلاب.

جدول **(18-2)**

درجة الرياضيات X	60	63	60	39	35	60	59	51
مستوى درجة الذكاء Y	عالٍ	عالٍ	متوسط	ضعيف	ضعيف جداً	عالٍ	فوق الوسط	تحت الوسط

الحل :

نرتب القيم ترتيباً تصاعدياً أو تنازلياً على السواء، لكن قد يصادفنا تكرار نفس القيمة أو التقدير عدة مرات؛ ولذلك من أجل تحديد الترتيب نأخذ المتوسط الحسابي للرتب المفروض أن تعطى للقيم المتكررة، فمثلاً نجد أن الدرجة (60) مكرره ثلاث مرات وتحتل الترتيب (5 و6 و7) نجمع الرتب الثلاث ونقسمها على ثلاثة فنحصل على متوسطها:

$$\frac{5+6+7}{3}=\frac{18}{3}=6$$

وأيضاً نجد عند ترتيب درجة الذكاء أن الصفة (عالٍ) مكررة ثلاث مرات عند الرتب (8, 7, 6) نحسب متوسط الرتب فنجد :

$$\frac{6+7+8}{3}=\frac{21}{3}=7$$

نكون الجدول (19-2)ومنه نحسب علاقة ارتباط الرتب :

$$r=1-\frac{6*8}{8*(64-1)}=0.905$$

التفسير: إن هناك علاقة قوية جداً ما بين تحصيل الطالب في مقرر الرياضيات ودرجة ذكائه.

جدول (19-2)

التسلسل	X_i	Y_i	d_x	d_y	$D=d_1-d_2$	D^2
1	51	تحت الوسط	3	3	0	0
2	59	فوق الوسط	4	5	-1	1
3	60	عالٍ	6	7	-1	1
4	35	ضعيف جداً	1	1	0	0
5	39	ضعيف	2	2	0	0
6	60	متوسط	6	4	2	4
7	63	عالٍ	8	7	1	1
8	60	عالٍ	6	7	-1	1
المجموع					-1	8

2-9-10- قياس شدة الارتباط بين الظواهر الوصفية:

في كثيرٍ من الدراسات الاقتصادية والاجتماعية تكون البيانات المتوفرة لدينا بيانات نوعية. مثلاً عند دراسة العلاقة ما بين نوع الجنس (ذكر أو أنثى) والحالة التعليمية لعينة ما، أو دراسة العلاقة ما بين الإنفاق وبين جنس رب الأسرة. وفيما إذا كانت البيانات نوعية فإنه يمكننا استخدام معامل الاقتران أو معامل التوافيق، وذلك حسب طبيعة الظاهرة المدروسة، وما يتناسب معها كمقياس لإظهار طبيعة العلاقة.

تمارين غير محلولة

1- تحدث عن أنواع المتغيرات.

2- كيف نميز العلاقة بين المتغيرات ؟

3- إذا علمت أن:

$$\sum XY = 6960 ; \sum X_i^2 = 1020 \quad n = 12 ; \sum Y_i = 756 ; \quad \sum X_i = 108$$

والمطلوب: احسب معامل الارتباط بين Y_i و X_i .

4- إذا علمت أن:

$$n = 12 \quad \sum Y_i^2 = 3812 ; \sum XY = 4114 ; \sum X_i^2 = 4812$$

والمطلوب: احسب معامل الارتباط بين Y_i و X_i .

5- ليكن لدينا الجدول التالي:

20	18	16	14	12	10	8	6	4	2	X_i
100	90	90	80	60	50	40	50	20	10	Y_i

والذي يبين العلاقة بين وحدة الأسعار X_i والكميات المعروضة لوحدة المنتج بالطن Y_i. والمطلوب: أ- ارسم شكل الانتشار لبيانات الجدول.

ب- أوجد معامل الارتباط بالطريقة المناسبة، وفسر النتائج.

5- الجدول أدناه يوضح العلاقة لعشرة طلاب بين رتب تأدية أعمال السنة الدراسية ونتائجهم في الامتحان النهائي.

J	I	H	G	F	E	D	C	B	A	الطلاب
64	85	62	65	60	80	90	73	75	87	نتائج الامتحان النهائي

والمطلوب: أوجد العلاقة بين تأدية الطلاب لأعمال السنة ودرجتهم في الامتحان النهائي.

6- لدراسة اختيار المستهلكين لاثني عشر صنفاً من معجون الأسنان من خلال مركزي تسويق معجون الأسنان، فكانت النتائج:

أصناف المعجون	A	B	C	D	E	F	G	H	I	J	K	L
المركز الأول	9	10	4	1	8	11	3	2	5	7	12	6
المركز الثاني	7	8	3	1	10	12	2	6	5	4	11	9

والمطلوب: احسب معامل الارتباط بين تصنيف المركزين .

7- ليكن لدينا الجدول التكراري للعلاقة ما بين الفئات العمرية (Y) وفئات الوزن (X) ،كما هو موضح أدناه:

فئات العمر فئات الوزن	-30	-30	-40	-50	60 وأقل من 70	المجموع
10وأقل من 14	1	2	2	0	0	5
14 وأقل من 18	3	8	10	4	0	25
18 وأقل من 22	1	7	12	11	4	35
22 وأقل من 26	0	3	6	5	1	15
المجموع	**5**	**20**	**30**	**20**	**5**	**80**

والمطلوب: أوجد معامل الارتباط للتوزيعات التكرارية في الجدول أعلاه وفسر النتائج.

8- ماذا يقيس معامل الارتباط (r) ؟ ما هو المدى الذي تقع فيه قيمته ؟ ما هي العلاقة بين معامل الارتباط وتحليل الانحدار ؟ اشتق الصيغ الآتية : $r^2 = \frac{SSR}{SST}$, $r = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2}\sqrt{\sum y_i^2}}$

9- كل البيانات الإحصائية سواء كانت مبوبة أو غير مبوبة لها توزيــع إحصائي خاص بها تختلف من بيانات لأخرى. ويمكن التعرف على التوزيعات الاحتمالية من خلال معلمتين (معيارين) أساسيين . أذكرهما وعلل ذلك ؟

10- البيانات التالية تمثل عدد العبوات الزجاجية التالفة المصنوعة بواسطة آلة صنع الزجاجات خلال كل ساعة من الساعات العشر الأخيرة في الإنتاج:

3, 7, 5, 6, 9, 11, 2, 4, 7, 7

والمطلوب :

1- أوجد كلاً من الوسط الحسابي ، الوسيط ، المنوال ، الوسط الهندسي ، الوسط التوافقي و الوسط التربيعي وقارن بين هذه الأوساط الحسابية.

2- أوجد كلاً من: المدى ، التباين ، الانحراف المعياري. حــدد وحــدة القياس الكل من المقاييس التي تقوم بإيجادها.

11- إذا كان هناك علاقة ارتباط بين الرقم القياسي للإبداع الفكري I ودرجة الحرارة مقاسة بالفهرنهايت F، وإذا علمت أن هناك علاقة بين قياس درجة الحرارة مقاسة بالفهرنهايت والدرجة المئوية على النحو: $C=\frac{5}{9}(F-32)$. والمطلوب : اثبت أن : $r_{I,C}=\frac{\mathbf{Cov}(I,C)}{\sqrt{\mathbf{Var}(I)\ \mathbf{Var}(C)}}=r_{I,F}$

12- على فرض أن للعلاقة الشكل التالي: C1D7 $Y=b_1+$ b_2Z

$$\text{Cov}(X, Y) = \text{Cov}(X, [b_1 + b_2Z])$$

إذا علمت أن لدينا المعلومات التالية: X3p16D

$$\text{Cov}(S,Y)= 10.5\ ,\ \text{Var}(S)= 5.6,\quad \text{Var}(Y)= 30.7$$

المطلوب : 1- احسب عند مستوى معنوية 95% فترات الثقة للمعلمة في علاقة أسعار التضخم على الأجور:

$$\hat{P} = \underset{(0.05)}{-1.21} + \underset{(0.10)}{0.82W}$$

2- وإذا علمت ان حجم العينة هو (n=20) . ماذا تستنتج من هذا الحساب؟

13- عدد قواعد التباين المشترك.

14- في بلد معين إن الضريبة التي تدفع من قبل الشركة هي T وهي محدودة بالعلاقة:

$$T = -1.2 + 0.2P - 0.1I$$

حيث إن: P هي الأرباح و I هو رأس المال للمستثمر. الطرف الثالث الموجود الذي يبين أثر توظيف الأموال في تحفيز (S) المبيعات وكل المتغيرات مقاسة بمليون وحدة نقدية كمعدلات سنوية. احسب التباين المشترك cov(S,T)، والتباين المشترك cov(S,I) وذلك لعينة مؤلفة من أربع شركات، كما هو مبين أدناه في الجدول وتأكد من صحة العلاقة التالية:

$$Cov(S,T) = 0.2Cov(S,P) - 0.1Cov(S,I)$$

اشرح بشكل تفصيلياً ، لماذا ينبغي أن تكون هذه الحالة.

Firm	*Sales*	*Profits*	*Investment*	*Tax*
1	100	20	10	1.8
2	50	9	4	0.2
3	80	12	4	0.8
4	70	15	6	1.2

15- باستخدام البيانات في التمرين السابق احسب:

cov (P,S) ، var (I) ، var (P) ، var (T)،

وتحقق من العلاقة التالية: $Var(T) = 0.04Var(P) + 0.01Var(I) - 0.04Cov(P,I)$

الفصل الثالث

بناء نموذج تحليل الانحدار الخطي البسيط

3-1- مقدمة:

هناك العديد من طرق الاقتصاد القياسي التي يمكن استخدامها للحصـــول على تقديرات لمعاملات العلاقات الاقتصادية من المشـــاهدات الإحصـــائية وأن أشهر هذه الطرق هي طريقة المربعات الصغرى(OLS)

The method of ordinary classical least squares

وتتميز هذه الطريقة عن غيرها من الطرق الأخرى بالمزايا التالية:

1- المعاملات المقدرة حسب (OLS) تتمتع بخصائص نموذجية، مثل عدم التحيز و أصغر تباين وغيرها من الخصائص الأخرى نستعرضها لاحقاً.

2- تتصف العمليات الحسابية لهذا النموذج بالبساطة والوضوح مقارنة مع الطرق الأخرى، ولا تحتاج إلى كم كبير من البيانات الإحصائية.

3- تم اختبار هذه الطريقة في مجالات عديدة، وأعطت نتائج مرضية، على الرغم من التطور التكنولوجي في مجال الحاسبات الإلكترونية، إلا أن هذه الطريقة بقيت واحدة من أسهل الطرق المتبعة في تكوين علاقات النموذج.

4- آلية تطبيق هذه الطريقة سهلة وواضحة.

5- أغلب الطرق الأخرى المتبعة تتضمن ضمن تطبيقها تطبيقاً لهذه الطريقة، مع أخذ التعديلات بما يتناسب والطريقة المتبعة، ما عدا طريقة التقدير بأعظم إمكانية وباستخدام بيانات كاملة.

لنستعرض مثال تطبيقي من الواقع لو أخذنا نظرية الطلـــب فـــي أبســـط صورها، حيث يوجد علاقة موجبة بين الكمية المطلوبة للسلعة وسعرها. وعلى

فرض بقاء جميع العوامل الأخرى ثابتة، فإذا ازداد الطلب على السلعة يزداد سعرها، والعكس صحيح.

نستخدم خطوات الاقتصاد القياسي التي استعرضناها سابقاً من أجل تحديد النموذج الرياضي للطلب. نحدد المتغير التابع وهو الطلب هنا، والمتغير المستقل وهو السعر. أما النظرية الاقتصادية فتزودنا بالمعلومات التالية حول دالة الطلب:

1- المتغير التابع هو كمية الطلب على السلعة (Y) والمتغير التفسيري هو متغير السعر (X) أي: شكل علاقة على النحو الآتي: $Y= f(X)$.

2- النظرية الاقتصادية لا تحدد طريقة دراسة الطلب هل من خلال نموذج بمتغير واحد أو باستخدام علاقة متعددة المتغيرات. للتبسيط سوف نختار نموذجاً بمتغير واحد، ثم نحاول تطويره إلى نموذج متعدد المتغيرات.

3- النظرية الاقتصادية لا تبين لنا ما هو شكل النموذج المزمع الوصول إليه ؟ هل نستخدم معادلة خطية أم معادلة لا خطية ؟ كتب التحليل الاقتصادي في كثير من الأحيان تمثل دالة الطلب من خلال مستقيم هابط إلى الأسفل. ولكن الاقتصاد القياسي عليه أن يقرر شكل هذه الدالة من خلال طريقته خاصة. وللتبسيط نفترض العلاقة بأبسط أشكالها الرياضية الممكنة هي علاقة خطية من الشكل: $\mathbf{Y_i = b_0 + b_1 X_i}$

هذا الشكل يدل على أن العلاقة سببية الاتجاه، أي أن السعر هو السبب فيما يطرأ على الكمية من تغيرات، والعكس ليس صحيحاً. ولا بد من تحديد معاملات هذه العلاقة وهي $(\hat{b}_1, \hat{b}_0)$. أما فيما يتعلق بإشارة ومستوى حجم $(\hat{b}_0)$ هو عبارة عن ثابت التقاطع الرأسي أو الجزء المحصور بين نقطة تقاطع المحني مع المحور (Y) ونقطة المبدأ. فإنه من المتوقع أن تكون قيمة المعامل إما صفر وإما موجبة. أي حتى لو كان السعر مساوياً الصفر فإنه، وهناك كمية مطلوبة من السلعة المعينة. ومن المعتاد أن تكون موجبة في دالة الطلب، فإذا

ظهرت إشارة سالبة فلا معنى للكميات السالبة في الاقتصاد ما لم يكن هناك مبرراً اقتصادياً لتوقع ظهور إشارة سالبة للثابت. وإشارة $(\hat{b}_1)$ تحدد مرونة الطلب بالنسبة للسعر. أما إشارة $(\hat{b}_1)$ فهي سالبة لأن ميل منحني الطلب بشكل عام متجه نحو الأسفل. أما فيما يتعلق بمستوى حجم $(\hat{b}_0)$ ، فيجدر ملاحظة أن $(\hat{b}_0)$ هو جزء من مرونة الطلب بالنسبة للسعر أي: $E_p = \frac{\delta Y}{\delta X} * \frac{X}{Y}$

ومن تابع الطلب نجد أن $(\hat{b}_1)$ سالبة لأن: $E_p = \frac{\delta Y}{\delta X} = -b_1$.

من تابع الطلب نجد أن $\hat{b}_1$ سالبة لان: $\frac{\delta Y}{\delta X} = -b_1$

في حال استخدام معادلة خط الانحدار في حساب المرونة فإننا نستخدم $(\hat{b}_1)$ ومتوسط السعر $(\overline{X})$ ومتوسط الكمية المعروضة $(\overline{Y})$ من النموذج نجد أن:

$$E_p = -\hat{b}_1 * \frac{\overline{X}}{\overline{Y}}$$

معلوم لدينا أن: $\overline{Y} = \hat{b}_0 + \hat{b}_1 \overline{X}$

نعوض في معادلة المرونة فنجد: $E_p = \frac{-\hat{b}_1 \overline{X}}{\hat{b}_0 + \hat{b}_1 \overline{X}}$

إذا كان $\hat{b}_1 > 0$ نستنتج أن:

1- الطلب يكون مرنا $E_p > 1$

2- الطلب يكون غير مرنا $E_p < 1$

3- الطلب عديم المرونة $E_p = 0$

بعد هذه المقدمة النظرية لننطلق إلى معالجة هذا النموذج من المعادلة الخطية البسيطة التالية :

$$Y_i = b_0 + b_1 X_i + U_i \qquad (1)$$

(i) رقم المشاهدة ($i = 1,2,3 \ ..... \ n$) وقد تكون المشاهدات في صـــور بيانات مقطعيةCross section data أو سلاسل زمنية Time series data

(Y_i) قيمة المشاهدة (i) الخاصة بالمتغير التابع

(X_i) قيمة المشاهدة (i) الخاصة بالمتغير المستقل

(U_i) قيمة المشاهدة (i) الخاصة بالمتغير العشوائي، وهذه القيمـــة غيـــر قابلة للقياس، وذلك بخلاف قيم (x_i) و (Y_i) القابلة للقياس في الواقع التطبيقي.

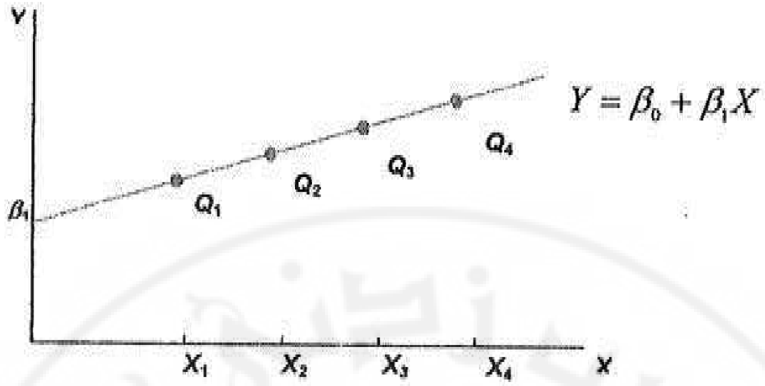
(b_1 , b_0) معالم المجتمع Population parameters وهمـــا يحـــددان أساساً العلاقة الحقيقية بين المتغيرين موضع البحث. فعلى سبيل المثال إذا كانت (X) تمثل دخل الأسرة و (Y) إنفاقها على المواد الغذائية فإن (b_0) تمثل معدل حالة الكفاف و (b_1) معدل الميل الحدي للاستهلاك بالنسبة إلى جميع الأسر في البلد المدروس.

وبما أن نموذج الانحدار الخطي البسيط يعتمد على المتغير العشـــوائي أو (الخطأ) فإن قيم (X) المحددة لا تقابل بقيم محددة مـــن (Y)، وإنمـــا بتوزيـــع احتمالي كامل.

وعليه فإن القيم (Y_i) تتحدد على أساس كل من قيمة (X_i) وتوزيع الخطأ (U_i)، ولهذا السبب فإن معادلة نموذج الانحدار (1) تعتبر غير كافيـــة، وإنمـــا يتطلب الأمر وضع معادلات تكميلية لتغطية الافتراضات الخاصة بتوزيع الخطأ وأسس تحديد قيم المتغير المستقل.

3-2- الانحدار البسيط:

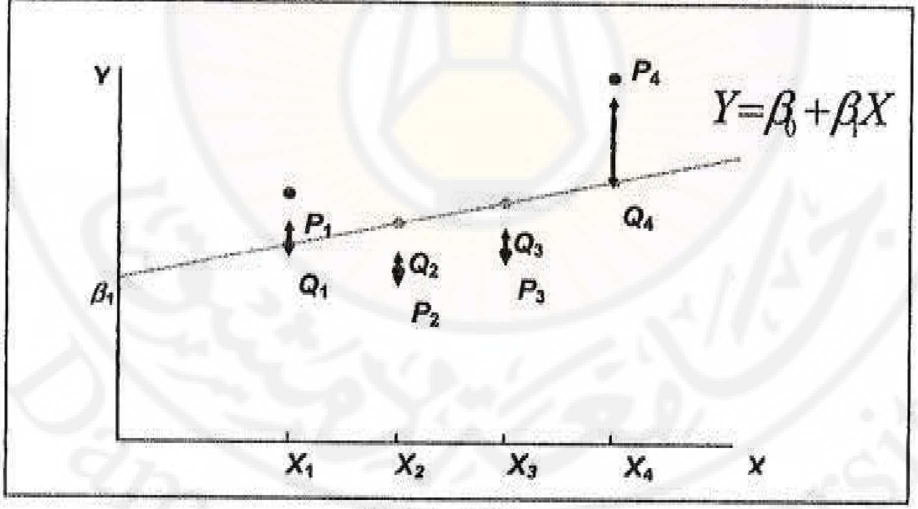
نفترض أن المتغير Y يشكل علاقة خطية مع المتغير الآخر X، ونريـــد تقدير المعلمات b_0 و b_1 وعلى فرض أن لدينا عينة مكونة من 4 مشاهدات كما هو موضح على المحور X. إذا كانت العلاقة دقيقة جداً، فإن المشاهدات سوف تقع على مستقيم، ونستطيع تقدير قيم b_0 و b_1 و دون أي مشكلة.



الشكل (3-1)

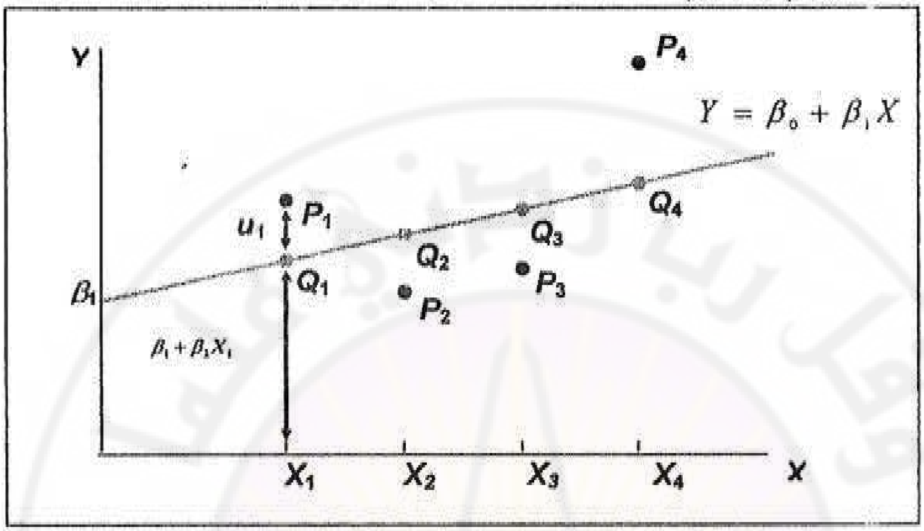
لكن التجربة أثبتت لمعظم الاقتصاديين أن هذه العلاقة ليست كذلك، وحساب قيم المتغير Y لا تقع على مستقيم. وهذا يجبرنا على كتابة النموذج على الشكل التالي: $Y = b_0 + b_1X + u$

حيث u المتغير العشوائي.



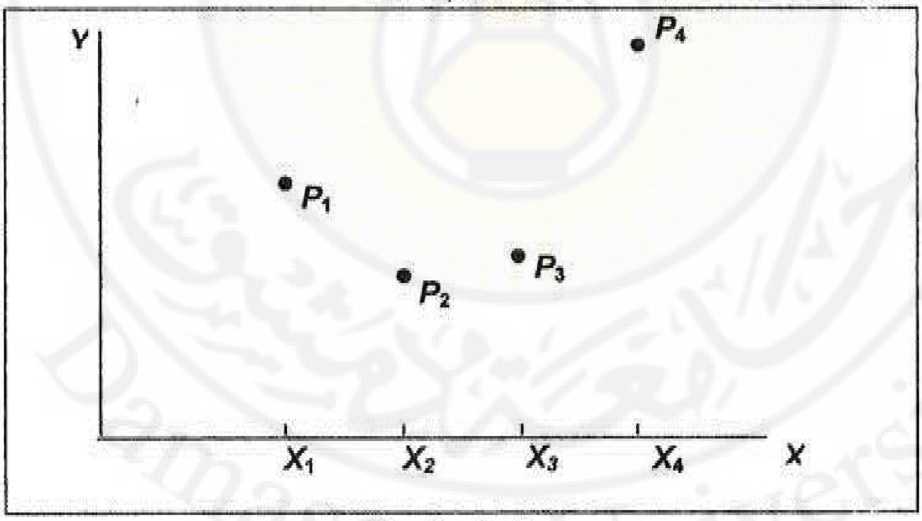
الشكل (3-2)

كل قيمة للمتغير قيمة غير عشوائية Y تحدد بالعلاقة : $b_0 + b_1X$ و قيمة عشوائية تحدد بالعنصر u . القيمة الأولى من المشاهدة يمكن تحليلها إلى عنصرين (مركبتين).



الشكل (3-3)

نحاول ملاحظة فقط النقط P على الرسم فنجد:

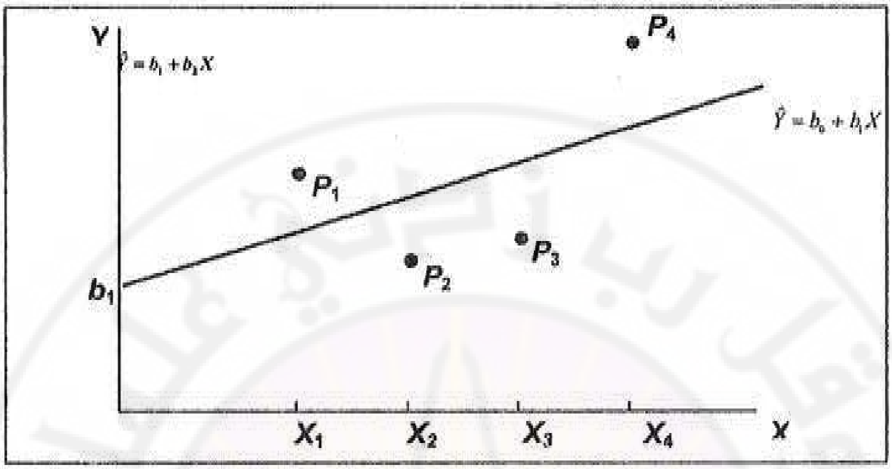


الشكل (4-3)

يمكننا توفيق مستقيم تقريبي لهذه النقاط يمثل بالعلاقة الخطية التالية:

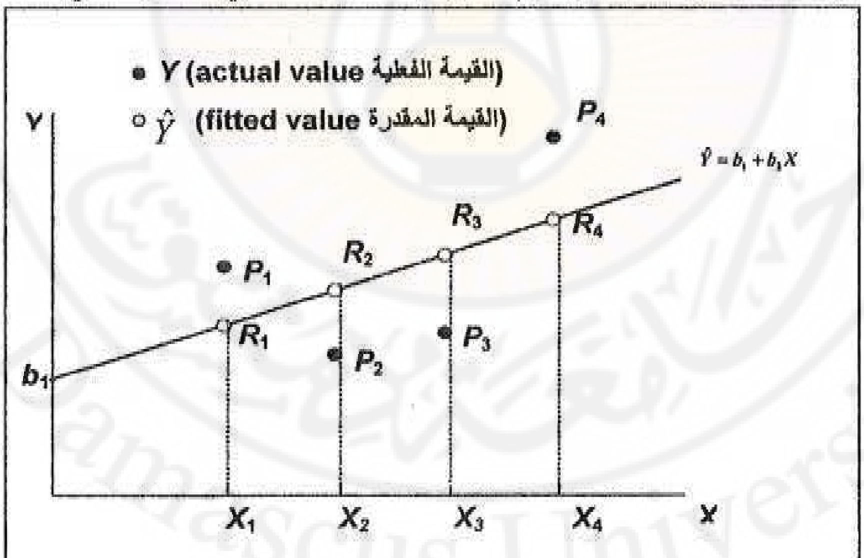
$. Y = \beta_0 + \beta_1 X$

إذا كتبنا علاقة المستقيم بهذا الشكل: $\hat{Y} = b_0 + b_1 X$ فعندئـــذ b_0 تكـــون تقديراً لــ β_0, b_1 تقديراً لــ β_1.



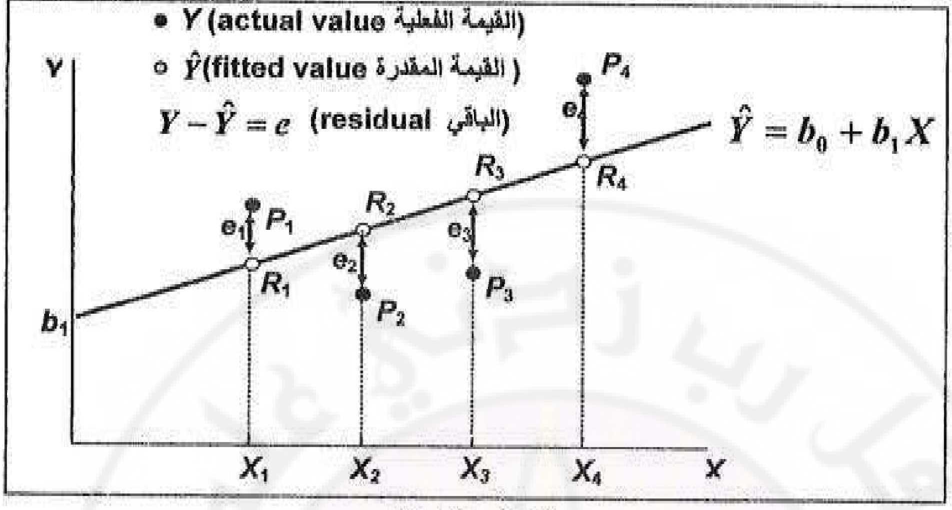
الشكل (3-5)

الخط يسمى النموذج المقدر ، وقيم Y المتنبأ بها بهذه العلاقة تسمى القيمة المقدرة لــ $\hat{Y}$. وهذا يعطينا قيم نمثلها بالنقاط R كما في الشكل التالي:



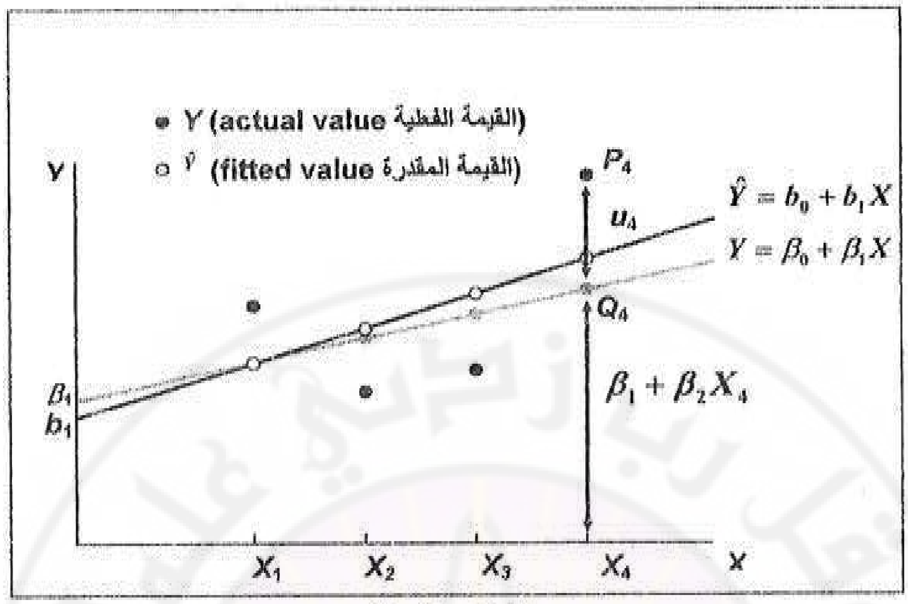
الشكل (3-6)

الاختلاف بين القيم الفعلية والقيم المقدرة لـــ Y تسمى البواقي.



الشكل (7-3)

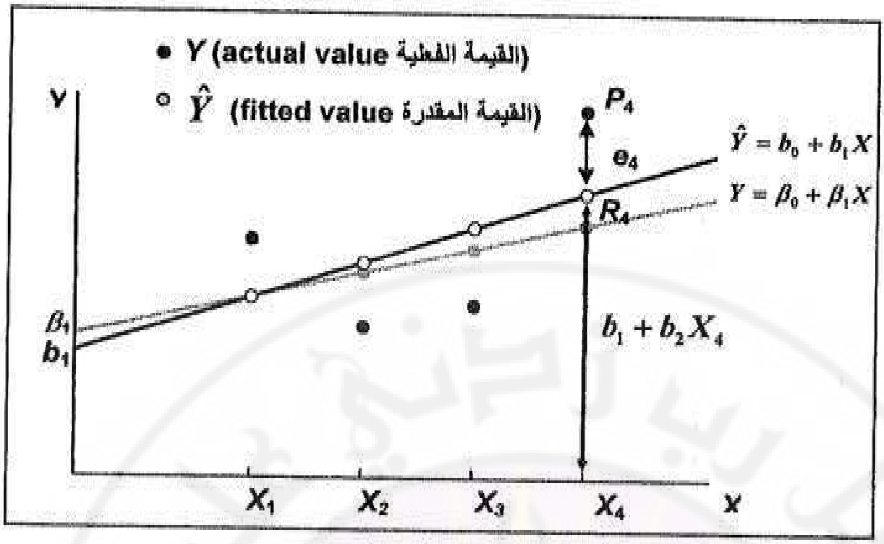
مع ملاحظة أن قيم البواقي ليست هي قيم الخطأ العشوائي نفسها. الشكل البياني يبين حقيقة العلاقة كمنحني التقدير. الخطأ العشوائي لكل مشاهدة ينقسم إلى مركبتين ، مركبة غير العشوائية للعلاقة الحقيقية و مركبة القيم الفعلية للمشاهدة. البواقي عبارة عن الفرق بين القيمة الفعلية والقيمة المقدرة. إذا كان التقدير ملائماً يكون جيداً، فالبواقي والخطأ العشوائي سوف يكونان متماثلان ، أي يجب أن يكونان محافظين على خصائصهما.



الشكل (3-8)

سوف نستخدم كلا المنحنيين في التحليل، أي كل تحاليل مركبات القيم Y. سنستخدم المثال المكون من أربع مشاهدات. نستخدم العلاقة النظرية Y نجد أنها مكونة من مركبتين مركبة متضمنة في العلاقة $b_0 + b_1X$ وأخرى مركبة الخطأ العشوائي؛ كما موضح في الشكل أعلاه .

هذا هو التحليل العملي الذي سوف نستخدمه.



الشكل (3-9)

والتحليل الآخر يكون مع الإشارة إلى منحني التقدير. كل حالة من المشاهدات، القيم الفعلية لـ Y تكون مساوية القيم التقديرية بالإضافة إلى البواقي. و هو التحليل العملي الذي سوف نستخدمه في هذه الدراسة. هذا التحليل يبقى نظرياً لأننا لا نعرف قيم المعاملات β_0 أو β_1 أو قيم الخطأ العشوائي، ولا بد من استخدام في تحليلنا خصائص معامل التحديد.

للبدء سوف نرسم المنحني المقدر الناتج عن الحد الأدنى لمربع البواقي RSS.

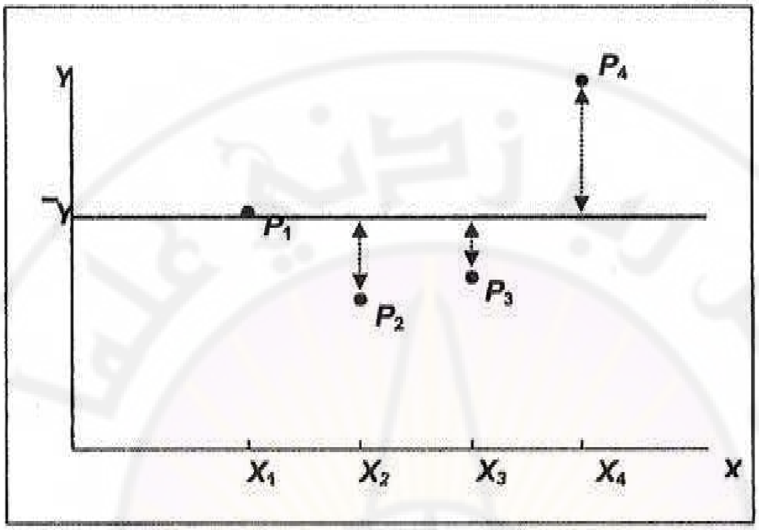
الشكل التالي يبين الحد الأدنى لمربع البواقي:

$$RSS = \sum_{i=1}^{n} e_i^2 = e_1^2 + ... + e_n^2$$

والسؤال الذي يطرح نفسه ، لماذا لا نأخذ مجاميع البواقي الأقل دون تربيع؟ لماذا مربع البواقي؟ لماذا ليس تماماً مجموع البواقي؟

الجواب سوف نحصل على المظهر الدقيق للمنحني من خلال رسم الخط الأفقي، الذي يمثل متوسط قيم المتغير Y . فعندئذ نجد أن مجموع البواقي سوف

يساوي الصفر. لأن مجموع القيم السالبة يساوي القيم الموجبة، ولذلك نستخدم مربع البواقي. و هناك طرق أخرى في حل هذه المشكلة. معيار القيمة الأقل للمربعات تملك جاذبية في اشتقاق التقدير، وهذه الجاذبة مبينة على خواص لها شروط مرضية.



الشكل (3-10)

لكن لتطبيق ذلك لا بد من تحقيق بعض الفروض الأساسية، وسوف نستعرض هذه الفروض في الفصل اللاحق.

3-3- تقدر معالم الانحدار Estimation of Regression coefficients

هناك نوعان من التقدير الإحصائي، الأول (تقدير النقطة Point Estimation) والثاني تقدير المجال Introvert Estimation وينصب تقدير النقطة على تحديد قيمة مفردة لكل معلمة من معالم المجتمع.

أما تقدير المجال فإنه يحدد المجال الذي تقع ضمنه المعلمة الحقيقية بدرجة ثقة معينة. وفي الواقع إن تقدير المجال عبارة عن عملية تكميلية لتقدير النقطة، وذلك لأن تقدير النقطة يتم على أساس العينة. الأمر الذي يسبب في

حصول انحراف في القيم التقديرية للمعالم عن القيم الحقيقية نتيجة لأخطاء المعاينة.

4-3 - تقدير الفترة Introvert Estimation

نرمز للقيم التقديرية لـ (b_1 , b_0) بـ $(\hat{b}_1 , \hat{b}_0)$ كما نطلق تسمية (خط انحدار العينة Sample regression line) على الخط المتكون بواسطة القيمتين $(\hat{b}_1 , \hat{b}_0)$. ذلك أنه يوفق البيانات الواردة في العينة على أفضل وجه . وهو بمثابة خط تقدير لخط انحدار المجتمع Population regression line ونورد أدناه صيغة خط انحدار العينة المقدر :

$$\hat{y}_i = \hat{b}_0 + \hat{b}_1 X_i + e_i$$

حيث إن $(\hat{Y}_i)$ عبارة عن القيمة الموفقة fitted value لقيمة المشاهدة الأصلية، علماً بأن القيمة الموفقة تنحرف عن قيمة المشاهدة بمقادير معينة يطلق عليها تسمية (البواقي residuals) ويرمز لها بالرمز (e_i) أي أن :

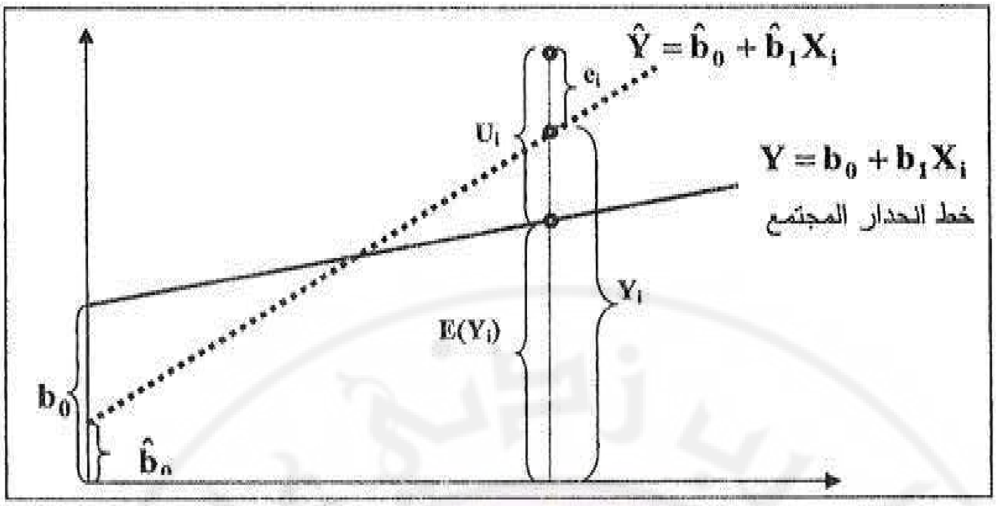
$$Y_i - \hat{Y}_i = e_i$$

مما تقدم يجب أن نميز بين الخطأ في المجتمع (U_i) والبواقي في العينة(e_i) كالآتي :

$$Y_i = b_0 + b_1 X_i + U_i \qquad \text{(القيم الفعلية)}$$

$$\hat{Y}_i = \hat{b}_0 + \hat{b}_1 X_i + e_i \qquad \text{(القيم التقديرية)}$$

بطبيعة الحال يعود الاختلاف بين البواقي والأخطاء إلى أخطاء المعاينة . وفي الواقع أن البواقي عبارة عن قيم تقديرية للأخطاء، والشكل البياني أدناه يبين ذلك :



الشكل (3-11)

3-5- إيجاد قيمة المعلمات (b_1, b_0) بطريقة المربعات الصغرى

The method of least squares

إن طريقة المربعات الصغرى LSO تشترط تصغير القيمة $\sum e_i^2$ إلى أدنى حد، وبالتالي فما هي إلا عبارة عن مشكلة النهايات الصغرى لعلاقة رياضية أي أن :

$$s = \sum e_i^2 \rightarrow \min$$

طريقة المربعات الصغرى لتقدير معلمات معادلة الانحدار الخطي البسيط، حسب نظرية غاوس-ماركوف، والتي تنص على أنه إذا كان المتغير المستقل والمتغير التابع عشوائيين في المجتمعات الجزئية المتساوية التباين، فإن طريقة المربعات الصغرى تعطي أكفأ تقديرات خطية غير متحيزة لمعالم علاقة الانحدار (b_1, b_0) في المجتمع الإحصائي. هذا يعني إنه يمكن على الاعتماد طريقة المربعات الصغرى في تقليل مجموع مربعات البواقي إلى حد النهاية الصغرى.

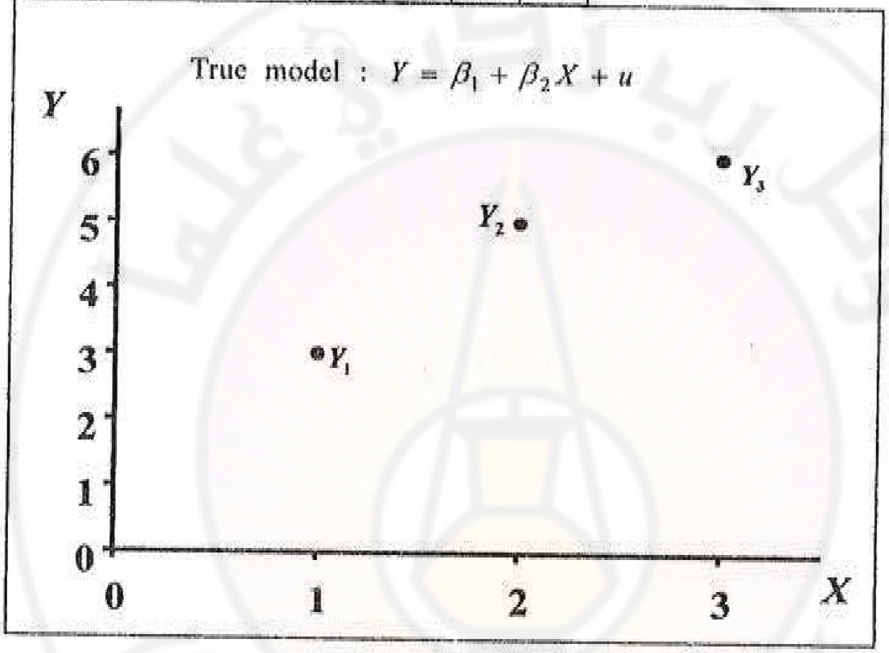
تأخذ معادلة الانحدار الخطي البسيط الصيغة التالية :

$$Y_i = b_0 + b_1 X_i + U_i \qquad (1)$$

سنحاول احتساب القيم التقديرية للمعالم على أساس (تصغير minimization) مجموع مربعات انحرافات قيم المشاهدات عن أوساطها باستخدام معيار القيمة الأقل ما يسمى بالمربعات الصغرى (OLS).

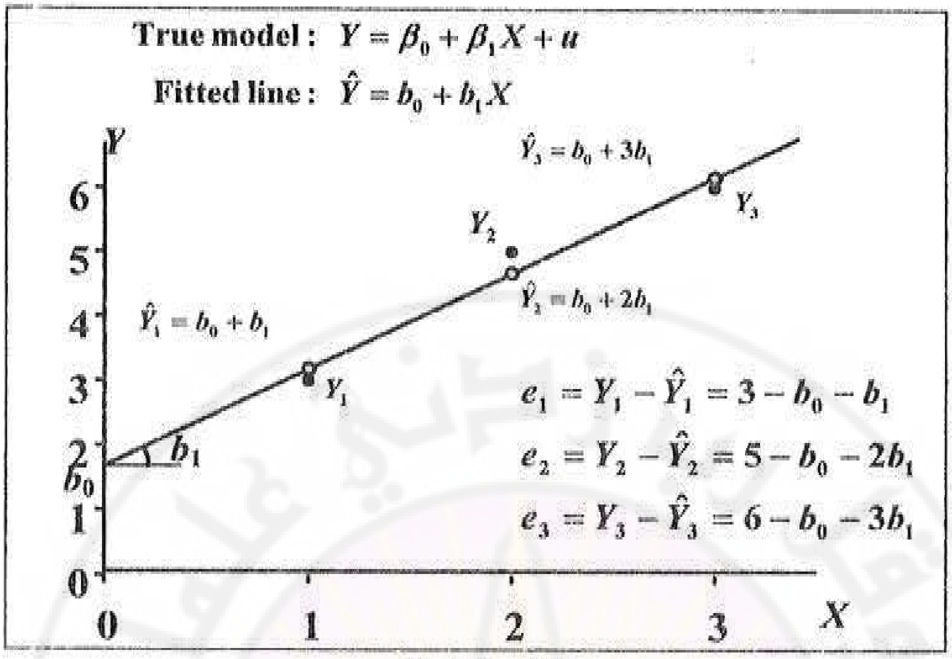
سوف نعرض مثال عددي لثلاثة مشاهدات هي:

X	1	2	3
Y	3	5	6



الشكل (3-12)

نكتب علاقة الانحدار المقدرة ، وسوف نحدد قيم (b_0)،(b_1)، حسب القيمة الدنيا ، لمجموع مربع البواقي RSS.



الشكل (3-13)

نكتب البواقي للمثال كما هو مبين أعلاه، و نطبق علاقة مجموع مربع البواقي، و نفك التربيع، و نجمع الحدود المتشابهة مع بعضها بعض فنحصل على:

$$RSS = e_1^2 + e_2^2 + e_3^2 = (3 - b_0 - b_1)^2 + (5 - b_0 - 2b_1)^2 + (6 - b_0 - 3b_1)^2$$
$$= 9 + b_0^2 + b_1^2 - 6b_0 - 6b_1 + 2b_0b_1$$
$$+ 25 + b_0^2 + 4b_1^2 - 10b_0 - 20b_1 + 4b_0b_1$$
$$+ 36 + b_0^2 + 9b_1^2 - 12b_0 - 36b_1 + 6b_0b_1$$
$$= 70 + 3b_0^2 + 14b_1^2 - 28b_0 - 62b_1 + 12b_0b_1$$

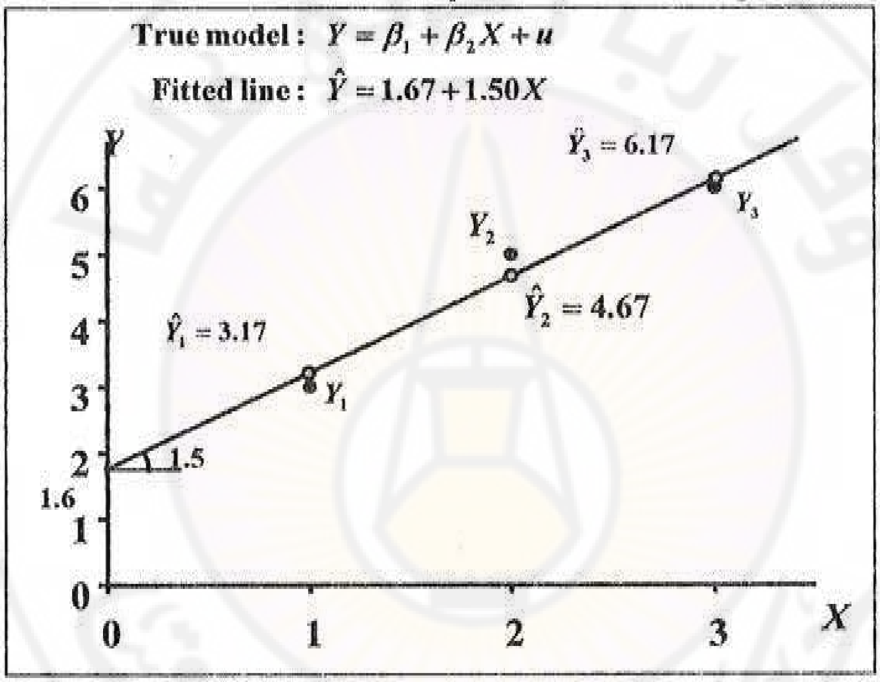
وللحصول على القيمة الدنيا للمعلمات b_0, b_1 يجب أن يتحقق شرطي الحالة الدنيا وهو الشرط اللازم والشرط الكافي . الشرط اللازم المشتق الأول يساوي الصفر. و المشتق الثاني أكبر من الصفر. نحسب المشتق الأول فنجد:

$$\frac{\partial RSS}{\partial b_1} = 0 \;\; \Rightarrow \;\; 6b_0 + 12b_1 - 28 = 0$$

$$\frac{\partial RSS}{\partial b_2} = 0 \quad \Rightarrow \quad 12b_1 + 28b_2 - 62 = 0$$

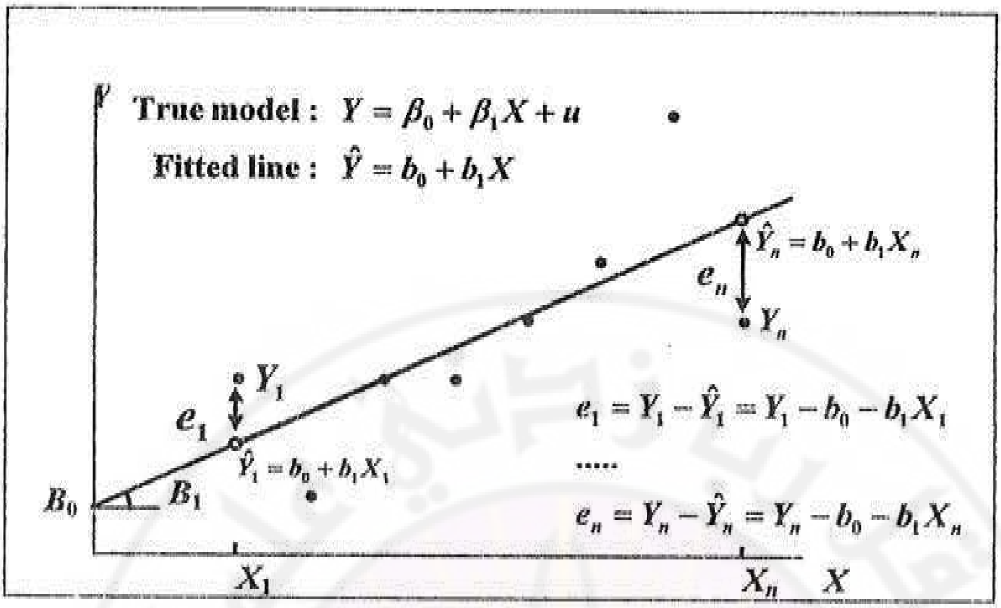
الشرط الأول يعطينا علاقتين بمجهولين. و بحل المعادلتين حلاً مشترك ، نحصل على قيم المعلمات b_0 و b_1 و هي: $b_1 = 1.50$, $b_0 = 1.67$ ، التي تحقق القيمة الدنيا.

وهنا نعود إلى الشكل البياني أعلاه ونرسم الخط المقدر و القيم المقدرة لـــ Y ، ونضع عليه المعادلات كما هي مبينة على الشكل.



الشكل (14-3)

الآن سوف نعمم ذلك على n حالة من المشاهدات كما هو مبين في الشكل البياني التالي:



الشكل (3-15)

فنجد:

$$RSS = e_1^2 + ... + e_n^2 = (Y_1 - b_1 - b_2 X_1)^2 + ... + (Y_n - b_1 - b_2 X_n)^2$$

$$= \quad Y_1^2 + \quad b_0^2 + \quad b_2^2 X_1^2 - \quad 2b_0 Y_1 - \quad 2b_1 X_1 Y_1 + \quad 2b_0 b_1 X_1$$

$$+ \; ...$$

$$+ \quad Y_n^2 + \quad b_0^2 + \quad b_1^2 X_n^2 - \quad 2b_0 Y_n - \quad 2b_1 X_n Y_n + \quad 2b_0 b_1 X_n$$

$$= \sum Y_i^2 + n b_0^2 + b_1^2 \sum X_i^2 - 2b_0 \sum Y_i - 2b_1 \sum X_i Y_i + 2b_0 b_1 \sum X_i$$

ومن هذه المعادلة التي تعتمد على قيم المشاهدات للمتغيرين X,Y والبيانات المتعلقة بهما. يتم تحديد المعلمات (b_0,b_1) للعلاقة الارتباطية باستخدام طريقة مربعات الصغرى ولذلك نقوم بما يلي:

أولاً: نقوم باشتقاق العلاقة بالنسبة للمتغيرb_0 . ونقوم بتبسيط الشكل الرياضي، فنحصل على قيمة b_1 .

$$\frac{\partial RSS}{\partial b_0} = 0 \quad \Rightarrow \quad 2nb_0 - 2\sum Y_i + 2b_1 \sum X_i = 0$$

$$nb_0 = \sum Y_i - b_1 \sum X_i$$

وبتقسيم الطرفين على n نحصل على :

$$b_0 = \overline{Y} - b_1 \overline{X} \qquad (1)$$

ثانياً: نقوم باشتقاق العلاقة بالنسبة للمتغير b_1 .

$$\frac{\partial RSS}{\partial b_1} = 0 \quad \Rightarrow \quad 2b_1 \sum X_i^2 - 2\sum X_i Y_i + 2b_0 \sum X_i = 0$$

نقسم العلاقة على الرقم 2 ، لذلك نحصل على الشكل التالي:

$$b_1 \sum X_i^2 - \sum X_i Y_i + b_0 \sum X_i = 0$$

و سنحاول الاستفادة من العلاقة الأولى للحصول على b_1 ، نعوض قيمة b_0 في المعادلة الثانية فنحصل على b_1 .

$$b_1 \sum X_i^2 - \sum X_i Y_i + (\overline{Y} - b_1 \overline{X}) n\overline{X} = 0$$

بالاستفادة من تعريف المتوسط الحسابي نجد:

$$\overline{X} = \frac{\sum X_i}{n} \quad \Rightarrow \quad \sum X_i = n\overline{X}$$

نحول الحدود التي لا تحتوي على b_1 إلى الطرف الآخر، ونقسم الطرفين على n .

$$b_1 \left(\sum X_i^2 - n\overline{X}^2 \right) = \sum X_i Y_i - n\overline{X}\overline{Y}$$

$$b_1 \left(\frac{1}{n} \sum X_i^2 - \overline{X}^2 \right) = \frac{1}{n} \sum X_i Y_i - \overline{X}\overline{Y}$$

هكذا نحصل على علاقة لحساب b_1 هي:

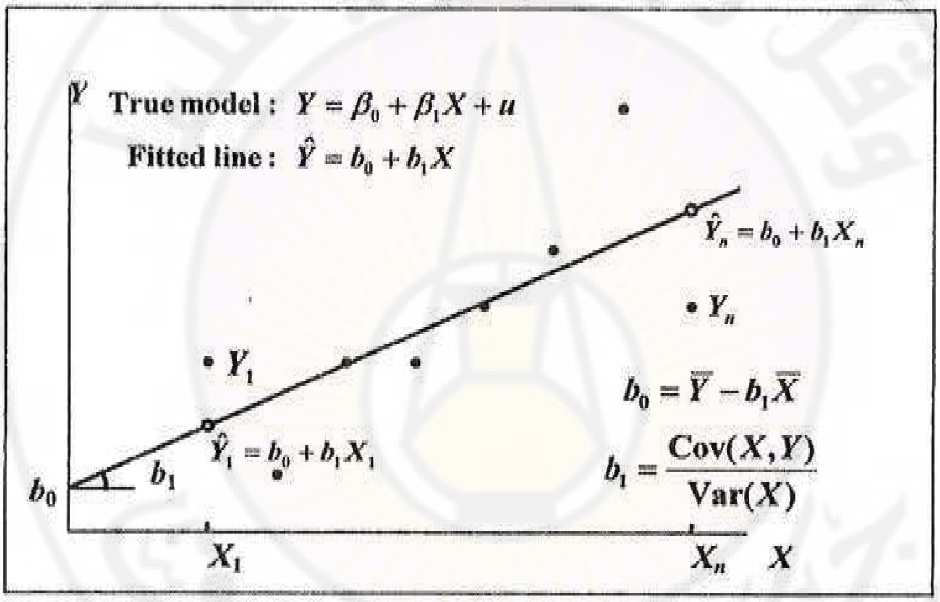
$$b_1 = \frac{\frac{1}{n} \sum X_i Y_i - \overline{X}\overline{Y}}{\frac{1}{n} \sum X_i^2 - \overline{X}^2} \qquad (2)$$

يمكننا كتابة هذه العلاقة من خلال علاقة التباين و التباين المشترك على النحو الآتي:

$$b_1 \mathbf{Var}(X) = \mathbf{Cov}(X,Y)$$

$$b_1 = \frac{\mathbf{Cov}(X,Y)}{\mathbf{Var}(X)} \qquad (3)$$

وإن شكل الانتشار، يلخص لنا ما عملناه . نفترض أن النموذج المعدل كما هو واضحاً على الشكل (3-16) نقوم بتقدير خط الانحدار. نحدد المعلمات لخط التقدير بطريقة مجموع المربعات الصغرى. بالاشتقاق نحصل على علاقات تحسب لنا المعلمات $b_0, \; b_1$



الشكل (3-16)

إن المعادلة الطبيعية لحساب b_1 تكون معيارية، ويمكن تعميمها بسهولة. ولكن هناك طرق عديدة لحساب المعلمة b_1 منها الخيارات التالية :

$$\hat{b}_1 = \frac{\mathbf{Cov}(X,Y)}{\mathbf{Var}(X)}$$

أخيراً نكون قد حصلنا على المعادلتين الضروريتين لتقدير المعلمتين b_1 , b_0 واللتين سوف يأتي استخدامها في المستقبل، وهما المعادلتان الأوليتان التاليتان:

$$\begin{aligned} \sum Y_i &= n\hat{b}_0 + \hat{b}_1 \sum X_i \\ \sum \hat{Y}_i &= \hat{b}_0 \sum X_i + \hat{b}_1 \sum X_i^2 \end{aligned} \qquad (12)$$

ومن معادلة الانحدار المقدرة وجدنا أن: $\overline{Y}_i = \hat{b}_0 + \hat{b}_1 \overline{X}_i + e_i$

نأخذ المجموع النوني للطرفين فنستنتج معادلة أخرى هي:

$$\sum \hat{Y}_i = n\hat{b}_0 + \hat{b}_1 \sum X_i + \sum e_i$$

وبمقارنة الطرف الأيمن من المعادلة الأخيرة أعلاه مع الطرف الأيمن من المعادلة (5) مع (12) نحصل على الآتي :

$$n\hat{b}_0 + \hat{b}_1 \sum X_i = n\hat{b}_0 + \hat{b}_1 \sum X_i + \sum e_i$$

ومنه نجد أن: $\sum e_i = 0$

وإذا ضربنا طرفي معادلة الانحدار بـ X_i ، وأخذنا المجموع النوني للطرفين نجد أن:

$$\sum Y_i X_i = \hat{b}_0 * \sum X_i + \hat{b}_1 \sum X_i^2 + \sum X_i e_i$$

و بمقارنة الطرف الأيمن مع الطرف الأيمن من المعادلة (12) نحصل على الآتي: $\sum X_i e_i = 0$

تقدير علاقة الانحدار عندما يكون ثابت التقاطع معدوماً: أي نموذج الانحدار من الشكل: $Y = b_1 X_i + U_i$

ولابد أن نلفت الانتباه إلى أنه عندما نكون في صدد البحث عن علاقة انحدار بسيطة ذات ثابت تقاطع صفري من الشكل : $Y_i = \hat{b}_1 X_i + U_i$ فإننا لا نقدر سوى معامل انحدار واحد هو $\hat{b}_1$

ويحسب عندما $\hat{b}_0 = 0$ من العلاقة: $b_i = \frac{\sum XY}{\sum X^2}$

3-6- إيجاد قيم المرونات من علاقة الانحدار المقدرة:

وجدنا أن علاقة الانحدار المقدرة لها الشكل الآتي: $\hat{Y}_i = \hat{b}_0 + \hat{b}_1 X_i$ هي معادلة الخط الذي له الجزء المحصور بين القيمة $\hat{b}_0$ ، وميل انحدار يساوي $\hat{b}_1$ ولكن $\hat{b}_1$ ما هي إلا عبارة عن مشتق $(\hat{Y}_i)$ بالنسبة لـــ X أي:

$$\hat{b}_1 = \frac{\delta \hat{Y}}{\delta X}$$

أي أنها تساوي معامل التغير في $(\hat{Y}_i)$ عندما (X) تأخذ مقــدار صــغير جداً، ويجب أن يكون واضحاً أنه إذا كان التابع المقدر تــابع طلــب، أو تــابع عرض، فإن $(\hat{b}_1)$ ليست هي المرونة السعرية وإنما هي جزء من هذه المرونة التي تحسب من العلاقة التالية: $E_p = \frac{\delta Y}{\delta X} * \frac{X}{Y}$

وبشكل آخر نكتبها بدلالة $(\hat{b}_1)$ على النحو الآتي: $E_p = \hat{b}_1 * \frac{\overline{X}}{\overline{Y}}$

ونشير أخيراً إلى أن دلالات استخدام $(\hat{b}_1)$في علاقات متعددة أخرى منها:

$$r = \frac{Cov(XY)}{S_x S_y} = \frac{Cov(XY)}{S_x} * \frac{S_x}{S_y} = \hat{b}_1 * \frac{S_y}{S_x} \qquad (10)$$

$$\hat{b}_1 = r * \frac{S_y}{S_x}$$

مثال (1-3): ليكن لدينا الجدول التالي الذي يمثل الكميــات المعروضــة بالطن لسلعة ما وسعرها . والمطلوب إيجاد معادلة الانحدار للعلاقة بين العرض Xو السعر Y.

6- أوضح الاختلاف في التعابير التالية :

أ- بين (β_1) ، (β_0) من جهة . وبينهم و $(\hat{\beta}_2 , \hat{\beta}_1)$ من جهة أخرى .

ب- بين $(e_i),(u_i)$.

ج- بين المعادلة الحقيقية للعلاقة بين (Y)،(X) ، والمعادلة التقديرية للعلاقة بين المتغيرين (Y)،(X) ، اكتب كلاً من المعادلتين .

7- عينة متكون من (200) زوج من المشاهدات ، تم الحصول منها على المعلومات الآتية : $\sum X^2 = 12.16 \quad ,\sum Y = 20.72,$

$$\sum X = 11.34 \ \ ,\sum XY = 22.13 \sum Y^2 = 84.96$$

أوجد القيم التقديرية للانحدار الخطي للمتغير (Y) على (X) .

8- عينة مكونة من (20) مشاهدة طبق عليها نموذج الانحدار التالي : $Y = \beta_0 + \beta_1 X + e$

وعلى افتراض أن (e_i) تخضع للتوزيع الطبيعي بوسط حسابي مساو للصفر وتباين ثابت مقداره (σ_u^2) ، وقد أعطيت هذه العينة المعلومات التالية :

$$,\sum(Y-\bar{Y})^2 = 86.9 \quad \sum X = 186 \quad ; \quad ,\sum(X-X)^2 = 215.4$$
$$;\sum Y = 21.9 ;,\sum(X-\bar{X})(Y-\bar{Y}) = 106.4$$

أ- أوجد القيمة التقديرية لكل من $(\hat{\beta}_0),(\hat{\beta}_1)$ ،

$$\sum Y^2 = 84.96 \ \ ,\sum XY = 22.13 \ \ ,\sum Y = 20.72$$
$$\sum X = 11.34$$

ب- أوجد القيم التقديرية للانحدار الخطي للمتغير (Y) على (X).

9- إذا كان لدينا البيانات التالية :

Y_i	2	3	5	4	5	3	4	6
X_i	6	5	4	2	3	2	1	1

أ- أوجد القيمة التقديرية لكل من $\hat{\beta}_0, \hat{\beta}_1$ من المعادلتين الطبيعيتين.

ب- أوجد القيمة التقديرية لكل من β_0, β_1 باستخدام صيغة معادلتين كريمر .

11- من البيانات المذكورة أدناه عن كمية سقوط الأمطار بالتغيرات (Y_i) ومحصول القمح (X_i) .

Y_i	125	80	100	140	160	135
X_i	44	36	40	48	60	56

أ- أوجد معامل الارتباط بين $(X_i),(Y_i)$.

ب- أوجد معامل انحدار (Y) على (X) .

جـ- ناقش بالتفصيل الفرق بين معامل الانحدار ومعامل الارتباط؟ أيهما أدق في شرح هذه العلاقة ؟

11- على فرض لدينا المثال التالي لشرح العلاقة بين الدخل والإنفاق على الرعاية الصحية في مجتمع ما من فترة 1985-2005 .

حيث تمثل (Y_i) الإنفاق ، (X_i) الدخل بوحدات نقدية معينة، ويوضح الجدول (2-3) أدناه هذين المتغيرين لعينة تضم 20 مشاهدة، وكانت حسابات كل من :

$$\sum X_i = 395.27; \ \sum X_i^2 = 9440.3 ; \ \sum X_i = 107.81,$$

$$\sum Y_i^2 = 945.5 \ ; \sum X_i Y_i = 2356.41; \ n = 20$$

أثبت أن معادلة انحدار الإنفاق على الدخل هي من الشكل:

$$\hat{Y}_i = 2.89 + 2.126X$$

وباستخدام هذه المعادلة المطلوب تقدير نقطة لكل من :

أ) الميل الحدي للإنفاق الصحي على الدخل (MPH) .

ب) المرونـــة الإنفاقيـــة (Elasticity of Health Expenditures) (EH) بالنسبة للدخل عندما يكون الدخل 20 ليرة .

الجدول(2-3): يوضح الدخل والإنفاق على الرعاية الصحية

	(Yi)الإنفاق	Xi (الدخل)
1	5.83	29.45
2	2.73	20.12
3	2.92	15.30
4	4.93	9.42
5	4.02	13.43
6	5.42	11.47
7	5.96	14.87
8	6.12	16.90
9	6.27	40.40
10	5.61	14.33
11	5.86	24.30
12	6.64	18.46
13	5.02	18.14
14	6.86	30.13
15	6.47	30.32
16	6.93	17.96
17	2.42	12.32
18	10.06	37.42
19	5.57	13.11
20	2.35	7.42

المصدر : بيانات افتراضية .

12- لنفترض وجود القيم التالية عن متغيرات النموذج الخطي البسيط. حيث إن:

$$\sum(Y-\bar{Y})(X-\bar{X})=20,\sum(Y-\bar{Y})^2=5,\sum(X-\bar{X})^2=150$$
$$\sum Y=165,\sum X=700,n=50$$

أ- أوجد القيم التقديرية لكل من (α)، (β) .

ب- أوجد معامل الارتباط بين المتغيرين $(Y_i),(X_i)$.

13- من النموذج الاقتصادي أدناه :

$$Y_i=\alpha+\beta\ X_i+U_i,(i=1,2,3,....n)$$

حيث إن: (X) متغير غير عشوائي وهو السعر ، وان (U_i) يخضع للتوزيع الطبيعي بوسط حسابي مساو للصفر، وتباين ثابت، وتباين مشترك مساوي للصفر ويمثل المتغيرات الأخرى غير السعر . اشتق صيغة التباين والتغاير من $(\hat{\alpha})$،$(\hat{\beta})$.

14- لو توفرت البيانات المذكورة أدناه عن النموذج أعلاه :

$$\sum(Y-\bar{Y})(X-\bar{X})=-8480,\sum(Y-\bar{Y})^2=180$$
$$\sum(X-\bar{X})^2=1060,\sum X=1200$$

حيث تشير (Y) إلى عدد الوحدات المشتراة ، و(X) تشير إلى سعر الوحدة المشتراة خلال الأسبوع، من (12) مخزن مختلف في مدينة معينة . والمطلوب:

أ- أوجد المرونة السعرية للطلب عند السعر (100) و.ن.

ب- اختبر معنوية $(\hat{\beta})$ مع تكوين 95% حدود ثقة لقيمة (Y_o) الوسطية عندما تكون قيمة $\bar{X}$ =100 .

جـ- ما المقصود بالمرونة السعرية، عدد أنواع المرونات ، ثم وضحها بيانيا، وما هي علاقة معامل المرونة بمعامل الانحدار في النموذج الخطي البسيط ؟

15- البيانات التالية تختص بمستويات الاستثمار (k) والمبيعات (S) بآلاف الليرات السورية من شركات خمس في فترة زمنية معينة :

K	20	50	80	30	10
S	60	40	150	100	10

أ- قدر معلمات الانحدار حسب العلاقة التالية:

$$K_i = \alpha + \beta S_i + u_i$$

ب- احسب بواقي الانحدار e_i وأوجد فترة ثقة 95% للمعلمة β .

جـ- كون جدول ANOVA ومنه أوجد F, r^2, r لمعلمة β .

16- البيانات التالية تختص بكمية النقود (X_i) والدخل القومي ببلايين الليرات السورية في اقتصاد معين خلال الفترة 1995-2004.

السنة	كمية النقود	الدخل القومي
1995	7.0	55.0
1996	7.5	55.5
1997	8.0	56.0
1998	8.5	57.0
1999	8.3	57.0
2000	9.0	58.9
2001	9.3	58.3
2002	9.5	59.0
2003	10.1	59.7
2004	10.1	60.0

أ- ارسم البيانات على شكل انتشار ، ومن ثم أوجد انحدار الدخل القومي Y على كمية النقود X.

ب- ارسم خط الانحدار المقدر على شكل الانتشار ، ماذا نستنتج من الشكل البياني .

جـ- حدد معاملات الانحدار وفسر النتائج.

الفصل الرابع

الفروض الأساسية (شروط ماركوف- غاوس) وعدم التحيز

4-1- الفرضيات الأساسية Basic Ammunitions

الفرض (1): إن (U_i) عنصر الخطأ هو متغير عشوائي حقيقي قد تكون قيمته موجبة أو سالبة أو تساوي الصفر، وهناك احتمال معين لكل قيمة يمكن أن يأخذها هذا المتغير.

الفرض (2): إن متوسط قيم (U_i) في أي فترة معينة وبالنسبة لأي قيمة محددة للمتغير هو الصفر أي: (الوسط الحسابي للخطأ مساوٍ للصفر أي أن:

$$E\ (U_i) = 0$$

هذا يعني أن مجموع قيم متغير الخطأ العشوائي المختلفة، والتي قد تكون أكبر من الصفر أو أصغر منه أو تساوي الصفر بالمتوسط.

وإذا كان: $E(u_i) = \mu \neq 0$

فإن: $v = u - \mu_u$,

وكذلك: $u = v + \mu_u$

فإن:

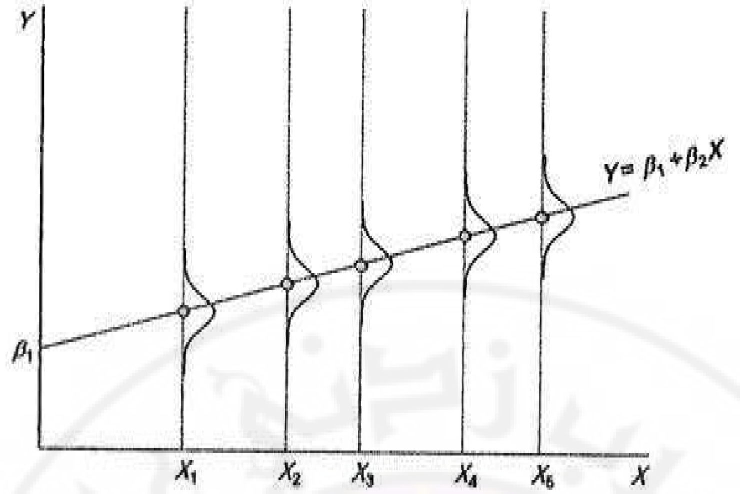
$$\begin{aligned} Y &= \beta_0 + \beta_1 X + v + \mu \\ &= (\beta_0 + \mu) + \beta_1 X + v \end{aligned}$$

$$E(v) = E(u - \mu_u) = E(u) - E(\mu) = 0$$

على سبيل المثال (هنا نبين انتظام المتغير التفسيري): على فرض علاقة الدخل بسنوات الدراسة S يحدد كأعلى درجة علمية يحصل عليها الفردHGC . وعلى فرض أن العينات الإحصائية الوطنية من أجل 1% من المجتمع لها الصفات التالية: $S = 8.3\%$ و $S = 9.5\%$ و $S = 10.7\%$ و $S = 11.43\%$ و $S = 12\ \%$ (المتخرجين من التدريب) وهكذا. وعلى فرض نريد تحديد مسح لعينة مكونة 1000 و نريد أن تكون العينة ممثلة للمجتمع إلى أبعد حد ممكن. نستطيع تقسيم العينة العشوائية إلى طبقات مكونة من 10 مشاهدات من $S = 8.30$ ومن $S = 9.5,$ وهكذا . أي حسب قيم S في العينة، والتي تم تحديدها مسبقاً بحيث تكون متجانسة، لأن التعليم والمتغيرات الديمغرافية الأخرى في المسوح الكبيرة تبدو مثل أي مجلس استشاري للمجتمع تحتوي كل الخصائص، وكذلك مثل المسح الوطني لدخل ونفقات الأسرة.

بالإضافة لشروط غاوس- ماركوف- إن الفروض الأساسية هي إخضاع الخطأ العشوائي للتوزيع الطبيعي، وهذا أمر هام لقانونية الاختبارات المعتادة.

تعني فرضية تجانس التباين بأن تباين الخطأ متساوٍ عند جميع قيم (X_i) ويبلغ (σ^2) . وهذه القيمة مجهولة، وتدل هذه الفرضية على ثبات تشتت قيم الخطأ عند قيم (X_i) المختلفة، فعلى سبيل المثال لا يمكن أن يكون التشتت كبيراً عند قيم (X_i) الكبيرة وصغيراً عند قيم (X_i) الصغيرة، وكما هو مبين في الشكل رقم (1).



الشكل رقم (1)

وبعبارة أخرى نعبر عن الفرض (6) بأن قيم (X_i) لا تتحدد عن طريق المصادفة، وإنما وفق إرادة الباحث، كما يشترط بأن تكون أرقام (X_i) مختلفة، أي أكثر من رقم واحد وإلا أصبح تباينها مساوياً للصفر. وأن قيم (X_i) المختلفة لا يمكن أن تتزايد أو تتناقص دون حدود نتيجة زيادة حجم العينة.

الفرض (7): إن عنصر الخطأ (U_i) مستقل عن المتغيرات التفسيرية أو المستقلة (X) أي أن التغاير $E(X, U) = 0$ أي ليس هناك اتجاه للتغير $E(X, U)$ بالوقت نفسه، أي أن (X) متغير عشوائي، فإن هذا الفرض يتطلب أن يكون توزيع قيم (X) مستقلاً عن توزيع قيم (U). من الناحية الحسابية نفرض أن قيم (X) هي مجموعة من القيم ثابتة أثناء أخذ عينات متكررة وهو إجراء أساسي في نماذج الانحدار الخطي. أي إذا أخذنا عدداً كبيراً من العينات بالنسبة للمتغيرات (X) و(Y)، فإن قيم (X) تبقى ثابتة في كل العينات في حين إن قيم (Y) سوف تتغير من عينة لأخرى وبالتالي تختلف قيم (U_i). على سبيل المثال نقيس الأسعار في بيانات السوق بشكل يومي $(x_1, x_2, x_3, \ldots, x_n)$ ونسجل الكميات المباعة (Y_i) في كل يوم عند هذه الأسعار. هذا يعني أن قيم

(X) بقيت ثابتة في حين قيم (Y_i) سوف تختلف حسب الأيام تبعاً لتأثيرات العشوائية. وفي ظل هذه الشروط سوف يكون تغاير القيم الثابتة (Y_i) وقيم (U_i) مساوياً الصفر. أي :

$$E([X_i - E(X_i)][U_i - E(U_i)])$$

على فرض أن :

$$Cov(XY) = \sum(X_i - \overline{X})(Y_i - \overline{Y})$$

إذاً هذه العلاقة تساوي الصفر فإنه لا يوجد علاقة تباين مشترك (تغـــاير) Covariance بين (XY) وكذلك الشيء نفسه بين (XU). يمكننا إثبات ذلـــك بعدة طرق:

1 – الطريقة أولى:

$$\begin{aligned}Cov(XU) &= E((X_i - E(X_i))(U_i - E(U_i)))\\ &= E((X_i - E(X_i))(U_i))\\ &= E((X_iU_i - E(X_i)E(U_i)))\\ &= X_iE(U_i) = 0\end{aligned}$$

علماً أن (X_i) ثابتة لأنها عينة من المجتمع.

2 – الطريقة ثانية:

انعدام التباين المشترك (التغاير) (covariance) بين المتغير العشوائي U والمتغير المستقل x ويمكن برهان هذا الافتراض على النحو الآتي:

نعوض عن (U) بقيمتها في العلاقة فنجد:

$$\sum \mathbf{XU} = \sum \mathbf{X}(\mathbf{Y} - \mathbf{b_0} - \mathbf{b_1X})$$

نصلح العلاقة فنجد:

$$\sum \mathbf{XU} = \sum \mathbf{XY} - b_0\sum X - b_1\sum X^2$$

نعوض عن (b_0) بقيمتها فنجد:

$$= \sum XY - \sum X\left(\bar{Y} - b_1\bar{X}\right) - b_1\sum X^2$$

$$= \sum XY - \bar{Y}\sum X + b_1\bar{X}\sum X - b_1\sum X^2$$

$$= \left[\sum XY - \frac{\sum Y\sum X}{n}\right] - b_1\left[\sum X^2 - \frac{\sum X\sum X}{n}\right]$$

$$= \sum xy - \frac{\sum xy}{\sum x^2}\cdot\sum x^2$$

$$\sum XU = 0$$

مع الأخذ بعين الاعتبار ما يلي:

$$\sum\left(X - \bar{X}\right)\left(Y - \bar{Y}\right) = \sum xy = \sum XY - \frac{\sum X\sum Y}{n}$$

$$\sum\left(X - \bar{X}\right)^2 = \sum x^2 = \sum X^2 - \frac{\left(\sum X\right)^2}{n}$$

$$b_1 = \frac{\sum xy}{\sum x^2}$$

نفترض انعدام التغاير بين المتغير العشوائي U والمتغير المستقل X والذي يمكن حسابه من خلال العلاقة $Cov(X,U) = \frac{\sum XU}{n}$ لأنه يؤثر في علاقة الارتباط بين المتغيرات (انظر معامل الارتباط).

4-2- الفرضيات الثانوية المساعدة :

الفرض (8): إن المتغير العشوائي U يمتص أثر المتغيرات التفسيرية التي لم تدخل في العلاقة، وربما أن هذا النوع من الخطأ يؤثر في قياس المتغير التابع(Y_i) . ولهذا سوف نفترض أن المتغيرات التفسيرية مقاسة دون أخطاء وإن المتغير التابع (Y_i) مقاس بأخطاء أو دون أخطاء أي حسبما اتفق.

الفرض (9): إن المتغيرات التجميعية الإجمالية قد تم تجميعها باستخدام أسلوب التجميع الصحيح. وغالباً ما تنطوي المتغيرات التفسيرية أو التابعة على متغيرات تجميعية. على سبيل المثال إذا استعرضنا دالة الاستهلاك من الشكل:

$$C = b_0 + b_1 Y + U_i$$

حيث نجد أن (C) هو عبارة عن مجموع نفقات المستهلكين و (Y) هو عبارة عن مجموع الدخل التصرفي عند الأفراد، هذا يعني أن المتغيرات تجميعية، ويفضل اختيار انسب الطرق في تجميعها تجنباً للوقوع في هذا النوع من الخطأ.

الفرض (10): إن المتغيرات التفسيرية ليس بينها ارتباط تام (Multicollinearity). أي إذا كان هناك أكثر من متغير مستقل يجب ألا يكون بينها ارتباط قوي؛ لنستطيع بسهولة التعرف على اثر كل منها على المتغير التابع بشكل منفصل.

الفرض (11): إن توصيف العلاقات قد تم بأسلوب سليم من حيث تحديد المتغيرات المفسرة، و اختيار الصيغة المناسبة للعلاقة الرياضية، أي : لا تحتوي على المتغيرات نفسها التي تحتويها علاقة أخرى في مجال البحث نفسه؛ لكي نكون على ثقة من أن القيم التي نحصل عليها هي معاملات فعلية للعلاقة المدروسة.

الفرض (12): العلاقة المراد تقديرها مميزة، بمعنى أن العلاقة فيها شكل إحصائي وحيد، وإنها تحتوي على المتغيرات نفسها الداخلة في أي علاقة أخرى مشابهة مرتبطة بالظاهرة محل البحث.

إن **الفرضيات (1-7)** تسمى Basic Assumptions الفرضيات الأساسية كما نطلق على معادلة الانحدار البسيط مع الفرضيات السبعة الأساسية تسمية (النموذج الكلاسيكي الطبيعي للانحدار الخطي The Classical Normal Liner Regression Model وهذا النموذج يشكل نقطة الانطلاق في تحليلات النظرية القياسية (Econometric Theory).

نوجز شرحاً مبسطاً للفرضيات الأساسية: توزع ووسط وتباين المتغير التابع (Y_i). مما تقدم يتبين من المعادلة الانحدار البسيط بأن المتغير (Y_i) عبارة

عن دالة خطية من المتغير العشوائي (U_i) وعليه فإن المتغير (Y_i) يكتسب الصفة العشوائية من المتغير (U_i) وكذلك صفة التوزيع الطبيعي وبأخذ التوقع الرياضي Expectation Mathematical لطرفي معادلة الانحدار البسيط نحصل على وسطه (Y_i) وكالآتي:

$$E(Y_i) = E(b_0 + b_1X_i + U_i) \qquad (2)$$

$$E(Y_i) = b_0 + b_1 X_i \qquad (3)$$

وذلك لأن كلاً من (b_0) و (b_1) و (x_i) عبارة عن ثوابت، وإن $E(U_i) = 0$ والآن نقوم باشتقاق صيغة تباين (Y_i)

$$Var(Y_i) = E[Y_i - E(Y_i)]^2 = E[(b_0 + b_1X_i + U_i) - (b_0 + b_1X_i)]^2 \qquad (4)$$

مما تقدم نبين بأن (Y_i) تتوزع توزيعاً طبيعياً بوسط مقداره $(b_0 + b_1X_i)$ وتباين مقداره $\sigma^2_{y_i}$ أي أن :

$$\mathbf{Y_i \sim N\left(b_0 + b_1X_i\ ;\ \sigma^2_{y_i}\right)}$$

تنحصر المسألة الأساسية في الاقتصاد القياسي في تحديد قيم (قياس قيم) معاملات الانحدار $\hat{b}_1, \hat{b}_0$ The Regression Coefficients والتي تعتبر تقديرات (Estimates) لقيم المعالم (Parameters) b_1 , b_0 في المجتمع الإحصائي المدروس.

4-3- خصائص تقديرات المربعات الصغرى العادية :

لنعتبر أن قيم Y_i , X_i هي محدودة، وتعود لمجتمع إحصائي كبير، نستطيع أن نسحب منه أعداداً كبيرة من العينات حجم كل منها n وحدة كل عينة تعطينا وسطاً حسابياً لـ Y قيمته مختلفة من عينة لأخرى . وبتعويض قيم X_I Y_I في المعادلتين الطبيعيتين (5) و (7) سوف يتولد لدينا سلسلة من القيم $\hat{b}_1, \hat{b}_0$ هي تابع إحصائي للمجتمع (b_1 , b_0). إن هذه القيم الناتجة $(\hat{b}_1, \hat{b}_0)$ تؤلف فيما بينها توزيعاً طبيعياً أو شبه طبيعي ، **وبالطبع بغض النظر فيما إذا كان المجتمع الأصلي بشكل توزيعاً طبيعياً أم لا ؟** كذلك فإن كل عينة يتم سحبها

تعطينا تبايناً معيناً تكون مجموعة التباينات الناتجة عن سحب عينات كثيرة، توزيعاً جديداً هو توزيع تباين العينات، يطلق على تلك التوزيعات للقيم $(\hat{b}_1, \hat{b}_0)$ وتباينات العينات اسم: توزيع المعاينة وهي توزيعات احتمالية، لها أهمية كبرى في الاستدلال الإحصائي، وتتصف بخصائص خاصة أهمها:

الخاصة الأولى: تتصف معلمات المقدرة بطريقة المربعات الصغرى بأنها توابع خطية بالنسبة للمشاهدات :

$$\hat{b}_1 = \frac{\sum x_i y_i}{\sum x_i^2} = \left(\frac{\sum XY - n\overline{X}\overline{Y}}{\sum X^2 - nX^2} \right)$$

يمكن إعادة كتابة الطرف الأيمن للمعادلة على الشكل التالي :

$$\hat{b}_1 = \frac{\sum \left(x_i \left(Y_i - \overline{Y} \right) \right)}{\sum x_i^2} = \left(\frac{\sum \left(X - \overline{X} \right)\left(Y - \overline{Y} \right)}{\sum \left(X - \overline{X} \right)^2} \right)$$

$$\hat{b}_1 = \frac{\sum X_i Y_i}{\sum X_i^2} - \frac{\overline{y} \sum X_i}{\sum X_i^2}$$

وبما أنه معلوم لدينا أن : $\sum x_i = 0$ ينتج لدينا :

$$= \frac{\sum x_i Y_i}{\sum x_i^2} + 0 \Rightarrow \frac{\sum x_i y_i}{\sum x_i^2} = \frac{\sum x_i Y_i}{\sum x_i^2} = \sum \left(\frac{x_i}{\sum x_i^2} \right) Y_i$$

نرمز للمقدار $\left(\frac{\sum x_i}{\sum x_i^2} \right) = \sum w_i$ نعيد كتابة المعادلة أعلاه على الشكل التالي :

$$\hat{b}_1 = \sum w_i Y_i$$

وبما أن X_i هي عبارة عن مجموع القيم المحدودة المعطاة، والتي تتكرر في عدد كبير من العينات ذات الحجم الثابت n هذا يعني أن W_i عبارة عن ثوابت لا تختلف من عينة لأخرى من العينات الافتراضية . وبالتالي فإن W_i عبارة عن أوزان الترجيح للقيم المختلفة لـ W_i أي أن :

$$\hat{b}_1 = w_1Y_1 + w_2Y_2 + ... + w_nY_n$$

أي أن تقدير المربعات الصغرى للمعلمة $\hat{b}_1$ هو دالة خطية للقيم والمشاهدات في العينة لقيم التابع .

وكذلك بالنسبة للمعلمة $\hat{b}_0$ التي تقدر من المعادلة :

$$\hat{b}_0 = \bar{y} - \hat{b}_1\bar{X}$$

حيث يمكن إعادة كتابة الطرف الأيمن على الشكل التالي :

$$\hat{b}_0 = \frac{1}{n}\sum Y_i - (\sum W_iY_i)\bar{X}$$

$$\hat{b}_0 = \sum\frac{1}{n}\sum Y_i - W_i\bar{X}Y_i\Big) = \sum\left(\frac{1}{n} - W_i\bar{X}\right)Y_i$$

وحيث إن $\bar{X}$ و W_i قيم ثابتة، هذا يعني أن قيمة $\hat{b}_0$ تتوقف على قيم المتغير التابع Y_i وبالتالي فإن $\hat{b}_0$ أيضاً دالة خطية في القيم المشاهدة في العينة لمتغير التابع (Y_i) أي $\hat{b}_0$ تقدير خطي للمعلمة b_0 .

- خواص W_i

$$1)\ \sum W_i = 0 \Rightarrow \sum\left(\frac{x_i}{\sum x_i^2}\right) = 0$$

$$2)\ \sum W_ix_i = \sum W_iX_i$$

$$\sum W_ix_i = \sum W_i(X_i - \bar{X}) = \sum W_iX_i - \bar{X}\sum W_i$$

وحيث $\sum W_i = 0 \Leftarrow$

$$\sum W_ix_i = \sum W_iX_i$$

نضرب الطرفين $\sum x_i$

$$3)\ \sum W_i = \sum\left(\frac{x_i}{\sum x_i^2}\right)$$

$$\sum W_iX_i = \frac{\sum x_i^2}{\sum x_i^2} = 1$$

$$4)\ \sum W_i^2 = \sum W_i\left(\frac{\sum x_i}{\sum x_i^2}\right) = \frac{\sum x_i}{\sum x_i^2}\cdot\frac{\sum x_i}{\sum x_i^2} = \frac{1}{\sum x_i^2}$$

الخاصية الثانية : إنها أفضل تقدير خطي غير متحيز Best Linear Unbiased Estimator (BLUE)

نسعى إلى تحقيق خصائص المربعات الصغرى للتقديرات بالبدء بمعامل الميل. الخطوة الأولى في سرعة التقدير للحدود من مركباتها الأساسية بتبديل Y بقيمتها الحقيقية في النموذج:

$$b_1 = \frac{\text{Cov}(X,Y)}{\text{Var}(X)} = \frac{\text{Cov}(X,[\beta_0 + \beta_1 X + u])}{\text{Var}(X)}$$

سوف نجري سلسلة من الإصلاحات ، على هذه العلاقة. هنا يمكننا استخدام القاعدة الأولى من التباين المشترك في تقسيم البسط:

$$b_1 = \frac{\text{Cov}(X,Y)}{\text{Var}(X)} = \frac{\text{Cov}(X,[\beta_0 + \beta_1 X + u])}{\text{Var}(X)}$$

$$= \frac{\text{Cov}(X,\beta_0) + \text{Cov}(X,\beta_1 X) + \text{Cov}(X,u)}{\text{Var}(X)}$$

الحد الأول يساوي الصفر اوجود القيمة الثابتة β_0. نستخدم العلاقة الثانية من التباين المشترك، نخرج خارج الحد الثاني لأنها قيم ثابتة β_1.

$$b_1 = \frac{0 + \beta_1 \text{Cov}(X,X) + \text{Cov}(X,u)}{\text{Var}(X)}$$

وهنا نحصلنا على القيمة الفعلية وحد الخطأ.

$$b_1 = \beta_1 + \frac{\text{Cov}(X,u)}{\text{Var}(X)}$$

من أجل تحقيق عدم التحيز، نأخذ التوقع الرياضي للطرفين، فنجد:

$$E(b_1) = E\left(\beta_1 + \frac{\mathrm{Cov}(X,u)}{\mathrm{Var}(X)}\right) = E(\beta_1) + E\left(\frac{\mathrm{Cov}(X,u)}{\mathrm{Var}(X)}\right)$$

باستخدام قواعد القيمة المتوقعة، يمكن كتابة العلاقة كما هو واضح أعلاه. مع العلم أن β_1 قيمة ثابتة توقعها يساويها أي β_1. وبالاعتماد على الفرض أن X متغير متجانس، وكذلك تباينه متجانس $\mathrm{Var}(X)$. وهكذا باستخدام القاعدة الثانية للقيمة المتوقعة نستطيع كتابة الحد الثاني كما هو مبين :

$$E(b_1) = E\left(\beta_1 + \frac{\mathrm{Cov}(X,u)}{\mathrm{Var}(X)}\right) = E(\beta_1) + E\left(\frac{\mathrm{Cov}(X,u)}{\mathrm{Var}(X)}\right)$$

$$= \beta_1 + \frac{1}{\mathrm{Var}(X)} E\big(\mathrm{Cov}(X,u)\big)$$

$$= \beta_1$$

لنثبت أن التقديرين $\hat{b}_1 , \hat{b}_0$ المقدرين بطريقة المربعات الصغرى هما تقديران غير متحيزين للمعلمتين (b_1 , b_0) ومن المعروف لدينا أن أفضل تقدير غير متحيز هو الذي تتوفر في صيغته التقديرية (Estimation Formula) الشروط التالية :

1) $\left(\hat{\hat{b}}_0\right)$ عبارة عن تشكيلة خطية (Linear Combination) من قيم مشاهدات العينة .

2) $\left(\hat{\hat{b}}_0\right)$ قيمة تقديرية غير متحيزة للمعلمة (b_0) أي أن:

$$= b_0 \left(\hat{\hat{b}}_0\right) E$$

3) تباين $\left(\hat{b}_0\right)$ أصغر تباين لأي تقدير خطي غير متحيز آخر وليكن * b_0 أي أن :

$$Var\left(\hat{b}_0\right) \le Var\left(b_0^*\right)$$

فترة الثقة ستكون أصغر مستوى دلالة. وبالتالي سيكون احتمال الحصول على نتائج جوهرية من الناحية الإحصائية كما هو مبين بالرسم البياني أعلاه نستنتج أنه كلما كان حجم العينة (n) كبيراً كلما اقتربت قيمة $\left(\left(\hat{b}_1\right)\right)$ من قيمة $\left(\left(b_1\right)\right)$.

ب- التقدير $\left(\hat{b}_0\right)$ هو تقدير غير متحيز للمعلمة b_0

من معادلة الانحدار وبالتعويض عن قيمة (Y_i) بقيمتها من العلاقة الأساسية نجد :

$$\hat{b}_0 = \sum\left(\frac{1}{n} - W_i\overline{X}\right)Y_i$$

$$\hat{b}_0 = \sum\left(\frac{1}{n} - W_i\overline{X}\right)\left(b_0 + b_1 X_i + U_i\right)$$

$$\hat{b}_0 = \sum\begin{bmatrix}\left(\frac{b_0}{n} - W_i\overline{X}b_0\right) + \left(\frac{b_1 X_i}{n} - b_1 X_i W_i \overline{X}\right) + \\ + \sum\left(\frac{1}{n} - W_i\overline{X}\right)U_i\end{bmatrix}$$

بالإصلاح نجد :

$$\hat{b}_0 = \frac{n \cdot b_0}{n} - b_0\,\overline{X}\sum X_i + b_1\frac{\sum X_i}{n} - b_1\,\overline{X}\sum W_i X_i$$
$$+ \sum\left(\frac{1}{n} - W_i\overline{X}\right)U_i$$

ومن خواص W_i معلوم لدينا أن $\sum W_i = 0$, $\sum W_i X_i = 1$

وبالتعويض نجد :

$$\hat{b}_0 = b_0 + \sum\left(\frac{1}{n} - W_i\overline{X}\right)U_i$$

نأخذ التوقع الرياضي لكلا الطرفين نجد أن :

$$E\left(\hat{b}_0\right) = E\left(b_0\right) + \sum\left(\frac{1}{n} - W_i\overline{X}\right)\mathrm{E}\left(\mathrm{U}_i\right)$$

وبما أن: $\sum(W_i) = 0$, $E(b_0) = b_0$ فإن : $E\left(\hat{b}_0\right) = b_0$

وهذا يعني أن $\hat{b}_0$ هو تقدير غير متحيز للمعلمة b_0

وباستخدام علاقة التباين المشترك سوف نستخدم القاعدة الثانية للقيمة المتوقعة في إخراج العامل المشترك خارج علاقة التوقع، و بإعادة كتابة العلاقة باستخدام القاعدة الأولى للتوقع كمجموع علاقتين لحدود فردية .

$$E(\mathrm{Cov}(X,u)) = E\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})(u_i - \overline{u})\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n}E\left((X_i - \overline{X})(u_i - \overline{u})\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})E(u_i - \overline{u}) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}) \times 0 = 0$$

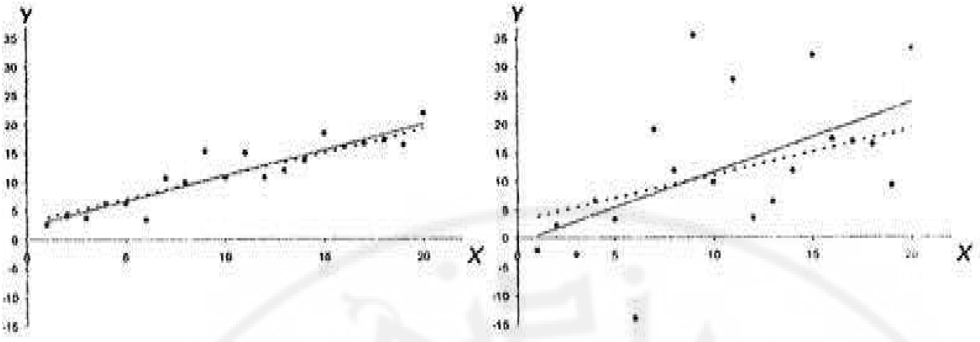
وهكذا تكون X متجانسة القيمة في جميع الحدود، وكذلك يمكننا إخـراج المتوسط كعامل مشترك خارج العلاقة.

القيمة المتوقعة لـــ u في كل مشاهدة تساوي الصـفر، وكـذلك القيمـة المتوقعة لمتوسط العينة يساوي الصفر. هكذا فإن العلاقة $\mathrm{Cov}(X,\ u)$ تسـاوي الصفر. هكذا يتبين لنا ان b_1 هي تقدير غير متحيز b_1.

$$E(b_1) = E\left(\beta_1 + \frac{\mathrm{Cov}(X,u)}{\mathrm{Var}(X)}\right) = E(\beta_1) + E\left(\frac{\mathrm{Cov}(X,u)}{\mathrm{Var}(X)}\right)$$

$$= \beta_1 + \frac{1}{\mathrm{Var}(X)}E(\mathrm{Cov}(X,u))$$

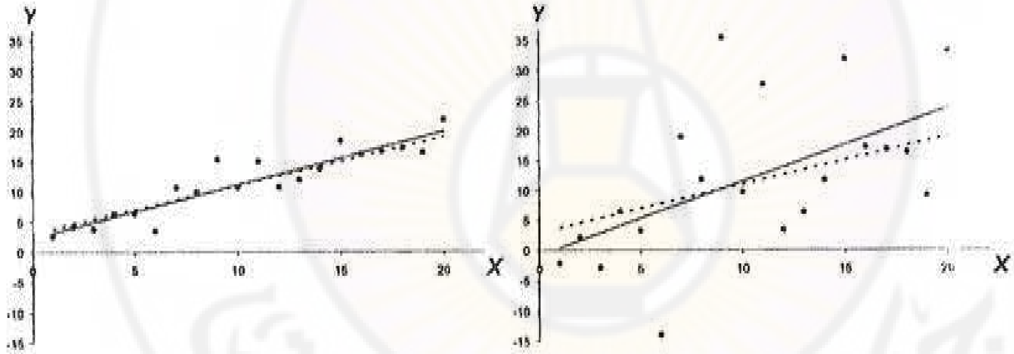
$$= \beta_1$$



$Y = 3.0 + 0.8X$

الشكل رقم (4-4)

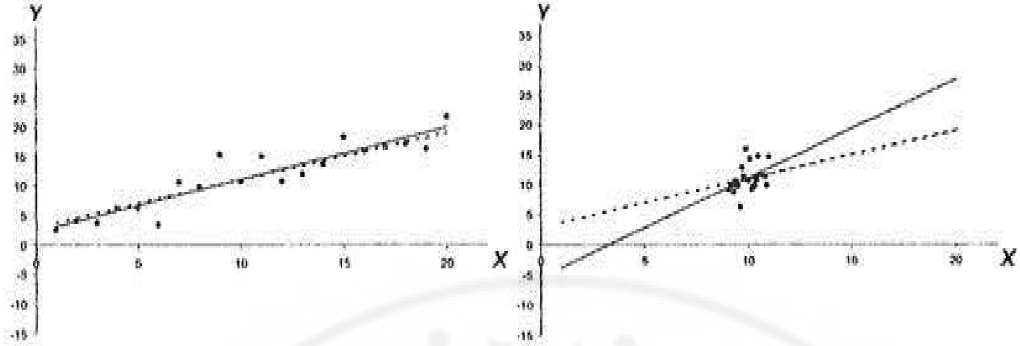
القيم X تكون نفسها والأرقام العشوائية نفسها يمكن استخدامها في توليد قيم حد الخطأ في 20 مشاهدة.



$Y = 3.0 + 0.8X$

الشكل رقم (4-5)

مع ذلك في الشكل اليميني الأرقام العشوائية العامل مضروب بـ 5 . وبالتالي فإن خط الانحدار بالخط الغامق يفتقر تقريباً لتجانس العلاقة. المعنى الثالث للعلاقة التباين هو عكس تباين X ،



$Y = 3.0 + 0.8X$

الشكل رقم (6–4)

مهما كان تباين $\mathrm{Var}(X)$ متناهياً في الصغر في الرسم اليميني لأن قيم X متقاربة جداً من بعضها بعض. يتبين من الرسم علاقة خط الانحدار أنه حساس جداً لحد الخطأ العشوائي، و بالتالي خط الانحدار لن يكون دقيقاً نسبياً.

بالطبع نستطيع أن نرى من تباين العلاقات، والتي هي نسب حقيقية لتباين X إلى تباين u والتي تعتبر هامة ، بالأحرى هي حجم مطلق أيضاً .

لا نستطيع حساب التباين بشكل دقيق؛ لأننا لا نعلم تباين حد الخطأ العشوائي، على أية حال تباين البواقي يشير إلى تباين حد الخطأ. وتباين البواقي أساسه مقدر. هذا يبين أن القيمة المتوقعة $n/(n-k)$ مرة من تباين حد الخطأ العشوائي، حيث K عدد المعلمات. وبالتالي فإن القيمة s_u^2 تعرف كما يظهر كتقدير غير متحيز: $$s_u^2 = \frac{n}{n-k}\mathbf{Var}(e) = \frac{n}{n-2}\mathbf{Var}(e)$$

في هذه الحالة لتحليل التباين البسيط فإن k تساوي 2. نستطيع أن نحصل على تقديرات الانحراف المعياري لتوزيع b_0 و b_1 بالتبديل s_u^2 من أجل s_u^2 تباين العلاقة، وأخذ الجذر التربيعي لها. للحصول على الخطأ المعياري لـ b_0 و b_1 "تقديرات الانحرافات المعيارية" صحيحة من أجل جزء من كل.

$$\text{s.e.}(b_0) = \sqrt{\frac{s_u^2}{n}\left\{1 + \frac{\overline{X}^2}{\text{Var}(X)}\right\}}$$

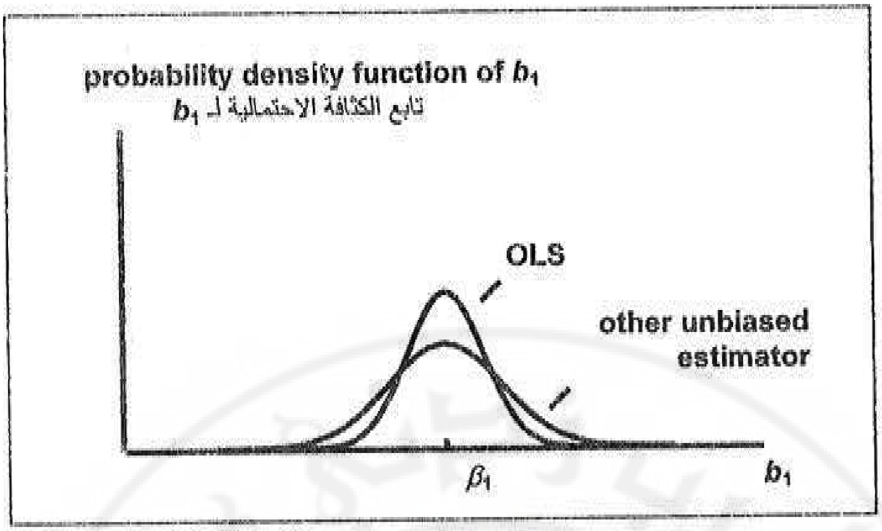
$$\text{s.e.}(b_1) = \sqrt{\frac{s_u^2}{n\text{Var}(X)}}$$

```
    Source |       SS       df       MS              Number of obs =     877
-------------+------------------------------           F(  1,   875) =    7.18
     Model |  202541871     1  202541871             Prob > F      =  0.0075
  Residual | 2.4685e+10   875  28211888.1            R-squared     =  0.0081
-------------+------------------------------           Adj R-squared =  0.0070
     Total | 2.4888e+10   876  28410894.9            Root MSE      =  5311.5
------------------------------------------------------------------------------
     incom |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  Eduction |   287.9866   107.4808     2.68   0.008     77.03627    498.9368
     _cons |   9039.462   310.6875    29.10   0.000     8429.682    9649.241
------------------------------------------------------------------------------
```

الأخطاء المعيارية للمعلمات تظهر دائماً جزءاً من مخرجات الانحدار. هنا لانحدار للأجر الذي يتقاضاه الفرد مقارنة مع المستوى الدراسي الذي وصل إليه للمثال الذي ناقشناه سابقاً. الأخطاء المعيارية تظهر في العمود .Std. Err

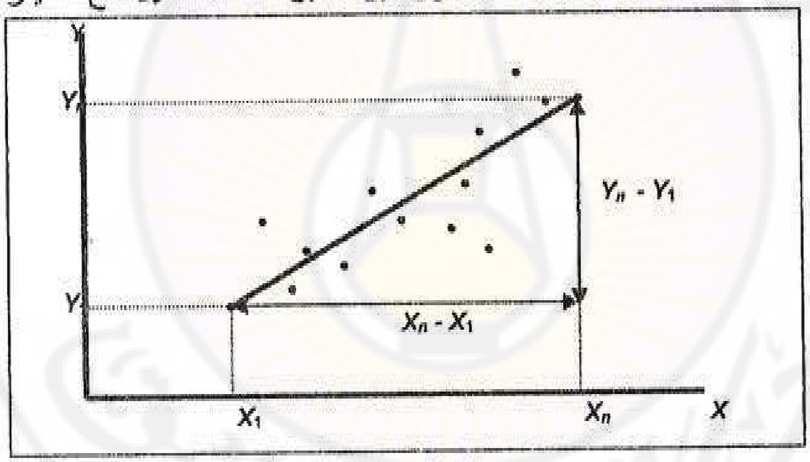
4-6- الكفاية(الفعالية) Efficiency:

إن نظرية ماركوف-غاوس تظهر أنها حددت بشكل صحيح، الشروط الصحيحة من أجل تقديرات OLS لأن لها تباينات صغيرة بالمقارنة مع أي تقديرات أخرى غير المتحيزة، و بالتالي فهي تقديرات كفوء.



الشكل رقم (4−7)

إن برهان هذه النظرية ليس بالصعب ، ولكن ليس هاماً نسبياً، وسوف نعتبره بديهياً، بدلاً عن ذلك سوف نقارن بين تقدير OLS للميل مع الخيارات.



الشكل رقم (4−8)

يفترض أننا قدرنا الميل بأخذ المشاهدة الأولى والمشاهدة الأخيرة للعينـــة، وقـــدرنا الميـــل بتقســـيم الاخـــتلاف بـــــ Y علـــى الاخـــتلاف فـــي X :

$b_1 = \dfrac{Y_n - Y_1}{X_n - X_1}$ وسوف نبحث عن خصائص هذا التقدير بالطرق المألوفة.

$$b_1 = \frac{Y_n - Y_1}{X_n - X_1} = \frac{(\beta_0 + \beta_1 X_n + u_n) - (\beta_0 + \beta_1 X_1 + u_1)}{X_n - X_1}$$

$$= \frac{\beta_1 (X_n - X_1) + (u_n - u_1)}{X_n - X_1} = \beta_1 + \frac{u_n - u_1}{X_n - X_1}$$

أولاً: نبدل قيم Y_n و Y_1 من خلال العلاقة الحقيقية. الحد β_0 يختصر، ونستطيع إعادة كتابة العلاقة لحد b_1 كما هو مبين أعلاه؛ لذلك يمكننا تحليل التقدير المثبت إلى القيم الحقيقية و حد الخطأ. و نأخذ توقع الطرفين للعلاقة ونحقق فيما بعد بالتقدير المتحيز، فنجد:

$$E(b_0) = E\left\{\beta_0 + \frac{u_n - u_1}{X_n - X_1}\right\}$$

نطبق القاعدة الأولى من قيمة المتوقعة على b_0 مقيمة ثابتة ، وتوقعها يساوي القيمة نفسها. بعد ذلك نجد X متجانسة لذلك نخرجها خارج الحد الثاني.

$$E(b_0) = E\left\{\beta_0 + \frac{u_n - u_1}{X_n - X_1}\right\} = E(\beta_0) + E\left\{\frac{u_n - u_1}{X_n - X_1}\right\}$$

$$= \beta_0 + \frac{1}{X_n - X_1} E(u_n - u_1) = \beta_0$$

ويبقى داخل قوس توقع القيمتين $E(u_1)$ و $E(u_n)$ وكلاهما يساوي الصفر. وهذا يعني ان التقدير غير متحيز.

السبب المعقول ان نرقى بالتقدير OLS إلى التقدير البسيط نسبياً؛ لأن التقدير OLS له أصغر تباين للمجتمع، وكذلك أكثر كفاءة .

$b_1 = \frac{Y_n - Y_1}{X_n - X_1}$	$\sigma_{b_1}^2 = \frac{2\sigma_u^2}{(X_n - X_1)^2}$

$b_1^{\text{OLS}} = \dfrac{\text{Cov}(X,Y)}{\text{Var}(X)}$	$\sigma^2_{b_1^{\text{OLS}}} = \dfrac{\sigma_u^2}{n\text{Var}(X)}$

نحن نعلم بهذا القدر تكون الحقيقة، بفضل نظرية ماركوف – غاوس. ويمكننا أيضاً إثبات ذلك بشكل مباشر بمقارنة تباينات المجتمع للعلاقات . كما هو مبين أعلاه . لسوء الحظ لم نتعرض لبرهان تباين المجتمع لـ b_0 يكون أكبر من تباين المجتمع لـ b_0^{OLS} لأن ذلك لا يدخل ضمن مقرر طلاب كلية الاقتصاد؛ لذلك تجاوزناه.

4-7- أهمية تقدير تباين الخطأ العشوائي :

تبين لنا من الفقرات السابقة أنه لأجل تقدير قيمة التباين للمعلمتين $(\hat{b}_1, \hat{b}_0)$ لابد لنا من تقدير σ_u^2 أولاً ، وبما أن (U_i) متغير غير مشاهد وبالتالي يصعب حساب تباينه؛ إلا أنه من الممكن تقدير قيمة له باستخدام البواقي ذلك في تقدير تباين (U) ويكون هذا على الشكل التالي:

$$\hat{Y}_i = \hat{b}_0 + \hat{b}_1 X_i$$

نأخذ المجموع لطرفي العلاقة ونقسم على (n)

$$\frac{\sum \hat{y}_i}{n} = \hat{b}_0 + \hat{b}_1 \frac{\sum X_i}{n}$$

$$\bar{\hat{y}} = \hat{b}_0 + \hat{b}_1 \bar{X}$$

بطرح العلاقتين أعلاه من بعضها بعض نحصل على :

$$\left(\hat{Y}_i - \bar{\hat{Y}}\right) = \hat{b}_1 \left(X_i - \bar{X}\right)$$

$$\hat{Y} = \hat{b}_1 X_i$$

والبواقي تكتب على الشكل التالي :

$$e_i = y_i - \hat{y}_i$$

بإضافة $\bar{Y}$ إلى الطرف الأيمن وطرحها لا يغير شيء بالنسبة للمعادلة فنجد :

$$e_i = \left(Y_i - \bar{Y}\right) - \left(\hat{Y}_i - \bar{Y}\right)$$

وكذلك يمكننا أن نكتب : $\hat{b}_0 = \bar{Y} - \hat{b}_1 \bar{X}$

نعوض بقيمة b_1

$$\bar{\hat{Y}} = \bar{Y} - \hat{b}_1 \bar{X} + \hat{b}_1 \bar{X} \Rightarrow \bar{\hat{Y}} = \bar{Y}$$

وهذا شيء منطقي لأن الوسط الحسابي للتقديرات يساوي الوسط الحسابي للقيم الأصلية.

$$e_i = (y_i - \bar{y}) - (\hat{y}_i - \bar{\hat{y}})$$

يمكن كتابتها على الشكل التالي :

$$e_i = y_i - \hat{y}_i \qquad\qquad e_i = y_i - \hat{b}_1 X_i$$

لنثبت أن تطوير التباين σ_u^2 المحسوب بالصيغة التالية :

$$S = \hat{\sigma}_u^2 = \frac{\sum e_i^2}{n-2}$$

هو تقدير غير متحيز للمعلمة σ_u^2

المراد إثباته هو :

$$E(\hat{\sigma}_u^2) = \sigma_u^2$$

فيما لو أخذنا عدداً كبيراً من العينات حجم كل منها (n) وقدرنا خط الانحدار، وحسبنا البواقي المرتبطة بكل خط، ثم حسبنا تباين البواقي ، يتولد لدينا عدد من التقديرات لتباين عنصر الخطأ العشوائي تتوزع حول الوسط الحسابي هو المعلمة σ_u^2 ونفترض بأن لها التوزيع الطبيعي نأخذ توقع المعادلة:

$$E(S) = E(\hat{\sigma}_u^2) \quad , E\left(\frac{\sum e_i^2}{n-2}\right)$$

$$= \frac{1}{n-2} E(e_i^2)$$

نحسب القيمة $\sum e_i^2$

من العلاقة الطبيعية $Y_i = b_0 + b_1 X_i + U_i$

نأخذ مجموع الطرفين ونقسم على n

$$\frac{\sum Y_i}{n} = \frac{nb_0}{n} + b_1 \frac{\sum X_i}{n} + \frac{\sum U_i}{n}$$

$$\overline{Y} = b_1 + b_2 \overline{X} + \overline{U}$$

وبطرح المعادلتين من بعضهما بعض نجد :

$$Y_i - \overline{Y} = b_0 + b_1 X_i + U_i - \left(b_0 + b_1 \overline{X} + \overline{U}\right)$$

$$\left(Y_i - \overline{Y}\right) = b_0 \left(X_i - \overline{X}\right) + \left(U_i - \overline{U}\right)$$

$$y_i = b_1 x_i + \left(U_i - \overline{U}\right)$$

و بما أن : $e_i = y_i - \hat{y}_i$

بالتعويض نحصل على :

$$e_i = b_1 X_i + \left(U_i - \overline{U}\right) - \hat{b}_1 X_i$$

$$e_i = \left(U - \overline{U}\right) + X_i \left(b_0 - \hat{b}_0\right)$$

$$e_i = \left(U - \overline{U}\right) - X_i \left(\hat{b}_1 - b_1\right)$$

$$e_i^2 = \left(U - \overline{U}\right)^2 - 2\left(U - \overline{U}\right) X_i \left(\hat{b}_1 - b_1\right) + X_i^2 \left(\hat{b}_1 - b_1\right)^2$$

نأخذ مجموع الطرفين فنجد :

$$\sum e_i^2 = \sum \left(U - \overline{U}\right)^2 + \left(\hat{b}_1 - b_1\right)^2 \sum X_i^2 - 2\left(\hat{b}_1 - b_1\right) \sum \left(U - \overline{U}\right) X_i$$

وبأخذ التوقع الرياضي نجد :

$$E\left(\sum e_i^2\right) = E\left[\sum \left(U - \overline{U}\right)^2\right] + E\left[\left(\hat{b}_1 - b_1\right)^2 \sum X_i^2\right] - $$
$$-2\,E\left[\left(\hat{b}_1 - b_1\right) \sum \left(U - \overline{U}\right) x_i\right]$$

لنحسب الحد الأول من الطرف الأيمن :

$$E\left[\sum\left(\mathrm{U}-\overline{\mathrm{U}}\right)^{2}\right]=E\left[\sum\left(\mathrm{U}^{2}-2U\overline{U}+U^{2}\right)\right]$$

$$=\mathrm{E}\left(\sum \mathrm{U}^{2}-2\overline{U}\sum U+n\overline{U}^{2}\right)$$

$$=\mathrm{E}\left(\sum U^{2}+\frac{n\left(\sum UU\right)^{2}}{n^{2}}-2\frac{\sum U_{i}\sum U_{i}}{n}\right)$$

$$=E\left(\sum U_{i}^{2}-\frac{\left(\sum U_{i}\right)^{2}}{n}\right)=E\left(\sum \mathrm{U}_{i}^{2}\right)-\frac{1}{n}E\left(\sum \mathrm{U}_{i}\right)^{2}$$

$$E\left(\sum \mathrm{U}_{i}^{2}\right)=E\left(U_{1}^{2}+U_{2}^{2}+\ldots\ldots+U_{n}^{2}\right)=n\sigma_{u}^{2}$$

$$E\left(\sum \mathrm{U}_{i}\right)^{2}=E\left(\mathrm{U}_{1}+U_{2}+\ldots\ldots+U_{n}\right)^{2}$$

$$=\mathrm{E}\begin{pmatrix}\mathrm{U}_{1}^{2}+U_{2}^{2}+\ldots\ldots+U_{n}^{2}+2U_{1}U_{2}+\\+\ldots\ldots+2U_{2}U_{3}+\ldots\ldots+U_{n-1}U_{n}\end{pmatrix}$$

$$=\mathrm{E}\,\mathrm{U}_{1}^{2}+EU_{2}^{2}+\ldots\ldots+EU_{n}^{2}+2EU_{1}U_{2}\ldots\ldots=$$

$$+2E\left(U_{n-1}U_{n}\right)$$

$$=\sigma_{u_{1}}^{2}+\sigma_{u_{2}}^{2}+\ldots\ldots+\sigma_{u_{n}}^{2}+0\ldots\ldots+\ldots0\ldots\ldots+\;\;0)$$

$$=n\,\sigma_{u}^{2}$$

بالتعويض نحصل على :

$$E\left(\sum\left(\mathrm{U}-\overline{U}\right)^{2}\right)=n\sigma_{u}^{2}-\frac{1}{n}n\sigma_{u}^{2}=(n-1)\sigma_{u}^{2}$$

لنحسب الحد الثاني من الطرف الأيمن :

$$E\left(\left(\hat{b}_{1}-b_{1}\right)^{2}\sum \mathrm{X}_{i}^{2}\right)=\sum \mathrm{X}_{i}^{2}\,\mathrm{E}\left(\hat{b}_{1}-b_{1}\right)=\sum X_{i}^{2}\cdot\frac{\sigma_{u}^{2}}{\sum \mathrm{X}_{i}^{2}}=\sigma_{u}^{2}$$

لنحسب الحد الثالث من الطرف الأيمن

$$2E\left(\left(\hat{b}_{1}-b_{1}\right)\sum\left(\mathrm{U}_{i}-\overline{U}\right)\mathrm{x}_{i}\right)=2\mathrm{E}\left(\hat{b}_{1}-b_{1}\right)\left[\sum \mathrm{U}_{i}x_{i}-\overline{U}\sum \mathrm{x}_{i}\right]$$

بما أن :
$$\hat{b}_2 - b_2 = \sum W_i U_i$$

$$W_i = \frac{x_i}{\sum x_i^2} \quad و \qquad \sum x_i = 0 \quad و$$

$$2\,E\,(\sum W_i\, U_i)\,(\sum U_i\, X_i) =$$

$$2E \sum W_i\, U_i \sum U_i x_i =$$

$$= 2E\left(\frac{x_i}{\sum x_i^2}\right) U_i \sum U_i x_i = 2E\left(\sum \frac{(x_i U_i)^2}{\sum x_i^2}\right)$$

$$= \frac{2}{\sum x_i^2} E\left(\sum x_i U_i\right)^2 = \frac{2}{\sum x_i^2} E\left(x_1 U_1 + x_2 U_2 + \ldots + x_n U_n\right)^2$$

$$= \frac{2}{\sum x_i^2} E\begin{pmatrix} x_1^2 U_1^2 + x_2^2 U_2^2 + \ldots + x_n^2 U_n^2 + 2x_1 U_1 x_2 U_2 - \ldots + \\ + 2x_{n-1} U_{n-1} x_n U_n \end{pmatrix}$$

$$= \frac{2}{\sum x_i^2} \sigma_u^2 \sum x_i^2 = 2\sigma_u^2$$

$$2E\left[\left(\hat{b}_0 - b_0\right)\left(\sum U_i x_i - \bar{U} \sum x_i\right)\right] = 2\sigma_u^2$$

الآن نعوض قيم الحدود المحسوبة فنحصل على:

$$E\left(e_i^2\right) = \sigma_u^2 (n-1) + \sigma_u^2 - 2\sigma_u^2 = (n-1)\sigma_u^2 - \sigma_u^2 = (n-2)\sigma_u^2$$

ومنه نحصل على :

$$E\left(\hat{\sigma}_u^2\right) = \frac{1}{n-2}(n-2)\sigma_u^2 = \sigma_u^2$$

$$E\left(\hat{\sigma}_u^2\right) = \sigma_u^2$$

أي أن $\hat{\sigma}_u^2$ تقدير غير متحيز للتباين σ_u^2 و $\hat{\sigma}_u^2 = \frac{E(e_i)^2}{n-2}$

تمارين غير محلولة

1- إذا علمت أن لدينا المعلومات التالية:

$$\mathrm{Cov}(S,Y)=10.5\ ,\ \mathrm{Var}(S)=5.6,\quad \mathrm{Var}(Y)=30.7$$

المطلوب : 1- احسب عند مستوى معنوية 95% فترات الثقة للمعلمة في علاقة أسعار التضخم على الأجور:

$$\underset{}{\hat{P}} = \underset{(0.05)}{-1.21} + \underset{(0.10)}{0.82W}$$

2- إذا علمت أن حجم العينة هو (n=20) في السؤال السابق. ماذا تستنتج من هذا الحساب؟

3- I- أراد أحد الباحثين تحديد الطلب على النقل العام في إحدى المدن وفق النموذج: $E = b_1 + b_2W + b_3NW + u$

وجمع بيانات عن 100 فرد بشكل متعاقب خلال عام من الإنفاق على النقل العام E مقاس بالليرة السورية، وعدد أيام العمل W ، وعدد أيام العطل NW، وبالتحديد يساوي إلى: W-365 .

والمطلوب: I- اشرح لماذا هو غير قادر على توفيق معادلة الانحدار (أعط بديهيتين وتوضيح تقني لذلك) . كيف يمكن أن يحل هذه المشكلة؟

II- الباحث قرر تطبيق المثال السابق، ولكن عمل على تقسيم أيام العطل إلى قسمين : عطلة بسبب المرض أعطيت الرمز I وعطلة لأسباب أخرى أعطيت الرمز O. متوسط القيم ذات الرمز I في العينة يساوي 2.1 ومتوسط قيم O يساوي 120.2 . وضع نموذج الانحدار(تم وضع الأخطاء المعيارية بين قوسين تحت كل معلمة من المعلمات).

$$0.45O\ 2.10W + -9.6 + \ \hat{E} =$$
$$(0.25)\quad (1.06)\quad (8.3)$$

إذا علمت أن : $R^2 = 0.72$ وقيم t الحرجة من أجل مستوى دلالة 0.1 % و 0.05 على التوالي: $t = 1.77$ $t = 1.98$ والمطلوب :

4- اختبر المعنوية الإحصائية لهذا النموذج.

5- اختبر النموذج ككل، إذا علمت أن قيمة F الجدولية F(2.60) من أجل مستوى دلالة 0.1% يساوي 7.8 .

6- عدد الفرضيات الأساسية Basic Ammunitions.

7- تحدث عن الفرضيات الثانوية المساعدة .

8- تحدث عن خصائص تقديرات المربعات الصغرى العادية .

9- ماذا نعني بالكفاية(الفعالية) Efficiency؟

الفصل الخامس

الاختبارات الإحصائية لمعنوية تقديرات المربعات الصغرى (اختبارات الدرجة الأولى)

استطعنا توليف معادلة الانحدار في الفصول السابق من خلال تقدير معاملات العلاقة بطريقة المربعات الصغرى، وتعرفنا على خصائص هذه التقديرات. وبعد ذلك لابد من وضع معايير لقياس جودة هذه التقديرات. لقد صنفنا في الفصل الأول المعايير التي تحكم جودة النموذج إلى ثلاث مجموعات، معايير النظرية الاقتصادية، ومعايير النظرية الإحصائية، ومعايير نظرية الاقتصاد القياسي. و قد أشرنا إلى أن معايير النظرية الاقتصادية تتحدد من خلال مطابقة النموذج في إشاراته و قيم معاملاته للنظرية الاقتصادية، وهذه نستنتج من خلال التعامل مع البيانات الإحصائية في التوليفة الأولى للنموذج عند الصياغة. بينما المعايير الإحصائية للحكم على النموذج سوف نستعرضها في هذا الفصل. ونترك معايير الاقتصاد القياسي للاختبارات من الدرجة الثانية، والتي سوف نأتي عليها لاحقاً.

5-1- اختبارات الفروض الإحصائية:

1- اختبار الفروض Testing hypotheses

الاقتصاد القياسي هو شكل من أشكال البحث العلمي، ويشكل أساس أي بحث علمي يبدأ بعد أن يحدد الباحث المشكلة الأساسية لهذا البحث. و بعدئذ عليه وضع فروض بحثه، فالباحث يلاحظ ظاهرة ما ومن ثم يتوقع مسببات ونتائج لهذه الظاهرة، لذلك معروف عند الباحثين أن فروض البحث العلمي هي

في الحقيقة توقعات ورهان، وعلى الباحث العلمي أن يثبت صحة تلك التوقعات، إما يكسب الرهان وإما يخسره، ويعيد الكرة مرة أخرى متقيداً بأصول البحث العلمي .

صياغة الفروض هي من قواعد البحث العلمي التي تهدف إلى تقليل الأخطاء (Error)[1]. إذاً على الباحث أياً كان أن يبدأ بحثه بالافتراضات العامة والواسعة ومن ثم ينتقل إلى جزئيات الفرضية البحثية، ومن أجل صياغة الفروض الأساسية للبحث العلمي، يجب أن تتوفر الصفات التالية:

1- يفضل وضع الفرضيات البحثية على شكل جمل استفهامية للتنبيه إلى المسألة المبحوثة.

2- يفضل فصل علاقة السبب والنتيجة بين المتغيرات التي تربط فروض البحث، أي ربط فروض البحث بين متغيرين أو أكثر.

3- يجب أن تحتوي فروض البحث على مفهوم ضمني مؤداه إمكانية قياس المتغيرات، وإمكانية إجراء الاختبارات الإحصائية على العلاقات قيد البحث.

4- يفضل أن تكون الفروض محددة من أجل الإقلال قدر الإمكان من الفروقات الجوهرية في البحث نتيجة قوى الحظ والمصادفة.

لا بد من الإشارة إلى ملاحظة هامة ألا وهي على الباحث العلمي التمييز بين الفرضيات الإحصائية و فروض البحث العلمي، وهي عادة فروض عامة تكون مبنية على نظرية علمية أو نتائج أو دراسات وأبحاث سابقة أو على أسس تطبيقية. بينما الفروض الإحصائية هي تعبير عن واحد أو أكثر من معالم المجتمع الإحصائية التي سحبت منها عينة الدراسة.

---

[1] الاقتصاد القياسي التطبيقي. د/عبدالرزاق الشرجي(72-73).

تستخدم الفروض الإحصائية لاختبار المعنوية الإحصائية لفروض البحث الموضوعة. والفروض الإحصائية ما هي إلا تعبير عن واحد أو أكثر من معالم المجتمع المدروس. قبل الدخول في تفصيل اختبار الفروض الإحصائية، لا بد لنا من استعراض بعض من التعاريف الأساسية، وهي:

– **تعريف الاختبار** : الاختبار هو وسيلة إحصائية تسمح لنا بأن نختار بين فرضين : H_1 , H_0 بالاعتماد على عينة ذات حجم معين مسحوبة من المجتمع الإحصائي المدروس. والمبدأ العام للاختبارات مهما تعددت أنواعه يبقى نفسه إذ أنه يتعلق بتحديد معنوية التقدير لمؤشر ما.

– **الفرضية الإحصائية:** هي تحديد قيمة **الثابت الإحصائي**، أي: تعيين خاصة من خواص توزيع المتغير العشوائي حيث يتميز كل توزيع بخواص إحصائية معينة مثل القيمة الأكبر احتمالاً، والوسط الحسابي، الانحراف المعياري، ويمكن تحديد القيم الحقيقية لهذه الخواص بالاعتماد على المعلومات المتوفرة عن جميع وحدات المجتمع. تسمى كل قيمة من هذه القيم الحقيقية الثابتة (الثابت الإحصائي للمجتمع) وتتحدد الفرضية الإحصائية التي نريد اختبارها بتحديد قيمة هذا الثابت.

– **التابع الإحصائي** : يتميز توزيع كل عينة شأنه في ذلك شأن توزيع المجتمع الإحصائي بخواص إحصائية معينة مثل القيمة الأكثر احتمالاً، والوسط الحسابي، والانحراف المعياري، وتتخذ قيم هذه الخواص بالاستناد إلى المعطيات المشاهدة لمجموع مفردات العينة. تسمى كل قيمة من هذه القيم المشاهدة (التابع الإحصائي) ولا تكون هذه القيم ثابتة، وإنما تختلف من عينة لأخرى.

– **أخطاء المعاينة** : إذا وجدت فروق واختلافات بين النتائج التي حصلنا عليها من العينة، والنتائج التي يمكن الحصول عليها من جراء تعداد شامل للمجتمع الذي سحبت منه العينة تسمى عندئذٍ هذه الفروق والاختلافات و (أخطاء

المتباينة) وهي متغيرات عشوائية عبرنا عنها حين دراسة النماذج القياسية بـ u وأسميناها بالأخطاء العشوائية للمعادلة المدروسة.

فرضية العدم **The null hypothesis** ونرمز لها بالرمز (H_0) والتي تنص على عدم وجود فرق جوهري بين التابع الإحصائي $\hat{\beta}$ والثابت الإحصائي β وإن الفرق المشاهد مرده إلى قوى (أخطا) الحظ والمصادفة. يتم اختبار الفرضيات باستخدام فرضية الاستقلال (t ستونت، كاي مربع χ^2، وفيشر F).

الفرضية البديلة: نرمز لها H_1 (وهي عبارة عن الحل المقابل لفرضية لعدم).

أي: نكتب العلاقة رياضياً على الشكل التالي:

$$H_0 \;:\; E\left(\hat{\beta}\right)=\beta$$

$$H_1 \;:\; E\left(\hat{\beta}\right)\neq\beta$$

اختيار درجة الثقة وتحديد مناطق القبول والرفض:

يهدف تقدير التابع الإحصائي إلى إيجاد عدد ما يعد (أفضل) تقدير لقيمة الثابت المجهولة.

وعلى الرغم من ذلك قد لا ينطبق هذا التقدير على قيمة الثابت الإحصائي إذ كما أشرنا سابقاً، المجتمع الإحصائي مؤلف من k وحدة يمكن اختيار عدد لانهائي من العينات، التي تتكون من n وحدة وتقدير قيمة الثابت الإحصائي لكل من هذه العينات. يمكن اعتماد كل قيمة من هذه القيم كتقدير للثابت الإحصائي، أي: أن هنالك عدد لا متناه من التقديرات نستدل منها على قيمة واحدة هي قيمة الثابت الإحصائي. ونشر هذه التقديرات حول الثابت الإحصائي بأبعاد مختلفة، لو أخذنا عينة واحدة فقط مؤلفة من n وحدة وحددنا قيمة التابع الإحصائي، واعتمدناه كتقدير للثابت الإحصائي عندئذٍ، ما هو احتمال أن يكون هذا التقدير صحيحاً. لابد إذاً من اللجوء إلى استعمال طريقة المدى، حيث نحدد حداً أدنى وحداً أعلى يمكن أن يكون صالحاً لوقوع القيمة المقدرة ضمنه، بافتراض أن

العينة تتوزع توزيعاً طبيعياً. ويطلق على الحدين الأعلى والأدنى اسم حدي الثقة، وعلى احتمال وقوع التقدير ضمن هذين الحدين اسم درجة الثقة.

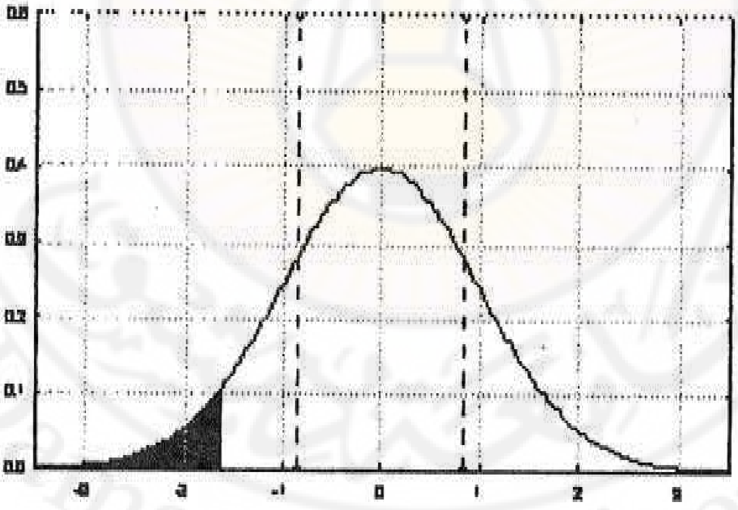
فلو افترضنا جدلاً أننا حددنا درجة الثقة بـ 95% والتي يقابلها مستوى دلالة (احتمال) A =5 % فهذا يعني أن مجال الخطأ المقبول في التقديرات هو 5 % فقط.

والاحتمال هذا يقسم المساحة الواقعة تحت المنحنى إلى منطقتين تتوافقان والفرضيتين H_0 , H_1 الذي سيتم الاختيار بينهما.

هاتان المنطقتان تسمحان لنا باتخاذ القرارات التالية :

1- إذا كان التقدير يقع في المنطقة التي تحقق H_0 قلنا: إن هذه المنطقة هي منطقة قبول بالنسبة للفرضية H_0 ومنطقة رفض بالنسبة للفرضية H_1 .

2- إذا كان التقدير يقع في المنطقة التي تحقق H_1 قلنا: إن هذه المنطقة هي منطقة قبول بالنسبة لـ H_1 ومنطقة رفض بالنسبة لـ H_0 .



الشكل رقم (1-5)

يحدد الحد الأدنى والحد الأعلى الذي يمكن أن يقع ضمنه ثابت المجتمع الأصلي.

$$P\left[-t_{0.025} < \frac{(\hat{\beta}-\beta)\sqrt{n\sum X_i^2}}{\sigma_u^2\sqrt{\sum X_i^2}} < t_{0.025}\right] = 0.95$$

$$P\left[\hat{b} - t_{0.025}\frac{\hat{\sigma}_u\sqrt{\sum X_i^2}}{\sqrt{n\sum X_i^2}} < b < \hat{b} + t_{0.025}\frac{\sigma_u^2\sqrt{\sum X_i^2}}{\sqrt{n\sum X_i^2}}\right] = 0.95$$

من المعروف إن اختبار Z للمعنوية الإحصائية يمكن استخدامه في الحالات التي تتوفر فيها الشروط التالي :

أ) أن يكون التباين حقيقياً للمجتمع معلوماً بصرف النظر عن حجم العينة.

ب) أما إذا كان التباين الحقيقي للمجتمع غير معلوم؛ فيشترط أن يكون حجم العينة المستخدم كبيراً، أي: $N > 30$ نلاحظ من خلال إيجاد قيمة تباين مجتمع بأنه مقدر وهو يحسب بعدة طرق منها الطريقة التالية $\sigma_u^2 = \frac{\sum e_i^2}{n-k}$ وبالاعتماد على ثابت الخطأ العشوائية يتم حساب تباين $\sigma_{b_1}^2 , \sigma_{b_0}^2$ في هذه الحالة فقط إذا كانت $n > 30$ نستطيع استخدام اختبار Z ولأجل العينات الصغيرة، وعدم معرفة التباين الحقيقي، يجب استخدام اختبارات أخرى مثل اختبار (t ستودنت) أو (χ^2) و فيشر (F) .

فمثلاً اختبار t يأخذ الشكل التالي : $t = \frac{X_i - \bar{X}}{\sigma_x}$

مع العلم أن : $\sigma_x^2 = \frac{\sum (X_i - \bar{X})^2}{n-1}$

(n−1) درجات الحرية . حيث إن: n عدد المشاهدات (حجم العينة)

إن توزيع ستودنت (t) دائماً متماثل بمتوسط يساوي الصفر وتباين $\frac{n-1}{n-3}$ وهو يقترب من الواحد الصحيح كلما كبر حجم العينة، ويكون توزيع (t) على الشكل التالي :

$$t_{n-k}=\frac{\hat{\beta}_1-\beta_1}{\sigma_{\beta}^{2}};\qquad t_{n-k}=\frac{\hat{\beta}_0-\beta_0}{\sigma_{\alpha}}$$

$$P\left[-t_{n-k}-\frac{\alpha}{2}\le t_{n-k}\le +t_{n-k}\frac{\alpha}{2}\right]$$

حيث إن df = n - k و (α) مستوى المعنوية (الدلالة level significance و (1−α) أي معامل الثقة (درجة الثقة coefficient of (confidence

$$P\left[\hat{\beta}-t_{n-k},\frac{\alpha}{2}\left(\sigma_{\hat{\beta}}\right)\le\beta\le\hat{\beta}+t_{n-k},\frac{\alpha}{2}\left(\sigma_{\hat{\beta}}\right)\right]=1-\alpha$$

$$\beta=\hat{\beta}\mp t_{n-k},\frac{\alpha}{2}\left(\sigma\hat{\beta}\right)$$

واختبار F فيشر يعطى بالمعادلة التالية :

$$F=\frac{\left[n\left(\hat{\beta}-\beta\right)^2+\left(\hat{\alpha}-\alpha\right)^2\sum X_i^2+2\left(\hat{\beta}-\beta\right)\left(\hat{\alpha}-\alpha\right)\sum X_i\right]}{\hat{\sigma}_u^2}$$

نقارن من أجل عدد درجات الحرية $\frac{\alpha}{2}$ قيمة F النظرية مع قيمة F المحسوبة، ويكون الحكم كالتالي :

1- إذا كانت F المحسوبة أصغر من القيمة الجدولية لـ $F(\frac{\alpha}{2})$; نقبل فرضية العدم H_0 ونرفض الفرضية البديلة H_1 أي أن التباينين متساويان وقيم

$\hat{\alpha}, \hat{\beta}$ المقدرة قريبة جداً من قيم β_0 , β_1 والأخطاء المشاهدة مرادها إلى قوى الحظ والمصادفة .

2- إذا كانت F المحسوبة أكبر من القيمة الجدولية لـ $F(\frac{\alpha}{2})$ ؛ نرفض فرضية العدم H_0 ونقبل الفرضية البديلة H_1 أي: أن قيم $\hat{\beta}_0 , \hat{\beta}_1$ المقدرة قريبة، ولكن لا تساوي قيم β_0 , β_1 والأخطاء المشاهدة مرادها إلى أن العينة مسحوبة من مجتمع إحصائي غير المجتمع الإحصائي المدروس، والفرق المشاهد هو فرق حقيقي .

نفترض أنه لدينا نموذج انحدار معياري، و نريد اختبار فرضية العدم H_0 لميل معامل الانحدار لبعض قيم $\beta_1{}^0$. نختبر فرضية العدم، ونبحث عن الفرضية البديلة H_1 والتي ببساطة تعني b_1 لا تساوي $\beta_1{}^0$. و سنقوم بوصف اختبار الفروض عند مستوى معنوية 5% و 1% .

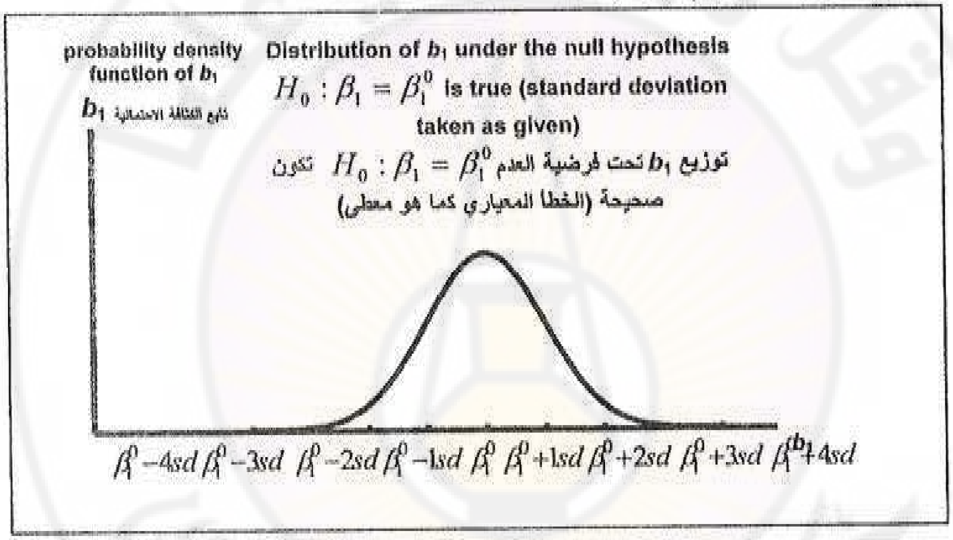
على فرض لدينا بيانات لوصف علاقة الأسعار بالأجور في حالة التضخم. حيث p معدل نمو الأسعار w معدل نمو الأجور ، كما هو مبين أدناه:

النموذج **(Model)**:	$Y = \beta_0 + \beta_1 X + u$
فرضية العدم (**Null hypothesis**):	$H_0 : \beta_1 = \beta_1{}^0$
الفرضية البديلة **(Alternative hypothesis)**:	$H_0 : \beta_1 \neq \beta_1{}^0$
مثال النموذج **(Example model)**	$p = \beta_0 + \beta_1 w + u$
فرضية العدم (**Null hypothesis**):	$H_0 : \beta_1 = 1.0$
الفرضية البديلة **(Alternative hypothesis)**:	$H_1 : \beta_1 \neq 1.0$

سوف نختبر الفرضية التي تنص على أن الأسعار في حالة التضخم تساوي معدلات الأجور في حالة التضخم. فرضية العدم هي : H_0: $b_1 = 1.0$ (وسوف نختبر أيضاً $b_1 = 0$).

إذا كانت فرضية العدم صحيحة فمعامل الانحدار b_1 له توزيع متوسطه 1.0 . لرسم التوزيع يجب معرفة الخطأ المعياري للتوزيع. سوف نفترض أيضاً معرفتنا للخطأ المعياري ويساوي 0.1 . بالطبع هذا الافتراض غير واقعي في التجارب يجب تقدير ذلك عملياً بطرق الإحصاء الرياضي المعروفة، انظر الشكل البياني رقم(3-5).

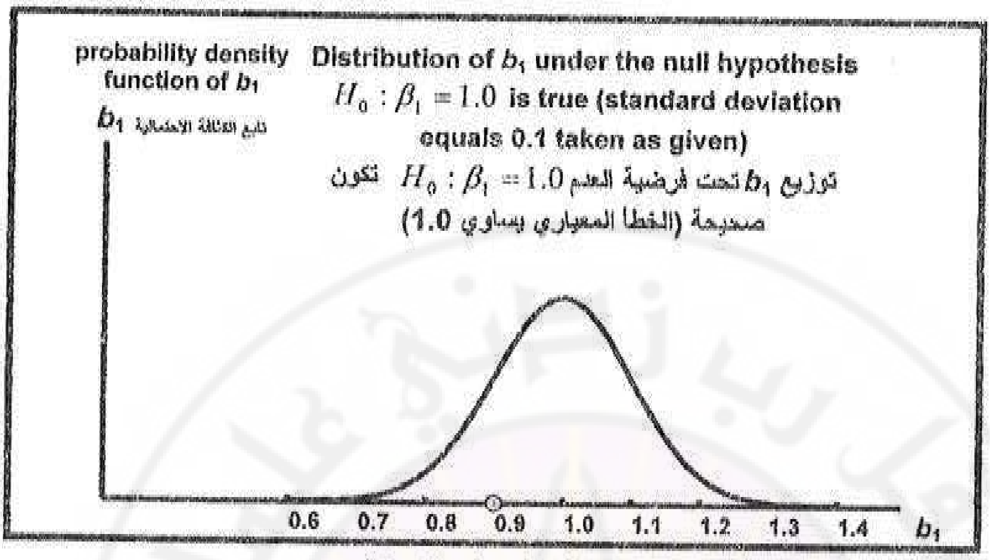
في هذا الشكل نبين التوزيع من أجل b_1 للحالة العامة ، و نفترض أننا حددنا الخطأ المعياري (sd).



الشكل رقم (3-5)

على فرض أن بيانات العينة لمثال الأسعار، والأجور في حالة التضخم للنموذج أعطت تقدير لمعامل الانحدار b_1 يساوي 0.9 . هل يقدم دليل ضد فرضية العدم $\beta_1 = 1$ ؟

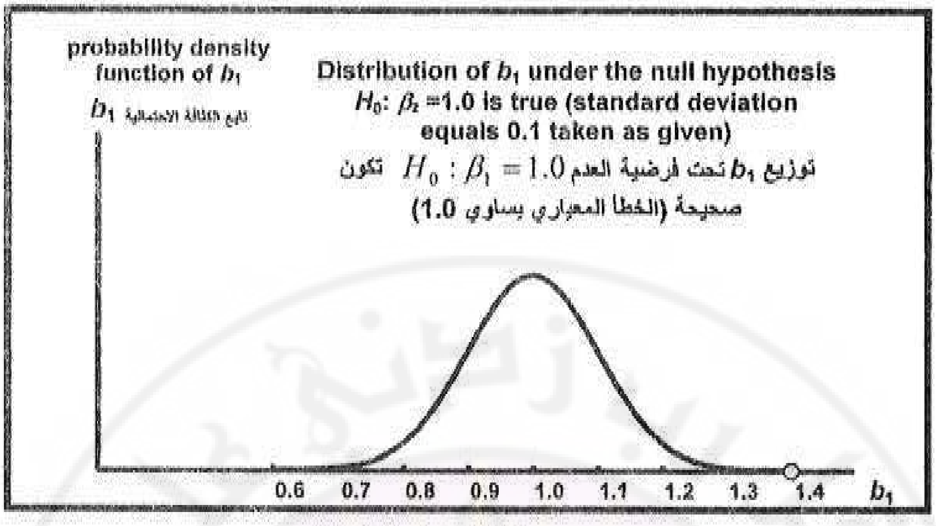
بالطبع هذه القيمة لا تساوي الواحد. بل هي أقل من الواحد ، ولكن بما أنه لا يوجد حداً للخطأ في النموذج لن نستطيع أن نجزم إن التقدير المتوقع يساوي 1 .



الشكل رقم (4-5)

إذا كانت فرضية العدم صحيحة ، سوف لن نتردد في تقديرها عند القيم 0.9 وكذلك لن يكون هناك تناقض حقيقي، في حدود الحالات العامة من أجل خطأ معياري واحد للتقدير .

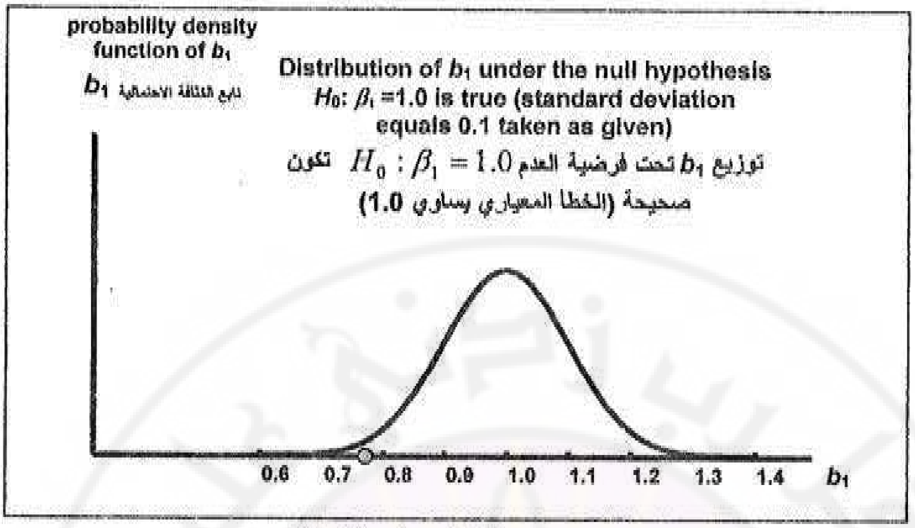
إذا كانت فرضية العدم صحيحة ، فإن احتمال الحصول على تقدير خطأ معياري واحد أكثر أو أقل من المتوسط هو 31.7% . الآن نفترض أن الأسعار في حالة التضخم و الأجور في حالة التضخم في النموذج أعطتنا تقديراً يساوي 1.4 ، وهذا يكون على تناقض واضح مع فرضية العدم.



الشكل رقم (5-5)

حيث إن 1.4 يعادل أربع أخطاء معيارية حول متوسط الفرضية، ومثل هذه الصدفة المتحققة غير مألوفة التقدير عند القيمة 0.05% ، ونرفض فرضية العدم.

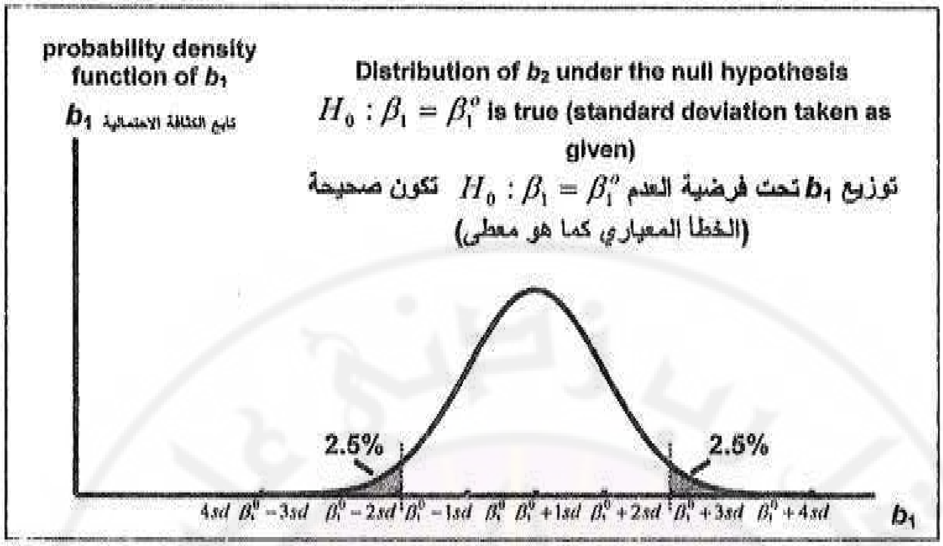
الآن نفترض من خلال أسعار التضخم وأجور التضخم في نموذج العينة قد أعطت تقديراً 0.77 ، و هذه النتيجة غير مقبولة مطلقاً؛ لأنه تحت الحد الأدنى لفرضية العدم، أي التقدير يجب أن يكون بين 2 و 3 أخطاء معيارية حول الوسط.



الشكل رقم (7-5)

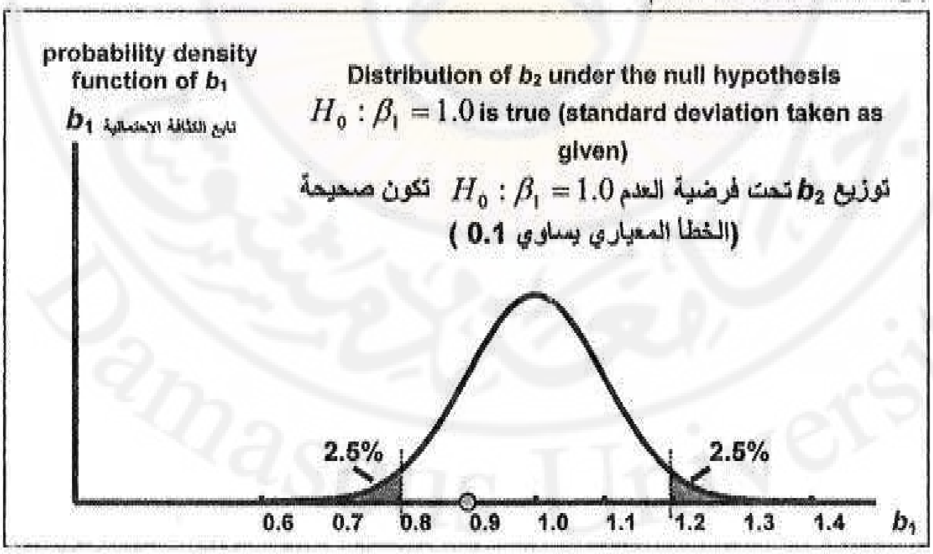
يوجد احتمالان : الأول فرضية العدم صحيحة ونتجاهل التقدير. و الثاني: فرضية العدم خاطئة أي معدل الأسعار في حالة التضخم لا يساوي معدل أجور في حالة التضخم.

الإجراء الاعتيادي أثناء المناقشة هو رفض فرضية العدم إذا دل ذلك على إمكانية الحصول على شيء غير مألوف في التقدير ذي احتمال قليل. على سبيل المثال، يمكن اختيار رفض فرضية العدم إذا تضمن احتمال التقدير غير المرغوب به أقل من (5%).



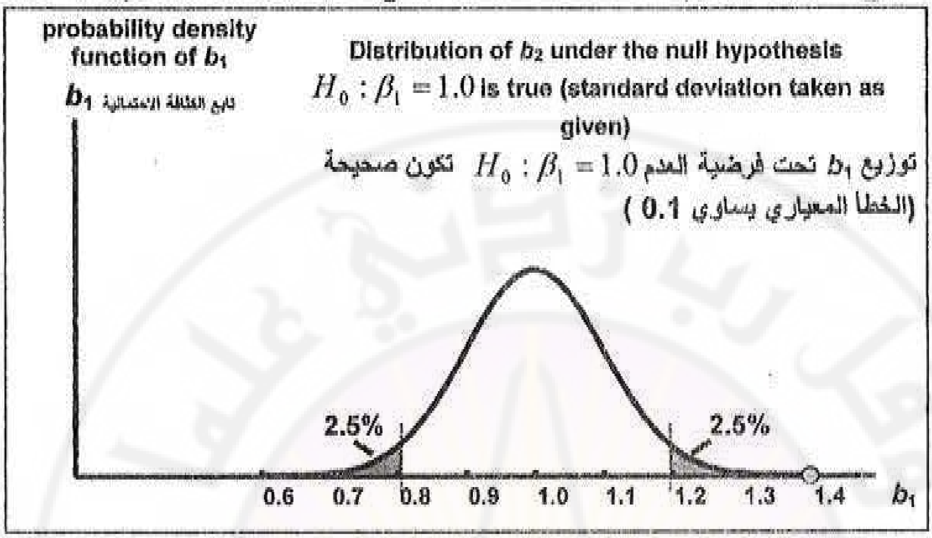
الشكل رقم (5-9)

وفقاً لهذه المناقشة المطروحة، سوف نرفض فرضية العدم إذا كان التقدير الحاصل أقل أو أكبر من القيمة 2.5% من الطرفين. و بالتطبيق على الأسعار في حالة التضخم والأجور في حالة التضخم، فالتقدير الأول b_1 سوف لن يقودنا إلى رفض فرضية العدم.



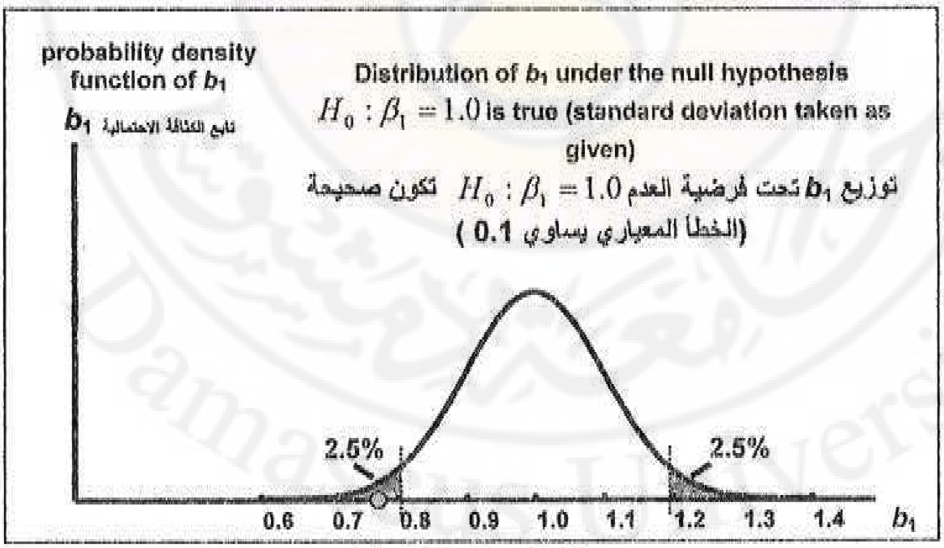
الشكل رقم (5-10)

الحالة الثانية كما هي محددة نرفض فرضية العدم لأن القيمة واقعة خارج منطقة القبول(ثلاثة أخطاء معيارية وتقع داخل منطقة الرفض).



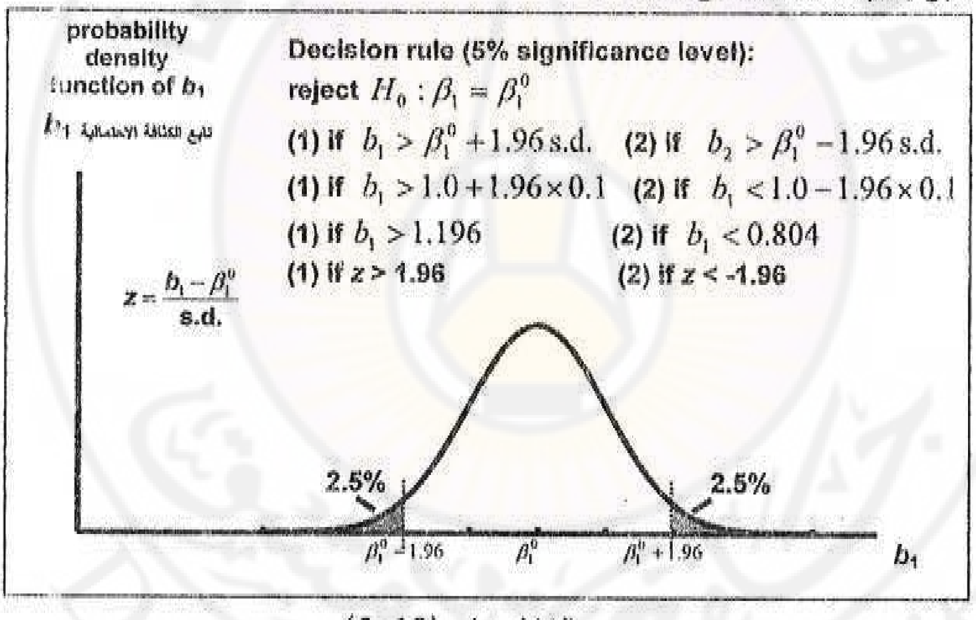
الشكل رقم (11-5)

أما الحالة الثالثة أيضاً كما هو واضح من الرسم البياني رقم(12-6) يقودنا إلى رفض الفرضية .



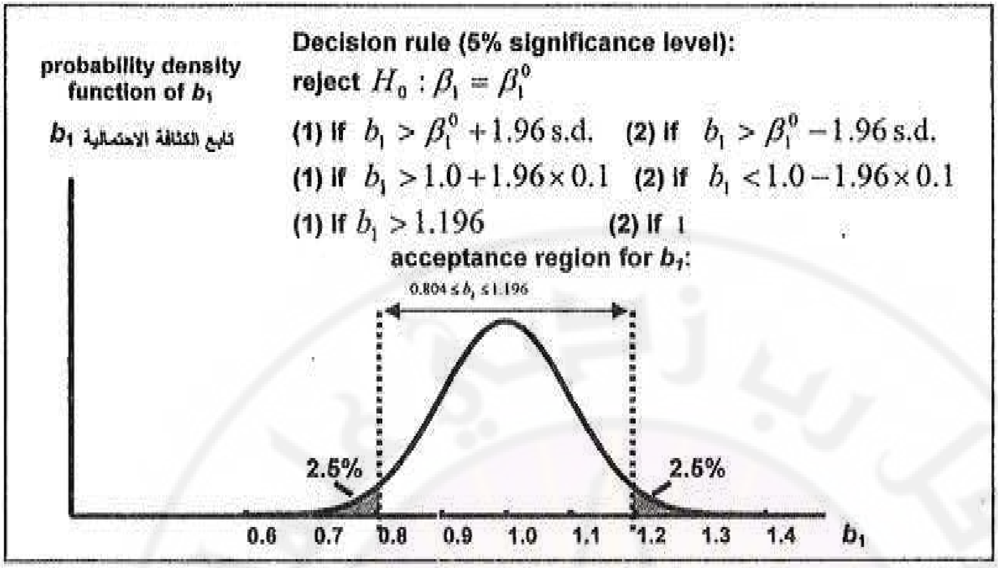
الشكل رقم (12-5)

تبدأ الـ2.5% من الجانبين للتوزيع الطبيعي من القيمة 1.96 للخطأ المعياري عن متوسطها. وهكذا سوف نرفض فرضية العدم إذا كان التقدير للخطأ المعياري 1.96 (أو أكثر) حول الوسط الافتراضي أو أقل. وسوف نرفض الفرضية H_0، إذا اختلفت بين تقدير العينة والقيمة الافتراضية كانت أكثر من 1.96 للخطأ المعياري. وكذلك نرفض فرضية العدم H_0، إذا كان الاختلاف واضح في حدود الخطأ المعياري، أو كان أقل من 1.96 بالقيمة المطلقة للحدود (سالب أو موجب). سنرمز لاختلاف الحدود الواضحة للخطأ المعياري بـ Z. و قاعدة الحكم في رفض فرضية العدم عندما قيمة z تكون أكبر بالقيمة المطلقة من 1.96.



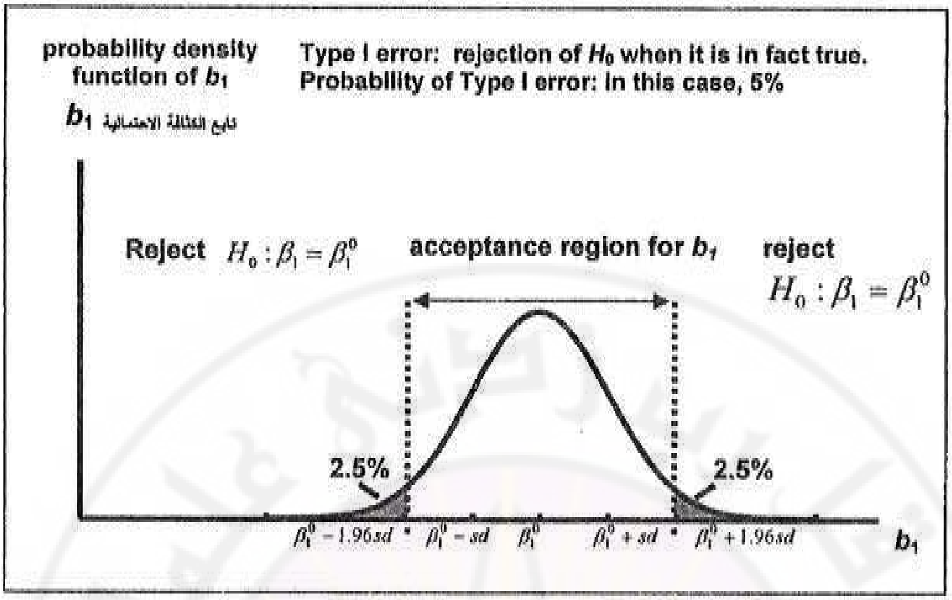
الشكل رقم (13-5)

منطقة القبول لـ b_1 هي المجال من القيمة 0.804 إلى القيمة 1.196، تقدير العينة لهذا المدى سوف لن يؤدي إلى رفض فرضية العدم.



الشكل رقم (5-14)

إن رفض فرضية العدم عندما تكون صحيحة يسمى الخطأ من النوع الأول. الاختبار الحالي، إذا كانت فرضية العدم صحيحة فإن الخطأ من النوع الأول يحدث عند 5% من المرات، أي سنحصل على تقدير أعلى من 5% من المرات أو أقل من 2.5 % من الطرفين.

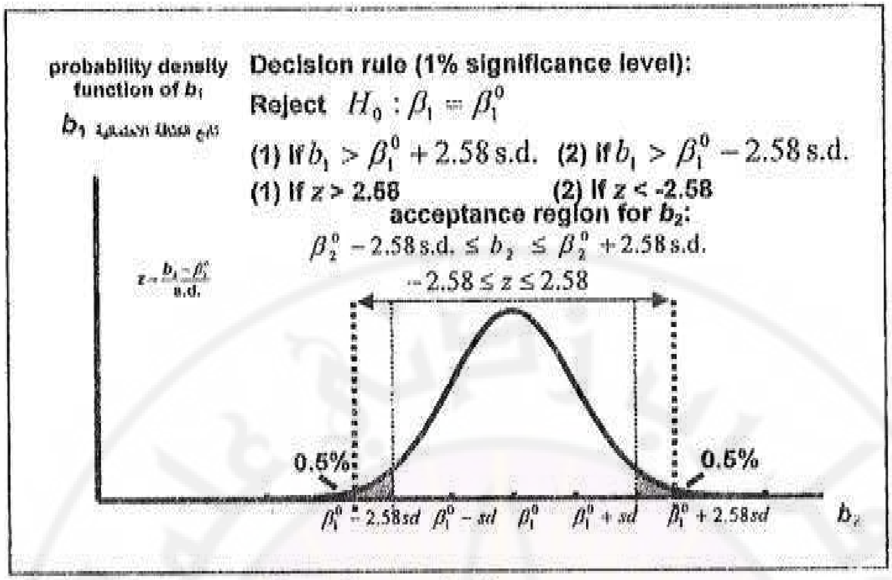


الشكل رقم (15-5)

مستوى المعنوية للاختبار يحدد احتمال الوقوع في الخطأ من النوع الأول إذا كانت فرضية العدم صحيحة. بالطبع نستطيع التقليل من خطر الوقوع في الخطأ من النوع الأول بتخفيض حجم منطقة الرفض.

على سبيل المثال: نستطيع تغير قاعدة الحكم من أجل رفض فرضية العدم إذا كان احتمال الحصول على تقدير العينة أقل من (1%). عندئذٍ منطقة الرفض أصبحت أعلى أو أقل من 0.5% من الجانبين.

يبدأ الـ0.5% من الجانبين في التوزيع الطبيعي من 2.58 خطأ معياري عن الوسط الحسابي، كذلك سوف نرفض فرضية العدم إذا كانت قيمة Z أكبر من 2.58 بالقيمة المطلقة للحدود.

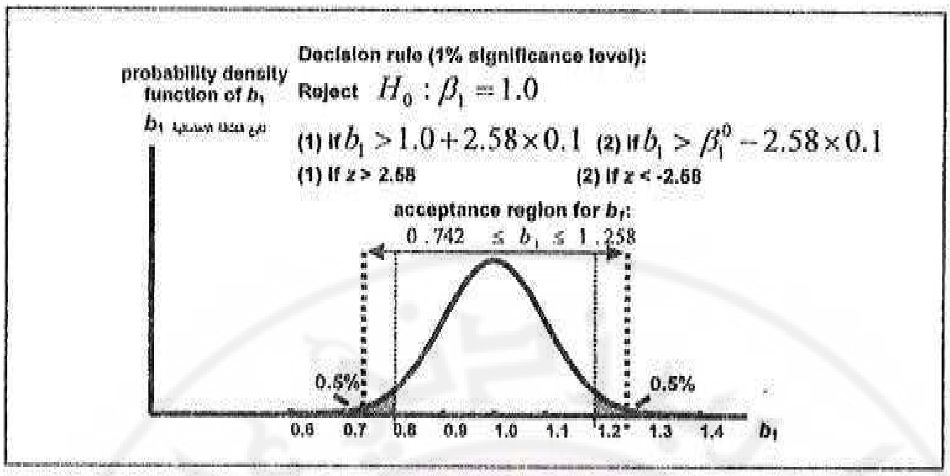


الشكل رقم (16-5)

إذاً احتمال الوقوع في الخطأ من النوع الأول ، إذا كانت فرضية العدم صحيحة يساوي 1% فقط أي الاختبار يكون معنوياً من أجل 1% مستوى معنوية الاختبار.

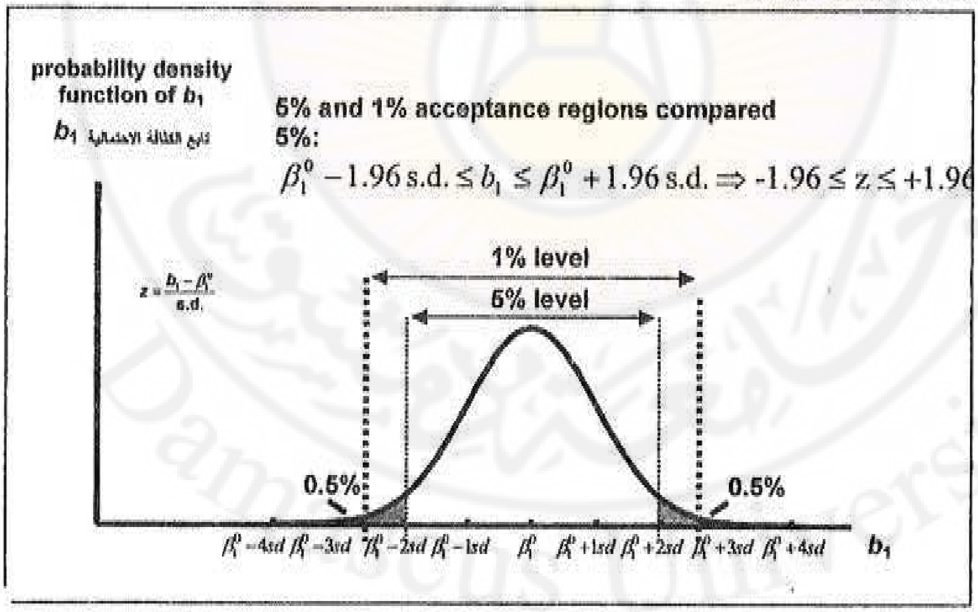
في حالة مثالنا أسعار في حالة التضخم والأجور في حالة التضخم للنموذج تعطي خطأ معيارياً 0.01 أي 0.5% من الجانبين تبدأ منطقة القبول من القيمة 0.758 إلى القيمة 1.258 حول الوسط الحسابي.

منطقة قبول b_1 تقع في المجال 0.742 إلى 1.58. أي أصبحت منطقة قبول أكثر اتساعاً مما كانت عليه عند اختبار 5 % ، مما يقلل من احتمال الوقوع بالخطأ من النوع الأول ، إذا كانت فرضية العدم صحيحة.

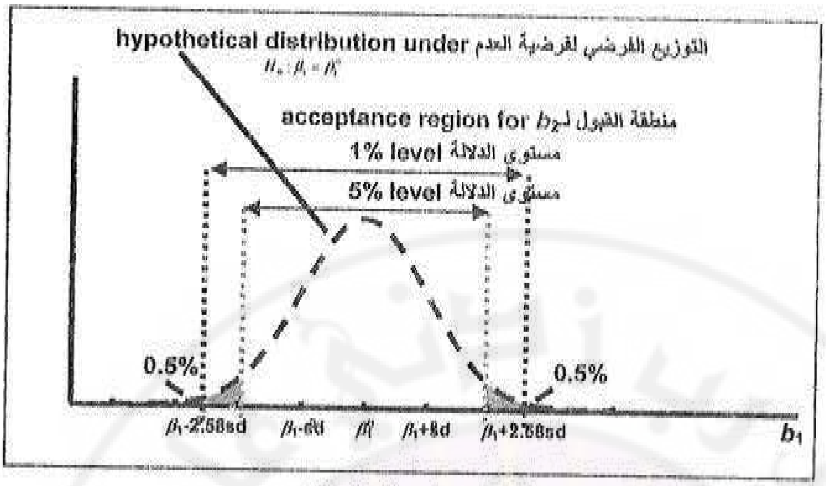


الشكل رقم (17-5)

من خلال الرسم البياني رقم (18-6) نعرض المقارنة بين قاعدتي الحكم من أجل 1% و 5% للاختبار. إذا رفضت فرضية العدم H_0 عند مستوى دلالة 1% ، علينا أيضا رفض فرضية العدم عند مستوى دلالة 5% حكماً، ولكن العكس ليس صحيحاً.

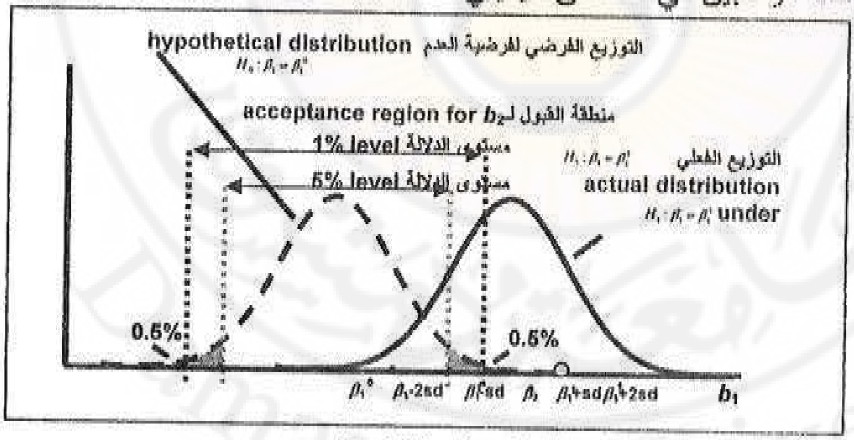


الشكل رقم (18-5)



الشكل رقم (5-21)

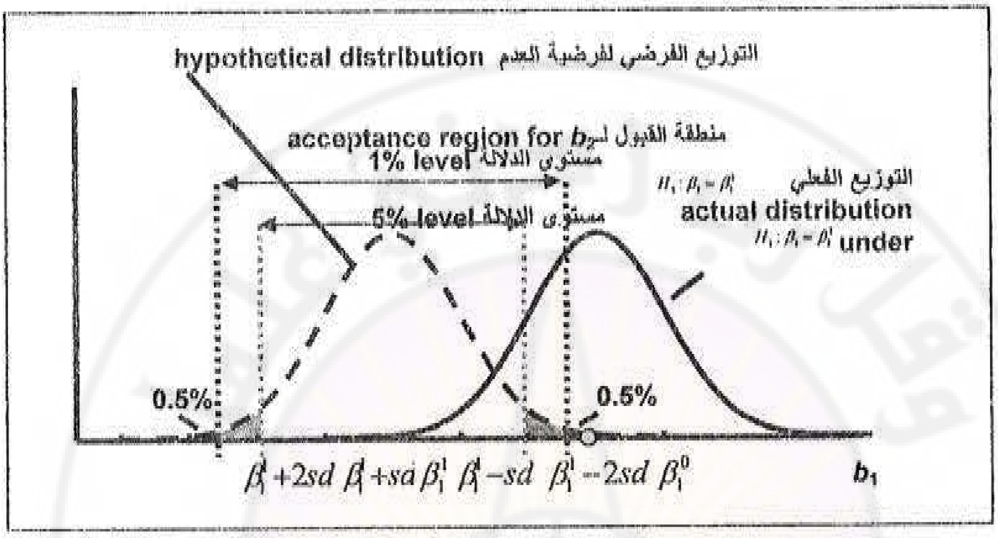
أما إذا أمعنا النظر في معاني ذلك أي اختيار مستوى معنوية للاختبار عندما تكون فرضية العدم خطأ، فإن الرسم البياني(6-22) يشرح كيفية صنع قرار الاختبار، ولكن لا يبين التوزيع الفعلي لـ b_1 . (والسبب في ذلك المنحني يخرج عن الخط) على فرض أن فرضية العدم فعلا صحيحة وتوزيع b_1 ، كما هو مبين في المنحنى اليميني للشكل.



الشكل رقم (5-22)

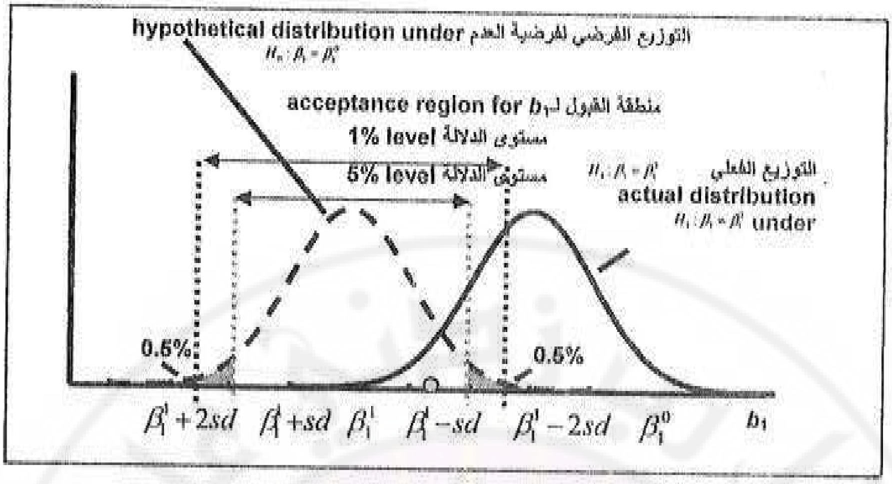
إذا حصلنا على بعض البيانات أثناء تنفيذ الانحدار يمكن أن يكون تقدير b_1 كما هو مبين في الشكل(5-23). في هذه الحالة سوف نتخذ قراراً صحيحاً

ونرفض H_0 و لا أهمية لأي اختبار نستخدم. نقوم بتقدير آخر مرة ثانية يمكننا اتخاذ قرار صحيح ونرفض فرضية العدم، ولا أهمية لاستخدام مستوى معنوية 5% أو 1% .



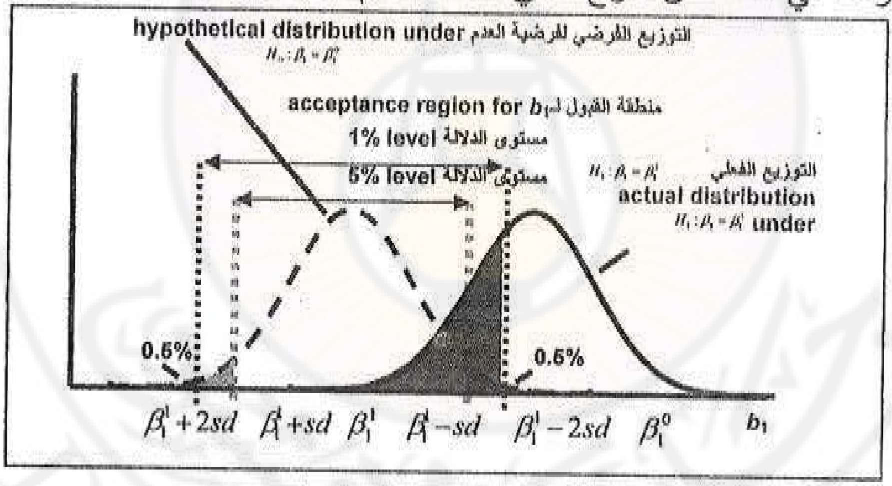
الشكل رقم (5-23)

إذا حصلنا على بعض البيانات أثناء تنفيذ الانحدار يمكن أن يكون تقدير b_1 كما هو مبين في الشكل(5-24). في هذه الحالة سنتخذ قرار صحيح ونرفض H_0 و لا أهمية لأي اختبار نستخدم. ولكن في الشكل(5-25) ، سنتخذ قرارا خاطئاً من النوع الثاني و نقع في رفض فرضية باستخدام مستوى معنوية آخر.



الشكل رقم (5-24)

نتخذ قراراً صحيحاً إذا استخدمنا مستوى معنوية 5% ، ولكن نكون قد وقعنا في الخطأ من النوع الثاني عند استخدام مستوى دلالة 1%.



الشكل رقم (5-25)

إمكانية الوقوع في الخطأ من النوع الثاني إذا استخدمنا 1% يعطينا إمكانية b_1 أن تقع داخل منطقة القبول، وهو المجال بين الخط الغامق والخط الفاتح.

إذاً بدلاً من أخذ مستوى معنوية للاختبار 5% واحتمال أن الوقوع في الخطأ من النوع الثاني، أي إذا كانت H_1 صحيحة وواقعة في توزيع H_1 في منطقة القبول من أجل 5% ، (تم عزل المنطقة الرمادية في الشكل البياني(25-5)). في هذه الحالة على وجه التدقيق نستخدم 1% بدلاً من 5% خشية الوقوع في خطر الخطأ من النوع الثاني. المشكلة بالطبع هنا تكون في معرفة H_0 أي الاثنان الصحيح من الخطأ. إذ H_0 فعلاً وقعت يكون صحيحاً استخدام مستوى معنوية 1% بدلاً من 5% من أجل تخفيض خطر الوقوع في النوع الأول (ولكن ألا نقع في الخطأ من النوع الثاني). على أية حال إذا كانت H_0 خطأ باستخدام 1% مستوى معنوية بدلاً من 5% هذا يزيد من احتمال الوقوع في الخطأ من النوع الثاني (وعدم الوقوع في الخطأ من النوع الأول).

ونخلص إلى القول : قد يحدث أن تؤدي النتيجة لرفض فرض العدم ،بينما هو في الواقع صحيح، ونسمي هذا خطأ من النوع الأول، كذلك قد يحدث الا نرفض فرض العدم و هو في الحقيقة خطأ، ونسمي هذا خطأ من النوع الثاني. عادة نرمز للخطأ من النوع الأول بـ α والخطأ من النوع الثاني بـ β. إذا كانت Ω فضاء عينة ذي n بعد وعرفنا فيه منطقة الرفض C ومنطقة القبول C* وهي متممة C في Ω ،وإذا كانت $(X_1, X_2, ..., X_n)$تمثل نتيجة العينة العشوائية للظاهرة المدروسة فإن حجم الخطأ من النوع الأول يمكن كتابته رياضياً على النحو الآتي:

$$\mathbf{Prob}(C / H_0) = \mathbf{Prob}[(X_1, X_2, ..., X_n) \in C / H_0] = \alpha$$

وحجم الخطأ من النوع الثاني يمكن كتابته رياضياً على النحو الآتي:

$$\mathbf{Prob}(C^* / H_1) = \mathbf{Prob}[(X_1, X_2, ..., X_n) \in C^* / H_1] = \beta$$

ملاحظة: الاختبار الذي يقلل من احتمالات الخطأ من النوعين في الوقت نفسه ولأقل قدر ممكن هو الاختبار الذي نفضله على سواه. ولكن حجم العينة

الثابت ، لا يمكن تخفيض احتمال أحد الخطأين دون زيادة حجم الآخر والطريقة الوحيدة لتخفيضهما معاً زيادة حجم العينة[2]

5-3- اختبار t للفرضيات المتعلقة بمعاملات الانحدار

t TEST OF A HYPOTHESIS RELATING TO A REGRESSION COEFFICIENT

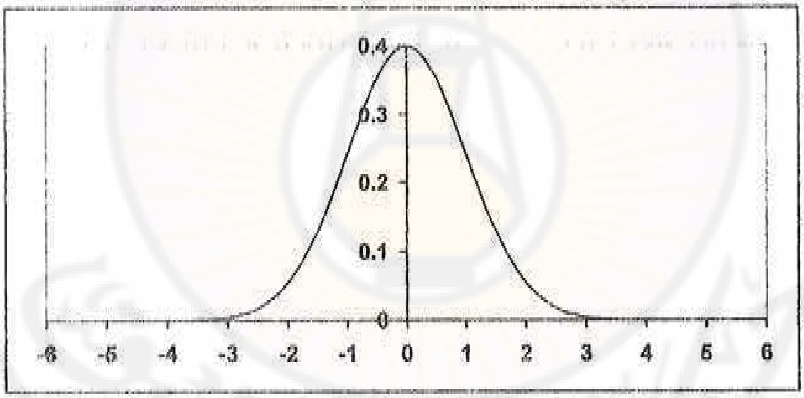
في هذه الفقرة سنلخص إنجاز الاختبار عند مستوى معنوية 5% على ميل معامل الانحدار على افتراض معرفتنا بالخطأ المعياري. هذا الافتـراض غيـر واقعي. عادة نملك تقدير الخطأ المعياري، ونستخدمه في الاختبارات الإحصائية عوضاً عن الانحراف المعياري؛ لأننا نملك تبديل الانحراف المعياري في المقام بالخطأ المعياري، وبالتالي استخدام الاختبار الإحصائي لتوزيع t بـدلاً مـن التوزيع الطبيعي.

نحن نبحث عن القيمة الحرجة لـ t وإذا كانت t الإحصائية أكبر منها ، بالإيجاب أو السالب، فنحن نرفض فرضية العدم، ونقبل الفرضية البديلة.

[2] كتاب الاستدلال الإحصائي، د.زين العابدين البشير-د. احمد عودة(172-173).

الخطأ المعياري لـ b_1 غير معروف s.d. of b_1 not known الاختلاف بين قيم الفروض وتقدير العينة في حدود الخطأ المعياري discrepancy between hypothetical value and sample estimate, in terms of s.e.: $t = \frac{b_1 - \beta_1^0}{\text{s.e.}}$ %5 مستوى الدلالة للاختبار (5% significance test) H_0: $\beta_1 = \beta_1$ (reject@ رفض $t > t_{crit}$ or $t < -t_{crit}$ (if) إذا كان:	الانحراف المعياري لـ b_1 معروف s.d. of $b_1$known الاختلاف بين قيم الفروض وتقدير العينة في حدود الخطأ المعياري discrepancy between hypothetical value and sample estimate, in terms of s.d.: $z = \frac{b_1 - \beta_1^0}{\text{s.d.}}$ %5 مستوى الدلالة للاختبار (5% significance test) H_0: $\beta_1 = \beta_1^0$(reject) رفض $z > 1.96$ or $z < -1.96$ (if) إذا كان:

يعرض الشكل البياني (26-5) التوزيع الطبيعي من أجل متوسط يساوي الصفر، وتباين يساوي وحدة واحدة.



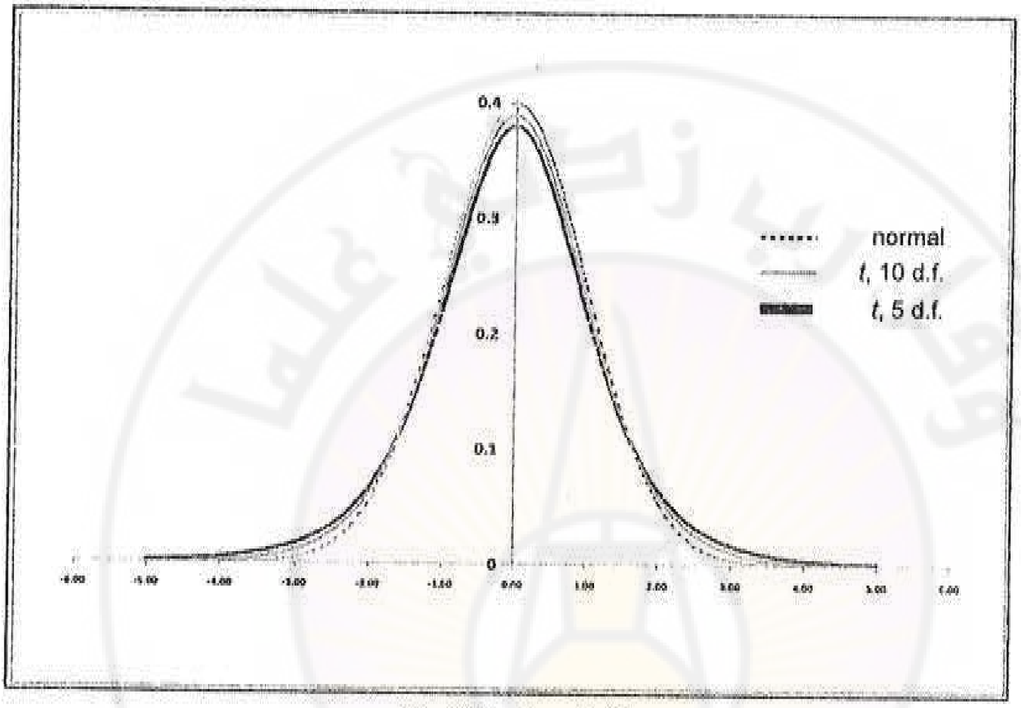
الشكل رقم (26-5)

لنضف إليه توزيعاً جديداً لـ t من أجل 5 درجات حرية ، ومن ثم نضف توزيعاً جديداً لـ t من أجل 10 درجات الحرية، انظر الشكل (27-5).

نلاحظ عندما يكون عدد درجات الحرية كبيراً ، فتوزيع t يكون مقارباً التوزيع الطبيعي (وكلما زدنا عدد درجات الحرية، كلما كان أكثر انطباقاً).

حتى عندما تكون درجات الحرية صغيرة - كما في هذه الحالة- فإن التوزيعين

متقاربان جداً. حتى توزيع t من أجل5 درجات الحرية، متشـابه جـداً مـع التوزيع الطبيعي.



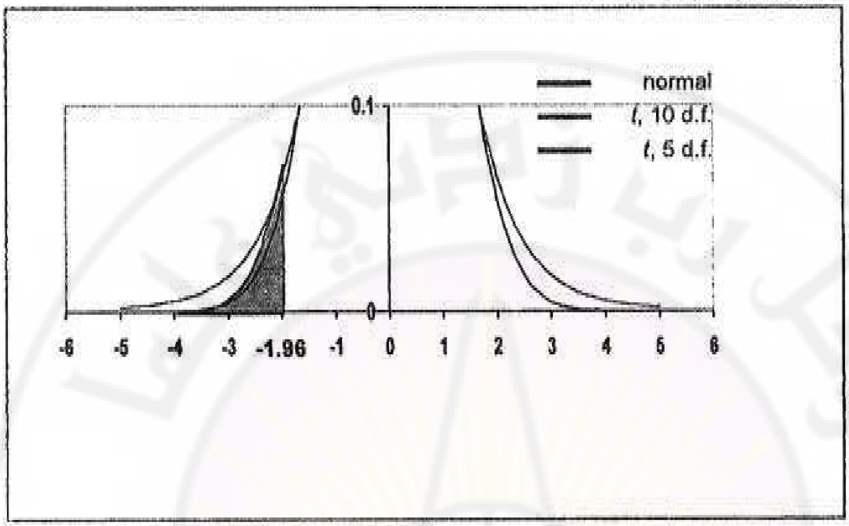
الشكل رقم (27-5)

والسؤال الذي يطرح نفسه: لماذا إذاً نلجأ إلى الاعتماد علـى التوزيـع الطبيعي؟ هل هناك في الواقع أهمية إذا استخدمنا 1.96 من أجل 5% مستوى دلالة أو 2.58 من أجل مستوى دلالة 1% ؟

الجواب على ذلك: هناك اختلاف؛ على الرغم من أن التوزيعات متماثلـة جداً ، إلا أن توزيع t يملك طرفين ممتدين أكثر من التوزيع الطبيعي، الاختلاف يبدأ كبيراً ، ويصغر حسب عدد درجات الحرية.

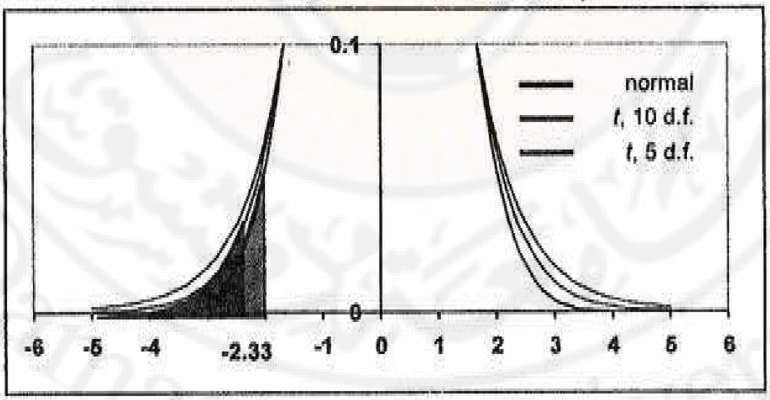
لذلك لضرورة الحصول على اختبار إحصائي عالي المعنوية يكون كبيراً مع استخدام توزيع t مقارنة مع التوزيع الطبيعي. هذا يعني أن منطقة الرفض

سوف تبدأ بانحراف معياري أكبر من أجل توزيــع t مقارنــة مــع التوزيــع الطبيعي. أي الــ2.5% من الطرف (الذيل) للتوزيع الطبيعي تبــدأ مــن 1.96 اختلاف معياري عن متوسطه.



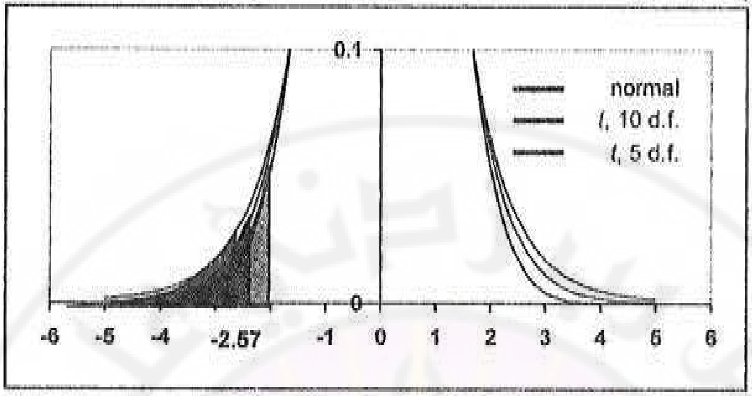
الشكل رقم (5-28)

والــ2.5% من طرف واحد لتوزيع t من أجل 10 درجات حرية تبدأ من 2.33 اختلاف معياري عن متوسطه.



الشكل رقم (5-29)

أما من أجل 5 درجات حرية لتوزيع t فإنه يبـدأ مـن 2.57 اخـتلاف معياري عن متوسطه.



الشكل رقم (30–5)

لهذا السبب إلى نحتاج جداول خاصة لقيم t المحسوبة (الحرجة) لإنجـاز اختبارات المعنوية الإحصائية على معاملات معادلة الانحدار . الجدول التـالي يبين توزيع t الإحصائية.

t Distribution: Critical values of t

Degrees of freedom درجات الحرية	Two-tailed test الاختبار من اتجاهين					
	10%	5%	2%	1%	0.2%	0.1%
	One-tailed test الاختبار من اتجاه واحد					
	5%	2.5%	1%	0.5%	0.1%	0.05%
1	6.314	12.706	31.821	63.657	318.31	636.62
2	2.920	4.303	6.965	9.925	22.327	31.598
3	2.353	3.182	4.541	5.841	10.214	12.924
4	2.132	2.776	3.747	4.604	7.173	8.610
5	2.015	2.571	3.365	4.032	5.893	6.869
...	...	...	...	...	...	...
...	...	...	...	...	...	...
18	1.734	[redacted]	2.552	2.878	3.610	3.922
19	1.729	2.093	2.539	2.861	3.579	3.883
20	1.725	2.086	2.528	2.845	3.552	3.850
...	...	...	...	...	...	...
...	...	...	...	...	...	...
120	1.658	1.980	2.358	2.617	3.160	3.373
∞	1.645	1.960	2.326	2.576	3.090	3.291

السطر الأول من الجدول يبين مستويات المعنوية الممكنة للاختبار من اتجاهين (في أغلب الأحيان يكون الاختبار من اتجاهين إلا في الحالات الواضحة المعالم، إن الاختبار من اتجاه واحد) ، ونتجاهل الاختبارات من اتجاه واحد إلا إذا كان واضحاً أن الاختبار من اتجاه واحد.

على سبيل المثال لإنجاز الاختبار من اتجاهين عند مستوى معنوية 5% ، نستخدم العمود الثالث كما هو مبين في الجدول. العامود اليساري من الجدول يعرض عدد درجات الحرية. عدد درجات الحرية في علاقة الانحدار يحدد بعدد المشاهدات ناقص عدد المعلمات المقدرة في العلاقة (n-k).

في الانحدار البسيط فإن عدد المعلمات يساوي 2 القيمة الثابتة ومعلمة الميل ، لذلك عدد درجات الحرية يساوي n-2 حيث n عدد المشاهدات.

إذا كان لدينا معادلة انحدار مع 20 مشاهدة ، مثل علاقة الأسعار في حالة التضخم مع الأجور في حالة التضخم ، فإن عدد درجات الحرية سوف يساوي 18 وعندئذ نبحث عن القيمة الحرجة تقاطع 18 درجة حرية مع مستوى دلالة 5% فنجدها تساوي 2.101 .

ملاحظة: إذا كان عدد درجات الحرية كبيراً فإن القيمة الحرجة تقع عند 1.96 المعادلة للتوزيع الطبيعي. هذا لأن توزيع t يلتقي مع التوزيع الطبيعي في العينات الكبيرة.

نعود و نلخص الاختبار، سوف نرفض فرضية العدم إذا كانت بالقيمة المطلقة t أكبر من 2.101 .

إذا رغبنا في الاختبار عند مستوى معنوية 1% ، سوف نستخدم الجدول(t) المبين أعلاه. **ملاحظة:** إذا كان عدد درجات الحرية كبير فإن القيمة الحرجة تساوي 2.58 تقريباً وهي القيمة الحرجة نفسها للتوزيع الطبيعي.

أما بالنسبة للانحدار البسيط لمثالنا السابق لـ20 مشاهدة فنجد أن القيمة الحرجة تساوي 2.878 عند مستوى معنوية 1% .

5-4- حدود (مجالات) الثقة CONFIDENCE INTERVALS

عند إنشاء حدي ثقة نميز بين الحالات التالية:

الحالة الأولى: إنشاء حدي ثقة حول الوسط الحسابي للعينة؛ بفرض أن تباين المجتمع معلوم σ^2 وباحتمال ثقة $[1 - \alpha]$ محدد وبفرض أن العينة عشوائية :

$X_1 , X_2 , \ldots , X_n$ تخضع للتوزيع الطبيعي $N \sim (\mu , \sigma^2)$:

$$\overline{X} \mp Z\,[1 - \frac{\alpha}{2}] * \frac{\sigma}{\sqrt{n}}$$

وحيث إن : $Z\,[1 - \frac{\alpha}{2}]$ – عبارة عن قيمة Z الجدولية في حال الاختبار من اتجاهين وتحدد قيمتها على الشكل التالي: $P\,(Z \le Z\,[1 - \alpha/2]) = 1 - \alpha/2$

عندما Z تخضع للتوزيع الطبيعي المعياري $N \sim (0.1)$.

الحالة الثانية: إنشاء حدي ثقة حول الوسط الحسابي للعينة بفرض أن تباين المجتمع غير معلوم σ^2 فإننا نستبدلها بالتقدير S^2 (تباين العينة) ومن أجل احتمال ثقة $[1 - \alpha]$ محدد، وبفرض أن العينة عشوائية، وإذا كان حجم العينة صغيراً فإننا نستخدم توزيع (t) ستودنت من أجل درجات حرية $(n - 1)$ واحتمال ثقة $(1 - \alpha/2)$. ونستخدم لإنشاء حدي الثقة العلاقة التالية:

$$\overline{X} \mp t\,[1 - \alpha/2\,;\, n - 1] * \frac{S}{\sqrt{n}}$$

حيث إن: $t[1-\alpha/2\,;n-1]$ نحددها من جداول توزيع t ستودنت، وهي القيمة التي تساوي إلى: $P\,(t \le t[1-\alpha/2\,;n-1]) = 1-\alpha/2$ وذلك عندما يكون التوزيع خاضعاً لتوزيع ستودنت (t).

الحالة الثالثة: إنشاء حدي الثقة حول متوسط النسبة المئوية للعينة ، عندما تكون العينة العشوائية $X_1\,, X_2\,, \ldots\,, X_n$ خاضعة لتوزيع ثنائي الحدين $b\,(1\,, p)$ ، فإن تقدير فترة الثقة أي إنشاء حدي الثقة يكون من خلال الصيغة التالي:

$$\overline{p} \mp Z\,[1-\alpha/2] * \sqrt{\frac{\overline{p}\,\overline{q}}{n}}$$

وباحتمال ثقة $(1-\alpha)$ للنسبة p عندما n تكون كبيرة بقدر كافٍ.

الحالة الرابعة: إنشاء حدي الثقة حول الفرق بين وسطين حسابين $(\mu_x - \mu_y)$ لعينتين عشوائيتين $X_1\,, X_2\,, \ldots\,, X_n$ و $Y_1\,, Y_2\,, \ldots\,, Y_n$ تخضعان للتوزيع الطبيعي $N \sim (\mu_1\,, \sigma_1^2), N \sim (\mu_2\,, \sigma_2^2)$ ومستقلتين عن بعضهما بعض؛ فإننا نستطيع إنشاء حدي ثقة باحتمال ثقة $(1-\alpha)$. ونميز حالتين:

أ- إذا كان $\sigma_1^2\,, \sigma_2^2$ معلومتين فإننا نطبق العلاقة التالية:

$$(\overline{X} - \overline{Y}) \mp Z\,[1-\alpha/2] * \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

ب- إذا كان قيمتي $\sigma_1^2\,, \sigma_2^2$ غير معلومتين ومتساويتين فإننا نستبدلهما بـ S_c^2 و هو تباين العينتين، حيث أن: $S_c^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$

وتصبح علاقة التقدير:

$$(\overline{X} - \overline{Y}) \mp t\,[1-\alpha/2\,; n_1+n_2-2] * Sc \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

الحالة الخامسة: إنشاء حدي ثقة حول الفرق ما بين نسبتين $p_1 - p_2$ لعينتين عشوائيتين تخضعان لتوزيع ثنائي الحدين $b(1, p_1)$ و $b(1, p_2)$ المقارب للتوزيع الطبيعي وكانتا مستقلين، ومن أجل احتمال ثقة فإننا نكون على ثقة $(1 - \alpha)$:

$$(\bar{P}_1 - \bar{P}_2) \mp Z\left[1 - \alpha/2\right] * \sqrt{\frac{\bar{p}_1\,\bar{q}_1}{n_1} + \frac{\bar{p}_2\,\bar{q}_2}{n_2}}$$

5-5- تمارين محلولة:

مثال (5-2): سحبت عينة حجمها 25 من توزيع طبيعي $N(\mu,16)$ تبين لنا أن وسط العينة (30)، والمطلوب: إنشاء حدي ثقة للوثوق في هذا الوسط باحتمال ثقة قدره 95% .

الحل :

1) العلاقة التي نطبقها هي : $\bar{X} \mp Z\left[1 - \alpha/2\right] * \frac{\sigma}{\sqrt{n}}$

2) نحسب قيمة Z الجدولية من أجل مستوى دلالة 5% فنجد أن:

$$Z[0.95] = 1.96$$

3) نحسب الخطأ المعياري للعينة: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{4}{\sqrt{25}} = 0.8$

4) نطبق العلاقة فنجد: $30 \mp 1.96 * 0.8 = [28.432\ ,\ 31.568]$

التفسير: إننا واثقون في هذا المتوسط بأن قيمته لن تقل عن 28.432 ولن تزيد عن 31.568 باحتمال ثقة قدره 95%.

مثال(3-5): سحبت عينة عشوائية حجمها 9 من توزيع طبيعي $N\sim(\mu,\sigma^2)$ وعلماً أن μ ,σ^2 مجهولتان. وإذا علمت أن متوسط العينة هو 25 وتباينها هو $S^2=1$. والمطلوب: أوجد حدي الثقة باحتمال ثقة 99.5% .

الحل :

1- في هذه الحالة نطبق العلاقة التالية: $\bar{X} \mp t\ [1-\propto/2; n-1]$

2- نحدد قيمة t الجدولية من أجل احتمال ثقة 99% و 8 درجات حرية فنجد أن: $t\ [\ 0.99,\ 8\ ] = 3.355$.

3- نطبق العلاقة فنجد: $25 \mp (3.355) * \frac{4}{\sqrt{9}} \quad \Rightarrow$

$$= 25 \mp 4.473 = [20.527\ ,\quad 29.473]$$

التفسير: إننا واثقون بمتوسط هذه العينة باحتمال ثقة 99 % إنه لن يقل عن 20.527 ولن يزيد عن 29.473 ، ومن أجل 8 درجات.

مثال(4-5): إذا علمت أنه تم سحب عينتين عشوائيين من مجتمع يخضع للتوزيع الطبيعي $n_1 = 16$, $n_2 = 25$ وتبين أن : $S_1^2 = 5.5 \quad S_2^2 = 5.25$

وأن متوسط العينة الأولى $\mu_1 = 20$, $\mu_2 = 22$

والمطلوب: أنشئ حدي ثقة للوثوق في الفرق ما بين المتوسطين باحتمال ثقة 95 %.

الحل :

1- نحسب أولاً: $S_C^2 = \frac{15 * 5.5 + 24 * 5.25}{16 + 25 - 2} = 5.346$

2- نحسب t الجدولية فنجد أن: $t[95\,\%\,; 39] = 1.642$

3- نطبق العلاقة فنجد:

$$(25-16)\mp 1.642 * \sqrt{5.346} * \sqrt{\frac{1}{16}+\frac{1}{25}}=[0.8\,,3.2]$$

التفسير: إننا على ثقة باحتمال 95 % إن الفرق ما بين المتوسطين لن يزيد عن 3.2 ولن يقل عن 0.8 .

مثال (5-5): لمقارنة نسبة المدخنين ما بين الذكور والإناث في الجامعة. أخذت عينة من الذكور حجمها 600 طالب، ووجد من بينهم 300 مدخن. وأخذت عينة من الإناث حجمها 100 طالبة فتبين أن 30 طالبة مدخنة. أوجد حدي الثقة باحتمال ثقة قدره 95% للفرق بين نسبتي المدخنين الطلاب والطالبات في هذه الجامعة.

الحل :

1- الدالة المطلوبة من الشكل التالية:

$$(\overline{P}_1-\overline{P}_2)\pm Z\,[1-\alpha/2]*\sqrt{\frac{\overline{p}_1\,\overline{q}_1}{n_1}+\frac{\overline{p}_2\,\overline{q}_2}{n_2}}$$

2- نحدد قيمة Z الجدولية : $Z\ [95\%]=1.96$

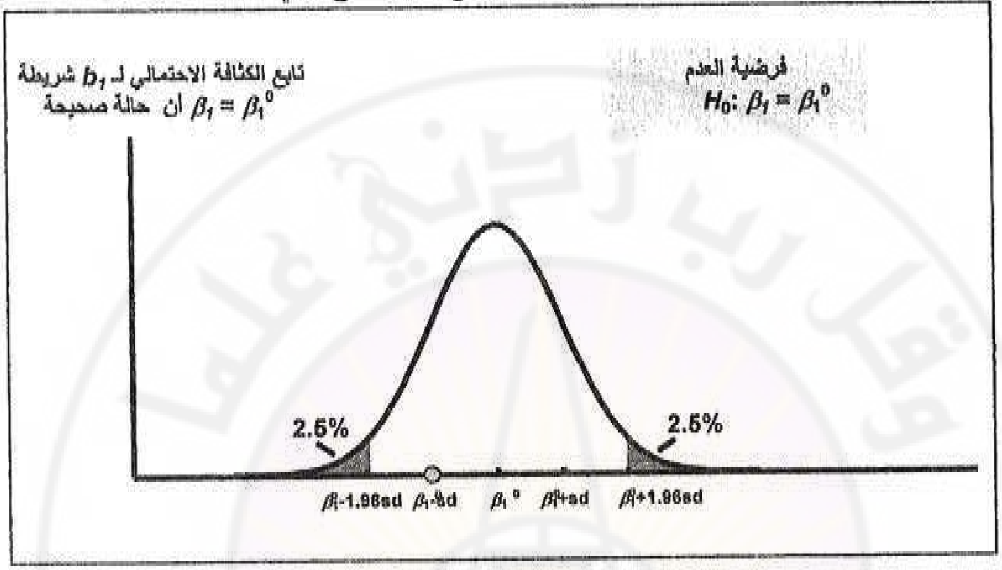
3- نطبق العلاقة فنجد أن:

$$(0.5-0.3)\mp 1.96 * \sqrt{\frac{0.5 * 0.5}{600}+\frac{0.3 * 0.7}{100}}=$$

$$=0.2\mp 0.098=[0.102\,,0.298]$$

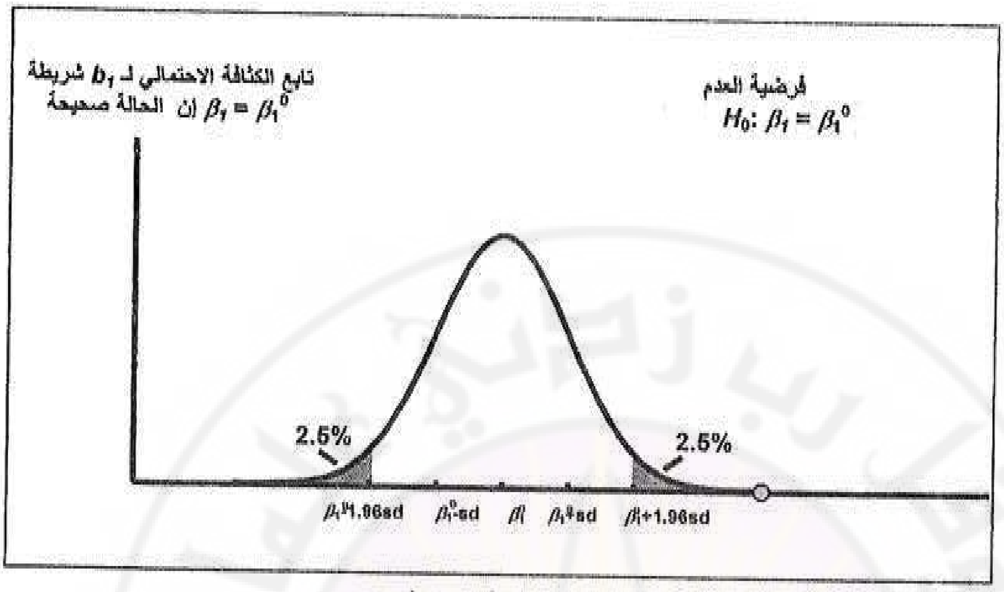
التفسير: إننا واثقون باحتمال ثقة 95 % إن الفرق ما بين نسبتي المدخنين من الذكور والإناث سوف تتراوح ما بين القيمتين: 10.2 % و 29.8% .

لنحاول في هذه الفقرة عرض اختبارات الفرضيات من خلال الرسم البياني. نبدأ من فرضية العدم على سبيل المثال : $H_0: \beta_1 = \beta_1{}^0$ وندرس تقدير b_1 المشتق من العينة لنبين فيما إذا كانت تقع أو لا تقع في منطقة الرفض.



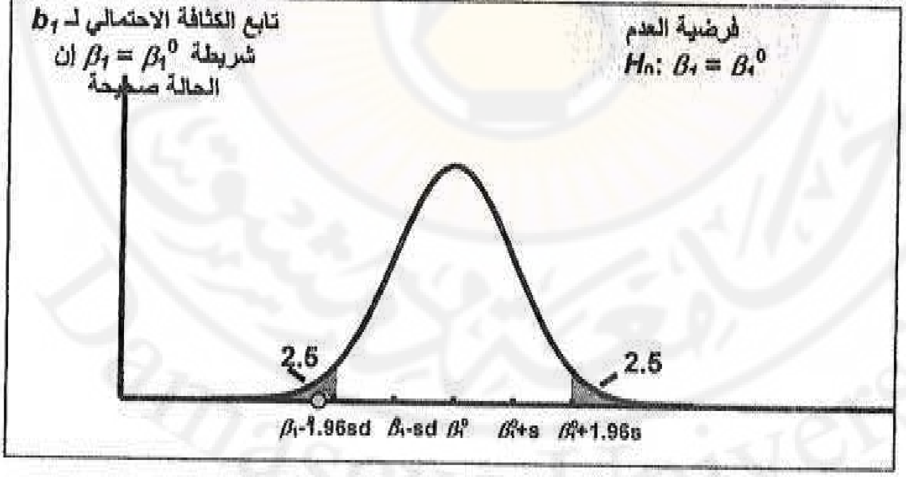
الشكل رقم (5-31)

من الشكل رقم(5-32) نجد أن قيمة b_1 وقعت في منطقة القبول داخل منحني الكثافة الاحتمالية لتوزيعb_1 عند استخدام مستوى دلالة 5% من اتجاهين . هذا التقدير لعدم الوقوع في منطقة الرفض لـ H_0. هذا التقدير يؤدي إلى وقوع H_0 في منطقة الرفض عند القيمة 2.5% لأن الاختبار من اتجاهين، ولأن القيمة المقدرة وقعت خارج المنحني.

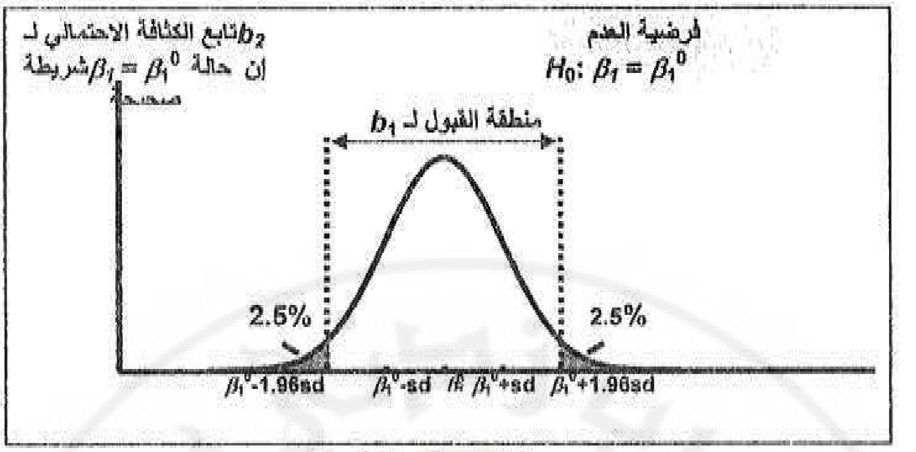


الشكل رقم (32–5)

أما في الحالة التالية فالتقدير سيؤدي إلى وقوع H_0 في منطقة الرفض عند القيمة 2.5% لأن الاختبار من اتجاهين ولأن القيمة المقدرة ضمن منطقة الرفض ، كما هو مبين بالشكل رقم(33–5) .

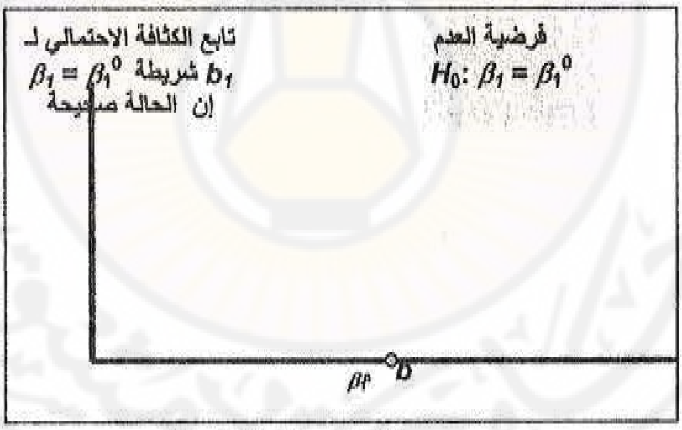


الشكل رقم (33–5)



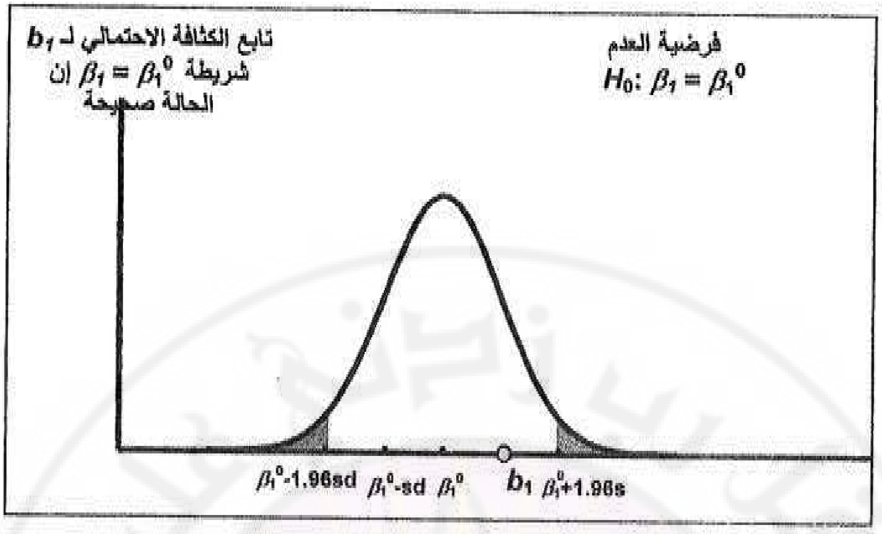
الشكل رقم (5-34)

لننهي تجارب تقديرات المجالات للحالات الملائمة والتي تقع ضمن منطقة القبول. لنعمل العكس تماماً، أي نحدد تقدير العينة ، ونبحث عن مجموعة الفرضيات التي سوف تتعارض مع ذلك، باستخدام مستوى دلالة 5% من اتجاهين. على فرض وجدنا قيمة واحدة تحقق الفرضية $H_0: \beta_1 = \beta_1^0$.



الشكل رقم (5-35)

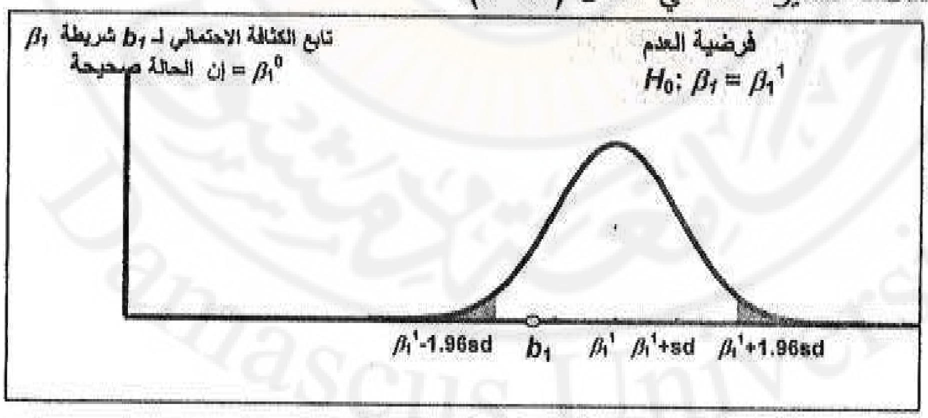
نحتاج إلى رسم التوزيع الاحتمالي لـ b_1 على شرط $H_0: \beta_1 = \beta_1^0$ صحيحة .



الشكل رقم (5-36)

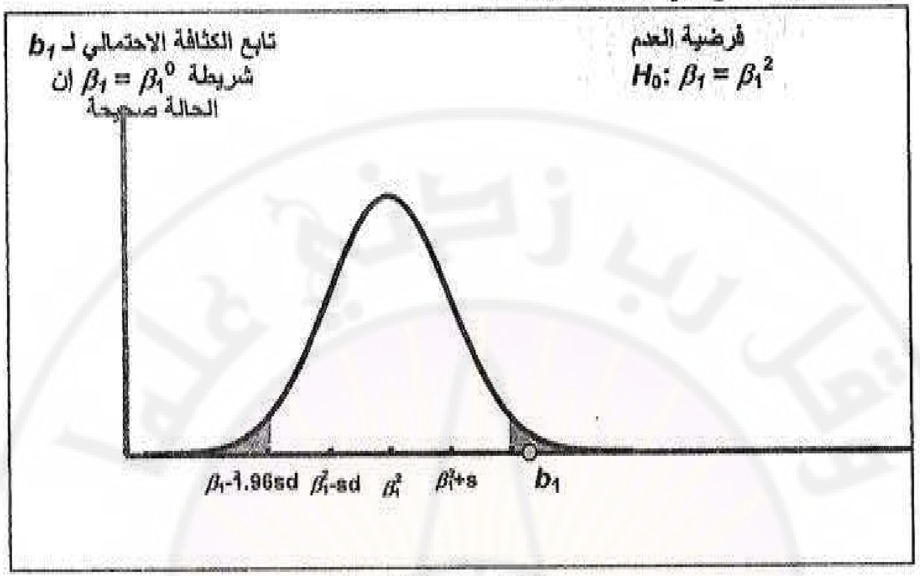
ولإنجاز ذلك نحتاج إلى الانحراف المعيار للتوزيع. وفي كل مرة سنعتبر ان هذا الافتراض محقق، مع انه في الواقع لدينا التقدير ذاته. نملك تصوراً عن التوزيع من خلال فرضية العدم وعندما تكون غير متعارضة مع تقدير عينتنا.

هنا فكرة أخرى لفرضية العدم، لدينا تصور عن التوزيع لـ b_1 شريطة أن تكون هذه الفرضية صحيحة، نستطيع أن نرى ذلك على أن هذه الفرضية ملائمة لتقديرنا كما في الشكل (6-37).



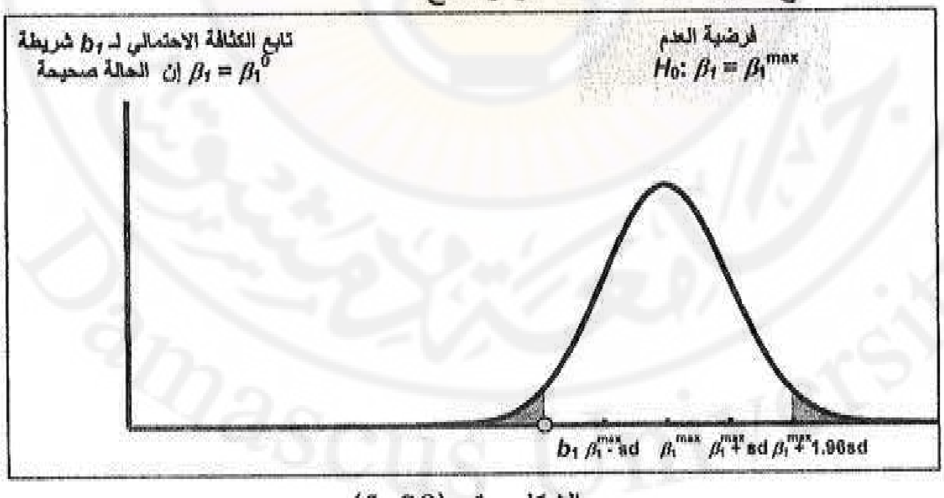
الشكل رقم (5-37)

أما في الشكل (38-5)، فتعتبر فرضية العدم غير ملائمة لتقديرنا؛ لأن قيمة التقدير ستقع في منطقة الرفض.



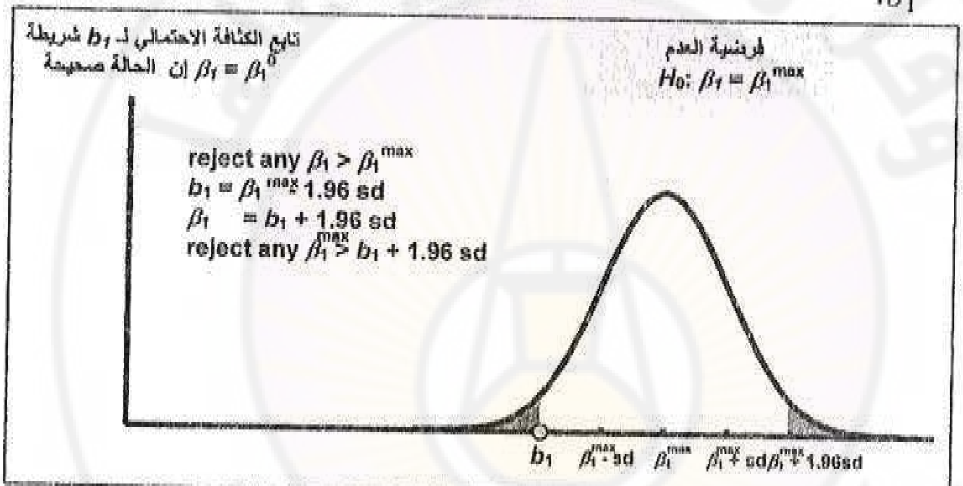
الشكل رقم (38-5)

ولكن الرسم البياني (39-5) يبين أن القيمة الافتراضية العليا لـــ β_1 غير متعارضة مع تقدير عينتنا ، كما هو واضح.



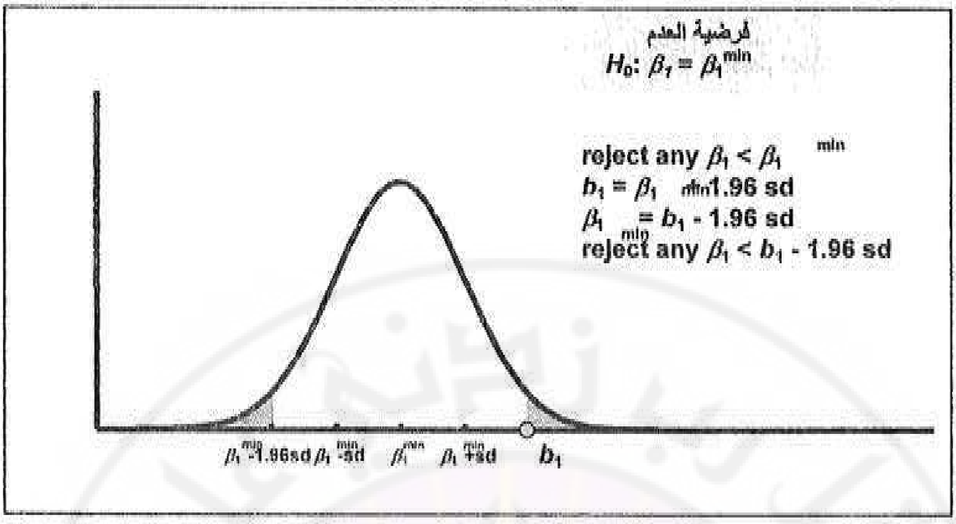
الشكل رقم (39-5)

عندما يكون التوزيع الاحتمالي لــ b_1 مشروطاً ، كما هو مبين على الرسم فإن b_1 تقع على الحافة اليسارية لــ 2.5% . نسمي هذه بالقيمة العظمى لــ b_1^{max} . إذا كانت القيمة الفرضية لــ b_1 في ارتفاع، سيكون مرفوضاً لأن b_1 تقع داخل الطرف الأيسر للتوزيع. لذلك b_1 تقع على الحافة اليسرى للذيل 2.5% ، ويجب أن يكون الانحراف المعياري أقل من 1.96 لــb_1^{max}. وهكذا أصبح معلوماً قيمة b_1 والانحراف المعياري، ونستطيع أن نحسب قيمة b_1^{max}.



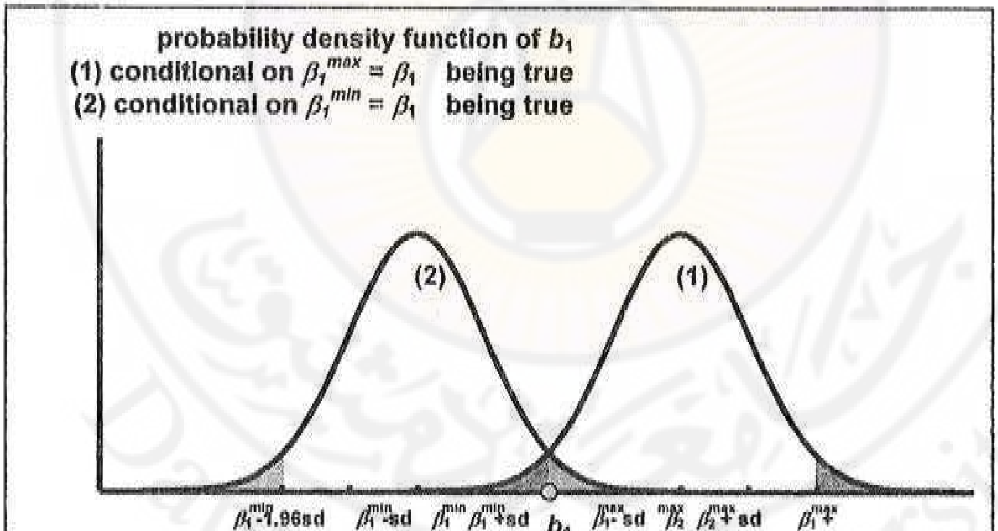
الشكل رقم (5-39)

بالطريقة نفسها نستطيع تحديد القيمة الفرضية لــ b_1 التي لا تكون متعارضة مع تقديرنا ، والتي تدلنا قيمتها b_1 على الحافة اليمنى للذيل 2.5% مترافقة شرطياً مع التوزيع الاحتمالي. و نطلق عليها تسمية b_1^{min}. أي القيمة الفرضية ستكون أقل تعارضاً مع تقديرنا. لذلك b_1 تقع في الحافة اليمنى للذيل 2.5% ، b_1 تساوي إلى b_1^{min} زائد 1.96 خطأ معياري. وكذا فإن b_1^{min} تساوي إلى b_1 ناقص 1.96 خطأ معياري.



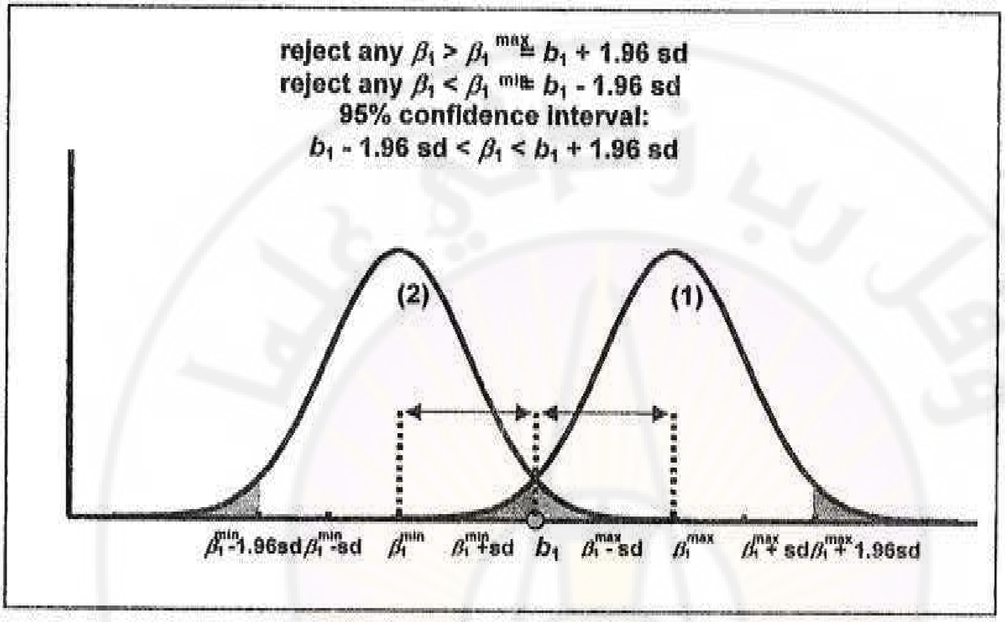
الشكل رقم (5-40)

يتبين لنا من الرسم رقم (41-6) نهاية القمم للقيم الافتراضية لـ β_1 مع بعضها بعض مترافقة التوزيع الاحتمالي لـ b_1.



الشكل رقم (5-41)

أي الفرضية تقع في المجال من β_1^{min} إلى β_1^{max} سوف تكون متوافقة مع تقدير العينة (ولن تكون مرفوضة)، ونسمي هذا المجال فترة الثقة حـــول 95%.



الشكل رقم (5-42)

التسمية الظاهرة في التطبيقات للمجال، يمكن أن يظهر القيمـــة الحقيقـــة لمجال ثقة باحتمال 95%، شريطة أن يكون فيه النموذج صحيح الخصـــائص. بالطريقة نفسها السابقة يمكن استخدام 1% مستوى معنويـــة اختبـــار لتحديـــد فرضية ملائمة لتقدير عينتنا، نستطيع إنشاء حدي ثقة بمستوى دلالة 99%.

سوف يكون المجال بين القيمتين β_1^{min} و β_1^{max} مساوياً 2.58 خطأ معياري من اليمين ومن اليسار حول b_1.

حتى الآن نعمل مفترضين، أنه معلوم لدينا الخطأ المعيـــاري للتوزيـــع، ولكن في الواقع العملي لا نملك إلا تقديرها فقط.

الانحراف المعياري معلوم Standard deviation known
95% درجة الثقة $b_1 - 1.96sd \le \beta_1 \le b_1 + 1.96sd$
99% درجة الثقة $b_1 - 2.58sd \le \beta_1 \le b_1 + 2.58sd$

بما أنه استخدمنا توزيع t بدلاً من التوزيع الطبيعي عندها يتحدد موضع حدي المجال β_1^{min} و β_1^{max} و الخطأ المعياري بدلالة مقدر الخطأ المعياري، وعندئذٍ نكتب:

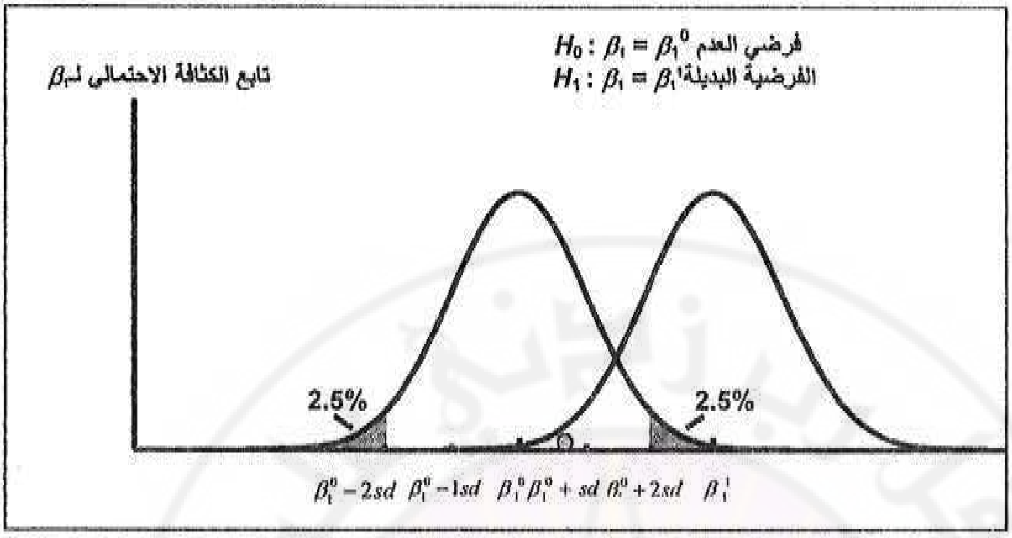
$$b_1 - t_{crit}(0.05)se \le \beta_1 \le b_1 + t_{crit}(0.05)se \quad \text{95\% درجة الثقة}$$

$$b_1 - t_{crit}(0.01)se \le \beta_1 \le b_1 + t_{crit}(0.01)se \quad \text{99\% درجة الثقة}$$

هذا يرشدنا إلى أنه يتم تحديد حدي المجال بعد بضرب الخطأ المعياري بالقيمة t الحرجة(النظرية)، من أجل مستوى المعنوية محدد وعدد درجات الحرية معروفة.

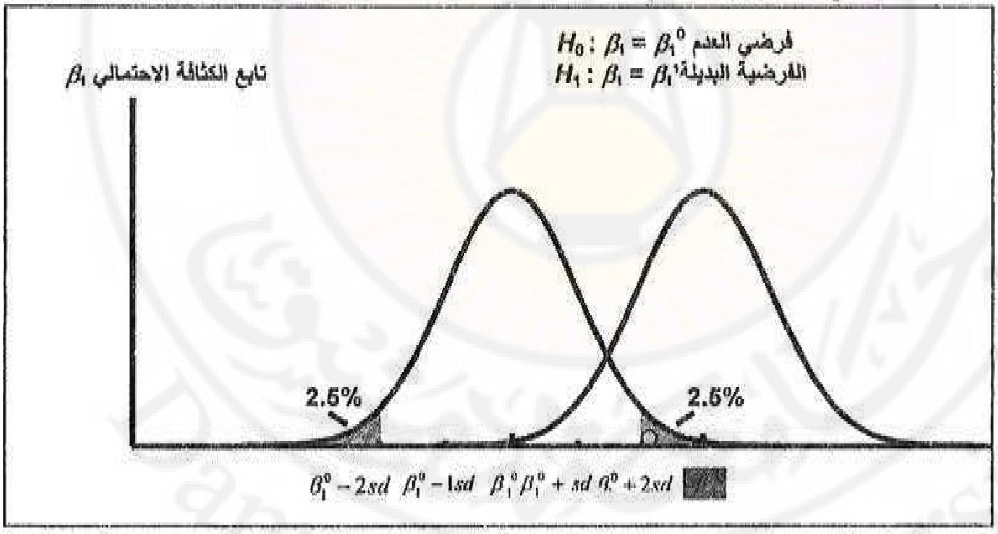
5-6- الاختبار من اتجاه واحد ONE-TAILED TESTS

تصادفنا في حياتنا العملية والتجريبية ضرورة تحديد اتجاه الاختبار، وذلك للأهمية. وقد عالجنا الاختبار من اتجاهين، وفي هذه الفقرة سنعالج الاختبار من اتجاه واحد باستخدام اختبار t . نبدأ بأخذ القيمة β_1 ، ونحدد لها قيمتين احتماليتين β_1^0 لفرضية العدم و β_1^1 للفرضية البديلة . سنعرض في هذا المقام تجربتين: التجربة الأولى نظامية، والأخرى طويل الأمد. نفترض أن العينة التي حصلنا عليها مبينة في الشكل رقم(5-43) . وسوف لن نرفض فرضية العدم لأن تقدير العينة يقع ضمن منطقة القبول لـ β_1.



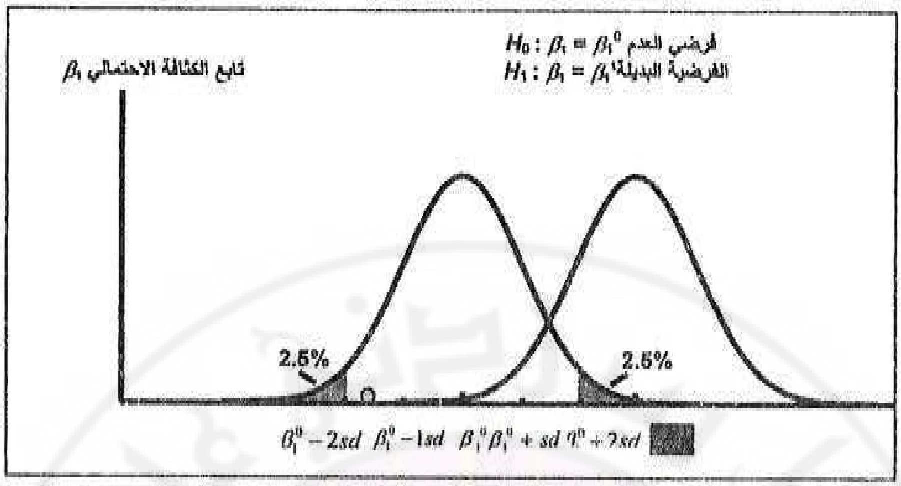
الشكل رقم (5-43)

على فرض سوف نرفض فرضية العدم لأن القيمة واقعة في المجال الأبعد، كما في الشكل(44-6).



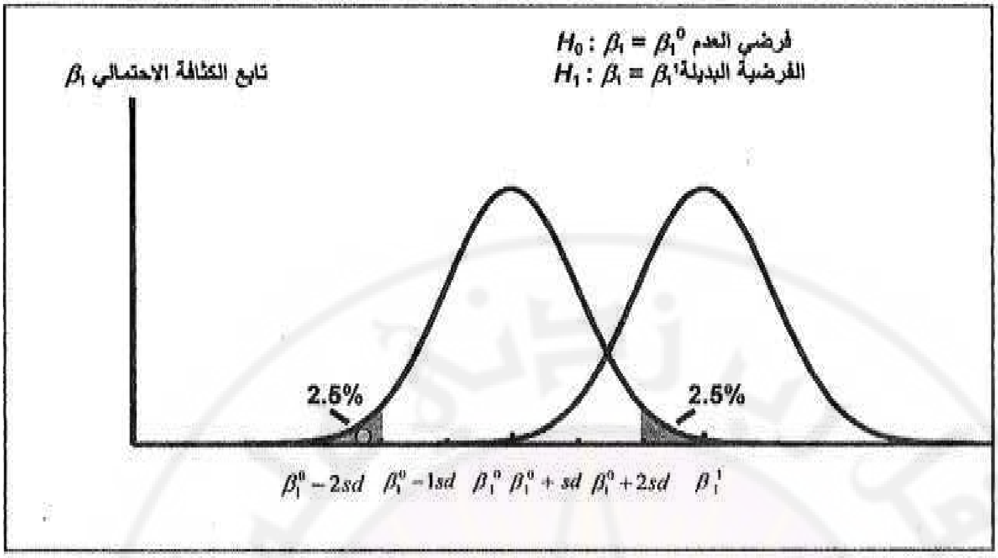
الشكل رقم (5-44)

أما بالنسبة للحالة المعروضة بالشكل رقم (5-45) سوف نقبل فرضية العدم .



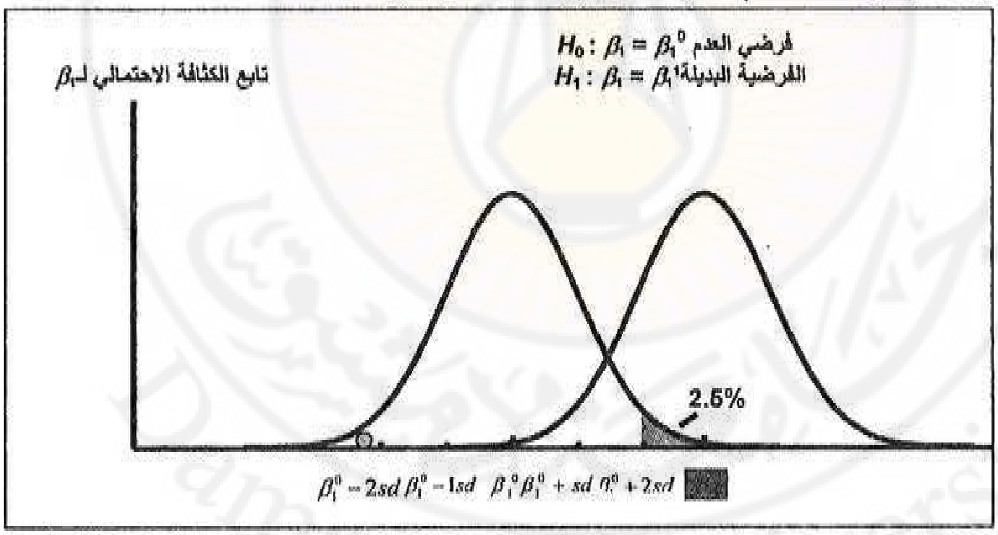
الشكل رقم (5-45)

إلا العينة هنا تشير إلى مسألة في غاية الأهمية. هي أن القيمة تقع في منطقة الرفض لـ β_1 ، و بالوقت نفسه يمكن أن تقع في منطقة β_1. ولكن منطقة الرفض لـ H_0 بالمقارنة مع H_1 لن يكون ممكناً. للخروج من ذلك المأزق نأخذ مخرجات العينة نفسها H_0 ، ولكن بشكل مخالف لـ H_1 بشكل حازم. نجد أن إمكانية الحصول على عينة كما هو حاصل صغير جداً للفرضية البديلة H_1 مقارنة مع فرضية العدم H_0.



الشكل رقم (6-46)

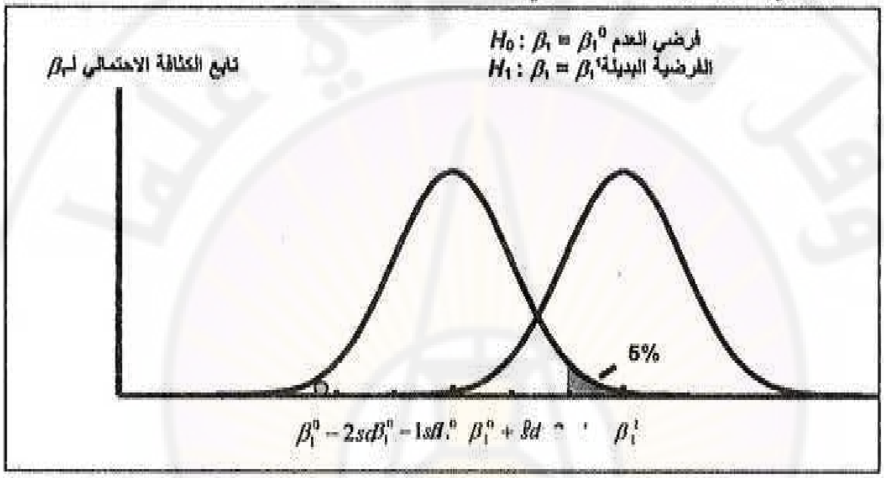
لهذا السبب الطرف الأيسر يجب أن يحدد منطقة الرفض لفرضية العدم H_0 . وسوف نستخدم الطرف الأيمن كمنطقة رفض فقط.



الشكل رقم (5-47)

إمكانية الوقوع في الخطأ من النـــوع الأول، إذا كانـــت فرضـــية العـــدم صحيحة من أجل 2.5% أحياناً إذاً يكون مستوى المعنوية لهذا الاختبار أيضـــاً 2.5%.

كما بينا سابقاً عند 5% مستوى معنوية الاختبار كما هو مبين في الطرف الأيمن حتى يتضمن 5% من الإمكانيات تحت المنحني. هذا يبـــدأ مـــن 1.645 خطأ معياري عن الوسط الحسابي.



الشكل رقم (48-5)

ماذا نريد أن نعمل هنا؟ للإجابة عن هذا السؤال نعود إلى الخلف للمقارنة ما بين الخطأ من النوع الأول و الخطأ من النوع الثاني.

عند 5% يوجد حظ كبير للوقوع في الخطأ من النـــوع الأول إذا كانـــت فرضية العدم صحيحة، ولكن خطر الوقوع في الخطأ من النوع الثاني أقـــل إذا كانت فرضية العدم غير صحيحة.

ملاحظة المنطق في انحدار الطرف الأيسر يعتمد فقط على β_1^1 فيمـــا إذا كانت أكبر من β_1^0. هذا لا يعني بالضرورة أن قيمة β_1^1 لها صفة محددة.

هكذا يمكننا أن نعمم أن الاختبار من اتجاه واحد في هذا الحالة يكون عندما الفرضية البديلة بكل بساطة للقيمة β_1^0 أكبر من β_1.

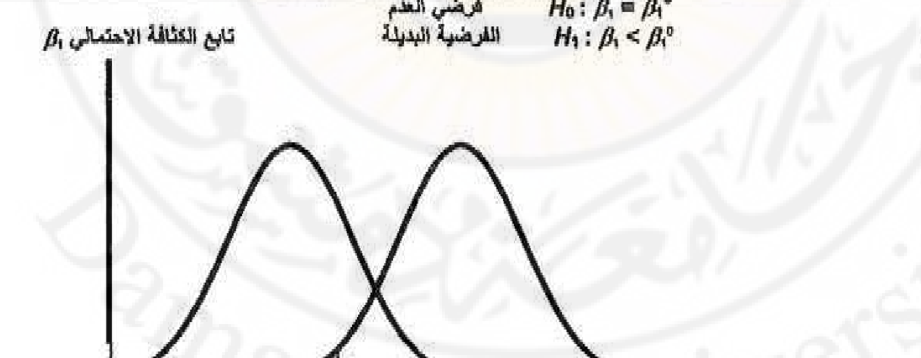
أي : فرضية العدم: $H_0 : \beta_1 = \beta_1^0$

والفرضية البديلة: $H_1 : \beta_1 > \beta_1^0$

إن مبرر استخدام الاختبار من طرف واحد لا يوجد قاعدة تحكمه، ونعتمد على النظرية الاقتصادية والتي بدورها تعتمد على التجربة في تقدير ذلك، أي β_1 أقل من β_1^0 . في بعض الأحيان تكون فرضية العدم على شكل معاكس لما عرض سابقاً بالاستناد إلى النظرية الاقتصادية أو التجربة ، تقع خارج القاعدة أي نقول أن β_1 أكبر من β_1^0. أي تصبح فرضية العدم والفرضية البديلة على النحو الآتي:

فرضية العدم: $H_0 : \beta_1 = \beta_1^0$ والفرضية البديلة: $H_1 : \beta_1 < \beta_1^0$

في هذا الحالة أيضاً يكون الاختبار من اتجاه واحد ، أحياناً مع الاختبار من الطرف الأيسر نستخدم منطقة الرفض. مع هذا التغير الحالة المنطقية هي السابقة نفسها.

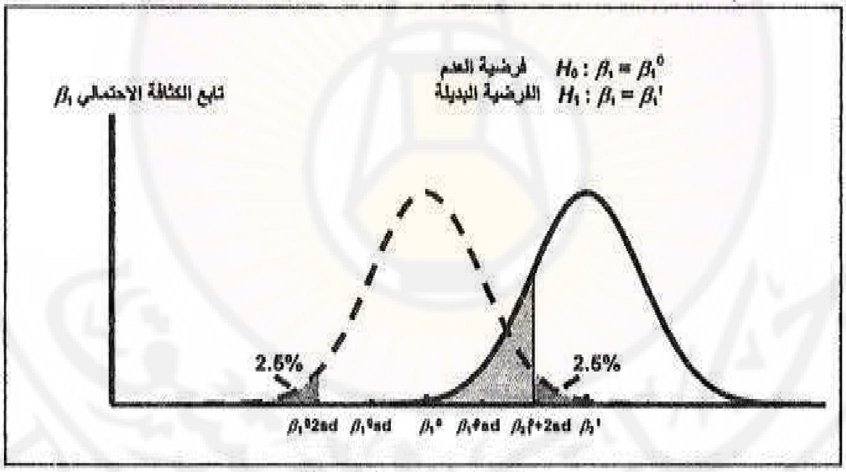


الشكل رقم (5-49)

نبين في الحالات اللاحقة كيف يمكن أن نتحسس الاختبار من اتجاه واحد، والمقارنة بين خطر الوقوع في الخطأ من النوع الأول والثاني.

نبدأ في العودة إلى الحالة عندما β_1 تأخذ إمكانية القيم β_1^0 و β_1^1 . على فرض أننا نستخدم الاختبار من اتجاهين بمستوى معنويـــة 5% للاختبـــار. إذا كانت فرضية العدم صحيحة H_0 هذا يعني خطر الوقوع في الخطأ من النـــوع الأول هو 5% .

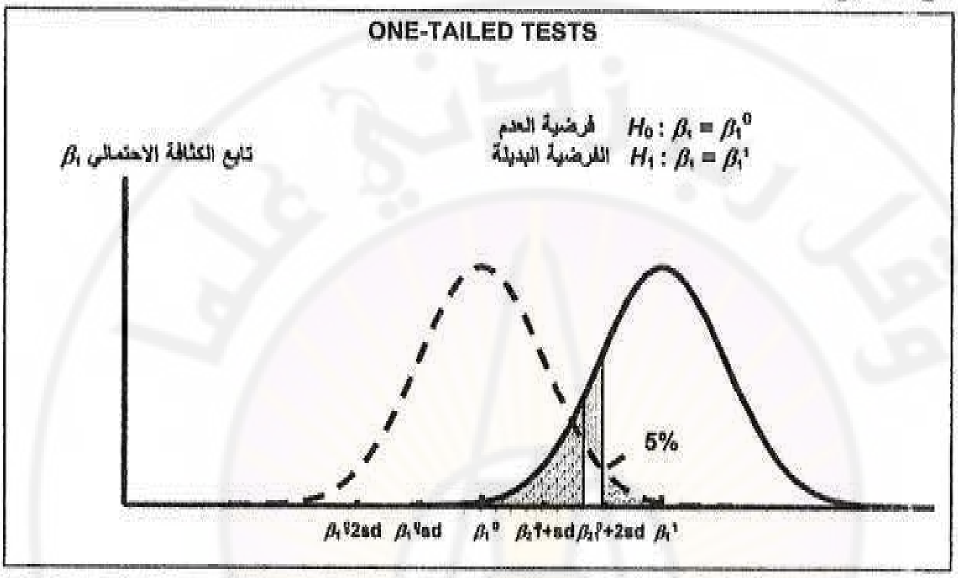
عندئذ مهما كانت H_0 يمكن أن تكون خطأ. في هذه الحالة احتمال عـــدم الرفض والوقوع في الخطأ من النوع الثاني؛ كما هو مبين في المنطقة الرمادية على الشكل (5-50). هذه المنطقة تحدد إمكانية التقدير، وتبين قبـــول منطقـــة فرضية العدم H_0. إذا كانت الفرضية البديلة صحيحة H_1.



الشكل رقم (5-50)

نفترض إننا نستخدم الاختبار من اتجاه واحد آخذين بعين الاعتبار مصلحة صحة ذلك، فمن غير المنطقي رفض فرضية العدم H_0 إذا كان التقـــدير مـــن

الطرف الأيسر. يمكننا التوسع للطرف الأيمن من أجل 5% مستوى معنوية الاختبار و خطر الوقوع في الخطأ من النوع الأول مازال 5% إذا كانت H_0 صحية . ولكن إذا كانت خطأ فإننا نقع في الخطأ من النوع الثاني باحتمال أقل من السابق.



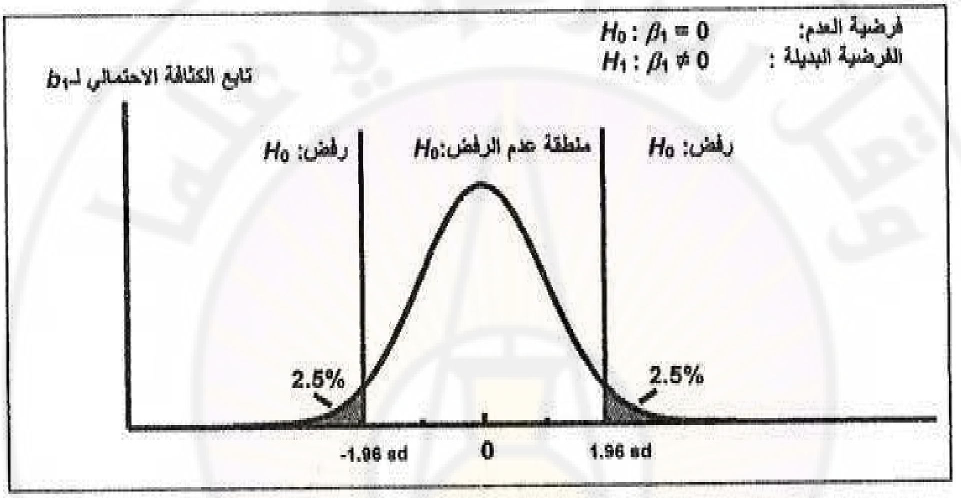
الشكل رقم (5-51)

إمكانية وقوع التقدير في منطقة القبول لـ H_0 يكون حالياً في المنطقة ذات الشكل هكذا مع ازدياد إمكانية الوقوع في الخطأ من النوع الأول (إذا كانت الفرضية صحيحة) ، نملك إمكانية أقل للوقوع في الخطأ من النوع الثاني (إذا كانت الفرضية خطأ). عندما الفرضية البديلة تساوي:

$H_1: \beta_1 > \beta_1^0$ أو $H_1: \beta_1 < \beta_1^0$ بعمومية أكثر (ونموذجية أكثر) للحالة ، سوف لا نستطيع رسم المنحي البياني. ومع ذلك نستطيع التأكد من استخدام الاختبار من اتجاه واحد، عندما نخفض إمكانية الوقوع في الخطأ من النوع الثاني (إذا كانت H_0 خطأ).

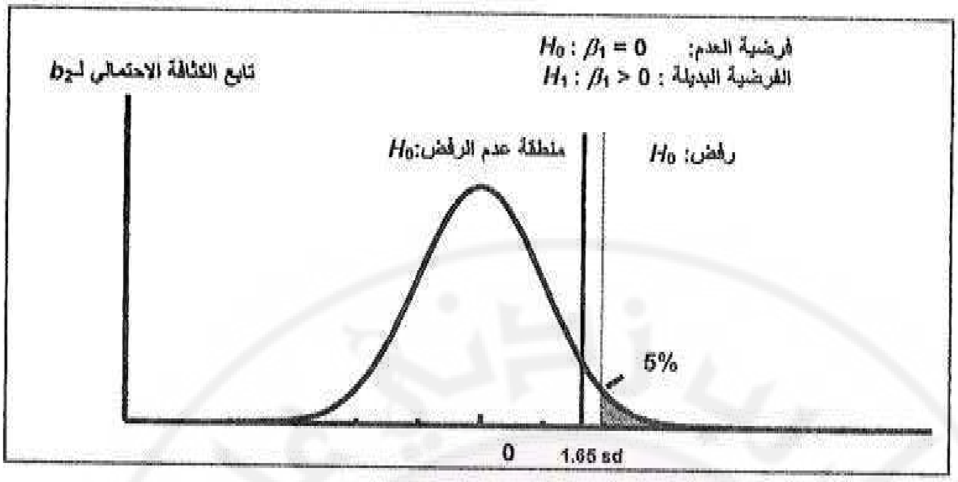
نناقش الحالة الأكثر شيوعياً: H_0: $\beta_1 = 0$. يكون الحدث عندما نرغب في عرض تأثير المتغير X على المتغير الآخر Y . نبدأ من فرضية العدم على صحة وتأثير رفض الفرضية H_0.

إذا كان الاختبار من اتجاهين بمستوى معنوية الاختبار 5% ، فإن تقديرنا يجب أن يكون حول 1.96 خطأ معياري، أو أقل من الصفر إذا كنا سنرفض H_0 .



الشكل رقم (5-52)

على أي حال إذا استطعنا تبرير استخدام الاختبار من اتجاه واحد على سبيل المثال : H_0: $\beta_1 > 0$ يكون تقديرنا فقط هو 1.65 خطأ معياري حول الصفر.



الشكل رقم (5-53)

هذا يجعلنا نرفض فرضية العدم، وبهذا الشكل تكون Y فعلاً متأثرة بالمتغير X (على افتراض النموذج صحيحاً).

سابقاً، افترضنا أن الخطأ المعياري للتوزيع لـ b_1 معروف، و يخضع للتوزيع الطبيعي . وفي التجربة طبعاً الخطأ المعياري يكون مقدر للانحراف المعياري، و يخضع لتوزيع t ذو العلاقة بالتوزيع. عند أي مستوى معنوية القيمة الحرجة لـ t من اتجاه واحد أقل منها من اتجاهين. هكذا إذا كانت H_0 خطأ فإن خطر رفض الفرضية سيؤدي إلى الوقوع في الخطأ من النوع الثاني وسيكون صغيراً.

ملاحظات هامة في اختبار t ستودنت : معلوم أن التقدير الدقيق يعتمد على حجم العينة بالدرجة الأولى وعدد المتغيرات في النموذج (المعلمات) التي ينبغي تقديرها ، إلا أنه يستخدم عادة قواعد حسابية عند النظر إلى معادلة الانحدار المقدرة . فإذا كانت قيمة المعلمة المقدرة أكبر من ضعف حجم الخطأ المعياري المقدر للمعلمة، نستنتج في حال كان الاختبار من اتجاهين اختلاف

تقدير المعلمة معنوياً عن الصفر من أجل مستوى دلالة 5% مقابل الفرضية البديلة للاختبار أنها تختلف عن الصفر، والنتيجة رفض فرضية العدم. أما إذا كان من ثلاثة أمثال حجم الخطأ المعياري المقدر للمعلمة نرفض فرضية العدم عند مستوى دلالة 1% .

برهان ذلك: عند اختبار فرضية العدم $H_0 : b_1 = 0$ مقابل الفرض البديل : $H_1 : b_1 \neq 0$ نعتمد في ذلك على إنشاء حدي ثقة $\left(\hat{b}_1 \pm t_{n-2;0975}\hat{\sigma}_{\hat{b}_1}\right)$ فإذا كان مستوى الدلالة 5% ، و لم يتضمن مجال حدي الثقة القيمة صفر فإننا نرفض فرضية العدم. قيمة المعلمة كما هو معلوم من الممكن أن تكون موجبة أو سالبة، ولكن المقدار $t_{n-2;0975}\hat{\sigma}_{\hat{b}_1}$ دائماً موجب . إذاً يحتوي مجال حدي الثقة على الصفر إذا كان: $0 > \hat{b}$ عندما $0 < \left(\hat{b} + t_{n-2;0975}\hat{\sigma}_{\hat{b}_1}\right)$

أو : $\hat{b} > 0$ عندما $0 < \left(\hat{b} - t_{n-2;0975}\hat{\sigma}_{\hat{b}_1}\right)$

وهذه الشروط تكتب على النحو الآتي:

$$\hat{b} < 0 \quad \text{عندما} \quad 0 < \left(\frac{\hat{b}}{\hat{\sigma}_{\hat{b}}} < t_{n-2;0975}\right)$$

أو : $$\hat{b} > 0 \quad \text{عندما} \quad 0 < \left(\frac{\hat{b}}{\hat{\sigma}_{\hat{b}}} > t_{n-2;0975}\right)$$

يمكننا اختصار هذه الشروط بالعلاقة التالية: $\left|\frac{\hat{b}}{\hat{\sigma}_{\hat{b}}}\right| > t_{n-2;0975}$

هذا يعني إذا كانت القيمة المطلقة للمقدار $\left|\frac{\hat{b}}{\hat{\sigma}_{\hat{b}}}\right|$ أكبر من قيمة t من أجل $t_{n-2;0975}$ فإننا نرفض فرضية العدم . مع الأخذ بعين الاعتبار أن الاقتصاديين يتعاملون عادة مع عينات حجمها لا يقل عن 15 . فإذا كان n=15 فإن: $\left(t_{15-2;0975}\right) = 2.16$ أما إذا كان حجم العينة لانهائي أي: n=∞ فإن

$1.96 = (t_{\infty-2;0975})$. أي بنظرة سريعة إلى جدول t يتبين أن القيمة $t_j=0.975$ وبالتالي : $13 < j < \infty$ وبالتالي قيمة t الحرجة تقع بين 1.96 والقيمة 2.16 .

والقاعدة تصبح على النحو الآتي: كلما زادت النسبة $\left|\frac{\hat{b}}{\hat{\sigma}_{\hat{b}}}\right|$ عن 2 بالقيمة المطلقة نرفض فرضية العدم . و بالطريقة نفسها نثبتها لثلاثة أمثال.

وبالطريقة نفسها نثبت ذلك على الاختبار من اتجاه واحد، حيث إن الشرط الضروري لرفض فرضية العدم في أي الحالتين هو: $\left|\frac{\hat{b}}{\hat{\sigma}_{\hat{b}}}\right| > t_{n-2;095}$

5-7- اختبار جودة التوفيق TEST OF GOODNESS OF FIT

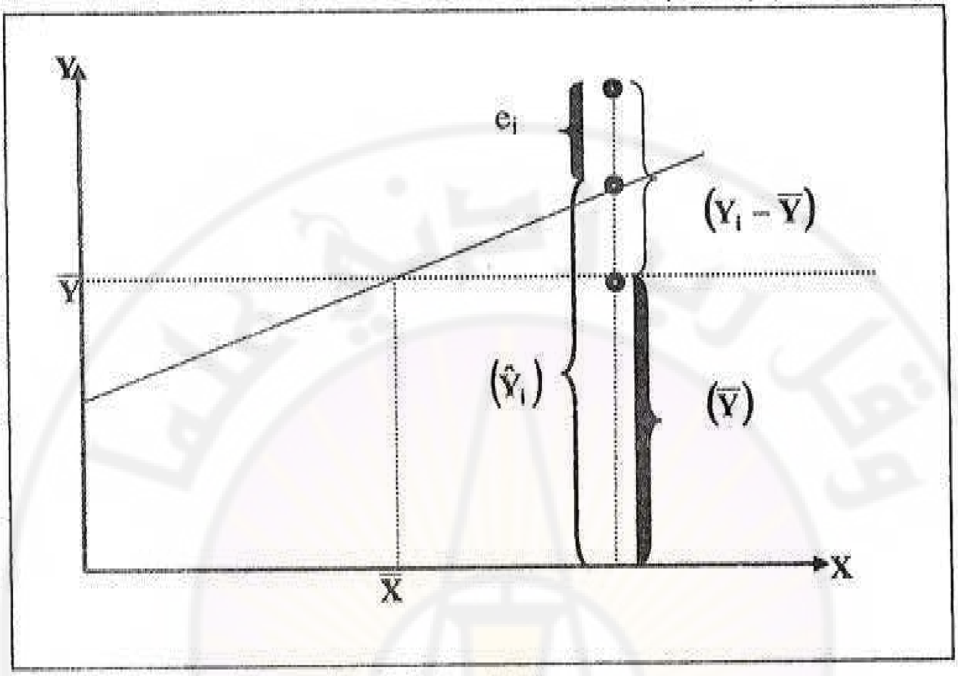
أسباب انحرافات مشاهدة العينة الخاص بالمتغير التابع (Y_i)

Decomposition of the sample variation of (Y_i)

من المعروف تغير المتغير (Y_i) يتبع تغيرات المتغير (X_i) ، فالتباين يحتسب على أساس تشتت قيم Y_i عند ثبات قيمة X_i من عينة لأخرى بينما يتحقق الانحراف في قيم المتغير التابع عندما تتغير قيم (X_i) من مشاهدة إلى أخرى ضمن العينة الواحدة . ويعود عدم ثبات قيم (Y_i) عند مستوى معين إلى سببين رئيسيين الأول هو تأثير المتغير المستقل في القيمة $[E(Y_i)]$ ، والثاني تأثير الخطأ الذي يتخذ قيماً مختلفة من مشاهدة إلى أخرى. إذاً لابد من تحليل الانحرافات لمعرفة مدى مساهمة المتغير المستقل (X_i) والخطأ المعياري (u_i) في الانحراف الحاصل.

عادة تقاس قيمة انحرافات Y_i على أساس $\bar{Y}$ ، وهو عبارة عن ذلك المستقيم الموازي للمحور الأفقي لـ (X) وهو يمثل قيم Y_i في حالة انعدام التأثيرات المختلفة، وعندئذٍ تكون قيم Y_i متساوية وتساوي متوسطها $(\bar{Y})$.

وبالتالي فإن الانحراف لكل مشاهدة من المشاهدات (i) عبارة عن $\left(Y_i - \overline{Y}\right)$ كما في الشكل رقم (5-54).



الشكل رقم(5-54)

من الشكل (5-54) نكتب: $\left(Y_i - \overline{Y}\right) = \left(\hat{Y}_i - \overline{Y}\right) + e_i$

لكن كما هو معلوم لدينا أن مجموع انحرافات القيم عن متوسطها يساوي الصفر من أجل الخروج من هذا المأزق نأخذ مجموع مربع الانحرافات، والذي سوف نطلق عليه الانحرافات الكلية Total Variation .

$$\sum\left(Y_i - \overline{Y}\right)^2 = \sum\left[\left(\hat{Y}_i - \overline{Y}\right) + e_i\right]^2$$
$$= \sum\left[\left(\hat{Y}_i - \overline{Y}\right)^2 + 2\sum\left(\hat{Y}_i - \overline{Y}\right)e_i + \sum e_i^2\right]$$

الآن نستطيع أن نبرهن أن : $\sum\left(\hat{Y}_i - \overline{Y}\right)e_i = 0$ على النحو الآتي:

$$\sum\left(\hat{Y}_i - \overline{Y}\right)e_i = \sum\left[\left(\hat{\beta}_1 + \hat{\beta}_2 X_i - \overline{Y}\right)e_i\right]$$
$$= \hat{\beta}_1 \sum e_i + \hat{\beta}_2 \sum X_i e_i - \overline{Y}\sum e_i$$

نعلم أن: $\sum X_i e_i = 0 \Rightarrow \sum e_i = 0$ وبالتالي: $\sum(\hat{Y}_i - \overline{Y}) = 0$

نستطيع أن نكتب :

$$\sum(Y_i - \overline{Y})^2 = \sum(\hat{Y} - \overline{Y})^2 + \sum e_i^2$$

الانحرافات الكلية TSS الانحرافات المفسرة ESS الانحرافات غير المفسرة RSS

TSS = ESS + RSS

يمكن أن نكتب العلاقة السابقة بالشكل التالي :

الانحرافات المفسرة

$$\sum(\hat{Y}_i - \overline{Y})^2 = \sum(\hat{\beta}_1 + \hat{\beta}_2 X_i - \overline{Y})^2$$

$$= \sum(\overline{Y} - \hat{\beta}_2 \overline{X} + \hat{\beta}_2 X_i - \overline{Y})^2$$

$$= \sum(\hat{\beta}_2 X_i)^2$$

$$\sum(\hat{Y} - \overline{Y}) = \sum(\hat{\beta} X_i)^2 = \hat{\beta}_2^2 \sum X_i^2$$

وتعتمد نتائج تحليل انحرافات (Y_i) من أجل تقييم جودة توفيق خط الانحدار للعينة (goodness of fit) .

- يعرف مجموع مربع الانحرافات الكلية لقيم المتغير التابع y عن الوسط الحسابي $\overline{Y}$ بالاختلافات الكلية في المتغير التابع ، والذي نحاول تفسيره عن طريقة معرفتنا بالاختلافات الكلية لمتغير أو أكثر من المتغيرات المستقلة، ونرمز له بالرمز TSS .

- الانحراف المفسر : فهو ذلك الجزء من الانحراف الكلي في المتغير التابع (Y_i) والذي تم تحديده بمعادلة الانحدار، ونرمز له بالرمز ESS ويقصد به مجموع مربع لانحرافات في المتغير التابع (Y) والمحددة بالانحدار .

أما الانحراف غير المفسر فهو ذلك الجزء من التباين لم نتمكن من تحديده عن طريق العلاقة بين المتغير التابع Y والمتغير المستقل X ، ويطلق عليه اسم البواقي (Residuals) ويرمز له عادة بـــ RSS .

5-8- معامل التحديد : **coefficient of determination**

وهو عبارة عن مقياس لقياس جودة توفيق منحنى الانحدار (خط الانحدار) ونرمز له بـــ R^2 وهو عبارة عن نسبة الانحرافات المفسرة من الانحرافـــات الكلية . وهو دائماً موجب، وتتراوح قيمته بين الصفر والواحد، أي أن :

$$R^2 = \frac{ESS}{TSS} = \frac{\hat{\beta}_2^2 \sum x_i^2}{\sum x_i^2} \qquad (142)$$

$$\hat{\beta}_2 = \frac{\sum x_i Y_i}{\sum x_i^2}$$

$$R^2 = \frac{\hat{\beta}_2 \sum x_i Y_i}{\sum y_i^2} \qquad 0 \le R^2 \le 1$$

كذلك يمكن التعبير عنه بشكل آخر :

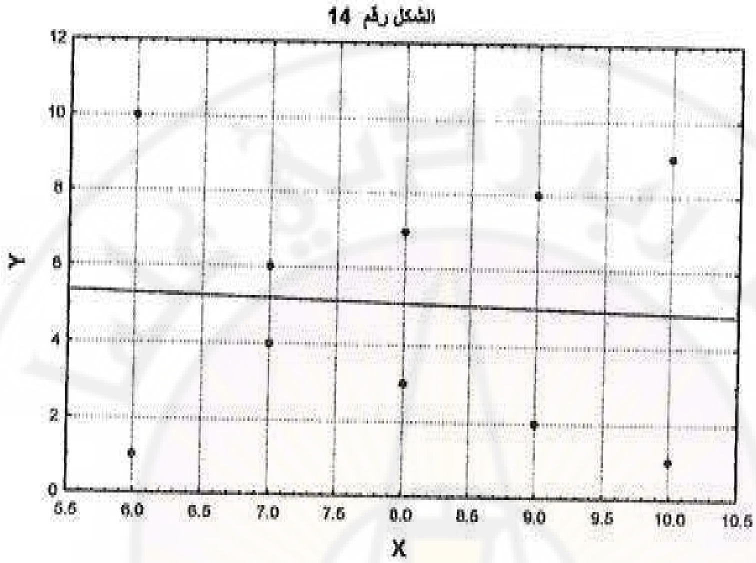
$$R = \frac{\sum(Y_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} = \frac{\sum(\hat{Y} - \bar{Y}_i)^2}{\sum(Y_i - \bar{Y})^2} + \frac{\sum e_i^2}{\sum(Y_i - \bar{Y})^2}$$

$$1 = R^2 + \frac{\sum e_i^2}{\sum y_i^2} \Rightarrow R^2 = 1 - \frac{\sum e_i^2}{\sum y_i^2} \qquad (143)$$

$$1 = R^2 + \frac{\sum e_i^2}{\sum y_i^2} \Rightarrow R^2 = 1 - \frac{\sum e_i^2}{\sum y_i^2}$$

من المعادلة أعلاه يتبين أنه في حالة انعدام تأثير (X_i)على (Y_i) وتصبح قيمة $(\hat{\beta}_2)$مساوية للصفر، وبالتالي تصبح قيمة (R^2) مساوية أيضاً للصـــفر . وهذا ما يمكن شرحه أيضاً بواسطة المعادلة أعلاه حيث إنه في حالة اتخاذ خط

انحدار العينة وضعاً أفقياً تصبح قيم (e_i) مساوية لقيم (y_i) وبالتالي تصبح قيمة (R^2) مساوية للصفر .



الشكل رقم (5-55)

2- وإذا علمت ان حجم العينة هو (n=20) . ماذا تستنتج من هذا الحساب؟

13- عدد قواعد التباين المشترك.

14- في بلد معين إن الضريبة التي تدفع من قبل الشركة هي T وهي محدودة بالعلاقة:

$$T = -1.2 + 0.2P - 0.1I$$

حيث إن: P هي الأرباح و I هو رأس المال للمستثمر. الطرف الثالث الموجود الذي يبين أثر توظيف الأموال في تحفيز (S) المبيعات وكل المتغيرات مقاسة بمليون وحدة نقدية كمعدلات سنوية. احسب التباين المشترك cov(S,T)، والتباين المشترك cov(S,I) وذلك لعينة مؤلفة من أربع شركات، كما هو مبين أدناه في الجدول وتأكد من صحة العلاقة التالية:

$$Cov(S,T) = 0.2Cov(S,P) - 0.1Cov(S,I)$$

اشرح بشكل تفصيلياً ، لماذا ينبغي أن تكون هذه الحالة.

Firm	*Sales*	*Profits*	*Investment*	*Tax*
1	100	20	10	1.8
2	50	9	4	0.2
3	80	12	4	0.8
4	70	15	6	1.2

15- باستخدام البيانات في التمرين السابق احسب:

cov (P,S) , var (I) , var (P) , var (T)،

وتحقق من العلاقة التالية: $Var(T) = 0.04Var(P) + 0.01Var(I) - 0.04Cov(P,I)$

الفصل الثالث

بناء نموذج تحليل الانحدار الخطي البسيط

3-1- مقدمة:

هناك العديد من طرق الاقتصاد القياسي التي يمكن استخدامها للحصول على تقديرات لمعاملات العلاقات الاقتصادية من المشاهدات الإحصائية وأن أشهر هذه الطرق هي طريقة المربعات الصغرى(OLS)

The method of ordinary classical least squares

وتتميز هذه الطريقة عن غيرها من الطرق الأخرى بالمزايا التالية:

1- المعاملات المقدرة حسب (OLS) تتمتع بخصائص نموذجية، مثل عدم التحيز و أصغر تباين وغيرها من الخصائص الأخرى نستعرضها لاحقاً.

2- تتصف العمليات الحسابية لهذا النموذج بالبساطة والوضوح مقارنة مع الطرق الأخرى، ولا تحتاج إلى كم كبير من البيانات الإحصائية.

3- تم اختبار هذه الطريقة في مجالات عديدة، وأعطت نتائج مرضية، على الرغم من التطور التكنولوجي في مجال الحاسبات الإلكترونية، إلا أن هذه الطريقة بقيت واحدة من أسهل الطرق المتبعة في تكوين علاقات النموذج.

4- آلية تطبيق هذه الطريقة سهلة وواضحة.

5- أغلب الطرق الأخرى المتبعة تتضمن ضمن تطبيقها تطبيقاً لهذه الطريقة، مع أخذ التعديلات بما يتناسب والطريقة المتبعة، ما عدا طريقة التقدير بأعظم إمكانية وباستخدام بيانات كاملة.

لنستعرض مثال تطبيقي من الواقع لو أخذنا نظرية الطلب في أبسط صورها، حيث يوجد علاقة موجبة بين الكمية المطلوبة للسلعة وسعرها. وعلى

فرض بقاء جميع العوامل الأخرى ثابتة، فإذا ازداد الطلب على السلعة يزداد سعرها، والعكس صحيح.

نستخدم خطوات الاقتصاد القياسي التي استعرضناها سابقاً من أجل تحديد النموذج الرياضي للطلب. نحدد المتغير التابع وهو الطلب هنا، والمتغير المستقل وهو السعر. أما النظرية الاقتصادية فتزودنا بالمعلومات التالية حول دالة الطلب:

1- المتغير التابع هو كمية الطلب على السلعة (Y) والمتغير التفسيري هو متغير السعر (X) أي: شكل علاقة على النحو الآتي: $Y= f(X)$.

2- النظرية الاقتصادية لا تحدد طريقة دراسة الطلب هل من خلال نموذج بمتغير واحد أو باستخدام علاقة متعددة المتغيرات. للتبسيط سوف نختار نموذجاً بمتغير واحد، ثم نحاول تطويره إلى نموذج متعدد المتغيرات.

3- النظرية الاقتصادية لا تبين لنا ما هو شكل النموذج المزمع الوصول إليه ؟ هل نستخدم معادلة خطية أم معادلة لا خطية ؟ كتب التحليل الاقتصادي في كثير من الأحيان تمثل دالة الطلب من خلال مستقيم هابط إلى الأسفل. ولكن الاقتصاد القياسي عليه أن يقرر شكل هذه الدالة من خلال طريقته خاصة. وللتبسيط نفترض العلاقة بأبسط أشكالها الرياضية الممكنة هي علاقة خطية من الشكل: $Y_i = b_0 + b_1 X_i$

هذا الشكل يدل على أن العلاقة سببية الاتجاه، أي أن السعر هو السبب فيما يطرأ على الكمية من تغيرات، والعكس ليس صحيحاً. ولا بد من تحديد معاملات هذه العلاقة وهي $(\hat{b}_1, \hat{b}_0)$. أما فيما يتعلق بإشارة ومستوى حجم $(\hat{b}_0)$ هو عبارة عن ثابت التقاطع الرأسي أو الجزء المحصور بين نقطة تقاطع المحني مع المحور(Y) ونقطة المبدأ. فإنه من المتوقع أن تكون قيمة المعامل إما صفر وإما موجبة. أي حتى لو كان السعر مساوياً الصفر فإنه، وهناك كمية مطلوبة من السلعة المعينة. ومن المعتاد أن تكون موجبة في دالة الطلب، فإذا

ظهرت إشارة سالبة فلا معنى للكميات السالبة في الاقتصاد ما لم يكن هناك مبرراً اقتصادياً لتوقع ظهور إشارة سالبة للثابت. وإشارة $(\hat{b}_1)$ تحدد مرونة الطلب بالنسبة للسعر. أما إشارة $(\hat{b}_1)$ فهي سالبة لأن ميل منحني الطلب بشكل عام متجه نحو الأسفل. أما فيما يتعلق بمستوى حجم $(\hat{b}_0)$ ، فيجدر ملاحظة أن $(\hat{b}_0)$ هو جزء من مرونة الطلب بالنسبة للسعر أي: $E_P = \frac{\delta Y}{\delta X} * \frac{X}{Y}$

ومن تابع الطلب نجد أن $(\hat{b}_1)$ سالبة لأن: $E_p = \frac{\delta Y}{\delta X} = -b_1$.

من تابع الطلب نجد أن $\hat{b}_1$ سالبة لان: $\frac{\delta Y}{\delta X} = -b_1$

في حال استخدام معادلة خط الانحدار في حساب المرونة فإننا نستخدم $(\hat{b}_1)$ ومتوسط السعر $(\overline{X})$ ومتوسط الكمية المعروضة $(\overline{Y})$ من النموذج نجد أن:

$$E_p = -\hat{b}_1 * \frac{\overline{X}}{\overline{Y}}$$

معلوم لدينا أن: $\overline{Y} = \hat{b}_0 + \hat{b}_1\overline{X}$

نعوض في معادلة المرونة فنجد: $E_p = \frac{-\hat{b}_1\overline{X}}{\hat{b}_0 + \hat{b}_1\overline{X}}$

إذا كان $\hat{b}_1 > 0$ نستنتج أن:

1- الطلب يكون مرنا $E_p > 1$

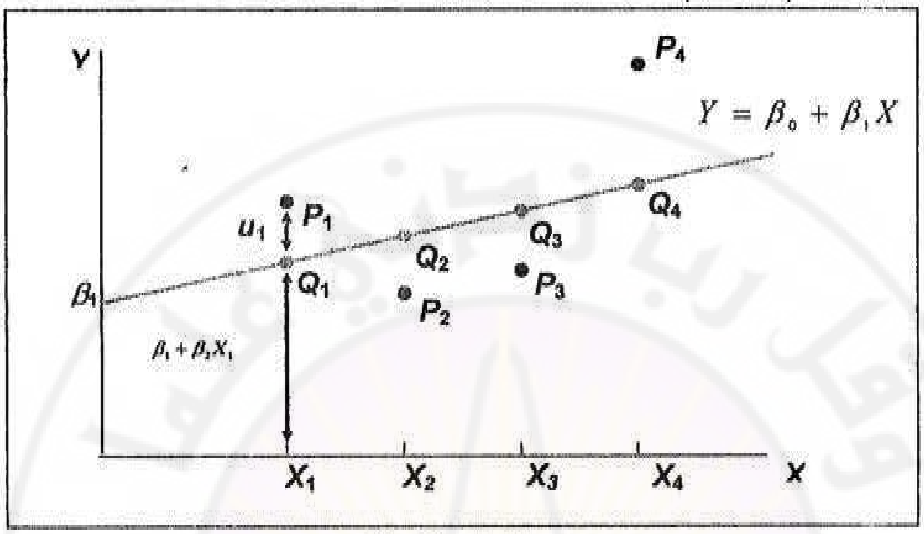
2- الطلب يكون غير مرنا $E_p < 1$

3- الطلب عديم المرونة $E_p = 0$

بعد هذه المقدمة النظرية لننطلق إلى معالجة هذا النموذج من المعادلة الخطية البسيطة التالية :

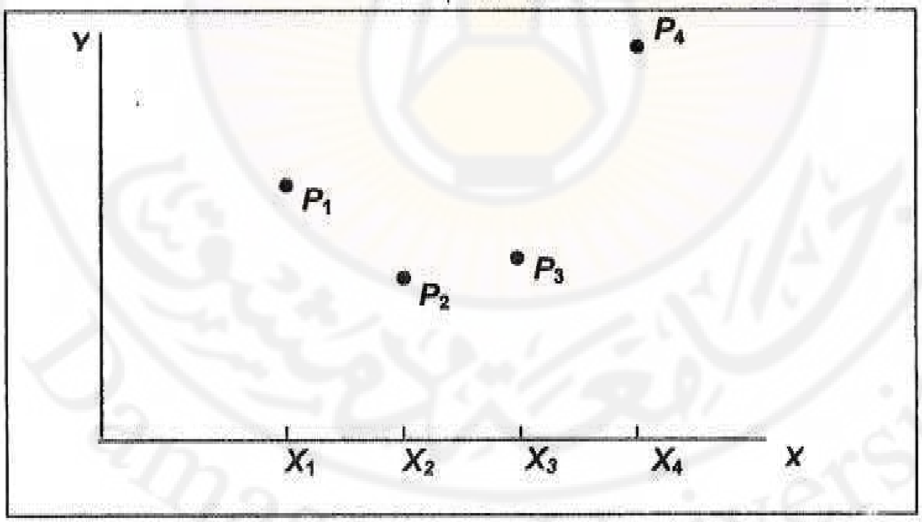
$$Y_i = b_0 + b_1 X_i + U_i \qquad (1)$$

كل قيمة للمتغير قيمة غير عشوائية Y تحدد بالعلاقة : $b_0 + b_1X$ و قيمة عشوائية تحدد بالعنصر u . القيمة الأولى من المشاهدة يمكن تحليلها إلى عنصرين (مركبتين).



الشكل (3-3)

نحاول ملاحظة فقط النقط P على الرسم فنجد:

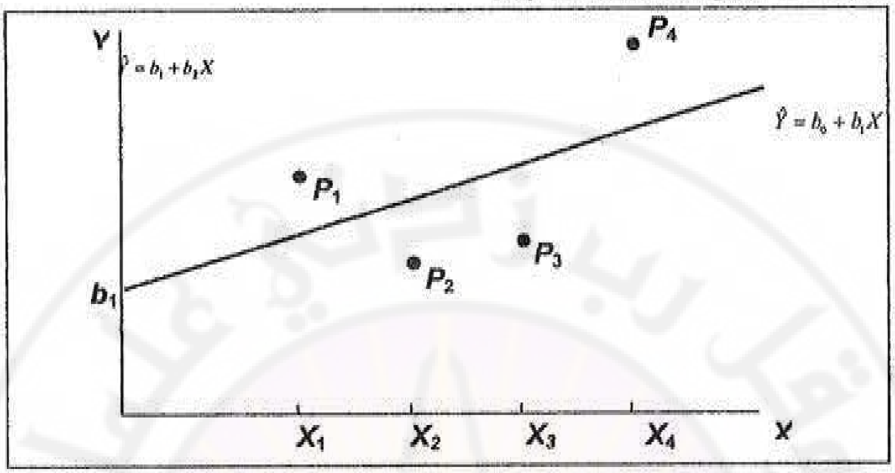


الشكل (3-4)

يمكننا توفيق مستقيم تقريبي لهذه النقاط يمثل بالعلاقة الخطية التالية:

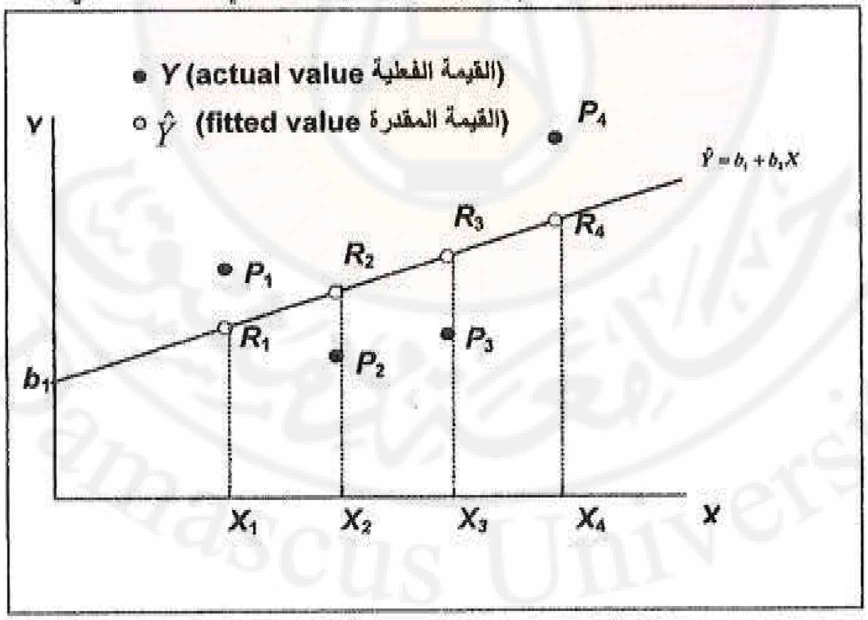
$.Y = \beta_0 + \beta_1 X$

إذا كتبنا علاقة المستقيم بهذا الشكل: $\hat{Y} = b_0 + b_1 X$ فعندئــذ b_0 تكــون تقديراً لـــ β_0, b_1 تقديراً لـــ β_1.



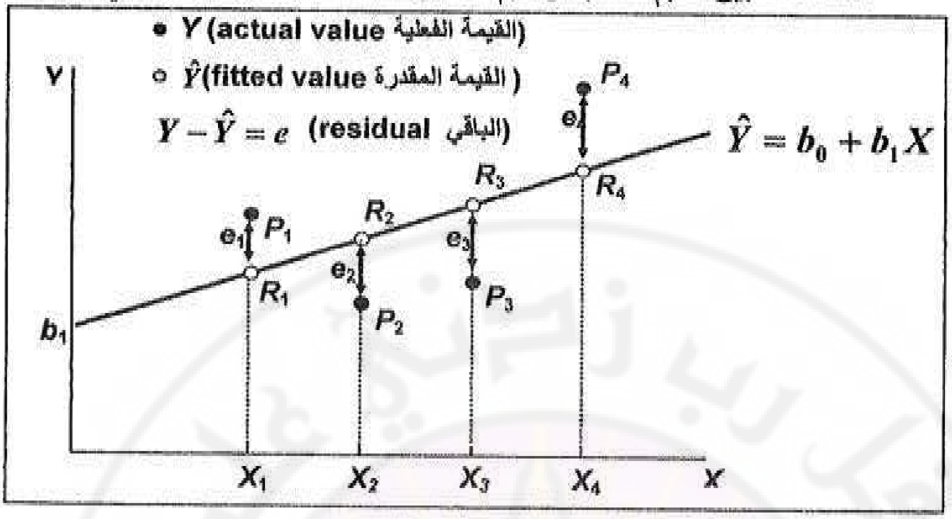
الشكل (3-5)

الخط يسمى النموذج المقدر ، وقيم Y المتنبأ بها بهذه العلاقة تسمى القيمة المقدرة لـــ $\hat{Y}$. وهذا يعطينا قيم نمثلها بالنقاط R كما في الشكل التالي:



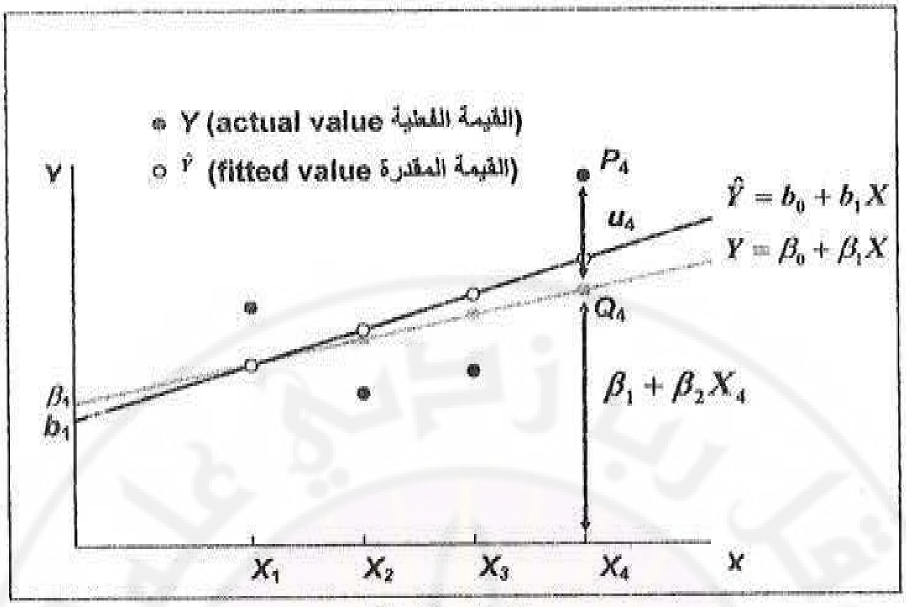
الشكل (3-6)

الاختلاف بين القيم الفعلية والقيم المقدرة لـ Y تسمى البواقي.



الشكل (7-3)

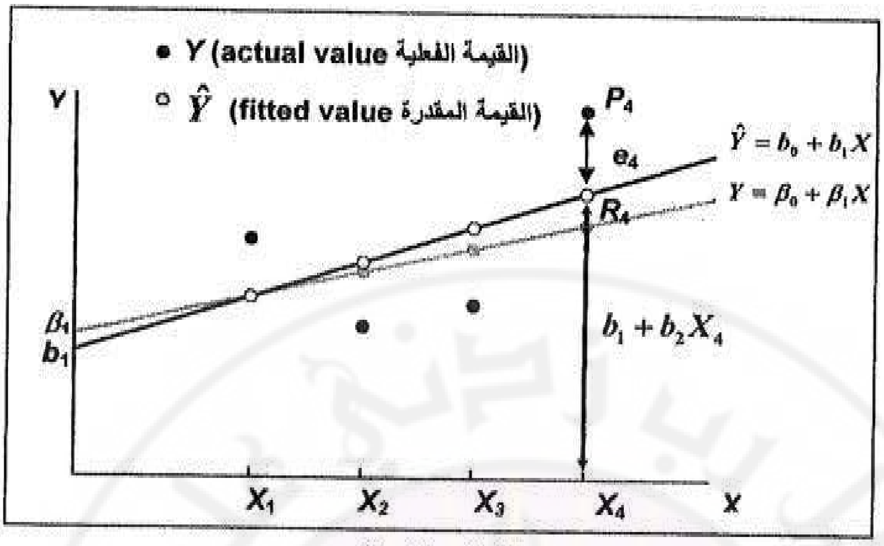
مع ملاحظة أن قيم البواقي ليست هي قيم الخطأ العشوائي نفسها. الشكل البياني يبين حقيقة العلاقة كمنحني التقدير. الخطأ العشوائي لكل مشاهدة ينقسم إلى مركبتين ، مركبة غير العشوائية للعلاقة الحقيقية و مركبة القيم الفعلية للمشاهدة. البواقي عبارة عن الفرق بين القيمة الفعلية والقيمة المقدرة. إذا كان التقدير ملائماً يكون جيداً، فالبواقي والخطأ العشوائي سوف يكونان متماثلان ، اي يجب أن يكونان محافظين على خصائصهما.



الشكل (3-8)

سوف نستخدم كلا المنحنيين في التحليل، أي كل تحاليل مركبات القيم Y. سنستخدم المثال المكون من أربع مشاهدات. نستخدم العلاقة النظرية Y نجد أنها مكونة من مركبتين مركبة متضمنة في العلاقة $b_0 + b_1X$ وأخرى مركبة الخطأ العشوائي؛ كما موضح في الشكل أعلاه .

هذا هو التحليل العملي الذي سوف نستخدمه.



الشكل (9-3)

والتحليل الآخر يكون مع الإشارة إلى منحني التقدير. كل حالة من المشاهدات، القيم الفعلية لـ Y تكون مساوية القيم التقديرية بالإضافة إلى البواقي. و هو التحليل العملي الذي سوف نستخدمه في هذه الدراسة. هذا التحليل يبقى نظرياً لأننا لا نعرف قيم المعاملات β_0 أو β_1 أو قيم الخطأ العشوائي، ولا بد من استخدام في تحليلنا خصائص معامل التحديد.

للبدء سوف نرسم المنحني المقدر الناتج عن الحد الأدنى لمربع البواقي RSS.

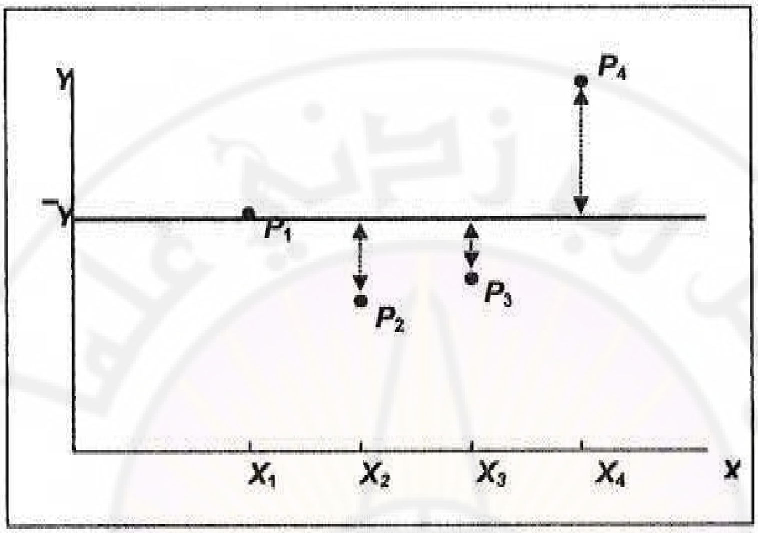
الشكل التالي يبين الحد الأدنى لمربع البواقي:

$$RSS = \sum_{i=1}^{n} e_i^2 = e_1^2 + ... + e_n^2$$

والسؤال الذي يطرح نفسه ، لماذا لا نأخذ مجاميع البواقي الأقل دون تربيع؟ لماذا مربع البواقي؟ لماذا ليس تماماً مجموع البواقي؟

الجواب سوف نحصل على المظهر الدقيق للمنحني من خلال رسم الخط الأفقي، الذي يمثل متوسط قيم المتغير Y . فعندئذ نجد أن مجموع البواقي سوف

يساوي الصفر. لأن مجموع القيم السالبة يساوي القيم الموجبة، ولذلك نستخدم مربع البواقي. و هناك طرق أخرى في حل هذه المشكلة. معيار القيمة الأقل للمربعات تملك جاذبية في اشتقاق التقدير، وهذه الجاذبة مبينة على خواص لها شروط مرضية.



الشكل (3-10)

لكن لتطبيق ذلك لا بد من تحقيق بعض الفروض الأساسية ، وسوف نستعرض هذه الفروض في الفصل اللاحق.

3-3- تقدر معالم الانحدار Estimation of Regression coefficients

هناك نوعان من التقدير الإحصائي، الأول (تقدير النقطة Point Estimation) والثاني تقدير المجال Introvert Estimation وينصب تقدير النقطة على تحديد قيمة مفردة لكل معلمة من معالم المجتمع.

أما تقدير االمجال فإنه يحدد المجال الذي تقع ضمنه المعلمة الحقيقية بدرجة ثقة معينة. وفي الواقع إن تقدير االمجال عبارة عن عملية تكميلية لتقدير النقطة، وذلك لأن تقدير النقطة يتم على أساس العينة. الأمر الذي يسبب في

تمارين غير محلولة

1- ماذا يقصد بالاختبار الإحصائي؟

2- تحدث عن الفرضية الإحصائية؟

3- ماذا يقصد بأخطأ المعاينة؟

4- تحدث عن خطوات اختبار معنوية المعلمات؟

5- تحدث عن الخطأ من النوع الأول والخطأ من الثاني؟

6- على فرض لدينا عينة مكونة من 70 مشاهدة لدراسة انحدار الدخل الساعي على التدريب طبقاً للعلاقة:

$$EARNINGS = B_0 + B_1\ TRAINING + U$$

لاختبار فرضية العدم $H_0: B_1 = 0$. مقابل الفرضية البديلة $H_1: B_1 \neq 0$ من أجل 5% و 1% مستوى دلالة ما هو قرارك بخصوص العلاقات التالية:

1. IF $B_2 = 0.30$, $S.E.(B_1) = 0.12$?
2. IF $B_2 = 0.55$, $S.E.(B_1) = 0.12$?
3. IF $B_2 = 0.10$, $S.E.(B_1) = 0.12$?
4. IF $B_2 = -0.27$, $S.E.(B_1) = 0.12$?

7- تحدث عن أهمية الاختبارات الإحصائية.

8- لنفترض أن إنتاج محصول المشمش (Y_i) مقاس بالطن يعتمد على متغير واحد (X_i) عدد ساعات العمل المبذولة للعناية بالأشجار، وقد توفرت لدينا البيانات التالية لعدة مزارع:

Yi	Xi	Xy
3	3	9
4	5	20
6	6	36
11	6	66

المطلوب :

1- قدر معلمات النموذج $\hat{\beta}_0,\hat{\beta}_1$.

2- استخدم الانحراف المعياري (Se) لاختيار مدى صلاحية كل من $\hat{\beta}_0,\hat{\beta}_1$ لاختيار معنوية النموذج .

3- استخدم اختبار t لتحديد معنوية النموذج ، اشتق فترة ثقة لكل من معلمات النموذج التقديري.

4- أوجد معامل التحديد ومعامل الارتباط r,r^2 .

5- أوجد اختبار F .

6- ما هو المقصود بجدول تحليل التباين(ANOVA) ، كون جدول ANOVA

9- توفر لدينا المعلومات الآتية عن إنتاج المشمش:

$$SSE = \sum\left(\hat{Y}-\bar{Y}\right)^2 = \sum \hat{Y}_1^2 = 19.44, \quad n=4$$

$$SSE = \sum e_i^2 = 17.84 \quad ,k=2$$

$$SST = \sum\left(Y_i-\bar{Y}\right)^2 = \sum \hat{Y}_1^2 = 38.00$$

$$k-1=1 \qquad \sum X_i Y_i = 11$$

$$n-2=2 \qquad \sum X_i^2 = 6;$$

$$n-1=3 \qquad \sum Y_i^2 = 38$$

من هذه المعلومات كون جدول ANOVA .

10- لنفترض أن $\hat{\beta}$ =29.48 ، وان $\hat{\sigma}$ =36.9 ، والمطلوب اختبار كون β الحقيقة تساوي 25.0 وبمستوى معنوية 5%.

11- لنفترض بأن دالة العرض لعينة حجمها (700) مشاهدة مقدرة كما يلي :

$$\hat{Y} = 100 \ \ +4.00X,$$
$$S.E : (20) \quad (1.5)$$

والمطلوب: اختبر معلمات النموذج بالاختلاف عن الصفر حسب فرضي العدم ، والفرضية البديلة :

$$H_0 : \beta_1 = 0 ; \quad H_1 : \beta_1 \neq 0$$
$$H_0 : \hat{\beta}_0 = 0 ; \quad H_0 : \hat{\beta}_0 \neq 0$$

12- نفترض أن لدينا عينة حجمها (20) وكان تقدير دالة الاستهلاك هو :

$$C=100 + 0.70Y_i$$
$$S.D(57.5) \ (0.21)$$

حيث إن: Y= الدخل ، C= الاستهلاك

والمطلوب: اختبر معلمات النموذج.

13- نفترض أن لدينا عينة مكونة من (12) مشاهدة ، وكانت تقديراتنا لدالة العرض كالآتي :

$$Y=33.75+3.25X$$
$$S.E(8.28) \ (0.89)$$

والمطلوب: اختبر معلمات النموذج.

14- من النموذج الخطي البسيط الآتي :

وان : $Y_i = \alpha + \beta X_i + U_i , i = 1,2,3.......,n$

وحيث إن (U_i) تخضع للتوزيع طبيعي وان :

$$E(U_i) = 0; \ E(U_i)^2 = \sigma_u^2 ; E(U_i U_j) = 0, \quad i \neq j$$

والمطلوب هو :

أ- أوجد حدود الثقة لقيمة (Yi) الوسطية والمرتبطة مع القيمة المعطاة للمتغير (U_i).

ب- نفترض أن عينة مقدارها (19) مشاهدة أخذت لهذا النموذج، وقد تم الحصول منها على المقادير التالية :

$$\sum(X-\bar{X})(Y-\bar{Y})=180;\quad \sum(X-\bar{X})^2=240;$$
$$\sum\left(Y-\bar{Y}\right)^2=169;\quad \sum X=190;\quad \sum Y=380$$

قدر قيمة كل من (α) (β) ، ومتوسط قيمة (X) عندما $X_0=16$.

جـ- احسب 95% حدود ثقة للوسط الحسابي المشروط والمذكور في (ب).

د- أوجد معامل الارتباط بين (X_i)و(Y_i) ، ومعامل التحديد $\left(r^2\right)$.

15- من بيانات الجدول التالي :

Y_i	4	3	0	4	4	4	3	0	4	3	1	3
X_i	3	3	0	2	2	3	0	3	2	1	3	2

أثبت أن :

$$\hat{Y}_i=0.22+1.14X_i$$
$$(0.39)\quad(4.56)\quad ,r^2\,0.68$$

أيضا أثبت أن $\left(\hat{\beta}\right)$ معنوية إحصائيا بمستوى معنوية قدره 1% و 5% حيث $\left(\hat{\alpha}\right)$ لم تكن معنوية إحصائيا بنفس المستويين .

16- أوجد معدلات النمو الخاصة بالمتغيرات التالية في اقتصاد معين في الفترة 1993-2004م، حيث إن :

GDP = إجمالي الناتج المحلي ببلايين الليرات

C = الاستهلاك الخاص ببلايين الليرات

I = الاستثمار ببلايين الليرات

G – الإنفاق الحكومي ببلايين الليرات .

السنة	الناتج المحلي GDP	الاستهلاك C	الاستثمار I	الإنفاق الحكومي G
1993	40	8	6	5
1994	100	10	8	10
1995	140	18	18	16
1996	170	24	34	30
1997	200	35	51	40
1998	230	55	67	47
1999	250	64	78	57
2000	3.90	102	97	78
2001	520	115	106	82
2002	530	127	122	130
2003	415	137	116	126
2004	370	145	110	120

والمطلوب:

أ- اختبر دقة هذه التقديرات باستخدام مستوى معنوية 5%.

ب- أوجد العلاقة بين المتغير المستقل التابع ، كذلك أوجد معامل التحديد r^2 .

جـ- كون جدول ANOVA ومنه أوجد اختبار F .

17- من البيانات المذكورة أدناه:

80	100	90	50	30	75	Y_i
20	30	18	10	6	14	X_i

المطلوب:

أ- أوجد معامل التحديد والارتباط بين X_i, Y_i

ب- أوجد معامل انحدار (Y_i) على (X_i) .

جـ- استخدام اختبار t بمستوى معنوية مقدارها 95% ، احسب درجة الثقة لمعلمات $\hat{\beta}$.

18- فيما يلي بيانات بالمساحة المزروعة قمحاً وسعر الطن منه في 11 منطقة من مناطق قطر .

المساحة (بالمائة فدان) :

الكميات بالطن	50	55	50	55	55	57	49	52	54	56	53	52
سعر الطن (بالليرة)	10	8	10	12	10	6	10	10	10	9	7	7

المطلوب:

أ- احسب مرونة العرض (معبراً عن العرض بالمساحة المزروعة) بالنسبة للسعر بفرض أن العلاقة خطية.

ب- أوجد معامل التحديد، علق على قيمته .

د- اختبر ما إذا كان معامل انحدار المساحة على السعر المحسوب في (أ) أعلاه مختلفاً عن الصفر عند مستوى المعنوية 5% .

هـ- هل من المعقول افتراض أن معامل انحدار المساحة على السعر يساوي 0.7 مقابل الفرض البديل القائل بأن هذا المعامل أقل من 0.7 عند مستوى المعنوية 1% .

و- تنبأ بالمساحة عندما يكون السعر (20) ليرة للطن ، هل تعتقد انه يمكن الوثوق كثيراً بصحة مثل هذا التنبؤ؟ ولماذا؟

19- قدرت العلاقة بين الدخل (Y) والثروة (W) من عينة مكونة من (15) أسرة فوجد أن هذه العلاقة هي :

$$\hat{Y}_i = 36.9 + 0.47W_i$$
$$S.E:(4.98)\,(0.117)$$

حيث تمثل القيم الموضوعة داخل الأقواس الأخطاء المعيارية للمعلمات، بفرض أن العنصر العشوائي لهذه العلاقة موزع توزيعًا طبيعياً .

أ- أوجد ثقة 95% لمعامل تزايد الدخل بالنسبة للثروة $\left(\frac{dy}{dw}\right)$.

ب- اختبر الفرض القائل بأن $(\hat{\alpha})$ يساوي (50) وحدة ، مقابل الفرض البديل أن $(\hat{\alpha})$ أقل من (50) عند مستوى المعنوية 1% .

20- ما هو المقصود بمعامل التحديد $\left(r^2\right)$ ؟ وضح كيفية اشتقاقه ، وما هي علاقته بمعامل الارتباط (r) البسيط ؟

الفصل السادس

النموذج الخطي العام

(The General linear Model)

6-1- مقدمة:

لقد اقتصرت دراستنا في الأبحاث السابقة على دراسة أبسط النماذج القياسية، وأسهلها، وهي النماذج الخطية ذات المتحول التفسيري الواحد.

أن واقع الحياة الاقتصادية عادة يكون أعقد من ذلك بكثير. ومعلوم أنه من أجل أن نعكس علاقة اقتصادية ما قد نحتاج لأكثر من معادلة خطية، إذاً لابد لنا من تحليل العلاقة الخطية لأكثر من متحولين تفسيريين كمثال على ذلك من أجل تحليل علاقة الطلب على سلعة استهلاكية معينة باعتبارها دالة في السعر المفروض على هذه السلعة، وفي سعر المادة البديلة (أو أسعار المواد البديلة) وكذلك الدخل، نحتاج إلى معادلة خطية بمتحولين تفسيريين، إذاً نحتاج إلى الانحدار الخطي المتعدد والجزئي؛ لذلك سوف ندرس في هذا الفصل الانحدار الخطي العام (General linear regression). فالنموذج الخطي العام ما هو إلا عبارة عن امتداد لنموذج الانحدار الخطي البسيط، وهذا يساعدنا كثيراً لأجل دراسة وتحليل النموذج العام، حيث بإمكاننا الاعتماد على الافتراضات الأساسية ذاتها المعتمدة في تحليل النموذج الخطي البسيط .

ومن أجل تبسيط النموذج العام سوف نبدأ من الحالة الخاصة للنموذج الخطي العام، وهي الحالة التي يكون فيها المتغير التابع دالة في متغيرين مستقلين .

6-2- معاملات الانحدار الجزئية (Partial regression coefficient) :

سابقاً بينا أن معامل الانحدار البسيط β تعين لنا معدل التغير المتوقع في المتغير التابع لها نتيجة تغيير المتغير المستقل X بوحدة قياس واحدة .

أما معامل الانحدار الجزئي β_i في تحليل الانحدار المتعدد؛ فيقيس معـدل التغير المتوقع في المتغير التابع Y نتيجة تغير المتغير المستقل X_i بوحدة قياس واحدة، مع بقاء أثر بقية المتغيرات المستقلة الأخـرى ثابتـاً (The effect of other independent variables Holding constant) . ونستطيع كتابة معادلة الانحدار المتعدد للمتغير Y على المتغيرين X_1 و X_2 كما يلي :

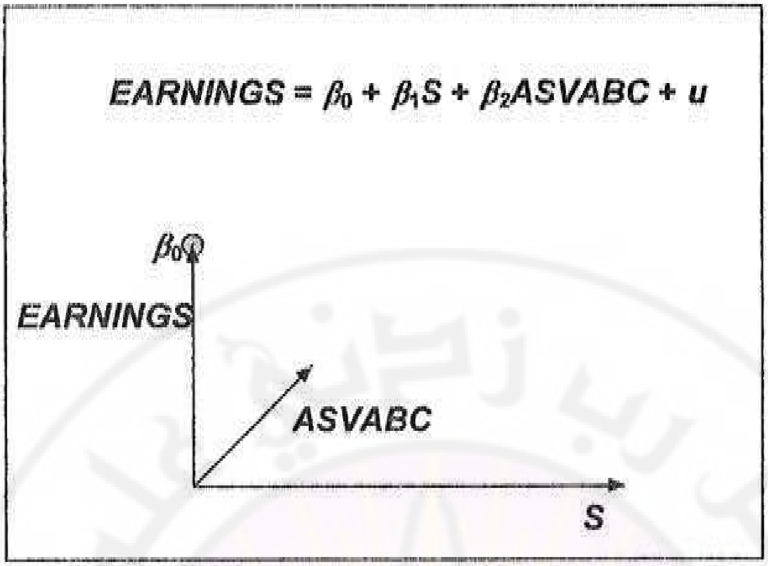
$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$$

بينما نستطيع كتابة معادلة الانحدار المتعدد لـ k متغير مستقل كما يلي :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots\ldots + \beta_k X_k + u$$

سنحاول في البداية إثبات العلاقة هندسياً لمعادلة الانحدار بمتغيرين تفسيريين ، وسنتعرض لمثال تطبيقي نبحث من خلاله عن نموذج الدخل، أي نبين مستوى الدخل في الساعة، ومستوى التعليم (المرحلة التعليمة العليا التي أنهاها العامل) ونرمز له S ، و مستوى المهارة المعرفية .

يدخل في النموذج ثلاثة متغيرات، أي ثلاثة أبعاد : الدخل في الساعة، والمستوى التعليمي، و المهارة المعرفية . وسنبدأ من نقطة تحديد التقاطع بين متغير الدخل $EARNINGS$ ومستوى التعليم S عن طريق دراسة المعامل β_1 .



الشكل رقم(1-6)

أي هذه العلاقة تحدد تقاطع متغير الدخل مع مستوى التحصيل العلمي، واعتبار مستوى المهارة ثابت يساوي الصفر. وهكذا التمسك بالمعنى الحرفي لـــ *EARNINGS* سوف يكون غير مقصود.

في الحد الآخر اليمني من الرسم يعطي اختلاف التأثير في S . الزيادة سنة واحدة في المتغير S سوف يؤدي إلى تغير في الدخل بمقدار β_1 وحدة نقدية، واعتبار التغير في المهارة ثابتاً.

الدنيا لمستوى الماهرة المعرفية هو 22 . سوف نحصل على تقدير غير منطقي لأنه استنتاج خارج مدى البيانات.

لابد من التذكير بأن الفروض التي تم اعتمادها من أجل الخطاء العشوائي سابقاً تنطبق تماماً على حالة الانحدار المتعدد ، مع التأكيد على أن انعدام وجود العلاقة التامة والقوية) بين المتغيرات المستقلة There is on Perfect relationship وهذا يعرف بالاقتصاد القياسي Multicolinearity بالتعدد الخطي، أي الارتباط الذاتي بين المتغيرات المستقلة ، كما مرت معنا سابقاً بأنه يمكننا الحصول على معاملات الانحدار البسيط بعدة طرق ، هذه الطرق أيضاً تمكننا من الحصول على معاملات الانحدار المتعدد أو الجزئية بعدة طرق:

1- باستخدام المعادلات الطبيعة .

2- باستخدام المصفوفات.

3- باستخدام مصفوفة معاملات الارتباط المتعدد بين المتغيرات .

4- باستخدام طريقة البواقي .

الطريقة الأولى : الحصول على معاملات الانحدار الجزئية باستخدام المعادلات الطبيعة، لذلك نكتب معادلة الانحدار على الشكل التالي :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e$$

نحن نعرف بأنه للحصول على معاملات الانحدار الجزئية $\beta_0 , \beta_2 , \beta_1$ والتي تجعل مربع البواقي $\sum e^2 = \sum (y - \hat{y})^2$ عند نهايتها الصغرى .

يمكن الحصول على المعادلات الطبيعة إما بطريقة التفاضل الجزئي، وإما بطريقة أخرى، وهي أن نقوم بضرب (Multiply) طرفي المعادلة السابقة بـX_1 فنحصل على :

$$\sum X_1 Y = \beta_1 \sum X_1^2 + \beta_2 \sum X_1 X_2$$

و نضرب المعادلة السابقة مرة أخرى بالحد بـ X_2 فنحصل على:

$$\sum X_2Y = \beta_1\sum X_1X_2 + \beta_2\sum X_2^2$$

أما بطريقة التفاضل

$$\sum e_i^2 = \sum \left(Y - \hat{Y}\right)^2$$

$$= \sum \left(Y - \beta_0 - \beta_1X_1 - \beta_2X_2\right)^2$$

$$\frac{\delta \sum e^2}{\delta \beta_0} = -2\sum \left(Y - \beta_0 - \beta_1X_1 - \beta_2X_2\right) = 0$$

$$\frac{\delta \sum e^2}{\delta \beta_1} = -2\sum \left(Y - \beta_0 - \beta_1X_1 - \beta_2X_2\right)\left(x_1\right) = 0$$

$$\frac{\delta \sum e^2}{\delta \beta_2} = -2\sum \left(Y - \beta_0 - \beta_1X_1 - \beta_2X_2\right)\left(X_2\right) = 0$$

ومنه نحصل على المعادلات الطبيعة التالية:

$$\sum Y = n\,\beta_0 + \beta_1\sum X_1 + \beta_2\sum X_2$$

$$\sum X_1Y = \beta_0\sum X_1 + \beta_1\sum X_1^2 + \beta_2\sum X_1X_2$$

$$\sum X_2Y = \beta_0\sum X_2 + \beta_1\sum X_1X_2 + \beta_2\sum X_2^2$$

ومنه نجد أن:

$$\hat{\beta}_0 = \overline{Y} - \beta_1\overline{X_1} - \beta_2\overline{X_1}$$

$$\hat{\beta}_1 = \frac{\left(\sum YX_1\right)\left(\sum YX_2^2\right) - \left(\sum X_2Y\right)\left(\sum X_1X_2\right)}{\left(\sum X_1^2\right)\left(\sum X_2^2\right) - \left(\sum X_1X_2\right)^2}$$

$$\hat{\beta}_2 = \frac{\left(\sum X_2Y\right)\left(\sum X_1^2\right) - \left(\sum X_1Y\right)\left(\sum X_1X_2\right)}{\left(\sum X_1^2\right)\left(\sum X_2^2\right) - \left(\sum X_1X_2\right)^2}$$

وباستخدام طريقة كرايمر يتم الحصول على β_1 و β_2 مــن المعــادلتين الطبيعيتين أعلاه .

$$\beta_1 = \frac{\begin{vmatrix} \sum X_1 Y_1 & \sum X_1 X_2 \\ \sum X_2 Y & \sum X_2^2 \end{vmatrix}}{\begin{vmatrix} \sum X_1^2 & \sum X_1 X_2 \\ \sum X_1 X_2 & \sum X_2^2 \end{vmatrix}} =$$

$$= \frac{(\sum X_1 Y)(\sum X_2^2) - (\sum X_2 Y)(\sum X_1 X_2)}{(\sum X_1^2)(\sum X_2^2) - (\sum X_1 X_2)^2}$$

$$\beta_2 = \frac{\begin{vmatrix} \sum X_2^2 & \sum X_1 Y_1 \\ \sum X_2 Y & \sum X_2 Y \end{vmatrix}}{\begin{vmatrix} \sum X_1^2 & \sum X_1 X_2 \\ \sum X_1 X_2 & \sum X_2^2 \end{vmatrix}} =$$

$$= \frac{(\sum X_2 Y)(\sum X_1^2) - (\sum X_1 Y)(\sum X_1 X_2)}{(\sum X_1^2)(\sum X_2^2) - (\sum X_1 X_2)^2}$$

يمكن أن نهمل الحد الثابت مبدئياً ، وهذا لن يؤثر على نتائج السطر على β_2 , β_1

أما الثابت β_0 فيمكن الحصول عليه بأخذ مجموع طرفي معادلة الانحدار $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$ فنحصل على :

$\sum Y = n\,\beta_0 + \beta_1 \sum X_1 + \beta_2 \sum X_2$ وقسمة الناتج على n كما يلي :

$$\frac{\sum Y_i}{n} = \frac{\beta_0 n}{n} + \frac{\beta_1 \sum X_1}{n} + \frac{\beta_2 \sum X_2}{n}$$

$$\overline{Y} = \beta_0 + \beta_1 \overline{X_1} + \beta_2 \overline{X_2}$$

$$\beta_0 = \overline{Y} - \beta_1 \overline{X_1} + \beta_2 \overline{X_2}$$

الطريقة الثالثة: نستطيع بعد الحصول على المعادلات الطبيعية المتعددة استخدام المصفوفات بالحل للحصول على معاملات الانحدار الجزئية؛ علماً بأنها الطريقة الأكثر اتساعاً في مجال البرامج الكمبيوترية . لقد حصلنا على المعادلات الطبيعية بالشكل التالي:

$$\sum Y = n\beta_0 + \beta_1 \sum X_1 + \beta_2 \sum X_2$$
$$\sum X_1 Y = \beta_0 \sum X_1 + \beta_1 \sum x_1^2 + \beta_2 \sum X_2 X_1$$
$$\sum X_2 Y = \beta_0 \sum X_2 + \beta_1 \sum X_1 X_2 + \beta_2 \sum x_2^2$$

يمكن كتابتها بشكل مصفوفي كما يلي :

$$\begin{bmatrix} n & \sum X_1 & \sum X_2 \\ \sum X_1 & \sum X_1^2 & \sum X_1 X_2 \\ \sum X_2 & \sum X_1 X_2 & \sum X_2^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} \sum Y \\ \sum X_1 Y \\ \sum X_2 Y \end{bmatrix}$$

$$(X) \quad * \quad (\beta) \quad = \quad (Y)$$

علماً بأن المصفوفة (X) هي ناتجة من عملية ضرب المصفوفة X بمدورها X' بينما الشعاع (β) هو عبارة عن حاصل ضرب الشعاع β بالمصفوفة المدورة X' وكذلك بالنسبة للشعاع Y فهو ناتج عن ضرب مدور $X \leftarrow X'$ بالشعاع Y .

مثال(1-6): لنفترض أن لدينا بيانات فرضية التالية لأربعة مشاهدة للمتغيرات التالية X_2 , X_1 , Y

X_2	X_1	Y
X_{21}	X_{11}	Y_1
X_{22}	X_{12}	Y_2
X_{23}	X_{13}	Y_3
X_{24}	X_{14}	Y_4

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ X_{11} & X_{12} & X_{13} & X_{14} \\ X_{21} & X_{22} & X_{23} & X_{24} \end{bmatrix} \begin{bmatrix} 1 & X_{11} & X_{21} \\ 1 & X_{12} & X_{22} \\ 1 & X_{13} & X_{23} \\ 1 & X_{14} & X_{24} \end{bmatrix} =$$

$$= \begin{bmatrix} n & \sum X_1 & \sum X_2 \\ \sum X_1 & \sum X_1^2 & \sum X_1 X_2 \\ \sum X_2 & \sum X_1 X_2 & \sum X_2^2 \end{bmatrix}$$

$$X' \qquad X \qquad = \qquad X'X$$

علماً أن إضافة المتغير الوهمي، والذي يأخذ القيم واحد للمصفوفة X هو ضروري للحصول على قيمة الثابت b_0 .

كيف لنا الحصول على X'Y

$$X'Y = \begin{bmatrix} 1 & 1 & 1 & 1 \\ X_{11} & X_{12} & X_{13} & X_{14} \\ X_{21} & X_{22} & X_{23} & X_{24} \end{bmatrix} * \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{bmatrix} = \begin{bmatrix} \sum Y \\ \sum X_1 Y \\ \sum X_2 Y \end{bmatrix}$$

نستطيع كتابة المعادلات الطبيعة في شكل مصفوفي:

$$(X'X)\, b = (X'Y)$$

عندئذ نحصل على مصفوفة المعاملات :

$$b = \frac{X'Y}{X'X} = (X'X)^{-1} * (X'Y)$$

أما بالنسبة لتباين توزيع المعاينة لمعامل الانحدار الجزئي

$$S_{b0}^2 = C_{00} \cdot \frac{SSE}{n-k-1}$$

$$S_{b1}^2 = C_{11} \cdot \frac{SSE}{n-k-1}$$

$$S_{b2}^2 = C_{22} \cdot \frac{SSE}{n-k-1}$$

علماً أن القيم C_{22} , C_{11} , C_{00} هي القيم القطريــة Diagonal values في مقلوب المصفوفة (inverse) المصفوفة (X'X) :

$$(X'X)^{-1} = \begin{bmatrix} C_{00} & C_{01} & C_{02} \\ C_{10} & C_{11} & C_{12} \\ C_{20} & C_{21} & C_{22} \end{bmatrix}$$

ولابد من الإشارة أنه في حالة استخدام المصفوفات للمشاهدات الأولية فإن $\sum e^2 = e'e$ كما أن مجموع الانحرافات الكلية للمتغير التابع (Y) يصبح على الشكل التالي :

$$SST = (Y'Y) = n\bar{Y}$$

الطريقة الثالثة: استخدام مصفوفة معاملات الارتباط بين المتغيرات مــن أجل تحديد معاملات الانحدار الجزئي :

قد تعتبر هذه الطريقة من الطرق السهلة ومن الطرق المعقدة؛ لأنه يسبق تحليل الانحدار تحليل آخر هو إيجاد مصفوفة معاملات الارتباط الجزئية بــين المتغيرات المستقلة . من خلال هذه الطريقة نستخدم العلاقة التالية :

$$\beta = R^{-1} . V$$

حيث إن :

β – هي الشعاع العامودي الذي يحتوي على معاملات الانحدار الجزئية بالوحدات المعيارية (Beta Weights)

R^{-1} – هي مقلوب مصفوفة معاملات الارتباط بين المتغيرات المستقلة .

V – هي الشعاع العامودي الذي يحتوي على معاملات الارتباط بين المتغير التابع وبين المتغيرات المستقلة .

يمكن تحويل معاملات الانحدار الجزئية بالوحدات المعيارية ، إلى معاملات الانحدار الجزئية بالوحدات الأولية ، وذلك باستخدام العلاقة التالية :

$$b = \beta \frac{S_y}{S_x}$$

أما تباين توزيع المعاينة لمعامل الانحدار الجزئي β بالوحدات الأولية يحسب على الشكل التالي :

$$S_b^2 = \frac{\frac{SSE}{n-k-1}}{\sum R^2 \left(1 - R_j^2\right)}$$

حيث إن :

R_j^2 – معامل التحديد المتعدد بين المتغيرات المستقلة الذي يراد تقييم أهمية النسبية وبين بقية المتغيرات المستقلة، ويمكن الحصول عليه باستخدام العلاقة التالية :

$$R_j^2 = 1 - \frac{1}{r^j}$$

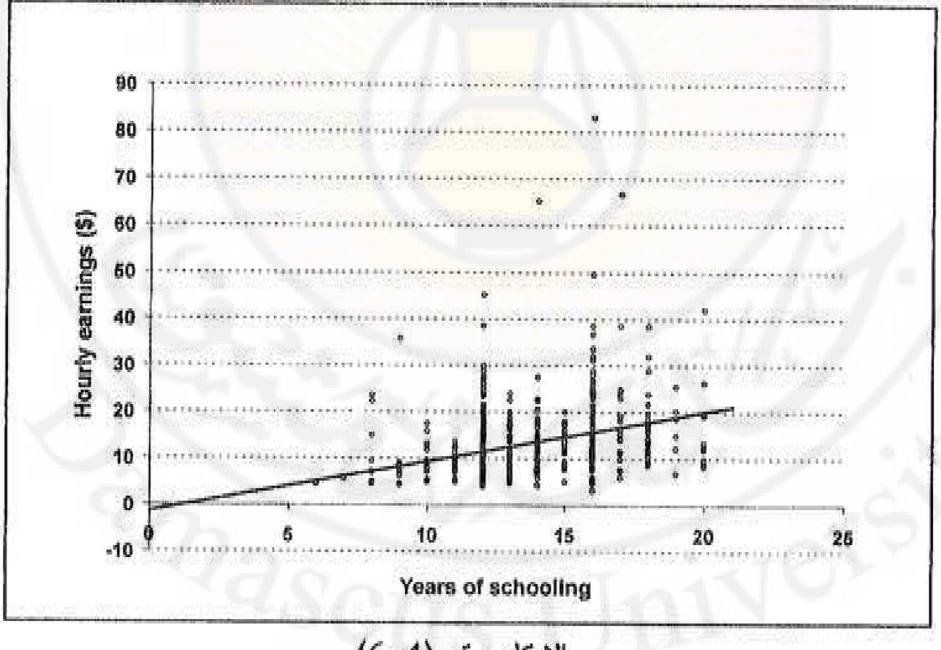
حيث إن : r^j – هي القيمة القطرية في مقلوب (inverse) المصفوفة معاملات الارتباط بين المتغيرات المستقلة R-1 والتي استخدمت في الحصول على معاملات الانحدار الجزئية بالوحدات المعيارية .

نتابع المثال التطبيقي السابق حول علاقـة الكسـب والتعلـيم والمـاهرة المعرفية:

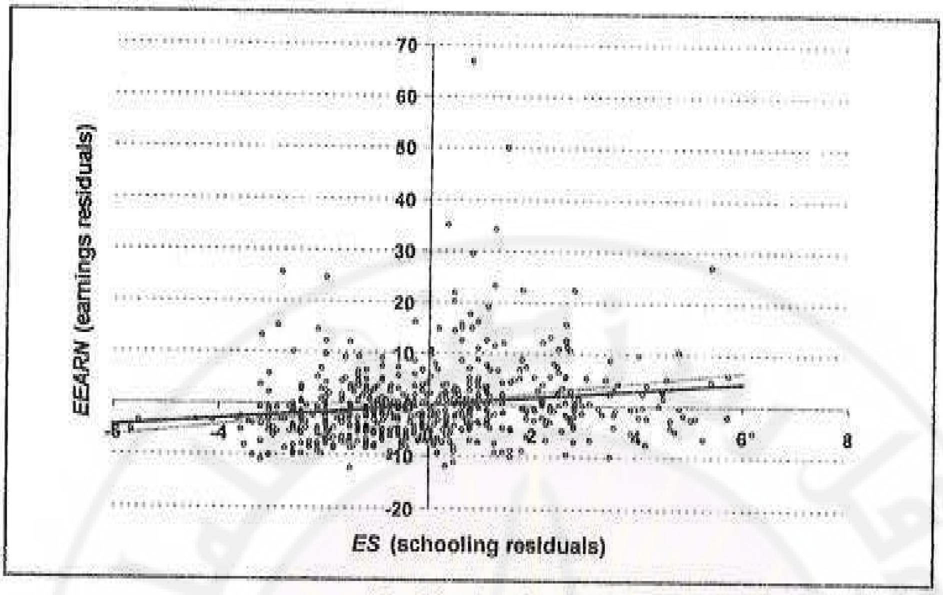
```
reg EARNINGS S ASVABC

   Source |       SS       df       MS              Number of obs =     570
----------+------------------------------           F(  2,   567) =   39.98
    Model |  4745.74965     2  2372.87483           Prob > F      =  0.0000
 Residual |  33651.2874   567  59.3497133           R-squared     =  0.1236
----------+------------------------------           Adj R-squared =  0.1205
    Total |  38397.0371   569  67.4816117           Root MSE      =  7.7039

------------------------------------------------------------------------------
 EARNINGS |Coef.     Std. Err.     t      P>|t|      [95% Conf. Interval]
----------+-------------------------------------------------------------------
        S |.7390366   .1606216    4.601   0.000     .4235506     1.054523
   ASVABC |.1545341    .0429486   3.598   0.000     .0701764     .2388918
    _cons |-4.624749  2.0132     -2.297   0.022    -8.578989    -.6705095
------------------------------------------------------------------------------
```

$$\widehat{EARNINGS} = -4.62 + 0.74S + 0.15ASVABC$$

يبين الشكل البياني أدناه علاقة المتغيرات، وشكل انتشار بيانات أجرة الساعة وسنوات التعليم في عينة الدراسة.



الشكل رقم (4-6)



الشكل رقم (6-6)

نبني نموذج الانحدار للبواقي $EEARN$ على ES .فنجد:

```
. reg EEARN ES
  Source |       SS       df       MS            Number of obs =     570
---------+------------------------------         F(  1,   568) =   21.21
   Model |  1256.44239     1  1256.44239         Prob > F      =  0.0000
Residual |  33651.2873   568  59.2452241         R-squared     =  0.0360
---------+------------------------------         Adj R-squared =  0.0343
   Total |  34907.7297   569  61.3492613         Root MSE      =  7.6971

------------------------------------------------------------------------------
   EEARN |    Coef.   Std. Err.      t     P>|t|     [95% Conf. Interval]
---------+--------------------------------------------------------------------
      ES |.7390366   .1604802    4.605   0.000      .4238296    1.054244
   _cons |-5.99e-09   .3223957   0.000   1.000     -.6332333    .6332333
------------------------------------------------------------------------------
```

للبرهان الرياضي نحتاج لجبر المصفوفات، وهذا ما استعرضناه سـابقاً. سوف نقنع أنفسنا تأكيد تقدير معامل الميل و أهمية التساوي ، باستخدام الخطأ المعياري sd و t الإحصائية تكون نفسها في الانحدار المتعدد.

```
. reg EEARN ES
  Source |       SS       df       MS              Number of obs =     570
---------+------------------------------           F(  1,   568) =   21.21
   Model |  1256.44239     1  1256.44239           Prob > F      =  0.0000
Residual |  33651.2873   568  59.2452241           R-squared     =  0.0360
---------+------------------------------           Adj R-squared =  0.0343
   Total |  34907.7297   569  61.3492613           Root MSE      =  7.6971
------------------------------------------------------------------------------
   EEARN | Coef.      Std. Err.     t      P>|t|      [95% Conf. Interval]
---------+--------------------------------------------------------------------
---------
      ES |.7390366    .1604802   4.605    0.000     .4238296     1.054244
   _cons |-5.99e-09   .3223957   0.000    1.000    -.6332333     .6332333
------------------------------------------------------------------------------
```

From multiple regression:

```
------------------------------------------------------------------------------
---------
EARNINGS |   Coef.     Std. Err.     t      P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
       S |.7390366    .1606216   4.601    0.000      .4235506      1.054523
  ASVABC |.1545341    .0429486   3.598    0.000      .0701764      .2388918
   _cons |-4.624749   2.0132    -2.297    0.022     -8.578989     -.6705095
------------------------------------------------------------------------------
```

6-3- الفرضيات الأساسية Basic Assumptions

يكون النموذج صحيحاً إذا حقق شروط ماركوف- غاوس في تقدير المربعات الصغرى، وهي عدم التحيز و الكفاءة والثبات، كما في الانحدار البسيط.

$$\hat{Y} = b_1 + b_2X_2 + b_3X_3 \qquad\qquad Y = \beta_1 + \beta_2X_2 + \beta_3X_3 + u$$

لقد قمنا بدراستها بشكل مفصّل في بحث الانحدار البسيط ، ونضيف إليها بعضاً من الفرضيات الجديدة .

1- (u_i) يخضع لتوزيع طبيعي $\quad u \sim N(0, \sigma_u^2)$

2- $E(u_i) = 0$

3- $E(u_i)^2 = \sigma_u^2 \; 0$

4- $E(u_i u_j) = 0 \quad (i \neq j)$

13- العلاقة المراد تقديرها مميزة، أي أن العلاقة لها شكل إحصائي وحيد، وأنها لا تحتوي على المتغيرات نفسها الداخلة في أي علاقة أخرى مشابهة مرتبطة بالظاهرة محل البحث .

نطبق طريقة المربعات الصغرى، حيث تحقق هذه الطريقة أصغر مجموع المربعات البواقي $\left(\sum e_i^2\right)$ بالشكل المصفوفي على الشكل التالي :

$$y = x\hat{\beta} + e$$

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} \qquad e = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

علماً بأن الشعاع العمودي $(\hat{\beta})$ ينطوي على قيم $\hat{\beta}_j$ التي من شأنها تصغير مجموع مربعات البواقي إلى أدنى حد .

$$[e_1\, e_2 \cdots e_n] \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} = e_1^2 + e_2^2 + \cdots e_n^2 = \sum e_i^2$$

فإن: $\sum e_i^2 = e'e$

وبما أن: $e = Y - X\hat{\beta}$

فإن:

$$\sum e_i^2 = e'e = \left(Y - X\hat{\beta}\right)'\left(Y - X\hat{\beta}\right) = \left(Y' - \hat{\beta}'X'\right)\left(Y - X\hat{\beta}\right)$$

$$e'e = Y'Y - Y'X\hat{\beta} - \hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta}$$

من المعادلة أعلاه يتبين بأن ($\beta' X' Y$) عبارة عن مدور(Transpose) القيمة $Y' X \hat{\beta}$. كما أن كلاً منهما عبارة عن قيمة ثابتة، وتساوي في الحالة الخاصة ($k = 2$)

$$\left(\hat{\beta}_1 \sum Y_i + \hat{\beta}_2 \sum X_2 Y_1 Y_i \right)$$

وعلية فإن :

$$Y' X \hat{\beta} = \beta' X' Y$$

وبالتالي فإن : $e'e = Y'Y - \hat{\beta}' X' Y - \hat{\beta}' X'X \hat{\beta}$

إن ($X'X$) عبارة عن مصفوفة متناظرة (Symmetric Matrix) بأبعاد (K * K) وأدناه صيغتها العامة:

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & & \vdots \\ X_{K1} & X_{K2} & \cdots & X_{Kn} \end{bmatrix} \begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{K1} \\ 1 & X_{12} & X_{22} & \cdots & X_{K2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{1n} & X_{2n} & & X_{kn} \end{bmatrix} =$$

$$\begin{bmatrix} n & \sum X_{11} & \sum X_{21} & \cdots & \sum XX_1 \\ \sum X_{1i} & \sum X_{1i}^2 & \sum X_{1i} X_{Ki} & \cdots & \sum X_{1i} X_{Ki} \\ \sum X_{2i} & \sum X_{2i} X_{1i} & \sum X_{2i}^2 & \cdots & \sum X_{2i} X_{K1} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \sum X_{K1} & \sum X_{1i} X_{21} & \sum X_i X_{31} & \cdots & \sum X^2 X_1 \end{bmatrix}$$

ولإيجاد مجموعة القيم ($\hat{\beta}$) التي من شأنها تصغير مجموع مربعات البواقي إلى الحد الأدنى . نشتق المعادلة السابقة جزئياً بالنسبة إلى ($\hat{\beta}$) كالآتي :

نشتق أولاً المقدار: ($\hat{\beta} X'X \hat{\beta}$) وذلك ينطبق على الحالة الخاصة

(k = 2)

$$\beta' X'X\beta = \left[\hat{\beta}_0 \hat{\beta}_1\right]\begin{bmatrix} n & \sum X_{ki} \\ \sum X_1 & \sum X_{1i}^2 \end{bmatrix}\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}$$

$$\hat{\beta}' X'X\hat{\beta} = \hat{\beta}_0 n + 2\hat{\beta}_0\hat{\beta}_1 \sum X_{1i} + \hat{\beta}_1^2 \sum X_{2i}^2$$

$$\frac{\partial\left(\hat{\beta}'X'X\hat{\beta}\right)}{\partial\,\hat{\beta}_1} = +\,2\hat{\beta}_0 n + 2\hat{\beta}_1 \sum X_{2i}$$

$$\frac{\partial\left(\hat{\beta}'X'X\hat{\beta}\right)}{\partial\,\hat{\beta}_1} = +\,2\hat{\beta}_0 \sum X_{1i} + 2\hat{\beta}_1 \sum X_{2i}^2$$

والطرف الأيمن في المعادلتين السابقتين يكتب على الشكل التالي :

$$2\begin{bmatrix} n & \sum X_{1i} \\ \sum X_{1i} & \sum X_{1i}^2 \end{bmatrix}\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = 2X'X\hat{\beta}$$

$$\frac{\partial(e'e)}{\partial\hat{\beta}} = -\,X'Y + \,2X'X\hat{\beta}$$

بمساواة هذه المعادلة بالصفر نحصل على الآتي :

$$X'X\hat{\beta} = X'Y \quad)$$

وهذه العلاقة تمثل المعادلات الطبيعية (Normal Equations)، وبما أن (X) تتمتع برتبة (Rank) مقدارها (k) فإنه رتبة (Rank) المصفوفة $(X'X)$ يبلغ (k) . وبأخذ الفرض (6) بعين الاعتبار نكتب:

$$\hat{\beta} = (X'X)^{-1}X'Y$$

نعود للمعادلة السابقة لنعرضها بصورة مفصلة على أساس الحالة الخاصة (k = 2) وذلك لإغناء التصور حولها . $(X'X)\,\hat{\beta} = X'Y$

$$\begin{bmatrix} n & \sum X_{1i} \\ \sum X_{1i} & \sum X_{1i}^2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum X_{1i} Y_i \end{bmatrix}$$

$$n\,\hat{\beta}_0 + \hat{\beta}_1 \sum X_{1i} = \sum Y_i$$

$$n\,\hat{\beta}_0 \sum X_{1i} + \hat{\beta}_1 \sum X_{1i}^2 = \sum X_{1i} Y_i$$

وهما هاتان المعادلتان ما همـا إلا المعادلاتـان الطبيعيتـان للمربعـات الصغرى . وفي الواقع أن زيادة المتغيرات توازيها في عدد المعادلات . فعلـى سبيل المثال للحالة (k = 3) نحصل على المعادلات التالية :

$$n\,\hat{\beta}_0 + \hat{\beta}_1 \sum X_{1i} + \hat{\beta}_2 \sum X_{2i} = \sum Y_i$$

$$\hat{\beta}_0 \sum X_{1i} + \hat{\beta}_1 \sum X_{1i}^2 + \hat{\beta}_2 \sum X_{1i} X_{2i} = \sum Y_i X_{1i}$$

$$\hat{\beta}_1 \sum X_{1i} + \hat{\beta}_1 \sum X_{2i} X_{1i} + \hat{\beta}_2 \sum X_{2i}^2 = \sum X_{2i} Y_i$$

وبطبيعة الحال إن إيجاد قيم $\hat{\boldsymbol{\beta}}_i$ يتم بحل المعادلات الطبيعية حلاً مشتركاً. أو من خلال العلاقة المصفوفية أعلاه بضـرب مقلـوب المصـفوفة $(X'X)$ بالشعاع العمودي $(X'Y)$. و لتبسيط هذه المعادلة نستخدم قـيم انحرافـات المشاهدات عن أوساطها .

إن هذا التبسيط يسهل عمليات الحساب بتقليل أبعاد المصـفوفة $(X'X)$ البالغة (k * k) بمقدار صف عامود، وهذا ما سنبينه لأن بشيء من التفصيل .

نبين من المعادلة (10) أن:

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \cdots + \hat{\beta}_k X_{ki} + e_i$$

$$\overline{Y} = \hat{\beta}_1 + \hat{\beta}_2 X_2 \ \hat{\beta}_3 \overline{X}_3 + \cdots + \hat{\beta}_k \overline{X_k}$$

$$(i = 1, \ldots.. n) \qquad \sum e_i = 0$$

$$\left(Y_i - \overline{Y}\right) = \hat{\beta}_2 \left(X_{2i} - \overline{X_2}\right) + \left(X_{3i} - \overline{X_3}\right)\hat{\beta}_3 + \cdots +$$

$$+ \hat{\beta}_k \left(X_{ki} - \overline{X_k}\right) + e_i$$

$$y_i = \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \cdots + \hat{\beta}_k X_{ki} + e$$

وبإعادة صياغة المعادلات السابقة في صورة مصفوفات نحصـــل علـــى الآتي :

$$Y = X\hat{\beta} + e_i$$

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \quad X = \begin{bmatrix} X_{1i} & X_{2i} & \cdots & X_{ki} \\ X_{12} & X_{22} & \cdots & X_{k2} \\ \vdots & \vdots & \ldots & \vdots \\ X_{1n} & X_{2n} & \cdots & X_{kn} \end{bmatrix} \quad \hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{bmatrix}$$

أبعاد المصفوفات الآنفة الذكر كالتالي :

$$Y \to n * 1 \ , \ \hat{\beta} \to (k-1) * 1 \quad X \to n * (k-1)$$

$$e'e = \left(Y * \hat{\beta}\right)' \left(Y - X\hat{\beta}\right) = \left(Y - \hat{\beta}'X'\right)\left(Y - X\hat{\beta}\right) =$$

$$e'e = Y'Y - Y'X'\hat{\beta} - \hat{\beta}X'Y + \hat{\beta}' \underline{X'} * X\hat{\beta} =$$

$$= Y'Y - 2\hat{\beta}' X'\underline{Y} + \hat{\beta}' X'X\hat{\beta}$$

ولإيجاد مجموعة $(\hat{\beta})$ التي من شأنها تصغير مجموع المربعات البـــواقي إلى الحد الأدنى؛ نشتق المعادلة أعلاه جزئياً بالنسبة إلى $(\hat{\beta})$ ونعادلها بالصفر .

$$\frac{\partial(e'e)}{\partial \underline{\hat{\beta}}} = -2X'Y + 2X'X\hat{\beta} = 0$$

$$X'X\hat{\beta} = X'Y$$

وبالأخذ بالفرض (6) بنظر الاعتبار نحصل على الآتي :

$$\hat{\beta} = (X'X)^{-1} X'Y$$

وبما أن ($\hat{\beta}$) لا تتضمن قيمة ($\hat{\beta}_1$) فإن الأخيرة تحسب بصورة مستقلة من وعلى الوجه التالي :

$$\hat{\beta}_1 = \overline{Y} - \hat{\beta}_2 \overline{X_2} - \hat{\beta}_3 \overline{X} - \cdots - \hat{\beta}_k \overline{X_k}$$

6-4- دقة معاملات الانحدار المتعدد

PRECISION OF THE MULTIPLE REGRESSION COEFFICIENTS

في هذه الفقرة نبحث في تباين المجتمع والخطأ المعياري لمعاملات الميل في نموذج انحدار مكون من متغيرين مستقلين:

$$\text{population variance of } b_1 = \sigma^2_{b_1} = \frac{\sigma^2_u}{n\text{Var}(X_1)} \times \frac{1}{1 - r^2_{X_1,X_2}}$$

علاقة تباين المجتمع لـ b_1 كما هي مبينة. العلاقة لـ b_2 تكون العلاقة الجزئية نفسها لـ 1 و 2 بالتبديل فقط.

العامل الأول في هذه العلاقة يتضمن تباين المجتمع لمعامل ميل الانحدار في نموذج الانحدار البسيط.

تباين المجتمع لـ b_1 يعتمد على تباين المجتمع للخطأ العشوائي، لعدد المشاهدات، وكذلك تباين المتغير X_2 للأسباب نفسها كما في الانحدار البسيط.

$$\text{population variance of } b_1 = \sigma^2_{b_1} = \frac{\sigma^2_u}{n\text{Var}(X_1)} \times \boxed{\frac{1}{1 - r^2_{X_1,X_2}}}$$

إن الاختلاف في تحليل الانحدار المتعدد للعلاقة هو المضروب الثاني للعلاقة، والذي يعتمد على معامل التحديد بين X_1 و X_2 . من البديهي الارتباط العالي بين المتغيرين سيؤدي إلى تباين مجتمع عالٍ . الارتباط العالي، يكون صعباً وغير مجدٍ والأفضل ألا يكون هناك علاقة ارتباط قوية بين المتغيرات المستقلة و المتغير Y وسينعكس ذلك على تقدير الانحدار مما يجعلها أقل دقة . نلاحظ أن تباين المجتمع في العلاقة يكون صحيحاً فقط للموذج من متغيرين مستقلين. عندما يوجد أكثر من متغيرين تصبح العلاقة معقدة، ويكون من الأفضل استخدام جبر المصفوفات المتقدم عندئذٍ.

الانحراف المعياري لتوزيع b_1 بالطبع هو الجذر التربيعي لتباين مجتمع b_1. ويحسب من العلاقة التالية:

$$\text{standard deviation of } b_1 = \sqrt{\frac{\sigma_u^2}{n\text{Var}(X_1)} \times \frac{1}{1-r_{X_1,X_2}^2}}$$

من علاقة تباين للخطأ العشوائي u يمكننا أن نحسب مركبات الانحراف المعياري لبيانات العينة على النحو الآتي:

$$E[\text{Var}(e)] = \frac{n-k}{n}\sigma_u^2$$

تباين المجتمع لـــ u يمكن تقديره. التباين البسيط للبواقي نحصل عليه من التقدير الثابت، ولكن يكون متحيز بالتوافق مع العامل $\frac{(n-k)}{n}$ المحددة في العينة، حيث k عدد المعاملات. من البديهي أن نحصل على تقدير غير متحيز لمضاعف تباين العينة للبواقي بالعامل المصحح $\frac{n}{(n-k)}$. وسوف نرمز لذلك بالرمز S_u.

$$s_u^2 = \frac{n}{n-k}\mathrm{Var}(e)$$

وهكذا فإن تقدير الانحراف المعياري للتوزيع الاحتمالي لـــ b_1 يعــرف على أنه الخطأ المعياري لـ b_1 بشكل تقريبي من خلال العلاقة:

$$\text{s.e.}\,(b_2) = \sqrt{\frac{s_u^2}{n\mathrm{Var}(X_2)} \times \frac{1}{1 - r_{X_2,X_3}^2}}$$

سوف نستخدم هذه العلاقة لتحليل الخطأ المعياري لعلاقة عــدد ســنوات الدراسة S مع علاقة أجرة الساعة لمعادلة الانحدار بشكل صغير ومتقطع لعينة جزئية من بيانات العينة الكلية؛ لمعادلة أجرة الساعة المستخدمة فيها بيانات مسح الشاب في الولايات المتحدة الأمريكية .

```
reg EARNINGS S ASVABC if COLLBARG==0

  Source |       SS       df       MS              Number of obs =     507
---------+------------------------------           F(  2,   504) =   40.31
   Model |  4966.96516     2  2483.48258           Prob > F      =  0.0000
Residual |  31052.2066   504  61.6115211           R-squared     =  0.1379
---------+------------------------------           Adj R-squared =  0.1345
   Total |  36019.1718   506   71.184134           Root MSE      =  7.8493

------------------------------------------------------------------------------
EARNINGS |     Coef.   Std. Err.      t     P>|t|      [95% Conf. Interval]
---------+--------------------------------------------------------------------
       S |  .8891909   .1741617    5.106   0.000       .5470186    1.231363
  ASVABC |  .1398727   .0461806    3.029   0.003       .0491425    .2306029
   _cons | -6.100961    2.15968   -2.825   0.005      -10.34404   -1.857877
------------------------------------------------------------------------------
```

لتحديد العينة الجزئية من خلال برنامج Stata يمكننـــا إضــــافة العبــارة (التعليمة if) بحيث يأخذ المتغير قيمة الواحد عند الاستجابة لمن مستوى الدفع عنده بشكل جماعي ، والصفر يعطى للحالات الأخرى.

```
. reg EARNINGS S ASVABC if COLLBARG==0

  Source |       SS       df       MS              Number of obs =     507
---------+------------------------------           F(  2,   504) =   40.31
   Model |  4966.96516     2  2483.48258           Prob > F      =  0.0000
Residual |  31052.2066   504  61.6115211           R-squared     =  0.1379
---------+------------------------------           Adj R-squared =  0.1345
   Total |  36019.1718   506  71.184134            Root MSE      =  7.8493

------------------------------------------------------------------------------
EARNINGS |   Coef.    Std. Err.     t     P>|t|     [95% Conf. Interval]
---------+--------------------------------------------------------------------
---------
       S |.8891909    .1741617    5.106   0.000   .5470186    1.231363
  ASVABC |.1398727    .0461806    3.029   0.003   .0491425    .2306029
   _cons |-6.100961   2.15968    -2.825   0.005  -10.34404   -1.857877
------------------------------------------------------------------------------
```

في هذه الحالة العينة الجزئية المستقلة يكون خطؤها المعياري S_u مساوياً 0.1742 . في هذه الحالة نجد الخطأ المعياري لعدد سنوات الدراسة S يساوي ضعف الخطأ السابق 0.3530 . سوف نحدد العناصر المكونة للخطأ المعياري.

$$\text{s.e.}(b_1) = s_u \times \frac{1}{\sqrt{n}} \times \frac{1}{\sqrt{\text{Var}(X_1)}} \times \frac{1}{\sqrt{1 - r^2_{X_1,X_2}}}$$

نكون الجدول التالي للمساعدة في تتبع التحليل المطلوب:

Decomposition of the standard error of S تحليل الخطأ المعياري لـS

Component	s_u	n	Var(*S*)	$r_{S,\,ASVABC}$	s.e.
Non-union					0.1742
Union					0.3530
Factor	s_u	$\frac{1}{\sqrt{n}}$	$\frac{1}{\sqrt{\text{Var}(X_1)}}$	$\frac{1}{\sqrt{1-r^2_{X_1,X_2}}}$	$\text{s.e.}(b_1)$
Non-union					
Union					

سوف نبدأ بـ s_u: يمكننا أن نبدل علاقة التباين بالشكل الرياضي $\text{Var}(e)$ معلوماً متوسط البواقي في طريقة المربعات الصغرى يساوي الصفر.

فإن:

$$s_u^2 = \frac{n}{n-k}\text{Var}(e) = \frac{n}{n-k} \times \frac{1}{n}\sum(e_i - \bar{e})^2$$
$$= \frac{1}{n-k}\sum e_i^2 = \frac{1}{n-k}RSS$$

هكذا يكون تقديرنا لتباين المجتمع كمجموع مربع البواقي مقسومة على n-k .

```
. reg EARNINGS S ASVABC if COLLBARG==1
  Source |       SS       df       MS              Number of obs =      63
---------+------------------------------           F(  2,    60) =    2.58
   Model |  172.902083     2  86.4510417           Prob > F      =  0.0844
Residual |  2012.88504    60  33.5480841           R-squared     =  0.0791
---------+------------------------------           Adj R-squared =  0.0484
   Total |  2185.78713    62  35.2546311           Root MSE      =  5.7921
------------------------------------------------------------------------------
EARNINGS |    Coef.   Std. Err.    t     P>|t|    [95% Conf. Interval]
---------+--------------------------------------------------------------------
       S |-.3872787   .3530145   -1.097   0.277   -1.093413     .3188555
  ASVABC | .2309133   .1019211    2.266   0.027    .0270407     .4347858
   _cons | 8.291716   4.869209    1.703   0.094   -1.448152     18.03158
------------------------------------------------------------------------------
```

$$s_u^2 = \frac{1}{n-k}RSS$$

لنطبق ذلك على البيانات المتوفرة لدينا :

عدد مشاهدة 507 في العينة الجزئية و K تساوي 3 . وهكذا فإن n-k تساوي 504. هكذا فإن العلاقة $\frac{RSS}{(n-k)}$ تساوي 61.6115 . ونحصل على S_uبأخذ الجذر التربيعي. وهذا يساوي إلى 7.8493 .

وبالمثل، العينة الكلية نحسب S_u نجدها تساوي 33.54808 ، والتي يكون جذرها مساوياً 5.7921 . وأيضاً نجد أن عدد المشاهدات في العينة يساوي 63.

سوف نحسب تباين العينة S للعينتين الجزئيتين من عينة البيانات. إن معاملي الارتباط بين S و ASVABC يساويان إلى 0.5826 و 0.5380 للعينة الكلية والعينة الجزئية على التوالي.

```
. cor S ASVABC if COLLBARG==0
(obs=507)

         |        S   ASVABC
---------+------------------
        S|   1.0000
   ASVABC|   0.5826   1.0000

. cor S ASVABC if COLLBARG==1
(obs=63)

         |        S   ASVABC
---------+------------------
        S|   1.0000
   ASVABC|   0.5380   1.0000
```

هكذا أكملنا نصف مدخلات الجدول أعلاه . والآن سوف نبحث عن مفاهيم كل حد في علاقة الخطأ المعياري مستخدمين العلاقة الرياضية :

$$\text{s.e.}(b_1) = s_u \times \frac{1}{\sqrt{n}} \times \frac{1}{\sqrt{\text{Var}(X_1)}} \times \frac{1}{\sqrt{1 - r^2_{X_1,X_2}}}$$

s_u مكونة من العناصر المتكيفية، وهي كبيرة نسبياً للعينة الجزئية، وتملك تأثير معاكس على الخطأ المعياري. عدد المشاهدات كبير جداً للعينة الجزئية، وكذلك العامل الثاني يكون صغيراً جداً للعينة الكلية. ربما المفاجئة هي أن التباين لمستوى الدراسة لكلا العينتين متقارب جداً.

الارتباط بين المستوى التعليمي و مستوى خبرة العينة الكلية أكبر منه في العينة الجزئية، وهو ذو تأثير معاكس على الخطأ المعياري . ضرب العوامل مع بعضها البعض نحصل على الأخطاء المعيارية. يختلف قليلاً عما حصلنا عليه بالأرقام العشرية بسبب أخطاء التقريب. سوف نرى أن الخطأ المعياري صغير للعينة الجزئية، وهذا بسبب أن حجم المشاهدات أكبر في العينة الكلية . من جهة أخرى الخطأ المعياري سيكون غير معنوي.

Decomposition of the standard error of S تحليل الخطأ المعياري لـ S

Component	s_u	n	Var(S)	$r_{S,ASVABC}$	s.e.
Non-union	7.8493	507	6.0645	0.5826	0.1742
Union	5.7921	63	6.0136	0.5380	0.3530
	s_u	$\frac{1}{\sqrt{n}}$	$\frac{1}{\sqrt{\text{Var}(X_1)}}$	$\frac{1}{\sqrt{1-r^2_{X_1,X_2}}}$	s.e.(b_1)
Non-union	7.8493	0.0444	0.4061	1.2304	0.1741
Union	5.7921	0.1260	0.4078	1.1863	0.3531

5-6- معامل التحديد المتعدد (R^2)

The Coefficient of multiple determination

معامل التحديد كما عرفناه سابقاً هو عبارة عن نسبة التباين في الانحرافات الكلية في المتغير التابع ، والتي نستطيع تحديدها (تفسيرها) بانحدار Y على المتغيرات المستقلة X'S . علماً أنه يوجد العديد من الطرق لاحتساب قيمة R^2 :

1) $$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

2) لو كنا ندرس حالة انحدار Y على كل من X_1 , X_2 وعلى افتراض أن المتغيرات مقاسه في شكل انحرافات من الوسط الحسابي عندئذ نجد أن :

$$RSS = \sum e^2 = \sum e\left(Y - \hat{Y}\right) = \sum e\left(Y - b_1X_1 - b_2X_2\right)$$
$$= \sum eY - b_1\sum eX_1 - b_1\sum eX_2$$

يمكن استنتاج أن $0 = \sum e X_2 = \sum e X_1$ من المعادلات الطبيعة للحصول على مستوى الانحدار؛ لأن :

$$\frac{\delta e^2}{\delta b_1} = -2\sum\left(Y - b_1X_1 - b_2X_2\right)\left(X_1\right) = 0$$

إذاً : $0 = -2\sum e X_1$

$$\frac{\delta e^2}{\delta b_2} = -2\sum (Y - b_1X_1 - b_2X_2)(X_2) = 0$$

إذاً : $0 = -2\sum e X_2$

من النتائج السابقة : $RSS = \sum e Y$

نبدل e بقيمتها في معادلة الانحدار فنحصل على :

$$RSS = \sum Y(Y - b_1X_1 - b_2X_2)$$
$$= Y^2 - b_1 \sum X_1Y - b_2 \sum X_2Y$$

نبدل ESS بالقيمة التفصيلية لها في معادلة معامل التحديد فنحصل على:

$$R^2 = 1 - \frac{\sum Y^2 - b_1 \sum X_1Y - b_2 \sum X_2Y}{\sum Y^2}$$

$$R^2 = \frac{b_1 \sum X_1Y - b_2 \sum X_2Y}{\sum Y^2}$$

ملاحظة : ((من العلاقة الأخيرة نستنتج أن إضافة أي متغير مستقل جديد إلى معادلة الانحدار، ستزيد من مقدار قيمة معامل التحديد المتعدد (الجزئي) وذلك لأن المقام $\sum Y^2$ ثابت القيمة ، مهما كان عدد المتغيرات المستقلة . بينما سوف تزداد قيمة البسط بمقدار $\sum XY$ الجديدة عند إضافة متغير مستقل إلى معادلة الانحدار)) .

3) باستخدام طريقة المصفوفات :

$$R^2 = \frac{b'(X'Y) - n\overline{Y^2}}{(Y'Y) - n\overline{Y^2}}$$

علماً أن: $(Y'Y) = \sum Y^2$

كذلك يمكن أن نحصل على قيمة R^2 في حالة استخدام مصفوفة معاملات الارتباط بين المتغيرات كالآتي :

$$R^2 = V_1\beta_1 + V_2\beta_2 + \dots + V_k\,\beta_k$$

لكن كما ذكرنا سابقاً أن أية إضافة جديدة لمتغير مستقل إلى معادلة الانحدار ستزيد من قيمة معامل التحديد المتعدد ؛ ولذلك علينا تعديل (تصحيح) قيمة R^2 الحاصلة، وذلك باستخدام المعادلة الآتية :

$$\overline{R^2} = 1 - \left[1 - R^2\right]\frac{n-1}{n-k-1}$$

فنحصل على معامل التحديد المعدل Adjusted

ملاحظة: يتضح من معادلة معامل التحديد أنه عندما يزداد عدد المتغيرات المستقلة في معادلة الانحدار إلى أن تصبح $k = n - 1$ فإن قيمة $n - k - 1 = 0$ وبالتالي لا يوجد عندئذ درجات حرجة للانحرافات غير المفسرة، فتصبح الانحرافات غير المفسرة ($1 - R^2$) مساوية للصفر، وتصبح عندئذ قيمة ($R^2 = 1$) .

مثال(2-6): لو أخذ 9 متغيرات مستقلة، ولكل متغير 10 مشاهدات في معادلة الانحدار فيكون عدد درجات الحرية مساوياً الصفر؛ لأن : (N-k-1) فتصبح قيمة ($1 = R^2$) .

ملاحظة: إذا كانت قيمة n صغيرة و k كبيرة مقارنة بحجم العينة n فإن $\overline{R^2}$ ستكون أصغر بكثير من R^2 ، حتى وإن $\overline{R^2}$ في هذه الحالة قد تكون سالبة، ولكن نعلم أن $0 \le R^2 \le 1$

مثال(3-6): $n = 10 \quad k = 6 \quad R^2 = 0.50$

$$\overline{R^2} = 1 - \left[(1 - 0.5)\frac{10-1}{10-6-1}\right] = 1 - (1.50) = -50$$

الدخل القومي	كمية النقود	السنة
55.0	7.0	1990
55.5	7.5	1991
56.0	8.0	1992
57.0	8.5	1993
57.0	8.3	1994
58.9	9.0	1995
58.3	9.3	1996
59.0	9.5	1997
59.7	10.1	1998
60.0	10.1	1999

أ- ارسم البيانات على شكل انتشار ، ومن ثم أوجد انحدار الدخل القومي Y على كمية النقود X.

ب- ارسم خط الانحدار المقدر على شكل الانتشار ، ماذا نستنتج من الشكل البياني ؟

جـ- كم تبلغ قيمة القاطع ؟ وما ميل خط الانحدار ؟ وماذا يقيس؟

د- لبلوغ هدف الدخل القومي 65.0 بليون دولار، كم يجب أن يكون مستوى كمية النقود؟

هـ- كون جدول ANOVA وأوجد الإحصاءات المتعلقة به $\left(F, r, r^2\right)$

3- تُظهر مخرجات نتائج الانحدار النفقات على الطعام المُستهلك بالمنازل FDHO على إجمالي النفقات المنزلية، وعدد الأشخاص في الأسرة، باستخدام مجموعة بيانات حقيقية. قدّم تفسيراً لمعلمات الانحدار و اجري الاختبارات المناسبة.

```
. reg FDHO EXP SIZE if FDHO>0;
  Source | SS          df      MS                 Number of obs = 868
-------------+------------------------------      F( 2, 865) = 426.78
   Model | 1.4826e+09   2    741314291           Prob > F = 0.0000
Residual | 1.5025e+09  865  1736978.64          R-squared = 0.4967
-------------+------------------------------      dj R-squared = 0.4955
   Total | 2.9851e+09  867   3443039.33         Root MSE = 1317.9
------------------------------------------------------------------------------
FDHO | Coef.       Std. Err.    t      P>|t|    [95% Conf. Interval]
-------------+----------------------------------------------------------------
 EXP | .0372621   .0024547   15.18   0.000    .0324442     .04208
SIZE | 559.7692   30.85684   18.14   0.000    499.2061     620.3322
_cons | 884.5901  100.1537    8.83    0.000    688.0173     1081.163
------------------------------------------------------------------------------
```

4- يُظهر جدول مخرجات نتائج الانحدار حصة النفقات المنزليــة علــى الطعام المُستهلك في الأسرة من الدخل FDHOPC ، علــى حصـــة إجمـــالي النفقات المنزلية عند رأسمال محدد من الدخل EXPPC ، و عدد الأشخاص في الأسرة *SIZE* ، باستخدام مجموعة بيانات إحدى المسوح. قدّم تفسيراً لمعلمـــات الانحدار و اجري الاختبارات المناسبة.

```
. reg FDHOPC EXPPC SIZE if FDHO>0;

      Source |       SS       df       MS              Number of obs =     868
-------------+------------------------------           F(  2,   865) =  175.68
       Model |   142127276     2  71063638.2           Prob > F      =  0.0000
    Residual |   349895173   865   404503.09           R-squared     =  0.2889
-------------+------------------------------           Adj R-squared =  0.2872
       Total |   492022449   867  567499.942           Root MSE      =  636.01

------------------------------------------------------------------------------
      FDHOPC |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       EXPPC |   .0316606   .0026915    11.76   0.000     .0263779    .0369432
        SIZE |   -133.775   15.18071    -8.81   0.000    -163.5703   -103.9797
       _cons |   1430.123   67.10502    21.31   0.000     1298.413    1561.832
------------------------------------------------------------------------------
```

5- يُظهر جدول مخرجات نتائج الانحدار حصة الشخص الواحـــد مـــن النفقات المنزلية FDHOPC على حصة الفرد من إجمـــالي النفقـــات الغذائيـــة EXPPC النفقات ، وعدد الذكور البالغين SIZEAM ، وعدد الإناث البالغات

المقدرة بطريقة المربعات الصغرى أكبر من معامل الارتباط الكائن بين المتغير Y و أي تجميع خطي آخر.

على سبيل المثال قد تكون Y كمية الإنتاج الزراعي، وتعتمد على التغيير في كمية السماد وكمية المياه المستخدمة في الري، فإن للعلاقة بين هذه المتغيرات شكل الانحدار التالي: $\hat{Y} = b_0 + b_1X_1 + b_2X_2$

أما إذا كان شكل العلاقة بين المتغيرات دالة غير خطية عندئذ معادلة الانحدار تأخذ شكل التالي: $\hat{Y} = b_0 + b_1X_1^2 + b_2X_2^3$

يمكننا قياس الارتباط بين Y وبقية المتغيرات المستقلة مجتمعة، و ذلك باستخدام أحد مقاييس الارتباط المتعدد؛ لقياس الارتباط بين المتغير (Y) وواحد فقط من المتغيرات المستقلة، يتم ذلك باستبعاد أثر المتغيرات المستقلة الأخرى، وذلك بعد أن نثبتها عند مستوى معين، ويسمى مقياس الارتباط في هذه الحالة بالارتباط الجزئي. والارتباط الجزئي لا يهمل المتغيرات المستقلة الأخرى بل يأخذها في الحسبان، وما هو إلا تعبير عن " فرض بقاء المتغيرات الأخرى ثابتة على حالها عند مستوى معين ".

7-2- الارتباط المتعدد الخطي:

وسندرس الارتباط المتعدد بأبسط حالاته، وهي دراسة الارتباط بين ثلاثة متغيرات تربطها علاقة خطية :

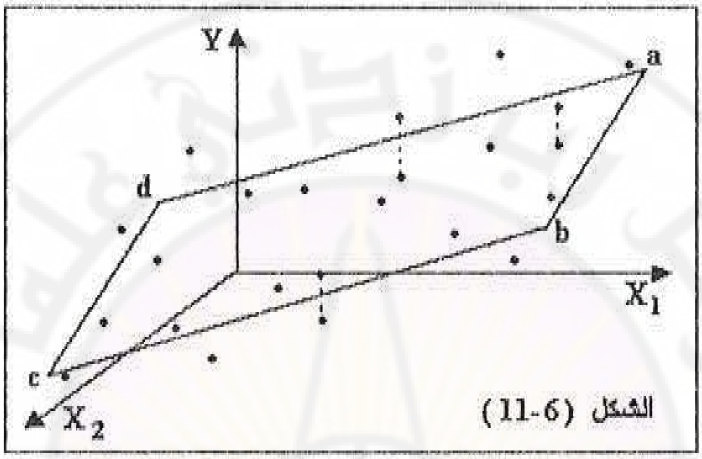
$$Y = b_0 + b_1 X_1 + b_2 X_2$$

وحيث إن : b_2 , b_1 , b_0 معالم ثابتة والمطلوب اختيارها، بحيث تجعل المعادلة أفضل وحدة تمثيل لقيم الظواهر .

من حيث المبدأ تعتبر هذه الدراسة مشابهة لدراسة الارتباط البسيط الذي استعرضناه سابقاً. ولتوفيق منحني الانحدار يمكننا استخدام طريقة المربعات الصغرى لكي نوفق مستوياً يصف العلاقة المتوسطة بين المشاهدات للظواهر

محل الدراسة، وبحيث يكون مجموع الانحرافات للمشاهدات عنه مساوية للصفر ويكون مجموع مربعات الانحرافات عن المستوي أقل ما يمكن.

أما تمثيل البيانات للعلاقة بين ثلاثة متغيرات فيكون على شكل انتشار على ثلاثة محاور إحداثية، انظر الشكل التالي:



الشكل (1-7)

من هذا الشكل نلاحظ أن المستوي (abcd) وهو مستوي الانحدار، والذي يشبه مستقيم الانحدار في العلاقة الخطية البسيطة. وباستخدام طريقة المربعات الصغرى لتقدير المعلمات b_2 , b_1 , b_0 ،نحصل على جملة المعادلات الطبيعية الثلاثة التالية :

$$\sum Y = nb_0 + b_1\sum X_1 + b_2\sum X_2$$
$$\sum YX_1 = b_0\sum X_1 + b_1\sum X_1^2 + b_2\sum X_1X_2$$
$$\sum YX_2 = b_0\sum X_2 + b_1\sum X_1X_2 + b_2\sum X_2^2$$

وباستخدام طرق الحل لجملة ثلاث معادلات خطية، والتي ذكرت سابقاً، نحصل على قيم المعلمات: b_0 , b_1 , b_2 .

ثم نحسب انحراف قيم المشاهدات عن المستوي الموفق للانحدار، ومنه نحسب الخطأ المعياري للتقدير الذي يستخدم في حساب معامل الارتباط المتعدد.

والخطأ المعياري للتقدير يحسب من العلاقة التالية :

$$S_{\hat{Y}} = \sqrt{\frac{\sum\left(Y - \hat{Y}\right)^2}{n}}$$

أو من هذه العلاقة :

$$S_{\hat{Y}}^2 = \frac{\sum Y^2 - b_0\sum Y - b_1\sum YX_1 - b_2\sum YX_2}{n}$$

نحسب معامل الارتباط (دليل الارتباط) المتعدد بين Y و كل من X_2 , X_1

من العلاقة:

$$r_{Y,X_1X_2} = \sqrt{1 - \frac{S_{\hat{Y}}^2}{S_Y^2}}$$

وتدعى هذه العلاقة بدليل الارتباط لاختلافها عن معامل الارتباط. وهذه العلاقة تقيس لنا علاقة الارتباط المتعدد بين التغير في المتغير Y والتغير في المتغيرات المستقلة X_2 , X_1 مجتمعة. وأيضاً يمكن القول: إن $-1 \le r_{Y,X_1X_2} \le 1$ (إن إشارة معامل الارتباط تتحدد من خلال إشارة ثابت معامل الانحدار)، كلما اقتربت قيمة معامل الارتباط المتعدد من الواحد الصحيح يكون الارتباط قوياً بين المتغيرات. وإذا كانت قيمته تساوي الواحد دل ذلك على أن الارتباط خطي تام، أي جميع المشاهدات تقع على المستوى الموفق. وإذا كان معامل الارتباط يساوي الصفر، فإن ذلك يجب ألا يفسر بعدم وجود علاقة بين الظواهر محل الدراسة، بل إن ذلك يعني فقط عدم وجود علاقة خطية بينهما.

ويمكن تبسيط العمليات الحسابية لإيجاد معادلة الانحدار المتعدد من خلال أخذ انحرافات القيم عن أوساطها الحسابية، و تصبح المعادلات الطبيعية بدلالة انحرافات القيم عن أوساطها الحسابية على الشكل التالي:

$$\sum yx_1 = b_1\sum x_1^2 + b_2\sum x_1x_2$$
$$\sum yx_2 = b_1\sum x_1x_2 + b_2\sum x_2^2$$

وبحل المعادلتين بالطرق الرياضية المعروفة نحدد قيمة المعلمتين b_2 , b_1. وحيث إن شكل معادلة الانحدار: $y = b_1x_1 + b_2x_2$

ونوجد قيمة المعلمة b_0 من العلاقة التالية: $b_0 = \bar{Y} - b_1\bar{X}_1 - b_2\bar{X}_2$

وبهذا الشكل نحصل على معادلة الانحدار والتي منها نستطيع حساب الخطأ المعياري للتقدير ومعامل الارتباط المتعدد .

مثال(1-7): احسب دليل الارتباط المتعدد $r_{y.x_1x_2}$ علماً بأن العلاقة بين المتغيرات من الدرجة الأولى.

جدول (1-7)

6	10	4	8	2	Y
20	12	8	10	10	X_1
12	8	8	12	10	X_2

الحل :

من أجل إيجاد قيم المعلمات b_2 , b_1 , b_0 نعوض في المعادلات الطبيعيـــة التالية:

$$\sum Y = nb_0 + b_1\sum X_1 + b_2\sum X_2$$
$$\sum YX_1 = b_0\sum X_1 + b_1\sum X_1^2 + b_2\sum X_1X_2$$
$$\sum YX_2 = b_0\sum X_2 + b_1\sum X_1X_2 + b_2\sum X_2^2$$

8	4	4	16	4	8	-2	-4	-2	8	8	4
0	16	4	0	-8	0	-2	0	4	8	12	10
16	0	4	64	0	0	2	8	0	12	20	6
20	40	16	88	0	12	0	0	0	50	60	30

نعوض بالمعادلتين الطبيعيتين المختصرتين :

$$12 = 88 * b_1 + 20 * b_2$$

$$0 = 20 * b_1 + 16 * b_2$$

وبحل المعادلتين حلاً مشتركاً نجد أن: $b_1 = 0.191$;

$$b_2 = - 0.238$$

أما قيمة b_0 فنحسبها من العلاقة التالية:

$$b_0 = \overline{Y} - \left(0.19 * \overline{X}_1\right) - \left(-0.238 * \overline{X}_2\right)$$
$$= 6 - \left(0.19 * 12\right) - \left(-0.238 * 10\right) = 6.1$$

ونحسب $S_{\hat{y}}^2$ بدلالة الانحرافات على الشكل التالي:

$$S_{\hat{y}}^2 = \frac{40 - 0 - \left(0.191 * 12\right) - \left(-0.238 * 0\right)}{5} = 7.542$$

و بدلالة الانحرافات نحسب S_y^2 :

$$S_y^2 = \frac{\sum y}{n} = \frac{40}{5} = 8$$

ونحصل على دليل الارتباط المتعدد من العلاقة التالية:

$$r_{y.x_1x_2} = \sqrt{1 - \frac{7.542}{8}} = \sqrt{0.0573} = 0.239$$

لقد حصلنا على النتيجة السابقة نفسها، كما في المثال (1-7) ولكن بطريقة الانحرافات. وتوخياً للسهولة يمكن تعميم هذه الطريقة عندما يكون عدد المتغيرات المستقلة أكثر من متغيرين. فإذا كان شكل العلاقة على النحو التالي:

$$\hat{Y} = b_0 + b_1X_1 + b_2X_2 + b_3X_3$$

عند حساب دليل الارتباط المتعدد ، نحسب انحرافات القيم عن أوساطها الحسابية. ومن ثم نحسب الخطأ المعياري للتقدير والانحراف المعياري للمتغير التابع، مع الأخذ بعين الاعتبار المتغير الجديد في علاقة الخطأ المعياري للتقدير، ثم نحسب معامل الارتباط المتعدد (دليل الارتباط) من العلاقة:

$$r_{Y,X_1X_2}\,(I_{Y,X_1X_2}) = \sqrt{1 - \frac{S_{\hat{Y}}^2}{S_Y^2}}$$

7-4- حساب دليل الارتباط المتعدد بدلالة معاملات الارتباط البسيطة :

يمكننا حساب دليل الارتباط المتعدد بدلالة معاملات الارتباط البسيطة إذا كان لدينا ثلاثة متغيرات فأكثر، فإن معامل الارتباط المتعدد يحسب على الشكل التالي:

$$r_{1,23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2\,r_{12}r_{13}r_{23}}{1 - r_{23}^2}}$$

حيث أن: $r_{1,23}$ - (1) - يرمز للمتغير التابع و (23) - ترمز للمتغيرين المستقلين X_1 و X_2.

r_{12} - معامل الارتباط البسيط بين المتغيرين الأول والثاني.

r_{13} - معامل الارتباط البسيط بين المتغيرين الأول والثالث.

r_{23} - معامل الارتباط بين المتغيرين الثاني والثالث.

كما يمكننا حساب معامل الارتباط المتعدد عندما نجعل المتغير الثاني هو المتغير التابع، والمتغيرين الأول والثالث متغيرين مستقلين، فيصبح معامل الارتباط المتعدد بدلالة معاملات الارتباط البسيطة على النحو التالي:

$$r_{2,13} = \sqrt{\frac{r_{21}^2 + r_{23}^2 - 2r_{21}r_{23}r_{13}}{1 - r_{13}^2}}$$

أما إذا جعلنا المتغير الثالث متغيراً تابعاً وبقية المتغيرات مستقلة؛ فإن العلاقة تصبح على الشكل التالي :

$$r_{3,12} = \sqrt{\frac{r_{31}^2 + r_{32}^2 - 2r_{31}r_{32}r_{12}}{1 - r_{12}^2}}$$

مثال(3-7): احسب دليل الارتباط المتعدد $r_{1,23}$, $r_{2,13}$, $r_{3,12}$ إذا علمت أن: $r_{13} = 0.70 \quad r_{23} = 0.55 \quad r_{12} = 0.65$

الحل:

$$r_{1,23} = \sqrt{\frac{(0.65)^2 + (0.70)^2 - 2*(0.65)*(0.70)*(0.55)}{1 - (0.55)^2}} = 0.769$$

$$r_{2,13} = \sqrt{\frac{(0.65)^2 + (0.55)^2 - 2*(0.65)*(0.55)*(0.70)}{1 - (0.7)^2}} = 0.663$$

$$r_{3,12} = \sqrt{\frac{(0.7)^2 + (0.55)^2 - 2*(0.7)*(0.55)*(0.65)}{1 - (0.65)^2}} = 0.711$$

من النتائج التي حصلنا عليها أعلاه ، نستنتج أن قيم دليل الارتباط المتعدد تزيد عن معاملات الارتباط البسيطة، أو تساويها على الأقل. وهذه النتيجة تفسير لنا لماذا عندما نضيف متغيراً مستقلاً جديداً إلى علاقة الانحدار يزيد في تفسر المتغير التابع، وإذا كانت هذه الإضافة لا تزيد في تفسيره ، فإن دليل الارتباط المتعدد يساوي معامل الارتباط البسيط على الأقل.

و لدراسة علاقة الارتباط المتعدد للعلاقات اللاخطية، فإننا نتركها للدراسات الإحصائية المتقدمة والمتخصصة.

7-5- الارتباط الجزئي Partial Correlation :

لاحظنا في الفقرات السابقة أنه يمكننا إيجاد معادلتي مستقيم انحدار بين المتغيرين Y و X. ويتوقف كل مستقيم من مستقيمات الانحدار على أي المتغيرين يعتبر متغيراً تابعاً وأيهما متغير مستقل. أما في حالة الارتباط المتعدد عندما يكون لدينا ثلاثة متغيرات على الأقل، فإنه يمكننا أيضاً تحديد ثلاثة مستويات للانحدار، ويتعلق هذا الأمر باختيار المتغير التابع من بين المتغيرات المحددة للعلاقة. ومن هنا تنبع أهمية تحديد المتغير التابع حسماً لمثل هذا الأشكال؛ حتى لا نضطر لحساب ثلاثة أشكال للانحدار. وحساب علاقات الانحدار الثلاث ليست المشكلة الوحيدة، ولكن المسألة الأكثر أهمية كيفية تحديد المتغيرات التي سيشتمل عليها التحليل؛ لذلك يجب النظر إلى المسألة بعناية لحساب شدة الارتباط بين المتغيرات. و يفترض أيضاً عند إنشاء علاقة انحدار متعدد أن تكون المتغيرات مستقلة عن بعضها بعض، وأن يكون الارتباط البسيط بين أزواجها شبه معدوم، وذلك للتخلص من مشكلة الارتباط الذاتي (الوهمي)، ومشكلة عدم ثبات التباين بين المتغيرات.

7-6- *طرق إيجاد معامل الارتباط الجزئي:

ويوجد عدة علاقات يمكن استخدامها لإيجاد قيم معاملات الارتباط الجزئية، ونذكر منها:

1 – حساب معامل الارتباط الجزئي بدلالة الجذر التربيعي لمعاملات الانحدار:

يمكننا الحصول على معامل الارتباط الجزئي إذا كان لدينا ثلاثة متغيرات على الأقل، متغير التابع Y ومتغيران مستقلان X_1 و X_2، وذلك بأخذ الجذر التربيعي لمعاملين مستقيمي الانحدار، انحدار Y على X_1 وانحدار X_1 على Y فنحصل على العلاقتين التاليتين:

$$\hat{X} = c_0 + c_1 Y_1 \quad و \quad \hat{Y} = b_0 + b_1 X_1$$

ونعتبر المتغير X_2 ثباتاً عند مستوى معين ، عندئذ يمكننا أن نحسب معامل الارتباط الجزئي على الشكل الآتي :

$$r_{YX_1/X_2} = \sqrt{b_1 * c_1}$$

حيث إن: r_{YX_1/X_2} – معامل الارتباط الجزئي بين X_1 , Y ومع افتراض ثبات X_2 . و بالطريقة نفسها يمكننا حساب معامل ارتباط (r_{YX_2/X_1}) أي العلاقة بين Y و X_2 مع إلغاء أثر التغير بالمتغير المستقل X_1 (أي تثبتها عند مستوى معين).

---

* يقرأ معامل الارتباط الجزئي من الشكل: (r_{YX_1/X_2}) معامل الارتباط الجزئي بين المتغيرين Y و X_1 مع استبعاد اثر المتغير X_2 (أي افتراض ثبات X_2 عند مستوى معين). ونميزه عن الارتباط المتعدد ($r_{Y,X_1/X_2}$) بوضع فاصلة بعد المتغير الأول ، بينما في الارتباط الجزئي توضع نقطة لفصل المتغيرات؛ التي يراد إظهار العلاقة فيما بينها عن المتغيرات الثابتة.

أما إذا كانت علاقة الارتباط بين أربعة متغيرات، ونريد أن نحسب معامل الارتباط الجزئي بين متغيرين. فإننا نتبع الخطوات السابقة نفسها، ولكن بعد استبعاد أثر متغيرين، وتكتب العلاقة على النحو التالي:

$$r_{12.34} = \sqrt{b_{12.34} * c_{21.34}}$$

حيث إن: $r_{12.34}$ – معامل الارتباط الجزئي بين المتغير الأول والثاني مع استبعاد اثر المتغيرين الثالث والرابع.

$b_{12.34}$ – معامل انحدار المتغير الأول على المتغير الثاني بفرض ثبات المتغيرين الأخريين.

$c_{12.34}$ – معامل انحدار المتغير الثاني على المتغير الأول بفرض ثبات المتغيرين الأخرين.

مثال(4–7): إذا كان لدينا علاقة انحدار Y على X_1 بعد استبعاد X_2 على الشكل التالي: $\hat{Y} = 10.2 + 0.6\ X_1$

وعلاقة انحدار X_1 على Y بعد استبعاد X_2 : $\hat{X}_1 = -3.5 + 1.6Y$

والمطلوب : احسب معامل الارتباط الجزئي $r_{YX_1X_2}$ بين المتغيرين .

الحل : $\mathbf{r_{YX_1/X_2} = \sqrt{0.6 * 1.6} = 0.9798}$

2 – حساب معامل الارتباط الجزئي بدلالة معاملات الارتباط البسيطة:

يمكننا حساب ذلك من خلال العلاقة التالية :

$$r_{12/3} = \frac{r_{12} - r_{13}\ r_{23}}{\sqrt{1 - r_{13}^2} * \sqrt{1 - r_{23}^2}}$$

حيث إن: $r_{12/3}$ – معامل الارتباط الجزئي بين المتغيرين الأول والثاني مع استبعاد أثر المتغير الثالث.

r_{12} ، r_{13} ، r_{23} – عبارة عن معاملات الارتباط البسيطة بين كل متغيرين على حدة.

مثال(5-7): احسب معاملات الارتباط الجزئي باستخدام انحرافات القيم عن أوساطها الحسابية من البيانات التالية:

جدول (4-7)

x_3^2	x_2^2	x_1^2	$x_2\,x_3$	$x_1\,x_3$	$x_1\,x_2$	x_3	x_2	x_1	X_3	X_2	X_1
0	1	1	0	0	1	0	-1	-1	7	7	5
4	1	1	2	2	1	-2	-1	-1	5	7	5
4	1	1	2	2	1	2	1	1	9	9	7
25	16	0	20	0	0	-5	-4	0	2	4	6
1	0	4	0	2	0	1	0	2	8	8	8
36	100	9	60	18	30	6	10	3	13	18	9
4	25	16	10	8	20	-2	-5	-4	5	3	2
74	144	32	94	32	53	0	0	0	49	56	42

الحل:

أولاً ، نحسب الأوساط الحسابية فنجد:

$$\overline{X}_1 = \frac{42}{7} = 6 \;, \quad \overline{X}_2 = \frac{56}{7} = 8 \;, \quad \overline{X}_3 = \frac{49}{7} = 7$$

ثانياً ، نحسب انحرافات القيمة عن أوساطها الحسابية كما هو موضح بالجدول (4-7).

ثالثاً ، نحسب معاملات الارتباط البسيطة بدلالة الانحرافات :

$$r_{12} = \frac{\sum x_1\, x_2}{\sqrt{\sum x_1^2} * \sqrt{\sum x_2^2}}$$

$$r_{12} = \frac{53}{\sqrt{32} * \sqrt{144}} = 0.781 \ ; \ r_{13} = \frac{32}{\sqrt{32} * \sqrt{74}} = 0.658$$

$$r_{23} = \frac{94}{\sqrt{144} * \sqrt{74}} = 0.911$$

رابعاً، نحسب معاملات الارتباط الجزئية:

$$r_{12/3} = \frac{0.781 - 0.658 * 0.911}{\sqrt{1-(0.658)^2} * \sqrt{1-(0.911)^2}} = 0.585$$

$$r_{13/2} = \frac{0.658 - 0.781 * 0.911}{\sqrt{1-(0.781)^2} * \sqrt{1-(0.911)^2}} = 0.208$$

$$r_{23/1} = \frac{0.911 - 0.781 * 0.658}{\sqrt{1-(0.781)^2} * \sqrt{1-(0.658)^2}} = 0.844$$

3 - حساب معاملات الارتباط الجزئية بدلالة معاملات الارتباط الجزئية من درجة أدنى:

نستطيع حساب معاملات الارتباط الجزئية لأربعة متغيرات بأخذ العلاقة بين متغيرين، مع بقاء بقية المتغيرات ثابتة؛ لحساب معاملات الارتباط الجزئية لأربعة متغيرات على سبيل المثال لحساب $r_{12.34}$ من العلاقة التالية:

$$r_{12.34} = \frac{r_{12.3} - r_{14.3} * r_{23.4}}{\sqrt{1 - r_{14.3}^2} * \sqrt{1 - r_{24.3}^2}} = \frac{r_{12.4} - r_{13.4} * r_{23.4}}{\sqrt{1 - r_{13.4}^2}\sqrt{1 - r_{23.4}^2}}$$

ونحتاج في هذا الحساب إلى معاملات الارتباط الجزئية بين ثلاث متغيرات من الشكل: $r_{24.3}$, $r_{14.3}$, $r_{12.3}$ أي من المرتبة الأدنى، حيث يوجد متغير واحد ثابت عند مستوى معين. إذا استخدمنا معاملات الارتباط البسيط بين متغيرين لحساب معاملات الجزئية من الرتبة الثانية، فإننا نستخدم معاملات الارتباط الجزئية من الرتبة الأولى. لحساب المعاملات من الرتبة الثالثة، نحتاج

$$F(k-1, n-k) = \frac{ESS/(k-1)}{RSS/(n-k)} = \frac{\frac{ESS}{TSS}\Big/(k-1)}{\frac{RSS}{TSS}\Big/(n-k)}$$

نعبر عن العلاقة بدلالة R^2 بالتقسيم البسط والمقام على TSS إجمالي مجموع مربع الانحرافات. وبما أن ESS / TSS يساوي إلى R^2 و RSS / TSS يساوي إلى $(1 - R^2)$ فنجد:

$$F(k-1, n-k) = \frac{R^2/(k-1)}{(1-R^2)/(n-k)}$$

وقد تم برهان ذلك في الفصول السابقة.

نعود إلى المثال الذي أوردناه سابقاً عن مسح الشباب في الولايات المتحدة الأمريكية، ونستخدم نموذج المستوى التعليمي، أي سنفترض أن المتغير S عدد سنوات الدراسة يعتمد على المتغير $ASVABC$ مقياس المهارة المكتسبة (سنوات الخبرة) والمتغير SM المستوى التعليمي للام و SF المستوى التعليمي للأب.

$$S = \beta_0 + \beta_1 ASVABC + \beta_2 SM + \beta_3 SF + u$$

إن فرضية العدم لاختبار F لتوفيق جودة التوفيق لكل المعلمات الثلاث يكون مساوي الصفر. الفرضية البديلة تنص على أن قيمة واحدة على الأقل لا تساوي الصفر.

$$H_0 : \beta_1 = \beta_2 = \beta_3 = 0$$

نعرض مخرجات الانحدار باستخدام بيانات مسح الشباب في الولايات المتحدة الأمريكية.

Source	SS	df	MS
Model	1278.24153	3	426.080508
Residual	2176.00584	566	3.84453329
Total	3454.24737	569	6.07073351

Number of obs	=	570
F(3, 566)	=	110.83
Prob > F	=	0.0000
R-squared	=	0.3700
Adj R-squared	=	0.3667
Root MSE	=	1.9607

S	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ASVABC	.1295006	.0099544	13.009	0.000	.1099486	.1490527
SM	.069403	.0422974	1.641	0.101	-.013676	.152482
SF	.1102684	.0311948	3.535	0.000	.0489967	.1715401
_cons	4.914654	.5063527	9.706	0.000	3.920094	5.909214

في هذا المثال k-1 عدد المتغيرات التفسيرية يساوي إلى ثلاث و n-k عدد درجات الحرية يساوي 566 .

$$F(k-1,n-k)=\frac{ESS/(k-1)}{RSS/(n-k)}$$

لتطبيق العلاقة F الإحصائية يجب تقسيم مجموع مربع التباينات التفسيرية مقسومة على $k-1$ $ESS/(k-1)$ نحصل على هذه القيمة من الجدول أعلاه العمود MS والسطر الأول . أما المخرج فهو السطر الثاني لعمود نفسه، والنتيجة المطلوبة مبينة في السطر الثاني من القسم الثاني من الجدول انظر الجدول أعلاه.

$$F(3,566)=\frac{1278/3}{2176/566}=110.8$$

نجد أن قيمة F الإحصائية تساوي 110.8 على العموم جميع البرامج الإحصائية تحسب لنا هذا المعيار من خلال مخرجات الانحدار.

نوجد القيمة الحرجة (الجدولية) من جدول F ، من أجل (k=3,n-k=120) عند مستوى 0.1% فتكون 5.78 وهي أصغر من القيمة المحسوبة ، ولهذا السبب نرفض فرضية العدم عند مستوى 0.1% . هذه النتيجة لها أهمية نسبية لأن كلاً من $ASVABC$ و SF معنويتها t الإحصائية عالية . وكذلك نعلم أن القيمتين b_2 و b_3 تختلفان عن الصفر. الشيء غير العادي أن تكون قيمة F الإحصائية ليست معنوية و بعض قيم t

Remaining Unexplained - مجموع مربع البواقي بعد إجراء التغيير.

Degrees of freedom remaining -عدد درجات الحرية السارية هو عدد درجات الحرية السائدة بعد إجراء التعديل.

سوف نوضح ذلك من خلال مثال الاختبار مع مستوى التعليم. نبني نموذج الانحدار لـ S على ASVABC باستخدام بيانات مسح الشباب.

الآن سنضيف إلى النموذج المستوى التعليمي للأهل . هل يتضمن المستوى التعليمي للأهل معنوية متضمنة؟ حسناً نرى أن اختبار t يبين أن SF لها معاملاً معنوياً ، ولكن سننجز اختبار F بأية طريقة. نأخذ RSS .

```
. reg S ASVABC SM SF

  Source |       SS       df       MS              Number of obs =     570
---------+------------------------------           F(  3,   566) =  110.83
   Model |  1278.24153     3  426.080508           Prob > F      =  0.0000
Residual |  2176.00584   566  3.84453329           R-squared     =  0.3700
---------+------------------------------           Adj R-squared =  0.3667
   Total |  3454.24737   569  6.07073351           Root MSE      =  1.9607

------------------------------------------------------------------------------
       S |   Coef.   Std. Err.      t     P>|t|     [95% Conf. Interval]
---------+--------------------------------------------------------------------
  ASVABC |.1295006    .0099544   13.009   0.000     .1099486    .1490527
      SM |.069403     .0422974    1.641   0.101    -.013676     .152482
      SF |.1102684    .0311948    3.535   0.000     .0489967    .1715402
   _cons |4.914654    .5063527    9.706   0.000     3.920094    5.909214
------------------------------------------------------------------------------
```

لتحسين التوفيق بعد إضافة متغيرات الأهل يكون تم الانخفاض في مجموع مربع البواقي.

التكلفة **cost** هي درجتان حرية؛ لأنه تم إضافة معلمتين للتقديرات . غير المفسر الساري هو مجموع مربع البواقي بعد إضافة SM و SF . عدد درجات الحرية السارية هو n-k أي أن 570-4 = 566 .

$$F(2,570-4)=\frac{(RSS_1-RSS_2)/2}{RSS_2/(570-4)}=\frac{(2300.4-2176.0)/2}{2176.0/566}=16.18$$

إذاً F الإحصائية تساوي إلى 16.18 . أما القيمة الحرجة أي الجدولية لـ

$F(2,120)$ لمستوى دلالة 0.1% هو 7.32 . القيمة الحرجة يجب أن تكون أقل لنرفض الفرضية، و نستنتج أن متغيرات المستوى التعليمي للوالدين تمتلك قوة معنوية في التفسير. سوف نبين أن اختبار t مكافئ للقيمة الحدية لاختبار F عندما تكون المجموعة الإضافية للمتغيرات تحتوي على متغير واحد.

على فرض أن النموذج الأصلي لـ Y هو معادلة من X_1 و X_2 وسيعدل النموذج بإضافة X_3 . فإن فرضية العدم لاختبار F تبين قوة تفسير المجموعة المضافة يكون كل ميل جديد مساوياً الصفر. هذا بالطبع يعني ميلاً متغيراً جديداً واحداً وهو المعامل b_3.

$$H_0 : \beta_3 = 0$$
$$H_1 : \beta_3 \neq 0$$

إن بنية اختبار F مألوفة. سوف نوضح ذلك بمثال مع نموذج المستوى التعليمي عندما S المتغير التابع ينحدر على ASVABC و SM في النموذج الأصلي، و على SF المضاف للنموذج.

$$F(\text{cost, d.f. remaining}) = \frac{\text{عدد المتغيرات الجديدة الداخلة في النموذج/التحسن في RSS}}{\text{عدد درجات الحرية للنموذج المحسن/ قيمة RSS في النموذج المحسن}}$$

$$F(\text{cost, d.f. remaining}) = \frac{\text{improvement} \;/\; \text{cost}}{\text{remaining unexplained} \;/\; \text{degrees of freedom remaining}}$$

$$F(2,570-4) = \frac{(RSS_1 - RSS_2)/2}{RSS_2/(570-4)}$$

هنا الانحدار للمتغير S على ASVABC و SM . نلاحظ مجموع مربع البواقي للنموذج. لنضف المتغير SF ومن جديد نحصل على مجموع مربع البواقي.

```
. reg S ASVABC SM SF

  Source |       SS       df       MS              Number of obs =     570
---------+------------------------------           F(  3,   566) =  110.83
   Model |  1278.24153     3  426.080508           Prob > F      =  0.0000
Residual |  2176.00584   566  3.84453329           R-squared     =  0.3700
---------+------------------------------           Adj R-squared =  0.3667
   Total |  3454.24737   569  6.07073351           Root MSE      =  1.9607

------------------------------------------------------------------------------
       S |     Coef.   Std. Err.       t     P>|t|     [95% Conf. Interval]
---------+--------------------------------------------------------------------
  ASVABC | .1295006    .0099544     13.009    0.000     .1099486    .1490527
      SM | .069403     .0422974      1.641    0.101     -.013676    .152482
      SF | .1102684    .0311948      3.535    0.000     .0489967    .1715401
   _cons | 4.914654    .5063527      9.706    0.000     3.920094    5.909214
------------------------------------------------------------------------------
```

التحسن بإضافة المتغير SF يكون في تخفيض مجموع مربع البواقي.

$$F(1,570-4)=\frac{(RSS_1-RSS_2)/1}{RSS_2/(570-4)}=\frac{(2224.0-2176.0)/1}{2176.0/566}=12.49$$

التكلفة هنا درجة حرية واحدة مفقودة عند تقدير b_3. غير المفسر الساري هو مجموع مربع البواقي بعد إضافة SF . يصبح عدد درجات الحرية السارية بعد إضافة المتغير SF، 570-4 =566 . وهكذا فإن قيمة F الإحصائية تساوي 12.49 . القيمة الحرجة لـ F عند مستوى دلالة 0.1% من أجل 120 درجة حرية يساوي 11.38 . القيمة الحرجة من أجل 566 درجة حرية يجب أن تكون أقل لنرفض فرضية العدم H_0 عند مستوى دلالة 0.1%

فرضية العدم تختبر تماما القيمة نفسها عند الطرفين لمعنوية اختبار t المتغير SF لننجز اختبار t ،حيث قيمة t الإحصائية تساوي 3.535 و القيمة الحرجة لاختبار t عند مستوى 0.1% ودرجات حرية 120 هو 3.376 . القيمة الحرجة عند 566 درجة حرية يجب أن تكون اقل. لنرفض فرضية العدم H_0 مرة ثانية.

نشاهد أن F الإحصائية لاختبار قوة تفسير مجموعة المتغيرات للمتغيـر المضاف هنا يجب أن تساوي مربع t الإحصائية لهذا المتغير.

نستطيع أيضاً أن نشاهد أن القيمة الحرجة لـ F يجب أن تساوي مربـع قيمة t . (القيم الحرجة المشاهدة عند درجات حرية 120 ، لكن هـذه أيضـاً صحيحة عند 566 درجة حرية.)

```
reg S ASVABC SM SF

  Source |       SS       df       MS              Number of obs =     570
---------+------------------------------           F(  3,   566) =  110.83
   Model |  1278.24153     3  426.080508           Prob > F      =  0.0000
Residual |  2176.00584   566  3.84453329           R-squared     =  0.3700
---------+------------------------------           Adj R-squared =  0.3667
   Total |  3454.24737   569  6.07073351           Root MSE      =  1.9607

------------------------------------------------------------------------------
       S |    Coef.   Std. Err.     t     P>|t|     [95% Conf. Interval]
---------+--------------------------------------------------------------------
  ASVABC | .1295006   .0099544   13.009   0.000      .1099486    .1490527
      SM | .069403    .0422974    1.641   0.101     -.013676     .152482
      SF | .1102684   .0311948    3.535   0.000      .0489967    .1715401
   _cons | 4.914654   .5063527    9.706   0.000      3.920094    5.909214
------------------------------------------------------------------------------
```

$$F(1,566)=\frac{(2224.0-2176.0)/1}{2176.0/(566)}=12.49$$

$$F_{crit,0.1\%}=11.38 \qquad t_{crit,0.1\%}=3.373$$

$$3.373^2=11.377 \qquad 3.535^2=12.496$$

هكذا نستنتج تطابق الاختبارين. هذا يعني أن النتائج اختبار t لمعامـل المتغير يكون اختبار للقيمة الحدية للقوة التفسيرية ، بعد إضافة كل المتغيـرات المتضمنة في المعادلة.

إذا كان المتغير على علاقة ارتباط مع واحد أو أكثر من المتغيرات تكون القيمة الحدية للقوة التفسيرية أقل تماماً، حتى إذا كان الانتماء حقيقياً في النموذج.

إذا كانت كل المتغيرات مرتبطة، يكون احتمال أن تملك قوة تفسيرية حدية و لعدم اختبار t يكون معنوياً حتى وان كانت F ذات معنوية عالية للقوة التفسيرية. أما إذا كان النموذج يعاني من مشكلة تعدد علاقات الارتباط سنناقشه لاحقاً.

إذا الهدف الرئيسي من تقدير الأهمية النسبية للمتغيرات المستقلة في تحديد الانحرافات الكلية في المتغير التابع هو تقييم النموذج ككل. علماً أنه يمكن تقدير ذلك باختبار جوهرية معامل الانحدار من الناحية الإحصائية . لكن نظراً لوجود البدائل في المتغيرات المستقلة التي تعطي القيمة نفسها لمعامل التحديد، وكذلك تعطي قيمة جوهرية لمعامل الانحدار β_1 , β_2 ، لذلك يفضل تقييم الأهمية النسبية للمتغيرات المستقلة باستخدام اختبار (F) الإحصائي . وذلك من أجل تقييم الانخفاض في قيمة معامل التحديد المتعدد من النموذج التام إلى النموذج المقيد .

لو كنا ندرس العلاقة التالية : $Y = \beta_0 + \beta_1X_1 + \beta_2X_2$ فإن معامل التحديد لهذه العلاقة نطلق عليه اسم معامل التحديد المتعدد . ونستطيع تقدير الأهمية النسبية لأي متغير مستقل ، وذلك بحذفه من معادلة الانحدار للنموذج العام، ثم تقييم الانخفاض في قيمة معامل التحديد المتعدد .

لو رغبنا في تقدير الأهمية النسبية للمتغير X_2 نحذفه من النموذج (العلاقة) ونحصل على النموذج المقيد $Y = \beta_0 + \beta_1X_1$ عندئذ نحصل على معامل التحديد للنموذج المقيد . ولتقدير الانخفاض النسبي لمعامل التحديد عند الانتقال من النموذج التام إلى النموذج المقيد نستخدم اختبار – F الإحصائي

$$F = \frac{\left(R_F^2 - R_R^2\right)/\left(k_1 - k_2\right)}{\left(1 - R_F^2\right)/\left(n - k_1 - 1\right)} = \frac{\sum \hat{y}_i^2 / (k-1)}{\sum e_i^2 / (n-k)} = \frac{R^2 / (k-1)}{\left(1 - R^2\right)/(n-k)}$$

حيث إن R_F^2 – معامل التحديد للنموذج التام ،

R_R^2 – معامل التحديد للنموذج المقيد،

k_1 – درجات الحرية في النموذج التام،

k_2 – درجات الحرية في النموذج المقيد.

ومن أجل تقدير الأهمية النسبية لـ X_1 نعيد X_2 ونحذف X_1 ونعيد العملية السابقة للنموذج المقيد .

لابد من الإشارة إلى أن قيمة معامل التحديد R^2 لا تختلف باختلاف ترتيب المتغيرات المستقلة ، لكن المتغير المستقل الأول الذي يدخل في المعادلة سيكون له أهمية نسبية أكبر من غيره من المتغيرات، وكذلك ستختلف أهميته النسبية فيما لو أدخلناه إلى النموذج كآخر متغير مستقل، ستكون عندئذ أهميته النسبية أقل من المتغيرات التي تسبقه.

لذلك يتوجب علينا كباحثين أن نعتمد المنطق والنظرية الاقتصادية بإدخال المتغيرات المستقلة في معادلة الانحدار .

علماً أنه تم اقتراح طريقة حل إحصائي ومنطقي طريقة تركيب معادلة الانحدار على مراحل Stepwise(الحل على مراحل) (Stepwise Solution) تساعد كثيراً في هذا المجال . ولقد تم إدخالها بعدة برامج كمبيوترية جاهزة منها: SPSS و BMD ، BMDP ، Eviews ، وSTATA .

أ – أوجد معادلة الانحدار وفسر المعلمات.

ب – استخدم المعادلة الناتجة في التنبؤ عن الضغط الممكن عند استعمال الغراء حول درجة حرارة 40 ونسبة رطوبة 65.

ج – أوجد مقدار الخطأ المعياري في التقدير.

د – احسب معامل الارتباط المتعدد وفسره.

هـ – احسب معاملات الارتباط الجزئية ، وفسرها.

الفصل الثامن

المتغيرات الوهمية (الصورية)
THE DUMMY VARIABLES

8-1- مقدمة:

تستخدم المتغيرات الوهمية تعبيراً عن المتغيرات النوعية (الوصفية)، سواء كانت متغيرات مستقلة أو تابعة .

وكقاعدة عامة : إذا كان المتغير النوعي k من المستويات والأقسام والمجاميع والفئات فإنه يمكن تمثيله بـ (k - 1) من المتغيرات الوهمية .

أمثلة على ذلك : لتمثل جنس رب الأسرة ذكر أو أنثى فإن :

أي يوجد متغير واحد متقطع $k - 1 = 2 - 1 = 1$

$$x_1 = \begin{cases} 1 & \text{ذكر} \\ 0 & \text{انثى} \end{cases}$$

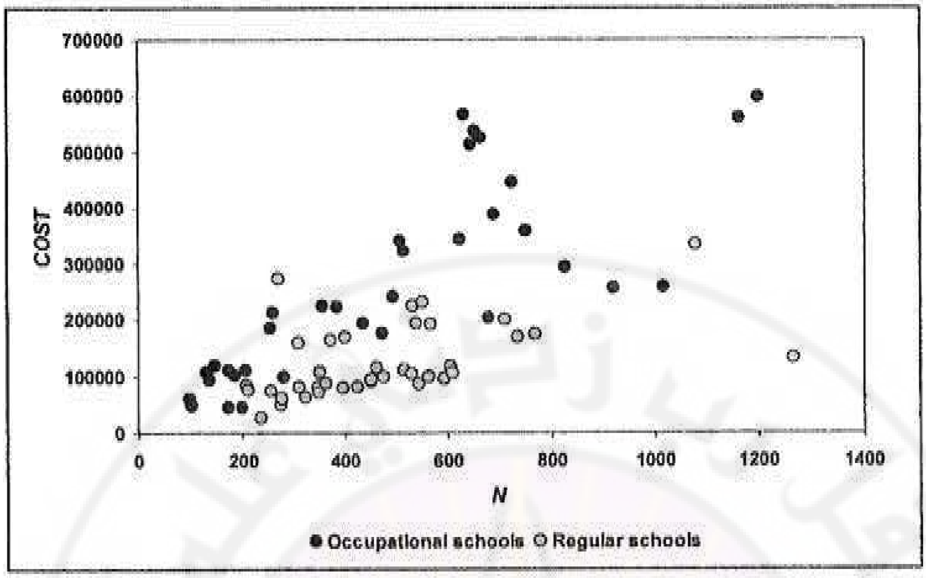
أو بالعكس

$$x_1 = \begin{cases} 1 & \text{انثى} \\ 0 & \text{ذكر} \end{cases}$$

2) أما عند دراسة حالة المدخنين من أرباب الأسرة وعلاقتها بتصلب الشرايين، أي تصنيف أرباب الأسرة إلى :

1 - لا يدخن 2- يدخن أحياناً 3 - يدخن كثيراً

في هذه الحالة فإن : $2 = k - 3 = k - 1$ = عدد المتغيرات أي :



الشكل (3-8)

سنعتمد المتغير OCC كمتغير وهمي لتميز نوع المدرسة.عند تنفيذ الانحدار لمتغير التكلفة على N و OCC ويعتبر المتغير OCC كأي متغير تفسيري، على الرغم من أنه متغير اصطناعي في الجوهر. مخرجات نتائج البيانات حسب برنامج Stata مبينة أدناه.

```
. reg COST N OCC

  Source |       SS       df       MS              Number of obs =      74
---------+------------------------------           F(  2,    71) =   56.86
   Model |  9.0582e+11     2  4.5291e+11           Prob > F      =  0.0000
Residual |  5.6553e+11    71  7.9652e+09           R-squared     =  0.6156
---------+------------------------------           Adj R-squared =  0.6048
   Total |  1.4713e+12    73  2.0155e+10           Root MSE      =   89248

------------------------------------------------------------------------------
    COST |      Coef.   Std. Err.       t     P>|t|     [95% Conf. Interval]
---------+--------------------------------------------------------------------
       N |   331.4493   39.75844     8.337    0.000     252.1732    410.7254
     OCC |   133259.1   20827.59     6.398    0.000     91730.06    174788.1
   _cons |  -33612.55   23573.47    -1.426    0.158    -80616.71    13391.61
------------------------------------------------------------------------------
```

نجد أن واحداً من التفسيرات لعدم منطقية القيمة السالبة للتكلفة الكلية للتعليم النظامي يكمن في عدم وجود قيمة كلية، و تقديرنا عبارة عن رقم عشوائي. حتى نكون أكثر واقعية نترجم الفرضية على ان قيمة الثابت تساوي قيمة موجبة ، ولكن صغيرة (أي كما لو أخذنا مستوى ثقة 95 % لفترة الثقة فإنه يتضمن قيم موجبة)، عندئذ يعزى التقدير السالب إلى الخطأ العشوائي.

وسنحاول فيما بعد بناء نموذجاً أكثر تحديداً باستخدام بعض الطرق القياسية.

8-3- حالة وجود متغير مستقل نوعي واحد في معادلة الانحدار المتعدد:

نفرض أن علاقة الانحدار تحتوي على أكثر من متغير، فيها متغير مستقل X و متغير نوعي واحد X_0^1 على النحو الآتي:

$$y = B_0 + B_1 X_1 + B_2 X_2 + e_i$$

حيث X_1 - متغير مستقل .

X_2 - متغير نوعي .

سنعالج الحالات التالية :

1- عندما يكون للمتغير المستقل النوعي فئتان أو مجموعتان فقط :

8-3-1- عند عدم وجود تداخل بين المتغير المستقل النوعي والمتغير المستقل الكمي:

مثال(8-2) : لدراسة العلاقة بين مقدار الدخل السنوي (Y) وعدد سنوات الخدمة (X_1) والخبرة الإدارية (X_2) لعينة من 20 موظفاً في إحدى الهيئات الدولية، وكانت النتائج على النحو الآتي:

	الدخل لسنوي (Y)	الخبرة الإدارية	(X_2)	عدد سنوات الخدمة (X1)	التداخل $X_3 = (X_1 X_2)$
1	3250	لا يوجد خبرة	0	5	0
2	1600	لا يوجد خبرة	0	1	0
3	7500	لا يوجد خبرة	0	16	0
4	13750	عنده خبرة	1	10	10
5	6000	عنده خبرة	1	5	5
6	8700	عنده خبرة	1	1	1
7	11350	عنده خبرة	1	3	3
8	3000	لا يوجد خبرة	0	4	0
9	15700	عنده خبرة	1	15	15
10	11350	عنده خبرة	1	6	6
11	12250	عنده خبرة	1	8	8
12	700	لا يوجد خبرة	0	2	0
13	10250	يوجد خبرة	1	4	4
14	3500	لا يوجد خبرة	0	8	0
15	4500	لا يوجد خبرة	0	10	0
16	16350	يوجد خبرة	1	13	13
17	3800	لا يوجد خبرة	0	6	0
18	9800	يوجد خبرة	1	3	3
19	10800	يوجد خبرة	1	2	2
20	2300	لا يوجد خبرة	0	3	0

لتفسير معلمات هذا النموذج نجد أن الخبرة الإدارية X_2 يمكن تمثيلها على النحو الأتي:

$$x_i = \begin{cases} 1 & \text{عنده خبرة} \\ 0 & \text{لا يوجد خبرة} \end{cases}$$

إن حدي الثقة للمكافأة التي تضاف إلى الراتب نتيجـة وجـود الخبـرة باحتمال ثقة قدرة 95 % هو :

$$6669.12 \leq \beta_2 \leq 9356.771$$

ب- حالة وجود تداخل بين عدد سنوات الخدمة X_1 والخبـرة X_2 فـي المثال السابق ، عندئذ إن علاقة الانحدار تصبح على الشكل التالي :

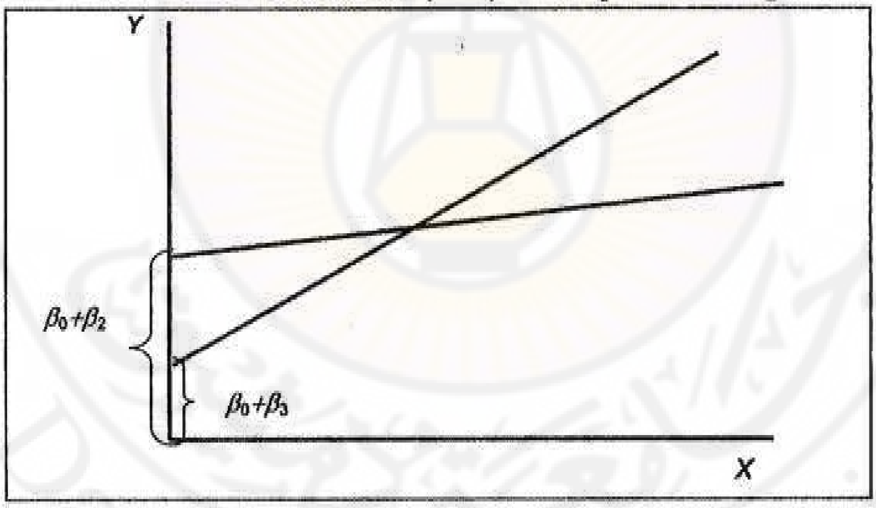
$$\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 (X_1 X_2)$$

ولتفسير هذه العلاقة : نلاحظ عندما لا يكون عند الموظف خبرة إداريـة أي $X_2 = 0$ فإن علاقة الانحدار لها الشكل :

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 (0) + \beta_3 (x_1 * 0)$$

$$\hat{y} = \beta_0 + \beta_1 x_1$$

أي أن دالة الاستجابة هي عبارة عن خط مسـتقيم بمعامـل انحـدار β_1 ونقطة تقاطع β_0 ، كما في الشكل (6-8) .



الشكل (6-8)

أما الموظف الذي عنده خبرة إدارية، فإن علاقة الانحدار تأخذ الشكل التالي:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 (1) + \beta_3 (1)$$
$$= \beta_0 + \beta_1 x_1 + \beta_2 x_1 + \beta_3 x_1$$
$$\therefore \ \hat{y} = (\beta_0 + \beta_2) + (\beta_1 + \beta_3) x_1$$

وبالتطبيق على مثالنا السابق بعد حساب المتغير الجديد لتحديد أمر التداخل X_3 حيث : $X_3 = (X_1 X_2)$ نحصل على النتائج التالية :

```
.reg y x1 x2 x3

  Source |       SS       df       MS              Number of obs =      20
---------+------------------------------           F(  3,    16) =   68.19
   Model |  414833238      3  138277746            Prob > F      =  0.0000
Residual |  32444137.1    16  2027758.57           R-squared     =  0.9275
---------+------------------------------           Adj R-squared =  0.9139
   Total |  447277375     19  23540914.5           Root MSE      =    1424
------------------------------------------------------------------------------
       y |      Coef.   Std. Err.       t     P>|t|     [95% Conf. Interval[
---------+--------------------------------------------------------------------
      x1 |   398.5388    107.678      3.70   0.002      170.2716    626.8059
      x2 |   7193.898   1108.603      6.49   0.000      4843.766     9544.03
      x3 |   131.5725   145.3785      0.91   0.379     -176.6162    439.7611
_   cons |   914.4854   811.3649      1.13   0.276     -805.5313    2634.502
------------------------------------------------------------------------------
```

$$\hat{Y} = 914.485 + 398.539 X_1 + 7193.899 X_2 + 131.572 (X_1 X_2)$$

S.OV	df	SS	MS	F
R (x_1 x_2 x_3)	3	414830000	138280000	68.19
R (x_1)	1	956050000		
R (x_1 x_2)	2	413170000		
R (x_2 x_3 / x_3)	2	319225000	159612500	78.71
R (x_3 / x_1 x_2)	1	1660000	1660000	0.8186
Erorr (x_1 x_2 X_3)	16	3244000	2027800	
Total	19	447280000		

لاختبار الفرضية القائلة بأن خطي الانحدار في الشكل السابق متطابقـان،

أي أن فرضية العدم: $H_0 : \beta_2 = \beta_3 = 0$

الفرضية البديلة أحدهما على الأقل: $H_1 : \beta_2 \neq 0$ أو $\beta_3 \neq 0$

نستخدم F الجزئية:

$$F = \frac{MS\,e\left(x_2\, x_3 / x_1\right)}{MSE\left(x_1\, x_2\, x_3\right)} = \frac{159612500}{2027800} = 78.71$$

وهذه القيمة أكبر من F الجدولية (3.63) عند مستوى دلالة 5 % هذا يعني أحدهما أو كليهما يختلفان عن الصفر، أي أن خطا الانحدار غير متطابقين ونختبر الآن إن كلاً من خطي الانحدار لهما معامل الانحدار نفسه، أي أنـه لا يوجد تداخل بينهما (ولكن لكل منهما نقطة تقاطع) .

$$H_0 : \beta_3 = 0$$
$$H_1 : \beta_3 \neq 0$$

$$F = \frac{MS1\left(x_3 / x_1\, x_2\right)}{MSE\left(x_1\, x_2\, x_3\right)} = \frac{1660000}{2027800} = 0.8186$$

وهذه القيمة أصغر من القيمة الجدولية (4.49) عند مستوى دلالة 5 % أي أنه لا يوجد تداخل بين خطي الانحدار، أي أن لهما معامل الانحدار نفسـه، وعليه فإن نتائج التحليل في الجدول أعلاه تؤكد اعتماد العلاقـة الأولـى مـن التحليل، وهي التي تمثل البيانات أفضل تمثيل :

$$\hat{Y} = 473.383 + 470.719\, X_1 + 8012.950 X_2$$

مثال : الجدول التالي يحتوي على بيانات لعشرين موظفـاً فـي إحـدى الهيئات الدولية تبين الراتب السنوي (Y) وعدد سنوات الخدمة (X_1) والحالة التعليمية (X_2)).

	الراتب (Y) السنوي بالدينار العربي	عدد سنوات الخدمة (X_1)	الحالة التعليمية	X_2	X_3	X4	X5
1	3250	5	جامعي	1	0	5	0
2	1600	1	عليا	0	1	0	1
3	7500	16	ثانوية	0	0	0	0
4	13750	10	جامعي	1	0	10	0
5	6000	5	ثانوية	0	0	0	0
6	8700	1	عليا	0	1	0	1
7	11350	3	جامعي	1	0	3	0
8	3000	4	جامعي	1	0	4	0
9	15700	15	عليا	0	1	0	15
10	11350	6	عليا	0	1	0	6
11	12250	8	عليا	0	1	0	8
12	700	2	ثانوية	0	0	0	0
13	10250	4	عليا	0	1	0	4
14	3500	8	ثانوية	0	0	0	0
15	4500	10	ثانوية	0	0	0	0
16	16350	13	جامعي	1	0	13	0
17	3800	6	جامعي	1	0	6	0
18	9800	3	عليا	0	1	0	3
19	10800	2	كلية	1	0	2	0
20	2300	3	كلية	1	0	3	0

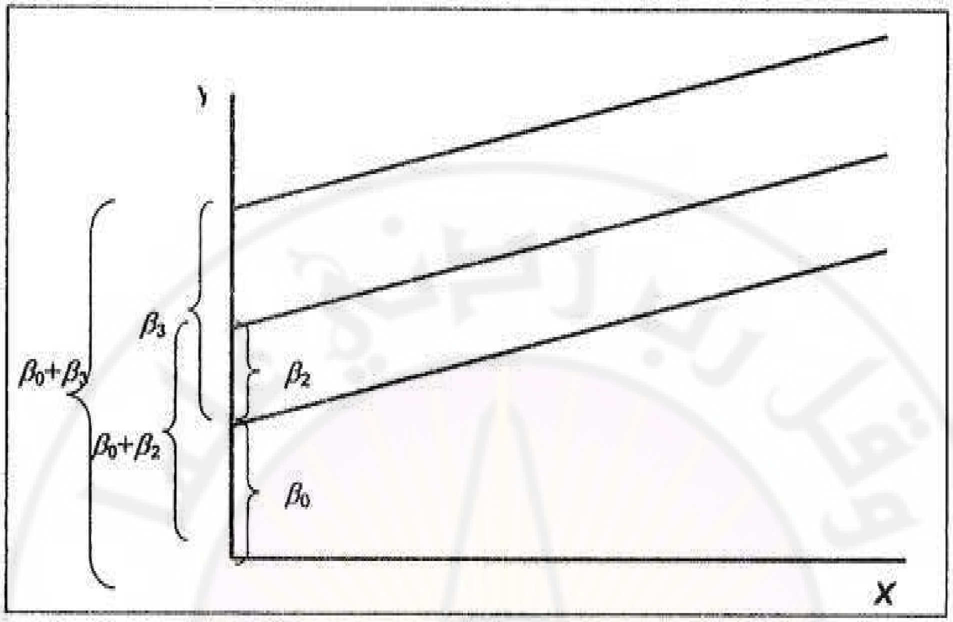
$$x_2 = \begin{cases} 1 & \text{جامعي} \\ 0 & \text{غير ذلك} \end{cases}$$

$$x_3 = \begin{cases} 1 & \text{دراسات عليا} \\ 0 & \text{خلاف ذلك} \end{cases}$$

شكل علاقة الانحدار حسب نوع الشهادة

نوع الشهادة العلمية	شكل علاقة الانحدار
الثانوية	$\hat{y} = \beta_0 + \beta_1 x_1$
الجامعية	$\hat{y} = (\beta_0 + \beta_2) + \beta_1 x_1$
العليا	$\hat{y} = (\beta_0 + \beta_3) + \beta_1 x_1$

لنمثل علاقات الانحدار على الرسم البياني(7-8) ، لإظهار التمايز فيما بينها.



الشكل (7-8)

نتائج تحليل البيانات:

```
.reg y x1 x2   x3

      Source |      SS     df      MS            Number of obs =      20
        -----------------+-------------          F(  3,    16) =    6.57
       Model |246870231   3  82290077.1         Prob > F      =  0.0042
    Residual |200407144 16  12525446.5         R-squared     =  0.5519
--+-------------------      --------------Adj R-squared =  0.4679
       Total |447277375 19  23540914.5         Root MSE      =  3539.1

------------------------------------------------------------------------------
           y |   Coef.   Std. Err.    t     P>|t|  [95% Conf. Interval[
-----------------------------------------------------+------------------------
          x1 | 659.532  186.0089     3.55    0.003  265.2107      1053.853
          x2 |5250.853    2068.442  2.54    0.022  865.9521      9635.755
          x3 |7337.846    2135.46   3.44    0.003  2810.873      11864.82
_       cons |-968.1623    2198.078 -0.44    0.665 -5627.88      3691.555
------------------------------------------------------------------------------
```

معادلة الانحدار :

$$\hat{Y} = 968.162 + 659.532\,X_1 + 5250.853\,X_2 + 7337.846\,X_3$$
$$(3.55) \qquad (2.54) \qquad (3.44)$$

S . O . V	d f	S S	ms	F
R ($x_1x_2x_3$)	3	246870000	25290000	6.57
Erorr ($x_1x_2x_3$)	16	200410000	12525000	
Total	19	447280000		

نجد من الجدول أعلاه أن قيمة F والبالغة 6.57 أكبر F الجدولية 3.24 عند مستوى دلالة قدرة 5 % . نرفض فرضية العدم ونقبل الفرضية البديلة أي أن الانحدار معنوي (على الأقل إحدى معاملات الانحدار β_1 , β_2 , β_3 لا تساوي الصفر). لذلك لابد من اختبار كل من الفرضيات التالية كلاً على حدة :

$$H_{01} : \beta_1 = 0$$
$$H_{02} : \beta_2 = 0$$
$$H_{03} : \beta_3 = 0$$

وبأخذ قيم t بعين الاعتبار نجد أن كلا من β_2 , β_3 يساهم في معنوية Y (في التنبؤ بـ Y) أي أن خطوط الانحدار غير متطابقة، أي أن خط الانحدار للموظف الذي يحمل الشهادة الثانوية أو الجامعية أو الشهادة العليا لها معامل الانحدار نفسه β_1 ذات قيمة معنوية، ولكن لكل منها نقطة تقاطع تختلف معنوياً عن بعضها بعض .

إن $\beta_1 = 659.532$ تعني بأن عند زيادة سنوات الخبرة بمقدار سنة واحدة سيزداد المرتب السنوي بمقدار 659.532 ديناراً مع بقاء X_2 , X_3 ثابتة.

أما $\beta_2 = 5250.8531$ فهي الفرق في المرتب السنوي لموظف خريج كلية مقارنة بالموظف الحاصل على الشهادة الثانوية .

أما $\beta_3 - \beta_2 = 2086.993$ فهو الفرق في الراتب السنوي لموظف حاصل على الشهادة العليا عن موظف خريج كلية .

8-3-2- عند وجود تداخل بين المتغير المستقل النوعي والمتغير المستقل الكمي:

نفرض أن هناك تداخلاً بين عدد سنوات الخدمة X_1 وبين الحالة التعليمية (X_3 , X_2) في المثال السابق، عندئذ فإن علاقة الانحدار تصبح

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5$$

حيث إن :

$$X_4 = X_1 X_2$$

$$X_5 = X_1 X_3$$

تأخذ نماذج الانحدار حسب الحالة التعليمية الشكل التالي:

الحالة التعليمية	شكل علاقة الانحدار
موظف حاصل على الثانوية	$\hat{y} = \beta_0 + \beta_1 x_1$
موظف حاصل على الشهادة الجامعية	$\hat{y} = (\beta_0 + \beta_2) + (\beta_1 + \beta_4) x_1$
موظف حاصل على الشهادة العليا	$\hat{y} = (\beta_0 + \beta_3) + (\beta_1 + \beta_5) x_1$

وبالتطبيق على المثال نجد أن :

```
.reg y x1 x2  x3 x4 x5

     Source |     SS     df     MS           Number of obs =      20
-----------------------------------------F(  5,    14) =    3.96
      Model |261954842  5  52390968.5        Prob > F      =  0.0190
   Residual |185322533 14  13237323.8        R-squared     =  0.5857
    ----------------------------------Adj R-squared =  0.4377
      Total |447277375 19  23540914.5        Root MSE      =  3638.3

------------------------------------------------------------------------------
          y |   Coef.    Std. Err.    t     P>|t|   [95% Conf. Interval[
------------------------------------------------------------------------------
         x1 |375.5319     342.567   1.10   0.291  -359.2013     1110.265
         x2 |1592.139    4052.342   0.39   0.700  -7099.271     10283.55
         x3 |4707.401    3886.695   1.21   0.246  -3628.731     13043.53
         x4 |515.2893    495.2262   1.04   0.316  -546.8653     1577.444
         x5 |339.5661    456.2856   0.74   0.469  -639.0691     1318.201
       cons |1360.638    3246.263   0.42   0.681  -5601.904      8323.18
------------------------------------------------------------------------------
```

ومنه نجد معادلة الانحدار لها الشكل التالي:

$$\hat{y} = 1360.638 + 375.532\,x_1 + 1592.140\,x_2 + 4707.401\,x_3 + $$
$$+ 515.289\,x_4 + 339.566\,x_5$$

S. O . V	df	S S	ms	F
R ($x_1x_2x_3x_4x_5$)	5	261950000	52391000	3.96
R (x_1)	1	95605000		
R ($x_1x_2x_3$)	3	246870000		
R ($x_2x_3x_4x_{5}/x_1$)	4	166345000	51586250	3.14
R ($x_4x_{5}/x_1x_2x_3$)	2	15080000	7540000	21
Erorr ($x_1x_2x_3x_4x_5$)	14	185330000	13237857	
Total	19	447280000		

نختبر مدى انطباق خطوط الانحدار حسب فرضية العدم

$H_0 : \beta_2 = \beta_3 = \beta_3 = \beta_4 = \beta_5 = 0$

$H_1 :$ يوجد قيمة واحدة على الأقل تختلف عن الصفر

$$F = \frac{m\,s\,R\left(x_2x_3x_4x_5 \mid x_1\right)}{m\,s\,R\left(x_1x_2x_3x_4x_5\right)} = \frac{41586250}{13237857} = 3.14$$

F المحسوبة = 3.14 > F الجدولية = 3.11 من أجل مستوى دلالة 5 %

هذا يعني أن خطوط الانحدار غير متطابقة .

نختبر فما إذا كان هناك تداخل بين المتغير المستقبل النــوعي والكمــي حسب فرضية العدم: $H_0 : B_4 = B_5 = 0$

الفرضية البديلة H_1 توجد قيمة واحدة على الأقل تختلــف عــن الصفر.

$$F = \frac{m\,s\,R\left(x_4 x_5 \mid x_1 x_2 x_3\right)}{m\,s\,R\left(x_1 x_2 x_3 x_4 x_5\right)} = \frac{7540000}{13237857} = 0.57$$

أي: أن F المحسوبة 0.57 > من F الجدولية .

أي لا يوجد تداخل بين المتغيرين X_1 و كلٍّ من X_2 , X_3 لذا فإن نتـائج التحليل السابقة تمثل البيانات خير تمثيل .

8-4- تصنيف المتغيرات الوهمية مع أكثر من تصنيفين

DUMMY CLASSIFICATION WITH MORE THAN TWO CATEGORIES

في هذه الفقرة سوف نشرح كيفية توسيع تقنية المتغير الـوهمي بحيـث يتضمن المتغير المستقل النوعي أكثر من فئتين. في الفقرة الـسابقة اسـتخدمنا المتغير الوهمي لنفرق بين المدارس العادية والمهنية عندما نوفق معادلة التكلفة. في القيم الفعلية لنوعين لقانونية الثانوية في شنغهاي. يوجد مدارس رئيسية التي تعلم بشكل أكاديمي ، ومدارس حرفية. تتضمن هذه التسمية المدارس المختصة، وهي تعنى بنقل المعرفة المهنية والمهارة كما في التعليم الأكاديمي.

على أي حال إن المدارس الحرفية ليكون لها منهاج دراسي صغير جداً، و المدارس المهنية تشابه المدارس الأساسية، حتى يكون للمـدارس النظاميـة ورشات عمل إضافية. بطريقة مماثلة يوجد فئتان من المدارس المهنية. يوجـد مدارس تقنية تدرب تقنيات ومهارات العمل ، ومدارس تدرب عاملين مهرة.

لذلك الآن نملك متغيراً وهمياً بأربع فئات . الإجراء النظامي يكون فـي اختيار فئة كمرجع، و تحديد المتغير الوهمي لكل واحدة أخـرى. فـي عينـة شنغهاي يكون دليل الاختيار في المدارس النظامية كمرجـع للفئـة ، يكـون معظمها الكثير، و المدارس الأخرى ، تختلف فيما بينها.

بناء على ذلك سوف نحدد المتغيرات الوهمية لثلاثـة أنـواع ، TECH متغير وهمي لتقنية المدارس : TECH المتغيـر يسـاوي الواحـد إذا كانـت

المدرسة تقنية وصفراً للمدارس الأخرى. وبالمثل سوف نحدد المتغير الـوهمي للعامل $WORKER$ و VOC لمدارس العامل الماهر والحرفي .

سوف يكون لكل المتغيرات الوهمية معامل يصور تكلفتها الكلية، وهـي قيمة قريبة نسبياً من القيمة المرجعية.

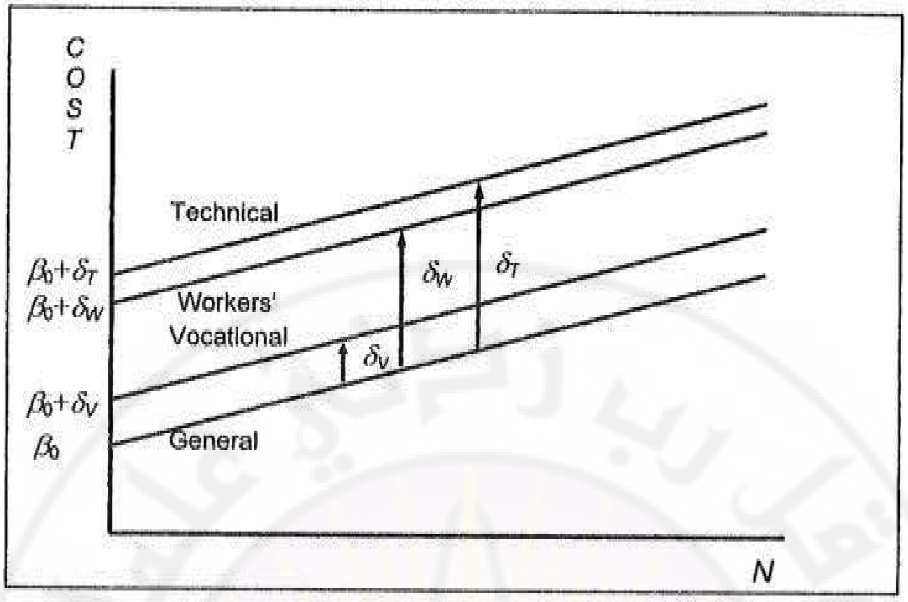
$\hat{COST} = \beta_0 + \delta_T TECH + \delta_W WORKER + \delta_V VOC + \beta_1 N + u$	
General School (*TECH = WORKER = VOC = 0*)	$\hat{COST} = \beta_0 + \beta_1 N + u$
Technical School (*TECH = 1; WORKER = VOC = 0*)	$\hat{COST} = (\beta_0 + \delta_T) + \beta_1 N + u$
Skilled Workers' School (*WORKER = 1; TECH = VOC = 0*)	$\hat{COST} = (\beta_0 + \delta_W) + \beta_1 N + u$
Vocational School (*VOC = 1; TECH = WORKER = 0*)	$\hat{COST} = (\beta_0 + \delta_V) + \beta_1 N + u$

ملاحظة: لا يتضمن المتغير الوهمي كمرجع للفئة أن يكون سبباً لمرجـع اعتيادي الوصف، ويلغي الفئة الأخرى.

إذا كانت المشاهدات تنتمي إلى المدارس النظامية المتغير الوهمي قيمتة الصفر و ينخفض في نموذج الانحدار أحد عناصر الأساس .

إذا كانت المشاهدات تنتمي إلى المدرسة التقنية TECH سـوف تعطـى الرقم 1 وغير ذلك القيمة 0 ، يوضح ذلك نموذج الانحدار كما هو مبين أعلاه.

نموذج الانحدار الموضح بشكل مماثل في حالة المشاهدات تنتمي إلـى مدارس العمال المهرة و المدارس الحرفية. والرسم البياني يبين هذا النمـوذج. المعامل δ يعبر عن التكلفة الكلية في المدارس التقنية، ومدارس العامل الماهر، و المدارس الحرفية، والتي تنتمي إلى التكلفة الكلية للمدارس النظامية .



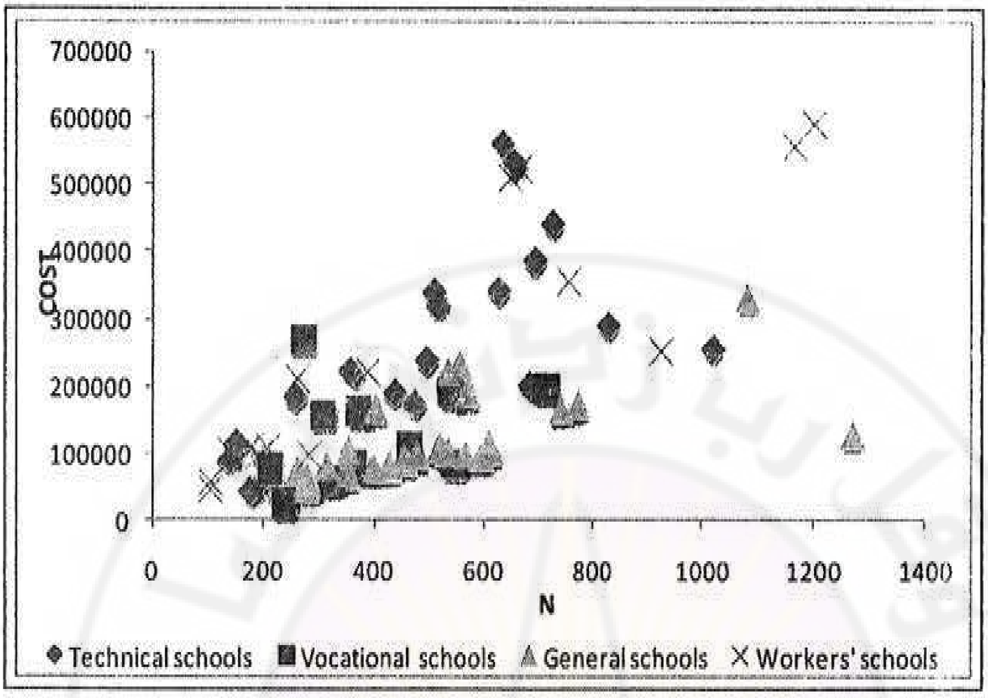
الشكل (8-8)

ملاحظة: يمكننا وضع افتراض مسبق حول حجم أو إشارة لتقدير المعلمات δ وحسب بيانات العينة.

سنبين في الجدول التالي المدارس العشرة الأولى من 74 مدرسة. لبيان كيفية تحديد قيم المتغيرات الوهمية $TECH, WORKER,$ و VOC حسب المدرسة في كل مشاهدة.

School	Type	*COST*	*N*	*TECH*	*WORKER*	*VOC*
1	Technical	345,000	623	1	0	0
2	Technical	537,000	653	1	0	0
3	General	170,000	400	0	0	0
4	Workers'	526.000	663	0	1	0
5	General	100,000	563	0	0	0
6	Vocational	28,000	236	0	0	1
7	Vocational	160,000	307	0	0	1
8	Technical	45,000	173	1	0	0
9	Technical	120,000	146	1	0	0
10	Workers'	61,000	99	0	1	0

شكل الانتشار يبين اختلاف المدارس حسب الفئات.



الشكل (9-8)

يبين الجدول التالي نتائج مخرجات Stata لنموذج الانحدار، حيث يتضمن معامل انحدار N التكلفة الحدية لكل طالب في السنة 343 يناً .

```
. reg COST N TECH WORKER VOC

  Source |     SS         df       MS              Number of obs =      74
---------+------------------------------           F(  4,    69) =   29.63
   Model |9.2996e+11   4  2.3249e+11               Prob > F      =  0.0000
Residual |5.4138e+11  69  7.8461e+09              R-squared     =  0.6320
---------+------------------------------          Adj R-squared =  0.6107
   Total |1.4713e+12  73  2.0155e+10              Root MSE      =   88578

------------------------------------------------------------------------------
    COST | Coef.     Std. Err.    t      P>|t|     [95% Conf. Interval]
---------+--------------------------------------------------------------------
---------
       N |342.6335  40.2195     8.519   0.000     262.3978     422.8692
    TECH |154110.9  26760.41    5.759   0.000     100725.3     207496.4
  WORKER |143362.4  27852.8     5.147   0.000     87797.57     198927.2
     VOC |53228.64  31061.65    1.714   0.091    -8737.646     115194.9
   _cons |-54893.09 26673.08   -2.058   0.043    -108104.4    -1681.748
------------------------------------------------------------------------------
```

المعلمات $TECH$ و VOC و $WORKER$ تكون 154000 و 143000 و 53000 على التوالي ويجب أن يكون التقاطع للتكلفة السنوية الإضافية ينتمي إلى للمدارس النظامية .

الحد الثابت 55000- يتضمن التكلفة الكلية السنوية للمدارس الأكاديمية يكون 55000- ين لكل سنة. ومن الواضح أن هذا غير منطقي، ويدل على عشوائية الخطأ في النموذج.

$\hat{COST} = -55000 + 154000TECH + 143000WORKER + 53000VOC + 343N$	
General School (*TECH* = *WORKER* = *VOC* = 0)	$\hat{COST} = -55000 + 343N$
Technical School (*TECH* = 1; *WORKER* = *VOC* = 0)	$\hat{COST} = -55000 + 154000 + 343N$ $= 99000 + 343N$
Skilled Workers' School (*WORKER* = 1; *TECH* = *VOC* = 0)	$\hat{COST} = -55000 + 143000 + 343N$ $= 88000 + 343N$
Vocational School (*VOC* = 1; *TECH* = *WORKER* = 0)	$\hat{COST} = -55000 + 53000 + 343N$ $= -2000 + 343N$

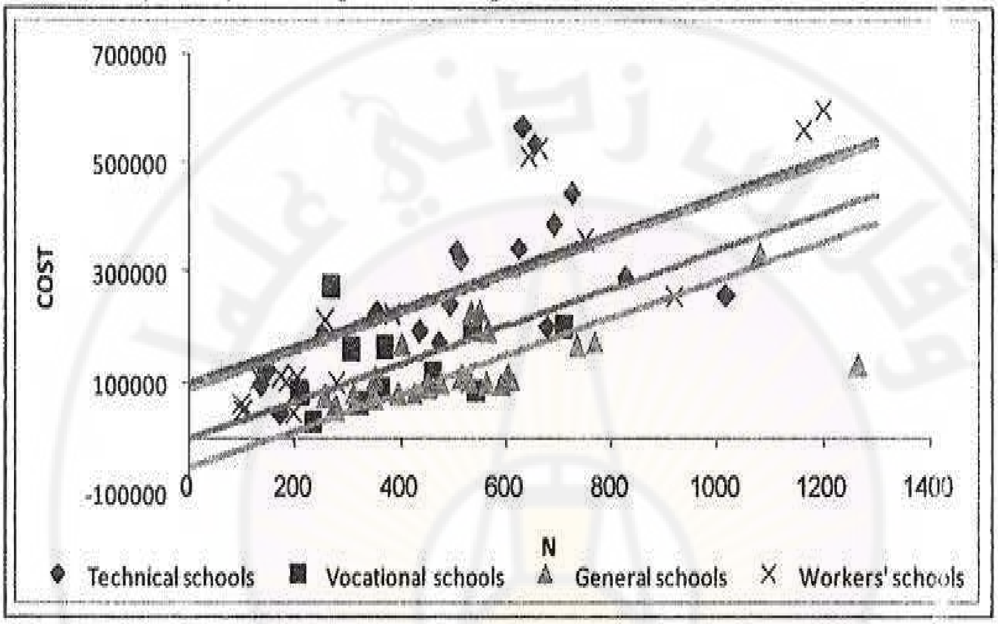
السطر الأعلى لجدول نتائج الانحدار يعرض شكل معادلة الانحدار كما هو مبين أعلاه. سوف نشتق تباعاً مضمون التكلفة للمعادلات لكل فئة من المدارس على حدة. في هذه الحالة المدرسة النظامية المتغير الوهمي للكل يساوي الصفر المعادلة تخفض إلى التقاطع و الحد N .

تساوي التكلفة الحدية السنوية لكل طالب 343 ين . تساوي التكلفة الكلية في كل مدرسة 55000- ين . واضح أن الإشارة السالبة غير منطقية، وكنا أشرنا سابقاً كيف يحصل ذلك.

التكلفة الكلية السنوية للمدارس التقنية المنتمية إلى المدارس النظامية تساوي 154000 ين . وهكذا نشتق مضمون معادلة التكلفة للمدارس التقنية.

وبالمثل التكلفة الكلية السنوية لمدارس العامل الماهر، والمدارس الحرفية المنتمية إلى المدارس النظامية؛ تكون على توالي 143000 و 53000 يناً .

ملاحظة : في كل حالة من الحالات التكلفة الحدية تقدر بـــ343 يــن . مواصفات الأشكال الأربعة للتكلفة، كما هي موضحة في الشكل (10-8).



الشكل (10-8)

نستطيع أن ننجز اختبار t على المعلمات بــالطرق الاعتياديــة. قيمــة t الإحصائية للمتغير N من جدول النتائج يساوي إلى 8.52 ولذلك التكلفة الحدية تكون معنوية جداً في اختلافها عن الصفر كما توقعنا. أما قيمـــة t الإحصـــائية للمتغير الوهمي للمدارس التقنية يساوي 5.76 متضمناً التكلفة السنوية الكليـــة للمدرسة التقنية يكون معنوياً بشكلاً واضحاً أكثر من المدارس النظاميـــة مـــرة ثانية كما توقعنا . وبالمثل مدارس العامل الماهر قيمة t الإحصـــائية تســـاوي 5.15 .

في هذه الحالة المدارس الحرفية كما هو مبين قيمة t الإحصائية تساوي 1.71 متضمنة التكلفة الكلية للمدرسة و معنويتها اقل بكثير من المدرسة النظامية . هذه ليس مفاجأة حيث المدارس الحرفية لا تختلف عن المدارس النظامية .

ملاحظة أن فرضية العدم لاختبارات المعلمات للمتغيرات الوهمية الذي يعطي التكلفة الكلية للمدارس الأخرى غير مختلف عن المدرسة النظامية .

في النهاية سوف ننجز اختبار F للربط بين قوة المتغيرات التفسيرية للمتغيرات الوهمية كمجوعة.

فرضية العدم هي : $H_0: \delta_T = \delta_W = \delta_V = 0$

والفرضية البديلة H_1 واحدة من قيم δ على الأقل تختلف عن الصفر.

مجموع مربع البواقي RSS المتضمن مواصفات المتغير الوهمي يساوي إلى $10^{11}\times5.41$.

```
. reg COST N TECH WORKER VOC

  Source |       SS       df       MS              Number of obs =      74
---------+------------------------------           F(  4,    69) =   29.63
   Model |  9.2996e+11     4  2.3249e+11           Prob > F      =  0.0000
Residual |  5.4138e+11    69  7.8461e+09           R-squared     =  0.6320
---------+------------------------------           Adj R-squared =  0.6107
   Total |  1.4713e+12    73  2.0155e+10           Root MSE      =   88578

------------------------------------------------------------------------------
    COST |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
       N |   342.6335    40.2195      8.519   0.000     262.3978    422.8692
    TECH |   154110.9   26760.41      5.759   0.000     100725.3    207496.4
  WORKER |   143362.4    27852.8      5.147   0.000     87797.57    198927.2
     VOC |   53228.64   31061.65      1.714   0.091    -8737.646    115194.9
   _cons |  -54893.09   26673.08     -2.058   0.043    -108104.4   -1681.748
------------------------------------------------------------------------------
```

مجموع مربع البواقي المتضمن مواصفات النموذج بعد استثناء المتغيرات الوهمية يساوي $10^{11}\times8.92$.

```
. reg COST N

  Source |       SS        df       MS           Number of obs =      74
---------+------------------------------         F(  1,    72) =   46.82
   Model |  5.7974e+11      1  5.7974e+11        Prob > F      =  0.0000
Residual |  8.9160e+11     72  1.2383e+10        R-squared     =  0.3940
---------+------------------------------         Adj R-squared =  0.3856
   Total |  1.4713e+12     73  2.0155e+10        Root MSE      = 1.1e+05

------------------------------------------------------------------------------
    COST |     Coef.   Std. Err.      t     P>|t|     [95% Conf. Interval]
---------+--------------------------------------------------------------------
       N |339.0432   49.55144     6.842   0.000     240.2642     437.8222
   _cons |23953.3    27167.96     0.882   0.381    -30205.04     78111.65
------------------------------------------------------------------------------
```

التخفيض الحاصل في RSS عندما أدخلنا المتغيرات الوهمية هو $(5.41-8.92)\times10^{11}$. لنفحص فيما إذا كان هذا التخفيض معنوياً أم لا باختبار F الاعتيادي. البسط في F النسبية يمثل التخفيض في RSS بالتقسيم على التكلفة التي تساوي ثلاث درجات حرية المعطاة عند إضافة المتغيرات الثلاثة الإضافية للمعلمات (معلمات المتغيرات الوهمية).

$$F(3,69)=\frac{(8.92\times10^{11}-5.41\times10^{11})/3}{5.41\times10^{11}/69}=14.92$$

الكسر في RSS المتضمن مواصفات المتغيرات الوهمية يقسم على عدد درجات الحرية بعد إضافة المتغيرات. قيمة F الجدولية لتحديد القيمة الحرجة من أجل ثلاث درجات حرية غير موجودة، ولكن يمكن أخذ القيمة الأقل عند ثلاث درجات حرية (3،60) التي تساوي إلى 6.17 عند مستوى دلالة 0.1% . $F(3,60)_{crit,0.1\%}=6.17$.

وهكذا سوف نرفض فرضية العدم عند مستوى معنوية عالية. هذه النتيجة غير مفاجئة لأن قيم اختبارات t للمتغيرات TECH و WORKER تملك معنوية عالية.

8-5- تأثيرات التغيير في التصنيف المرجعي:

THE EFFECTS OF CHANGING THE REFERENCE CATEGORY

في الفقرة السابقة اخترنا المدارس النظامية كمرجع (للحذف) للفئة، وحددنا المتغيرات الوهمية للفئات الأخرى. هذا يخولنا لمقارنة التكلفة الكلية للمدارس الأخرى مع المدارس النظامية، و اختبار -أي الاثنين - أكثر معنوية.

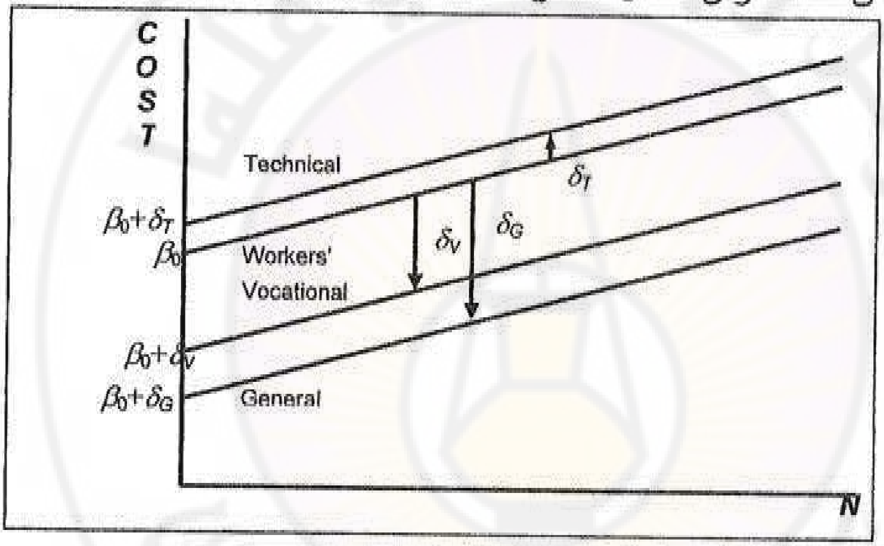
على أي حال نفترض أن الفائدة من الاختبار فيما التكلفة الكلية لمدارس العامل الماهر يكون مختلفاً عن الفئات الأخرى للمدارس. كيف نستطيع عمل هذا؟ إمكانية إنجاز اختبار t نستخدم مصفوفة التباين -التباين المشترك لمعلمات الانحدار لحساب علاقة الأخطاء المعيارية. ولكن يكون من المؤلم سهولة الوقوع بالأخطاء الحسابية. بشكل مبسط جداً إعادة تنفيذ الانحدار، وجعل مدارس العامل الماهر فئة مرجعية. الآن نحتاج إلى تحديد المتغير الوهمي GEN للمدارس النظامية . يأخذ النموذج الشكل التالي مع الأخذ بعين الاعتبار عدم وجود متغير وهمي طويل لمدارس العامل الماهر كمرجع فئة:

$$COST = 88{,}000 + 11{,}000TECH - 90{,}000VOC - 143{,}000GEN + 343N$$

$COST = \beta_0 + \delta_T TECH + \delta_V VOC + \delta_G GEN + \beta_2 N + u$	
Skilled Workers' School ($TECH = VOC = GEN = 0$)	$COST = \beta_0 + \beta_2 N + u$
Technical School ($TECH = 1$; $VOC = GEN = 0$)	$COST = (\beta_0 + \delta_T) + \beta_2 N + u$
Vocational School ($VOC = 1$; $TECH = GEN = 0$)	$COST = (\beta_0 + \delta_V) + \beta_2 N + u$
General School ($GEN = 1$; $TECH = VOC = 0$)	$COST = (\beta_0 + \delta_G) + \beta_2 N + u$

في هذه الحالة لمعالجة المشاهدات التي تنتمي إلى مدارس العامل الماهر في كل المتغيرات الوهمية تساوي الصفر. ومواصفات النموذج للتقاطع و الحد

N . في هذه الحالة المشاهدات التي تنتمي للمدارس التقنية TECH يساوي الواحد و التقاطع يضاف إليه القيمة δ. ملاحظة أن δ_T يجب أن يساوي تقاطع القيمة الكلية المضافة للمدرسة التقنية القريبة من مدرسة العامل الماهر. وبالمماثلة يمكننا اشتقاق معادلة المدارس الحرفية والمدارس النظامية، مع الأخذ بعين الاعتبار إضافة قيمة معاملات δ إلى التقاطع لتحديد التكلفة الكلية لمدارس العامل الماهر. يوضح الرسم البياني (11-8) النموذج مع ملاحظة أن δ تقيس من خط مدارس العامل الماهر .



الشكل (11-8)

نورد فيما يلي بيانات القيم العشرة الأولى من 74 مدرسة مع الأخذ بعين الاعتبار أن الفئة المرجعية مدارس العامل الماهر.

School	Type	COST	N	TECH	VOC	GEN
1	Technical	345,000	623	1	0	0
2	Technical	537,000	653	1	0	0
3	General	170,000	400	0	0	1
4	Workers'	526.000	663	0	0	0
5	General	100,000	563	0	0	1
6	Vocational	28,000	236	0	1	0
7	Vocational	160,000	307	0	1	0

8	Technical	45,000	173	1	0	0
9	Technical	120,000	146	1	0	0
10	Workers'	61,000	99	0	0	0

يعرض الجدول التالي نتائج المخرجات من برنامج Stata للانحدار .
سوف نشاهد أولاً معلمات الانحدار.

```
. reg COST N TECH VOC GEN

  Source |       SS       df       MS              Number of obs =      74
---------+------------------------------           F(  4,    69) =   29.63
   Model |  9.2996e+11     4  2.3249e+11           Prob > F      =  0.0000
Residual |  5.4138e+11    69  7.8461e+09           R-squared     =  0.6320
---------+------------------------------           Adj R-squared =  0.6107
   Total |  1.4713e+12    73  2.0155e+10           Root MSE      =   88578

------------------------------------------------------------------------------
    COST |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
       N |   342.6335    40.2195     8.519   0.000      262.3978    422.8692
    TECH |   10748.51   30524.87     0.352   0.726     -50146.93    71643.95
     VOC |  -90133.74   33984.22    -2.652   0.010     -157930.4   -22337.07
     GEN |  -143362.4    27852.8    -5.147   0.000     -198927.2   -87797.57
   _cons |   88469.29   28849.56     3.067   0.003      30916.01    146022.6
------------------------------------------------------------------------------
```

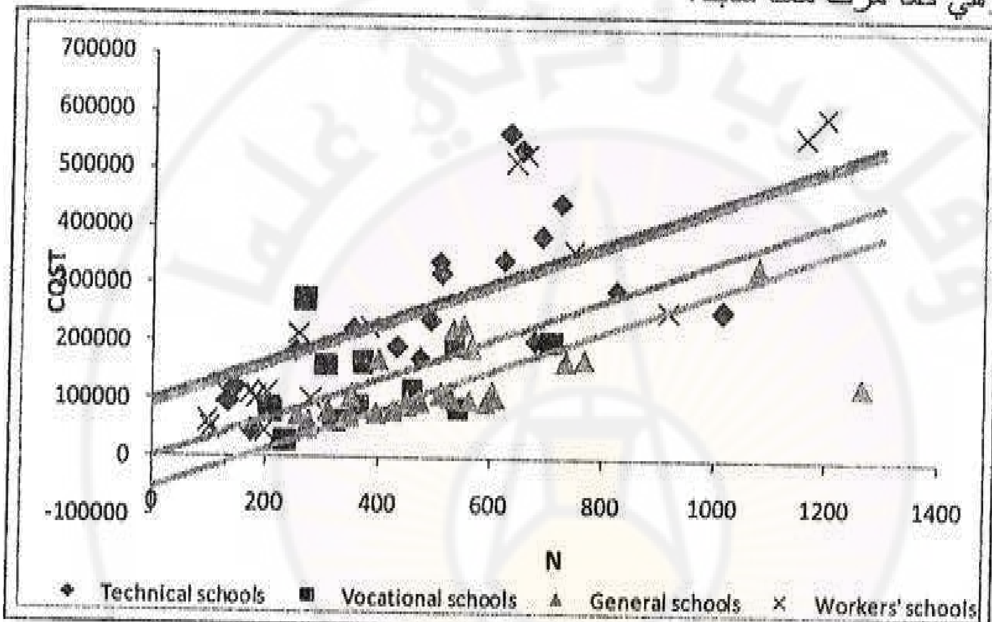
نستخلص من نتائج الانحدار معادلة النموذج كما في الجدول التالي:

$\hat{COST} = 88{,}000 + 11{,}000TECH - 90{,}000VOC - 143{,}000GEN + 343N$	
Skilled Workers' School (*TECH = VOC = GEN* = 0)	$\hat{COST} = 88{,}000 + 343N$
Technical School (*TECH* = 1; *VOC = GEN* = 0)	$\hat{COST} = 88{,}000 + 11{,}000 + 343N = 99{,}000 + 343N$
Vocational School (*VOC* = 1; *TECH = GEN* = 0)	$\hat{COST} = 88{,}000 - 90{,}000 + 343N = -2{,}000 + 343N$
General School (*GEN* = 1; *TECH = VOC* = 0)	$\hat{COST} = 88{,}000 - 143{,}000 + 343N = -55{,}000 + 343N$

نجعل جميع المتغيرات الوهمية تساوي الصفر، لنحصل على المعادلة للفئة المرجعية لمدارس العامل الماهر. ونجعل قيمة TECH تساوي الواحد و VOC و GEN تساوي الصفر؛ نحصل على معادلة المدارس التقنية. وبالمماثلة نحصل

على معادلات المدارس الحرفية والنظامية بإعطاء VOC و GEN القيمة واحد عند التحويل.

ملاحظة: معادلات التكلفة تعطينا بالضبط القيم نفسها عند استخدام المدارس النظامية كفئة مرجعية. فيما يلي شكل الانتشار لخطوط الانحدار، وهي كما مرت معنا سابقاً.



الشكل (8-12)

جودة التوفيق المقاسة بـ R^2 و RSS أو بالخطأ المعياري للانحدار (تقدير الانحراف المعياري لـ u وهنا نسميه MSE) بطريقة مماثلة لا تختلف نتيجة التغيير.

```
. reg COST N TECH VOC GEN

  Source |       SS       df       MS              Number of obs =      74
---------+------------------------------           F(  4,    69) =   29.63
   Model |  9.2996e+11     4  2.3249e+11           Prob > F      =  0.0000
Residual |  5.4138e+11    69  7.8461e+09           R-squared     =  0.6320
---------+------------------------------           Adj R-squared =  0.6107
   Total |  1.4713e+12    73  2.0155e+10           Root MSE      =   88578

------------------------------------------------------------------------------
    COST |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
       N |   342.6335    40.2195     8.519   0.000      262.3978    422.8692
    TECH |   10748.51   30524.87     0.352   0.726     -50146.93    71643.95
     VOC |  -90133.74   33984.22    -2.652   0.010     -157930.4   -22337.07
     GEN |  -143362.4    27852.8    -5.147   0.000     -198927.2   -87797.57
   _cons |   88469.29   28849.56     3.067   0.003      30916.01    146022.6
------------------------------------------------------------------------------
```

ولكن تخلف اختبارات t وتكون متأثرة. في هذه التجربة تختلف معاني فرضية العدم لمعلمات المتغيرات الوهمية المساوية للصفر. على سبيل المثال نجد أن التكلفة الكلية للمدارس التقنية نفسها كما في مدارس العامل الماهر حسب قيمة t الإحصائية لمعلمة المدرسة التقنية عند تطبيق فرضية العدم. معدل t في المعادلة يساوي فقط 0.35 لذلك نقبل فرضية العدم.

نجدد أن t الإحصائية لمعلمة VOC يساوي 2.65 ، لذلك نستنتج أن معنوية وحدة التكلفة الكلية للمدارس الحرفية أخفض مما هي عليه في مدارس العامل الماهر عند مستوى دلالة 1% .

المدارس النظامية واضح أنها تملك قيمة أقل في التكلفة الكلية من المدارس العامل الماهر، وفقاً للانحدار.

ملاحظة: يوجد بعض الاختلاف في الأخطاء المعيارية، حيث نجد أن الخطأ المعياري لمعلمة المتغير N غير متأثرة. هذا يتضمن اختباراً واحداً للمتغيرات الوهمية لأنه لا نستطيع إنجازه ، فيما إذا كانت التكلفة الكلية للمدارس النظامية ومدارس العامل الماهر مختلفة. اختيار مواصفات النموذج تجعل

الاختلاف خارج هذا الاختبار، و يؤدي الاختلاف فقط إلى حقيقة معامل الانحدار، الذي أصبح سالباً في المواصفات الثانية. ولكن الخطأ المعياري يكون نفسه ولذلك t الاحصائية لها بالقيمة المطلقة ذات أهمية، و الحصيلة أن الاختبار نفسه. على أي حال الخطأ المعياري لمعلمات المتغيرات الوهمية تتجاهل الضخامة في المواصفات الثانية.

هذا لأن مدارس العامل الماهر انتشارها أقل من المدارس النظامية، ولذلك عددها قليل في العينة (فقط 17 مقابل 28). وبالنتيجة الدقة تكون أقل في قياس الاختلاف بين التكلفة لهذه المدرسة والمدارس الأخرى أكثر مما هو عليه الحال عندما المدارس النظامية؛ أي الفئة المرجعية.

8-6- مجموعتان من المتغيرات الوهمية

TWO SETS OF DUMMY VARIABLES

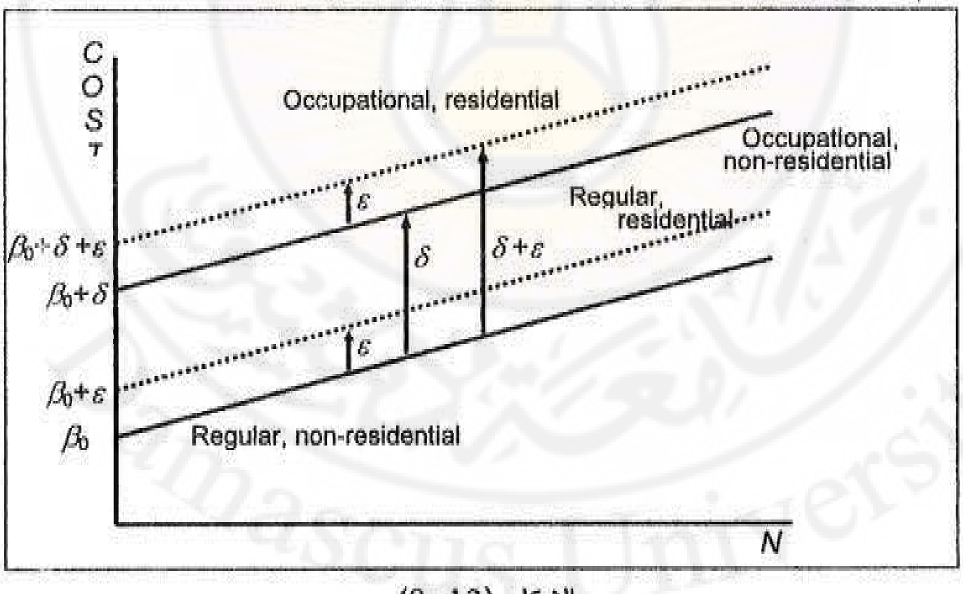
يمكن أن تتضمن علاقات الانحدار على عدد من المتغيرات التفسيرية تكون في طبيعتها مجموعات متعددة من المتغيرات الوهمية . في هذه الفقرة نعرض مثال لنموذج بنوعين. سنستمر بعرض معادلة تكلفة المدارس، ونوسع ذلك بأخذ حسابات حقيقية لبعض المدارس التي تقدم سكناً ملائماً. سندخل المتغير الوهمي RES الذي يساوي 1 للمدارس التي تقدم السكن و 0 للمدارس ذات السكن غير الملائم في نموذج التكلفة الكلية. ε يكون للتكلفة الكلية السنوية الإضافية للمدارس؛ التي تقدم السكن قريبة نسبياً من المدارس التي لا تقدم سكناً.

وهنا سوف نميز بين المدارس المهنية والعادية باستخدام المتغير الوهمي OCC المحدد في أول السلسلة . (من الأفضل استخدام أربع فئات مرجعية لتصنيف المدارس. إذا كانت المدرسة لها برنامج نظامي، و لا يوجد سكن ملائم فإن كلا المتغيرين الوهميين يساوي الصفر، ومعادلة التكلفة بمواصفات مكونات الأساس. تقديم السكن الملائم في المدارس النظامية RES تساوي 1 و

التقاطع يزداد بمبلغ ε . في هذا الحالة لا يوجد سكن ملائم في المدارس المهنية RES يساوي 0 و OCC يساوي 1 لذلك التكلفة الكلية تزيد بمقدار δ .
نلخص ذلك في الجدول التالي:

$\hat{COST} = \beta_0 + \delta OCC + \varepsilon RES + \beta_1 N + u$	
Regular, non-residential (*OCC* = *RES* = 0)	$\hat{COST} = \beta_0 + \beta_1 N + u$
Regular, residential (*OCC* = 0; *RES* = 1)	$\hat{COST} = (\beta_0 + \varepsilon) + \beta_1 N + u$
Occupational, non-residential (*OCC* = 1; *RES* = 0)	$\hat{COST} = (\beta_0 + \delta) + \beta_1 N + u$
Occupational, residential (*OCC* = *RES* = 1)	$\hat{COST} = (\beta_0 + \delta + \varepsilon) + \beta_1 N + u$

و يوضح الرسم البياني (13-8) النموذج، مع ملاحظة أن تأثيرات الاختلاف في مكونات النموذج يفترض أن تكون متمايزة، و مضافة لهذه المواصفات. نفترض في التجربة أن الإضافة في التكلفة الكلية للمدارس التي تقدم السكن سيكون كالمدارس النظامية والحرفية نفسها.

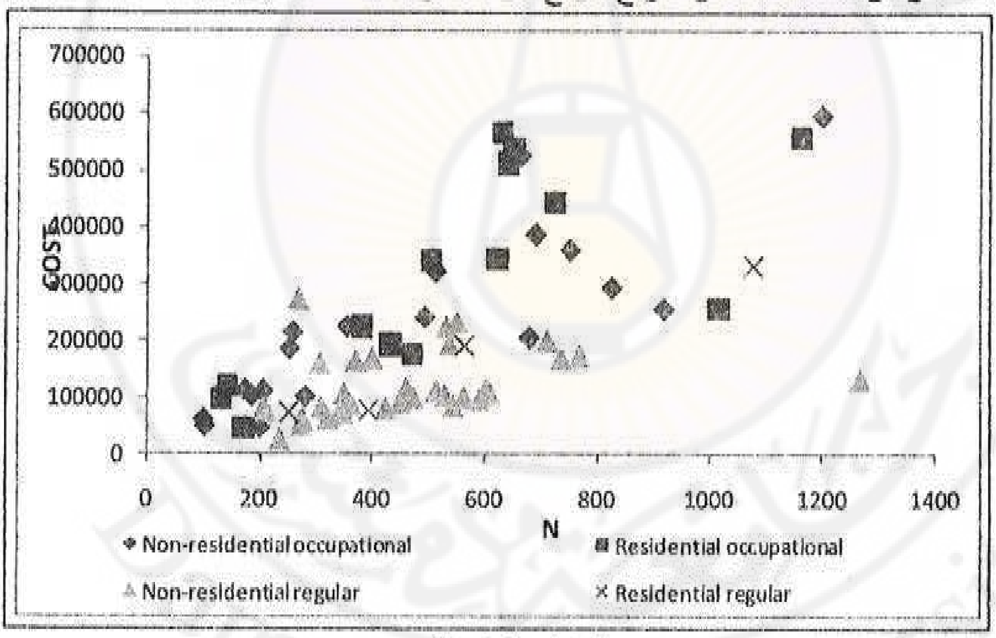


الشكل (13-8)

نظراً لكبر حجم العينة سوف نعرض البيانات العشر الأولى للعينة لإظهار كيفية توزيع المتغيرات الوهمية. قيم المتغيرات الوهمية وفق الخصائص الجديدة لصفات المدرسة. هنا نعرض البيانات العشر الأولى .

School	Type	Residential?	*COST*	*N*	*OCC*	*RES*
1	Occupational	No	345,000	623	1	0
2	Occupational	Yes	537,000	653	1	1
3	Regular	No	170,000	400	0	0
4	Occupational	Yes	526.000	663	1	1
5	Regular	No	100,000	563	0	0
6	Regular	No	28,000	236	0	0
7	Regular	Yes	160,000	307	0	1
8	Occupational	No	45,000	173	1	0
9	Occupational	No	120,000	146	1	0
10	Occupational	No	61,000	99	1	0

هنا نعرض شكل الانتشار لأربع أنواع من المدارس.



الشكل (14-8)

الجدول التالي يبين نتائج الانحدار حسب برنامج Stata ، وسوف نبدأ بتفسير معلمات الانحدار . المعلمة N تتضمن التكلفة الحدية للطالب، ويساوي 322 يناً في السنة.

```
. reg COST N OCC RES

  Source |       SS       df       MS              Number of obs =      74
---------+------------------------------           F(  3,    70) =   40.43
   Model |  9.3297e+11     3  3.1099e+11           Prob > F      =  0.0000
Residual |  5.3838e+11    70  7.6911e+09           R-squared     =  0.6341
---------+------------------------------           Adj R-squared =  0.6184
   Total |  1.4713e+12    73  2.0155e+10           Root MSE      =   87699

------------------------------------------------------------------------------
    COST |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
       N |   321.833    39.40225     8.168   0.000      243.2477    400.4183
     OCC |  109564.6    24039.58     4.558   0.000      61619.15      157510
     RES |  57909.01    30821.31     1.879   0.064     -3562.137    119380.2
   _cons | -29045.27    23291.54    -1.247   0.217     -75498.78    17408.25
------------------------------------------------------------------------------
```

الثابت يزودنا بتقدير التكلفة الكلية السنوية للفئة المرجعية ، وهي المدارس النظامية التي تقدم سكناً. مازالت قيمتة سالبة، وكما أشرنا سابقاً لا تعني شيئاً بسبب العشوائية.

معامل OCC يتضمن التكلفة السنوية الكلية للمدارس المهنية، ويساوي 110000 ين أكثر من المدارس النظامية. المعلمة RES تتضمن التكلفة السنوية الكلية للمدارس التي تقدم السكن يساوي إلى 58000 ين أكبر مما هو في المدارس التي لا تقدم السكن.

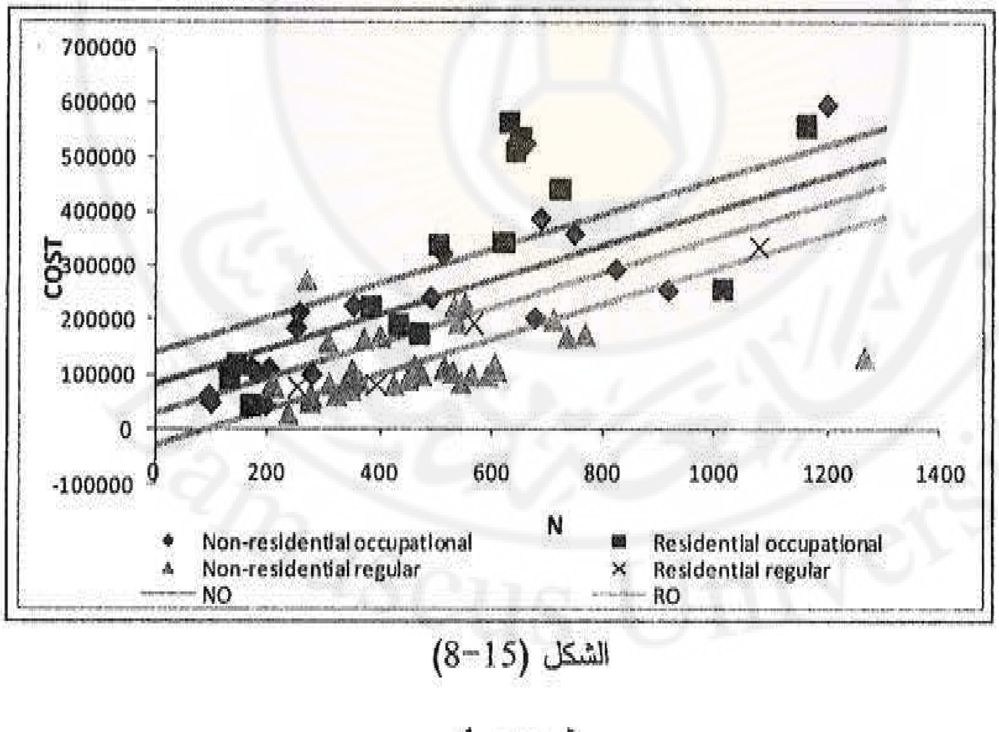
عند إعطاء القيمة صفراً لكلا المتغيرين الوهميين، نحصل على معادلة التكلفة المتضمنة للمدارس النظامي؛ التي لا تقدم السكن.

نعطي المتغير RES القيمة 1 ، ولكن نحفظ OCC مساوي 0 نحصل على معادلة التكلفة لمدارس النظامية التي تقدم السكن.

و بالمماثلة معادلة التكلفة لمدارس المهنية التي تقدم السكن، و تشتق المعادلة للمدارس المهنية التي لا تقدم السكن بوضع OCC تساوي 1 و RES يساوي 0 و 1 على التوالي.

$\hat{COST}$ = -29,000 + 110,000*OCC* + 58,000*RES* + 322*N*	
Regular, non-residential (*OCC* = *RES* = 0)	$\hat{COST}$ = -29,000 + 322*N*
Regular, residential (*OCC* = 0; *RES* = 1)	$\hat{COST}$ = -29,000 + 58,000 + 322*N* = 29,000 + 322*N*
Occupational, non-residential (*OCC* = 1; *RES* = 0)	$\hat{COST}$ = -29,000 + 110,000 + 322*N* = 81,000 + 322*N*
Occupational, residential 322*N* (*OCC* = *RES* = 1)	$\hat{COST}$ = -29,000 + 110,000 + 58,000 +322*N* = 139,000 + 322*N*

نبين فيما يلي شكل الانتشار لمعادلات التكلفة الأربعة لنتائج الانحدار.



الشكل (15-8)

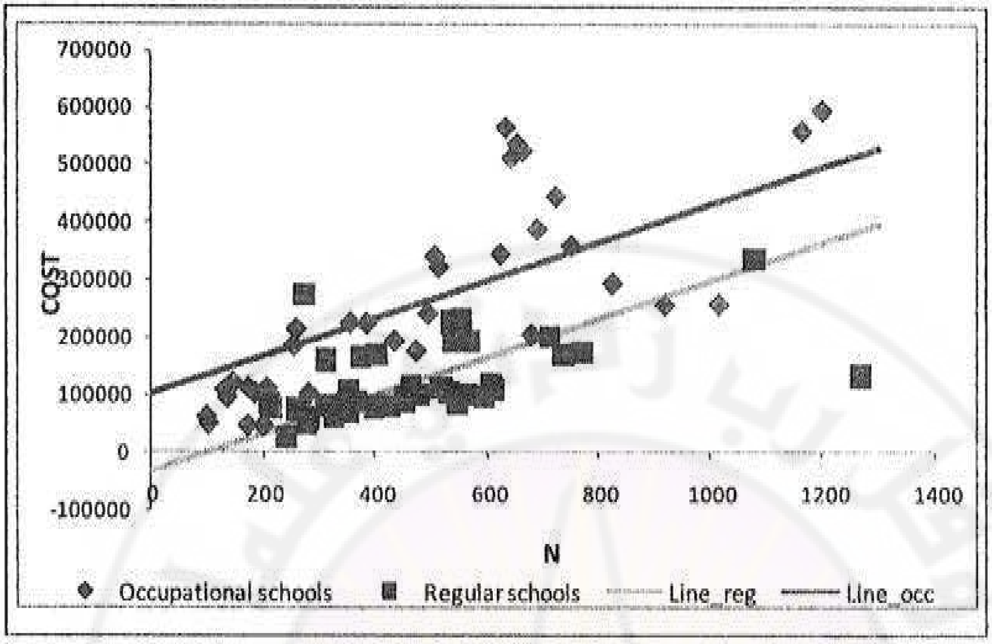
ننجز اختبار t الإحصائية بالطريقة الاعتيادية. نجد أن معلمة المدرسة المهنية بالنسبة للمتغير الوهمي معنوية ، فهي تختلف عن 0 عند مستوى 0.1%. وعلى أي حال t الإحصائية للمعلمة RES يساوي 1.87 . بالطبع يمكن إنجاز الاختبار من اتجاه واحد بنجاح (لماذا؟) لأن المعنوية تختلف عن الصفر عند 5% مستوى دلالة (ولكن لا تختلف عن الصفر عند 1% مستوى دلالة).

8-7- ميل المتغيرات الوهمية SLOPE DUMMY VARIABLES

يبين شكل الانتشار بيانات المدارس الـ74 في شنغهاي، و معادلات التكلفة المشتقة من الانحدار لـ COST على N و المتغير الوهمي لنوع التي لها مناهج دراسية (المهنية والعادية).

مواصفات النموذج الذي يدمج القيم الحدية لكل طالب له المدارس نفسها النظامية و المهنية ، وهكذا نجد معادلات انحدار لها تكلفة متوازية . على أي حال هذه ليست فرضية واقعية . معلوم أن المدارس المهنية تنفق على التدريب قيم مواد يتعلق بعدد الطلاب. أيضاً نسب هيئة التعليم المهني أعلى من التعليم الأكاديمي الكلاسيكي .

إذا دققنا في شكل الانتشار نستطيع أن نشاهد معادلة تكلفة المدارس المهنية شديد الانحدار، و بينما مدارس التعليم النظامي أقل انحداراً.



الشكل (19-8)

من جدول النتائج أدناه تشير الإشارة السالبة إلى أن التكلفة الكلية للمدارس المهنية تكون فعلياً أقل مما هي عليه في المدارس العادية. بالطبع هذا احتمال مستبعد. على أي حال t الإحصائية فقط 0.09- الأمر الذي يؤدي إلى عدم رفض فرضية العدم، لأن التكلفة الكلية لكلا النوعين متساوية. نستطيع أيضاً إنجاز اختبار F للدلالة على قوة المتغيرات التفسيرية للمتغير الوهمي مقارنة مع RSS عندما المتغيرات الوهمية تكون متضمنة مع RSS عندما لا تكون المتغيرات الوهمية مضمنة في النمو النموذج.

```
reg COST N OCC NOCC

  Source |       SS       df       MS              Number of obs =      74
---------+------------------------------           F(  3,    70) =   49.64
   Model |  1.0009e+12     3  3.3363e+11           Prob > F      =  0.0000
Residual |  4.7045e+11    70  6.7207e+09           R-squared     =  0.6803
---------+------------------------------           Adj R-squared =  0.6666
   Total |  1.4713e+12    73  2.0155e+10           Root MSE      =   81980

-------------------------------------------------------------------------------
-
    COST |   Coef.    Std. Err.    t     P>|t|    [95% Conf. Interval]
---------+---------------------------------------------------------------------
-
       N |  152.2982   60.01932   2.537   0.013   32.59349      272.003
     OCC |-3501.177   41085.46   -0.085   0.932  -85443.55     78441.19
    NOCC |  284.4786   75.63211   3.761   0.000   133.6351     435.3221
   _cons |  51475.25   31314.84   1.644   0.105  -10980.24     113930.7
-------------------------------------------------------------------------------
```

```
. reg COST N OCC NOCC

  Source |       SS       df       MS              Number of obs =      74
---------+------------------------------           F(  3,    70) =   49.64
   Model |  1.0009e+12     3  3.3363e+11           Prob > F      =  0.0000
Residual |  4.7045e+11    70  6.7207e+09           R-squared     =  0.6803
---------+------------------------------           Adj R-squared =  0.6666
   Total |  1.4713e+12    73  2.0155e+10           Root MSE      =   81980

-------------------------------------------------------------------------------
. reg COST N

  Source |       SS       df       MS              Number of obs =      74
---------+------------------------------           F(  1,    72) =   46.82
   Model |  5.7974e+11     1  5.7974e+11           Prob > F      =  0.0000
Residual |  8.9160e+11    72  1.2383e+10           R-squared     =  0.3940
---------+------------------------------           Adj R-squared =  0.3856
```

فرضية العدم لاختبار معاملات المتغيرات OCC و NOCC و كلاهما يساوي الصفر. الفرضية البديلة واحد على الأقل لا يساوي الصفر.

قياس التحسن في جودة التوفيق عند إضافة المتغيرات الوهمية حسب نسبة التخفيض الــ RSS .

$$F(2,70) = \frac{(8.92\times10^{11} - 4.70\times10^{11})/2}{4.70\times10^{11}/70} = 31.4$$

نظراً لإضافة متغيرين فإن عدد درجات الحرية 2 ، و كما ينخفض عدد درجات الحرية من 72 إلى 70 لمعاملات المتغيرات الوهمية. الحد الأول للبسط يساوي RSS بعد الإضافة. أما المقام فهو RSS بعد الإضافة مقسوم على عدد درجات الحرية المخفضة أي الـ 70 تصبح 74 مشاهدة و 4 ثوابت (معلمات) مقدرة. وهكذا نجد أن أقل معامل من معاملات المتغير الوهمي يختلف عن الصفر. ونعرف هذا تماماً من اختبار t ولذلك في هذه الحالة اختبار F لا يضيف أي شيء. لذلك نلجأ إلى اختبار أخر لحسم ذلك هو اختبار تشو CHOW TEST

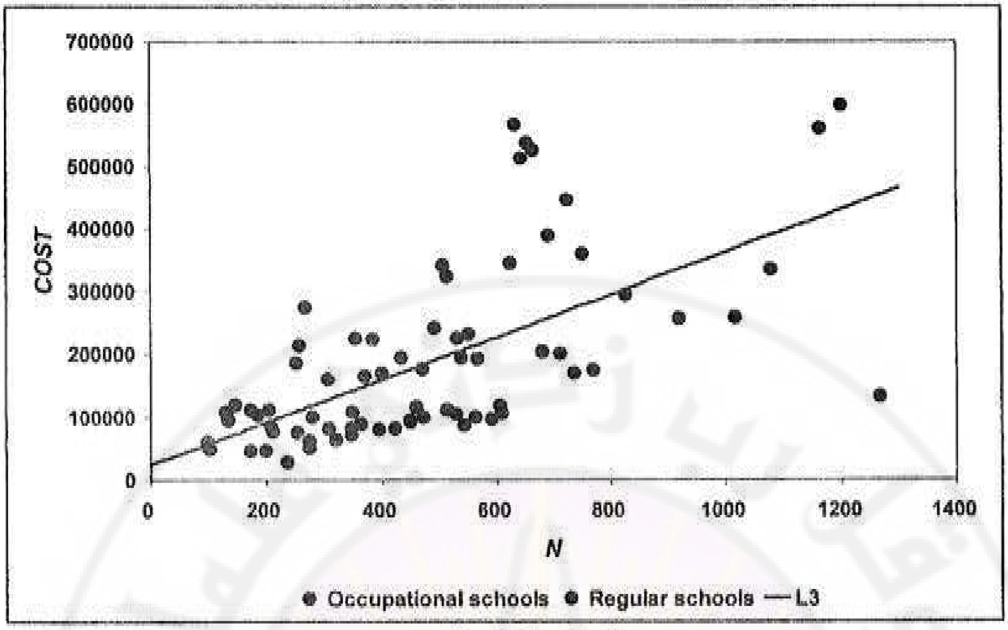
8-8- اختبار تشو CHOW TEST

للمفاضلة بين نوعين من المتغيرات في عينة البيانات ، أثناء استخدام نماذج الانحدار المطبقة على فئتين من المتغيرات لتميز نموذج عن الآخر؛ نلجأ إلى استخدام اختبار Chow .

سوف نوضح ذلك من خلال التطبيق على مثال عينة المدارس الثانوية 74 في شنغهاي. نعرض مخرجات الانحدار عندما نأخذ انحدار التكلفة على العدد N ونعمل على التمييز بين اختلاف نوع المدرسة.

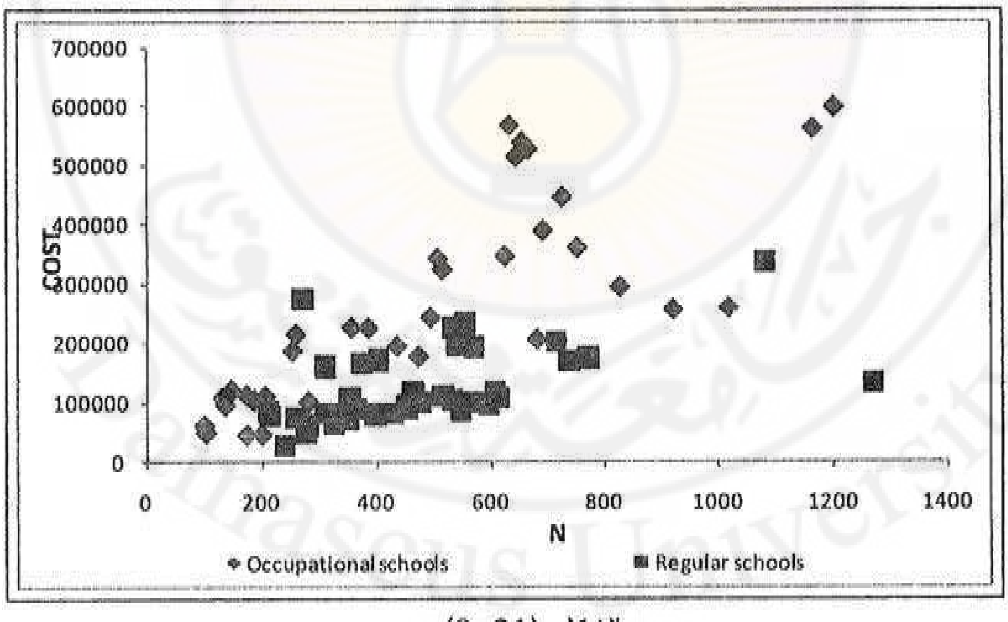
```
. reg COST N

  Source |       SS       df       MS              Number of obs =      74
---------+------------------------------           F(  1,    72) =   46.82
   Model |  5.7974e+11     1  5.7974e+11           Prob > F      =  0.0000
Residual |  8.9160e+11    72  1.2383e+10           R-squared     =  0.3940
---------+------------------------------           Adj R-squared =  0.3856
   Total |  1.4713e+12    73  2.0155e+10           Root MSE      = 1.1e+05

------------------------------------------------------------------------------
    COST |      Coef.   Std. Err.     t     P>|t|     [95% Conf. Interval]
---------+--------------------------------------------------------------------
       N |   339.0432   49.55144    6.842   0.000     240.2642    437.8222
   _cons |    23953.3   27167.96    0.882   0.381    -30205.04    78111.65
------------------------------------------------------------------------------
```

يبين شكل الانتشار (20-8)العلاقة بين الإنفاق السنوي، و عدد الطلاب، وخط الانحدار الممكن.



الشكل (8-20)

لنعمل على التمييز بين المدارس المهنية والمـدارس النظاميـة، و ننفـذ الانحدار لعينتين جزئيتين.



الشكل (8-21)

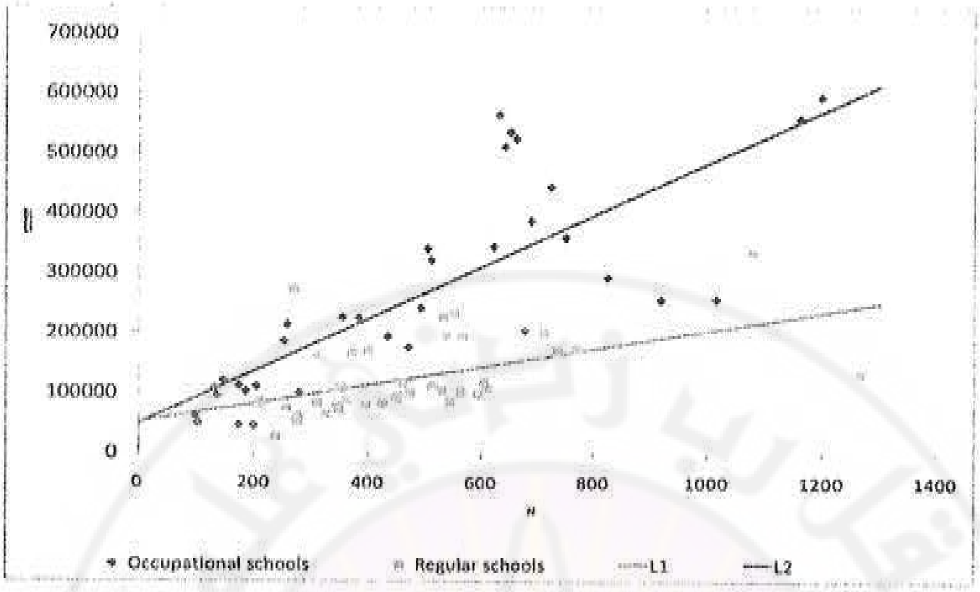
جدول نتائج مخرجات الانحدار للعينة الأولى عندما نأخذ انحدار التكلفة على N باستخدام عينة جزئية لمدارس التعليم المهني، وعددها 34 مدرسة.

```
. reg COST N if OCC==1

  Source |       SS       df       MS              Number of obs =      34
---------+------------------------------           F(  1,    32) =   55.52
   Model |  6.0538e+11     1  6.0538e+11           Prob > F      =  0.0000
Residual |  3.4895e+11    32  1.0905e+10           R-squared     =  0.6344
---------+------------------------------           Adj R-squared =  0.6229
   Total |  9.5433e+11    33  2.8919e+10           Root MSE      = 1.0e+05

------------------------------------------------------------------------------
    COST |     Coef.   Std. Err.      t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
       N |  436.7769   58.62085    7.451   0.000      317.3701    556.1836
   _cons |  47974.07   33879.03    1.416   0.166     -21035.26    116983.4
------------------------------------------------------------------------------
```

ونوفق نموذج آخر وهو انحدار التكلفة على N لعينة جزئية ثانية مكونة لـ 40 مدرسة نظامية.

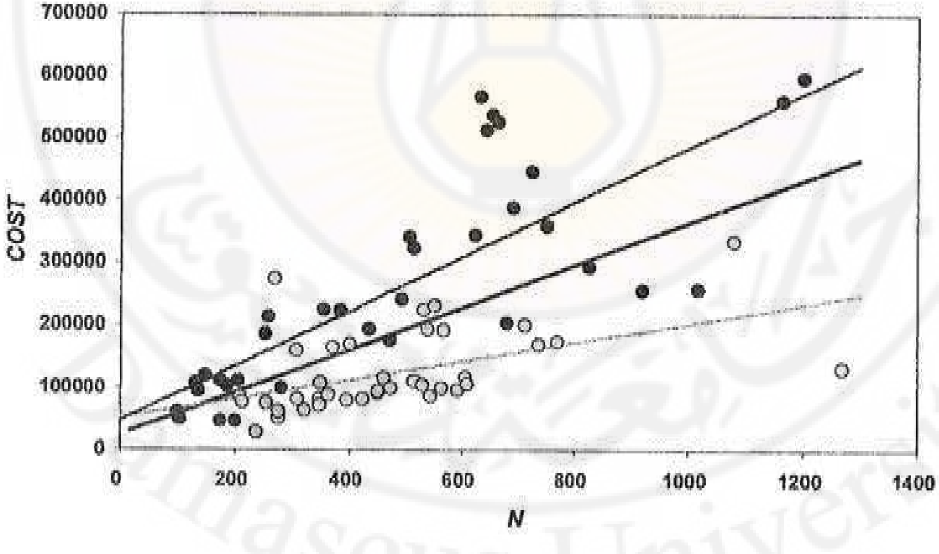
```
. reg COST N if OCC==0

  Source |       SS       df       MS              Number of obs =      40
---------+------------------------------           F(  1,    38) =   13.53
   Model |  4.3273e+10     1  4.3273e+10           Prob > F      =  0.0007
Residual |  1.2150e+11    38  3.1973e+09           R-squared     =  0.2626
---------+------------------------------           Adj R-squared =  0.2432
   Total |  1.6477e+11    39  4.2249e+09           Root MSE      =   56545

------------------------------------------------------------------------------
    COST |     Coef.   Std. Err.      t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
       N |  152.2982   41.39782    3.679   0.001      68.49275    236.1037
   _cons |  51475.25   21599.14    2.383   0.022      7750.064    95200.43
------------------------------------------------------------------------------
```

يبين الرسم البياني (22-8) خطي الانحدار للعينتين :



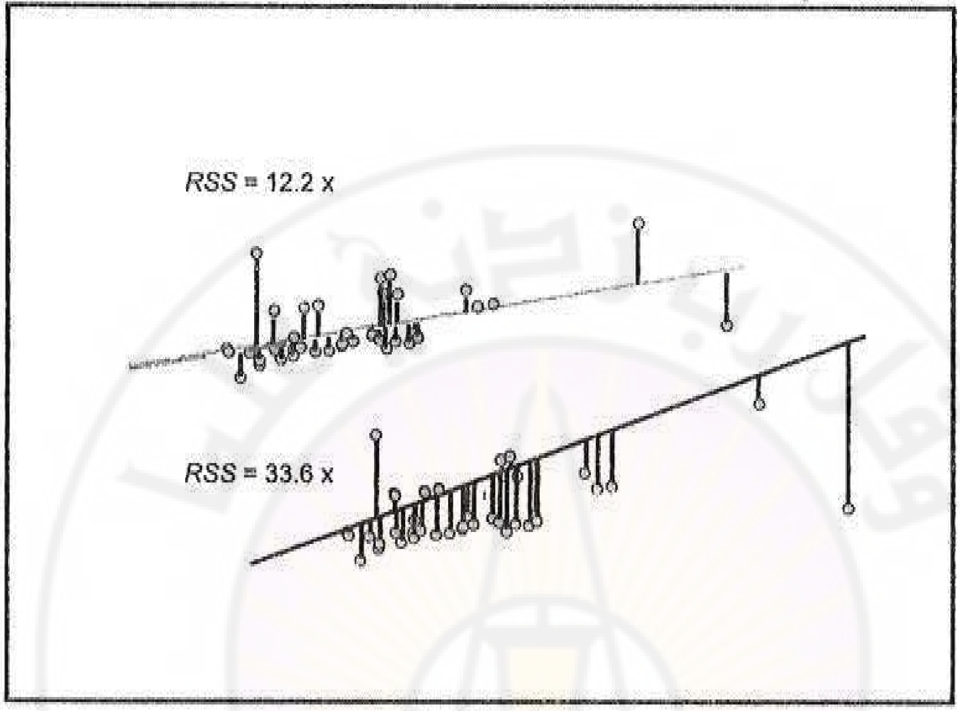
الشكل (22-8)

خط الانحدار للعينة المشتركة (إدخال العينة دون تمييز) ببين المقارنــة بين الخطوط الثلاثة.



الشكل (23-8)

نأخذ المجموعتين من البواقي لتظهر المقارنة . مرة ثانية RSS يجب أن تكون أعلى في العينة المشتركة للانحدار.



الشكل (27-8)

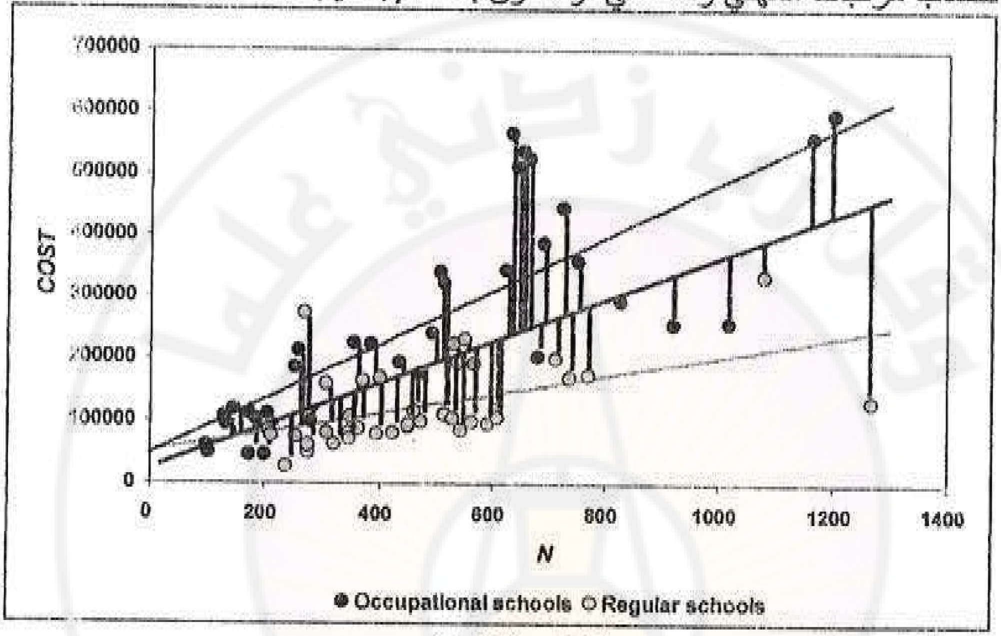
الجدول التالي يلخص بيانات RSS للنوعين من المدارس في الحالات المنفصلة والمشتركة الانحدار.

RESIDUAL SUM OF SQUARES ($x10^{11}$)			
Regression	Occupational	Regular	Total
	RSS_1	RSS_2	(RSS_1+RSS_2)
Separate	3.49	1.22	4.71
			RSS_P
Pooled 5.55	3.36	8.91	

سنرمز على التوالي RSS_1 و RSS_2 لمجموع مربعات البواقي المنفصلة للانحدار للمدارس المهنية والنظامية.

نجمعهما مع بعضها بعض لنحصل على مربع مجموع البواقي المنفصلة للانحدار العينتين الجزئيتين. ونقوم بمقارنة هذا المجموع مع RSSp مجموع مربع البواقي لعينة الانحدار المشتركة.

نعرض فما يلي الناتج مباشرة من الانحدار الأصلي المشترك، ولا داعي لحساب مركبات المهني والنظامي وستكون بشكلاً إجمالياً.



الشكل (28-8)

الفائدة المرجوة من ذلك المعنوية الإحصائية، فهي تنخفض عندما ننفذ الانحدارات المنفصلة للعينتين الجزئيتين.

الشكل العام لعلاقة اختبار تشو CHOW TEST :

$$F(k, n-2k) = \frac{\text{overall reduction in } RSS \text{ when separate regressions are run} \,/\, \text{cost in degrees of freedom}}{\text{total } RSS \text{ remaining when separate regressions are run} \,/\, \text{degrees of freedom remaining}}$$

$$= \frac{(RSS_p - [RSS_1 + RSS_2])/k}{(RSS_1 + RSS_2)/(n-2k)}$$

$$F(k, n-2k) = \frac{\text{الفرق بين مربع مجموع البواقي الناتج عن الانحدار للعينة ككل ناقص مجموع مربع البواقي للانحدار المنفصل للعينات الجزئية} \,/\, \text{تكلفة درجات الحرية } k}{\text{إجمالي مجموع مربع البواقي للانحدار المنفصل للعينات الجزئية} \,/\, \text{عدد درجات الحرية سارية المفعول}}$$

والأمر الذي يجب الانتباه إليه عند حساب F الإحصائية حسب هذه العلاقة هو قيمة k (عدد درجات الحرية) التكلفة في حدود درجات الحرية لتنفيذ الانحدارات المنفصلة. التكلفة تكون k لأنه يوجد مجموعتان من معلمات k المقدرة عندما الانحدارات تكون منفصلة، ويكون تنفيذها بدلاً من مجموعة واحدة مع الانحدار المشترك.

والحالة الثانية لـ F الإحصائية لهذه العلاقة هو أن n-2k إجمالي عدد درجات الحرية سارية المفعول لنماذج الانحدارات منفصلة التنفيذ ، يوجد n مشاهدة و k درجة حرية تكون مستخدمة لكل انحدار عندما الانحدارات المنفصلة تكون منفذة بشكل منفصل.

بسط علاقة F الإحصائية يتألف بشكل عام من التحسن فيRSS العينة مقسوماً على التكلفة في حدود درجات الحرية، عندما نميز الانحدارات.

أما المقام لـ F الإحصائية هو إجمالي RSS ساري المفعول بعد انقسام العينة على عدد درجات الحرية سارية المفعول.

$$RSS_p = 8.91;\ \ (RSS_1+RSS_2)=4.71\ ;$$

$$F(k,\ n-2k) = \frac{(RSS_p - [RSS_1 + RSS_2])/k}{(RSS_1 + RSS_2)/(n-2k)}$$

$$F(2{,}70) = \frac{(8.91\times10^{11} - [3.49\times10^{11} + 1.22\times10^{11}])/2}{(3.49\times10^{11} + 1.22\times10^{11})/70} = 31.2$$

يوجد فقط نوعين من المعلمات في النموذج الثابت ومعامل N وكذلك الحالة الأولى F الإحصائية تساوي 2 . مجموع مربع البواقي سارية المفعول بعد انقسام العينة هو: (RSS_1+RSS_2). يوجد 74 مشاهدة عندئذ يوجد 70 درجة حرية سارية المفعول بعد تقدير مجموعتين من المعلمات. إذاً تساوي F الإحصائية 31.2 . القيمة الحرجة لـ F(2,70) من جدول F نأخذ القيمة المقابلة لـ F(2,60) عند مستوى دلالة 0.1% فهي تساوي . $F(2{,}60)_{crit,\,0.1\%} = 7.8$

التخفيض بمجموع مربع البواقي من أجل مستوى معنوية 0.1% ، يلغي معادلة التكلفة المشتركة ويبين أنها ذو مواصفات غير كافية و يجب أن ننفذ الانحدارات المنفصلة لنوعي المدارس.

8-9- اختبار تشو و اختبار مجموعة المتغيرات الوهمية

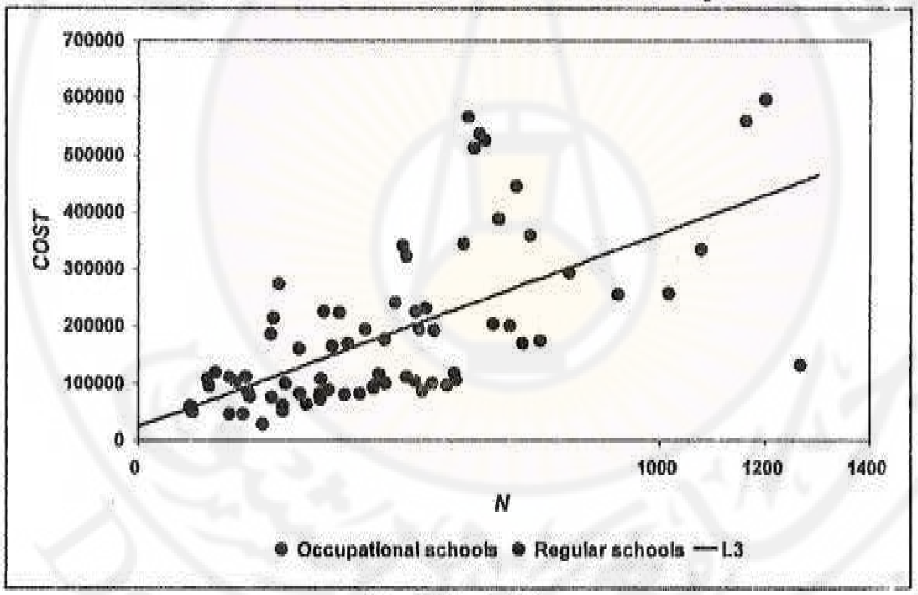
CHOW TEST AND DUMMY VARIABLE GROUP TEST

في الاختبارات السابقة المطبقة على المتغيرات الوهمية، و في اختبار Chow كنا نتحقق فيما إذا كانت معادلات التكلفة للتعليم المهني والأساسي مختلفة. نجد في كل حالة ننجز فيها الاختبار أن المعادلات معنوية الاختلاف. هل هذا الشيء الظاهر يعطي استنتاجات مختلفة ؟ الجواب بالطبع لا . اختبار Chow مساوياً اختبار F لقوة المتغيرات الوهمية المستقلة كمجموعة.

في كلا المقاربات نبدأ من نقطة يكون فيها الانحدار بسيطاً للإنفاق السنوي على عدد الطلاب المسجلين باستخدام العينة، ونركز على قيمة RSS .

```
. reg COST N

  Source |       SS       df       MS              Number of obs =      74
---------+------------------------------           F(  1,    72) =   46.82
   Model |  5.7974e+11     1  5.7974e+11           Prob > F      =  0.0000
Residual |  8.9160e+11    72  1.2383e+10           R-squared     =  0.3940
---------+------------------------------           Adj R-squared =  0.3856
   Total |  1.4713e+12    73  2.0155e+10           Root MSE      = 1.1e+05

------------------------------------------------------------------------------
    COST |    Coef.   Std. Err.       t     P>|t|      [95% Conf. Interval]
---------+--------------------------------------------------------------------
---------
       N | 339.0432   49.55144     6.842   0.000     240.2642    437.8222
   _cons | 23953.3    27167.96     0.882   0.381    -30205.04    78111.65
------------------------------------------------------------------------------
```

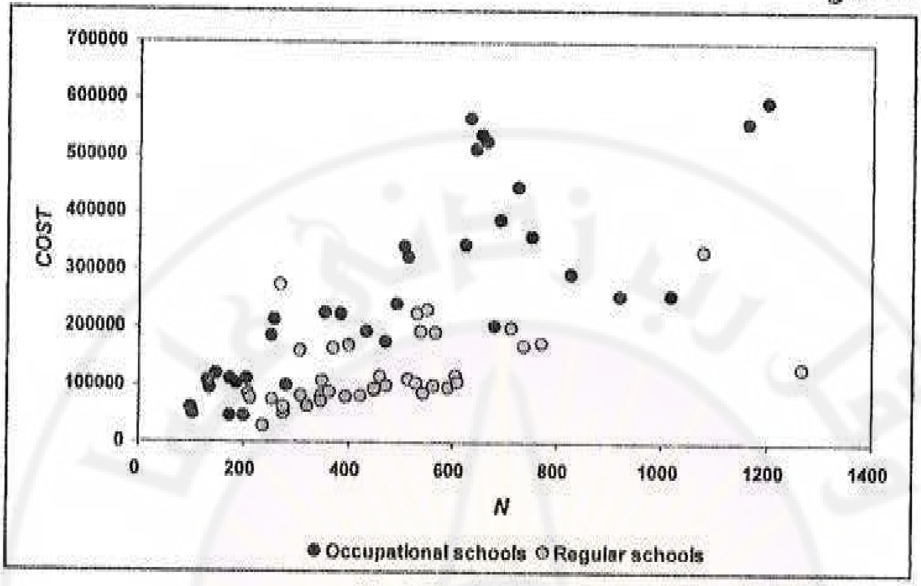
نبين فيما يلي خط الانحدار للعينتين الجزئيتين.



الشكل (29-8)

سوف نبين الاختلاف بين مدارس التعليم المهني والتعليم الأساسي ، حيث نضيف إلى تقاطع المتغير الوهمي ميل المتغير الوهمي؛ الذي يسمح بالقيمة

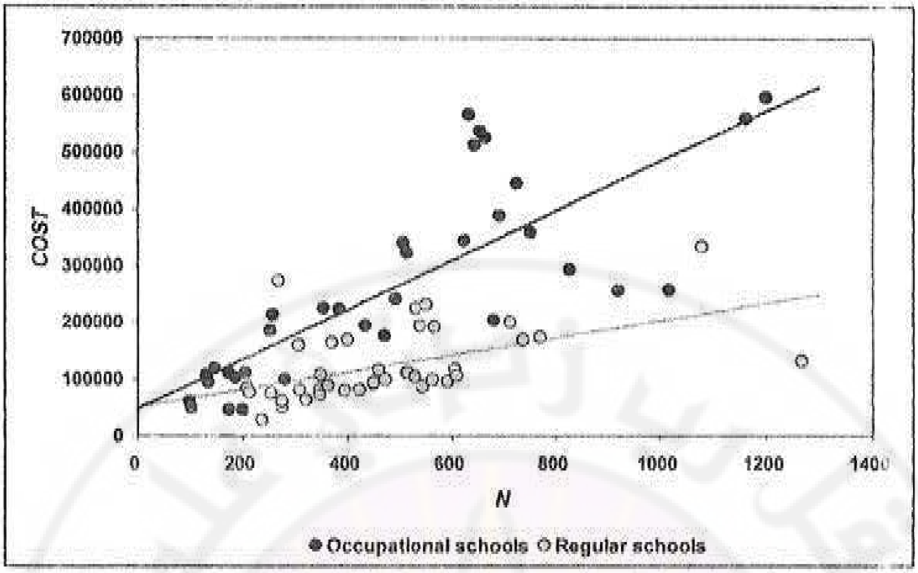
الكلية و التكلفة الحدية للتعليم المهني. مرة ثانية نركز أولاً على قيمة RSS الأصغر.



الشكل (8-30)

```
. reg COST N OCC NOCC

  Source |       SS       df       MS              Number of obs =      74
---------+------------------------------           F(  3,    70) =   49.64
   Model |  1.0009e+12     3  3.3363e+11           Prob > F      =  0.0000
Residual |  4.7045e+11    70  6.7207e+09           R-squared     =  0.6803
---------+------------------------------           Adj R-squared =  0.6666
   Total |  1.4713e+12    73  2.0155e+10           Root MSE      =   81980

------------------------------------------------------------------------------
    COST |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
       N |   152.2982   60.01932     2.537   0.013      32.59349     272.003
     OCC |  -3501.177   41085.46    -0.085   0.932     -85443.55    78441.19
    NOCC |   284.4786   75.63211     3.761   0.000      133.6351    435.3221
   _cons |   51475.25   31314.84     1.644   0.105     -10980.24    113930.7
------------------------------------------------------------------------------
```

نبين فيما يلي خط الانحدار للعينتين الجزئيتين.



الشكل (31-8)

إذا كانت معادلات التكلفة مختلفة المعنوية؛ عندئذ نبحث فيما إذا كان يوجد تخفيض معنوي في RSS عندما نضيف المتغيرات الوهمية.

ننجز اختبار F الموجود على ميل المتغيرات الوهمية. البسط في الاختبار الإحصائي عبارة عن التخفيض في RSS نتيجة إضافة المتغيرات الوهمية مقسومة على التكلفة بحدود درجات الحرية.

النموذج قبل إضافة المتغير الوهمي	
$COST = 24{,}000 + 339N$	$RSS = 8.91\times10^{11}$
النموذج بعد إضافة المتغير الوهمي	
$COST = 51{,}000 - 4{,}000OCC + 152N + 284NOCC$	$RSS = 4.71\times10^{11}$

$$F(2,70) = \frac{(8.91\times10^{11} - 4.71\times10^{11})/2}{4.71\times10^{11}/70} = 31.2$$

في المقام يكون RSS القيمة الحالية بعد إضافة المتغيرات الوهمية مقسومة على عدد درجات الحرية للقيمة الحالية(أي سارية المفعول). نجد أن

قيمة F الإحصائية تساوي 31.2 . أما القيمة الحرجة لـ F(2,70) من جدول F نأخذ قيمة F(2,60) عند مستوى دلالة 0.1% .

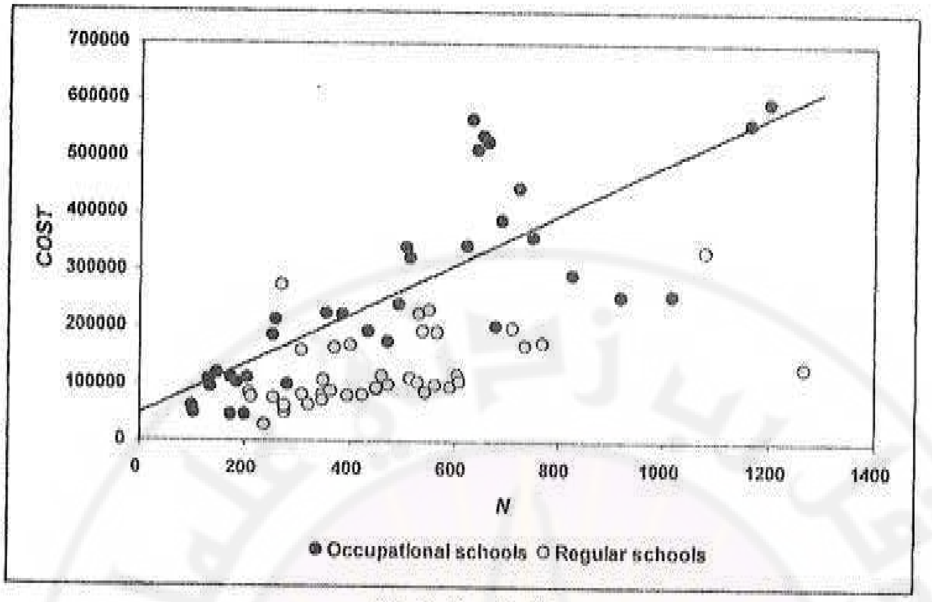
أما مع اختبار Chow أيضاً سنبدأ من تنفيذ الانحدار باستخدام العينة، ونلاحظ قيمة RSS . عندما نقسم العينة إلى مدارس مهنية ومداس نظامية و ننفذ الانحدار المقسم، مرة ثانية نحسب قيمة RSS . مخرجات الانحدار معروضة في الجدول التالي عندما نأخذ انحدار التكلفة على N لعينة جزئية من 40 مدرسة نظامية.

```
. reg COST N if OCC==0

  Source |       SS       df       MS              Number of obs =      40
---------+------------------------------           F(  1,    38) =   13.53
   Model |  4.3273e+10     1  4.3273e+10           Prob > F      =  0.0007
Residual |  1.2150e+11    38  3.1973e+09           R-squared     =  0.2626
---------+------------------------------           Adj R-squared =  0.2432
   Total |  1.6477e+11    39  4.2249e+09           Root MSE      =   56545

------------------------------------------------------------------------------
    COST |   Coef.    Std. Err.       t     P>|t|    [95% Conf. Interval]
---------+--------------------------------------------------------------------
       N |152.2982    41.39782     3.679    0.001    68.49275     236.1037
   _cons |51475.25    21599.14     2.383    0.022    7750.064     95200.43
------------------------------------------------------------------------------
```

ونبين في الجدول التالي مخرجات الانحدار عندما نأخذ انحدار التكلفة على N عندما نستخدم العينة الجزئية لـ 34 مدرسة مهنية.

```
. reg COST N if OCC==1

  Source |       SS       df       MS              Number of obs =      34
---------+------------------------------           F(  1,    32) =   55.52
   Model |  6.0538e+11     1  6.0538e+11           Prob > F      =  0.0000
Residual |  3.4895e+11    32  1.0905e+10           R-squared     =  0.6344
---------+------------------------------           Adj R-squared =  0.6229
   Total |  9.5433e+11    33  2.8919e+10           Root MSE      = 1.0e+05

------------------------------------------------------------------------------
    COST |   Coef.    Std. Err.      t     P>|t|    [95% Conf. Interval]
---------+--------------------------------------------------------------------
       N | 436.7769   58.62085    7.451    0.000    317.3701     556.1836
   _cons | 47974.07   33879.03    1.416    0.166   -21035.26     116983.4
------------------------------------------------------------------------------
```

يبين الشكل البياني(32-8) خطوط الانحدار للنموذجين المقترحين.



الشكل (34-8)

معادلة التكلفة المتضمنة في انحدار المتغير الوهمي تتطابق مع كلا الانحداريين المنفصلين، وسيكون لها البواقي نفسها. وهكذا المعنوية الإحصائية لاختبار F سيكون نفسه. البدء من كلا النقاط يقترب من مجموع مربع البواقي لأساس الانحدار، لا تمايزاً بين نوعي المدارس.

$COST = 51{,}000 + 152N$	المدارس النظامية فقط
	$RSS = 1.22 \times 10^{11}$
$COST = 47{,}000 + 436N$	المدارس المهنية فقط
	$RSS = 3.49 \times 10^{11}$
النموذج مع المتغيرات الوهمية	
COST = [illegible] 152N + 284NOCC	[illegible]
	$RSS = 4.71 \times 10^{11}$

$$F(2,70) = \frac{(8.91\times10^{11} - [3.49\times10^{11} + 1.22\times10^{11}])/2}{(3.49\times10^{11} + 1.22\times10^{11})/70} = 31.2$$

$$F(2,70) = \frac{(8.91\times10^{11} - 4.71\times10^{11})/2}{4.71\times10^{11}/70} = 31.2$$

$$F(2,60)_{crit,\,0.1\%} = 7.8$$

نجد أثناء اختبار Chow انخفاض RSS عند انقسام العينة، و مع المتغيرات الوهمية ينخفض RSS بإضافة التقاطع وميل المتغيرات الوهمية . سوف تكون قيمةRSS بعد التغيير نفسها لأن البواقي نفسه . هذا يعني أيضا الجزء الأول من المقام لـ F الإحصائية سيكون نفسه. تحسن التكلفة في التوفيق أيضاً نفسه، عند إدخال معلمتين إضافيتين للتقدير. و بطريقة أخرى عدد درجات الحرية سارية المفعول سيكون 70 عندما يكون عدد المشاهدات 74 و عدد المعلمات 4 في التقدير. وهكذا كل مركبات علاقة F الإحصائية تكون نفسها و ومخرجاتها للاختبار يجب أن يكون نفسه. في هذه الحالة تكون فرضية العدم مطابقة لمعادلات التكلفة لنوعي المدارس؛ سيكون الرفض عند مستوى دلالة 0.1% . إذاً ما هي الأفضلية في تطابق الاثنين عند ذلك؟

اختبار Chow سريع، فعندما يتم تنفيذ ثلاثة انحدارات، و نحسب الاختبار الإحصائي، ولكن لا تخبرنا كيف اختلفت المعادلات عند بنائها.

المتغير الوهمي يقترب ليتضمن تحضيراً وإعداداً؛ لأنه يحدد المتغير الوهمي للتقاطع، و لكل ميل من المعاملات. على أي حال تكون المعلومات أفضل عندما نستطيع إنجاز اختبار t على معاملات المتغيرات الوهمية منفردة، وتجد خارجاً اختلاف المعادلات إذا عملتها.

8-12- تطبيقات المتغيرات الوهمية في البيانات الطولية أي السلاسل:

إذا كانت الفترة الزمنية t تحتوي على فترات مختلفة، كأن تكون فصول السنة أو فترات الحصار أو الحرب والسلام؛ فإنه يمكن استخدام المتغيرات الوهمية لإيجاد نموذج رياضي لكل فترة على حدة .

وهذا النوع من الدراسة يستخدم بشكل واضح من أجل وضع النماذج القياسية لدراسة السلاسل الزمنية، وهي مرونة الدراسات والتطبيقات في الاقتصاد القياسي .

8-11- المتغير التابع الوهمي:

هناك كثير من الحالات يكون فيها المتغير التابع متغيراً وهمياً، وله قيمتان فمن الممكن في هذه الحالة استخدام المتغير الوهمي .

1- مثلاً كأن يكون هناك دراسة للعلاقة بين العمر (X) والتـامين علـى الحياة Y . فإن المتغير التابع هنا هو التأمين

$$y = \begin{cases} 1 & \text{المؤمن على حياته} \\ 0 & \text{خلاف ذلك} \end{cases}$$

2- أو كأن ندرس العلاقة بين سرعة صواريخ أرض جـو (X) علـى إصابة الهدف (الطيارة)

$$y = \begin{cases} 1 & \text{اصابة الهدف} \\ 0 & \text{عدم الاصابة} \end{cases}$$

3 – دراسة تأثير الدخل التصرفي (X) على صفة احتلال سيارة (Y)

$$y = \begin{cases} 1 & \text{يملك سيارة} \\ 0 & \text{لا يملك} \end{cases}$$

4- دراسة الحالة الاجتماعية وتأثير الدخل التصرفي على صفة الحالـة الاجتماعية .

إن المشكلة الرئيسية في تحليل الانحدار باستخدام المتغيرات الوهمية هنا هو كون تباين الخطأ (لا يتوزع توزعاً طبيعياً).

نجد أن Y في هذه الحالة يتوزع حسب برنورلي؛ لذا فإن احتمال = Y) (1 هو P_i أي: $P(Y = 1) = P_i$

واحتمال أن تأخذ و Y = 0 ، $P(Y = 0) = 1 - P_i$

لذلك فإن تباين الخطاء هو $Var(e_i) = P_i q_i$

$$Var(e_i / X_i) = Var(Y_i / X_i)$$

$$= P_i(1 - P_i)$$

فإذا عوضنا عن P_i بقيمتها : $(B_0 + B_1 X_i)$ نجد أن :

$$Var(e_i / X_i) = (B_0 + B_1 X_i)(1 - B_0 - B_1 X_i)$$

لذا فإن تباين الخطأ غير متجانس، ويجب استخدام المربعات الصغرى الموزونة، فيمكن اختيار W_i كوزن فإن نأخذ معكوس التباين أي :

$$W_i = \frac{1}{Var(y_i / x_i)} = \frac{1}{P_i(1 - P_i)} =$$

$$= \frac{1}{(B_0 + B_1 x_i)(1 - B_0 - B_1 x_i)}$$

يمكننا تقدير B_1, B_0 بطريقة المربعات الصغرى الاعتيادية .

$$\hat{W}_i = \frac{1}{\hat{y}_i(1 - \hat{y}_i)}$$

وسوف نستعرض حالتين

1) إذا كان النموذج خطياً

أراد أحد الباحثين دراسة العلاقة بين الدخل التصرفي (X_1) وحالة امتلاك المسكن لعينة من 20 موظفاً، وكانت النتائج على النحو الآتي:

$\hat{W}_i$	$1-\hat{y}_i$	$\hat{y}_i$	Y_i	امتلاك البيت	الدخل	
6.34217	0.8035	0.1962	0	لا يملك	2900	1
15.78045	0.0681	0.9305	1	يملك	7700	2
5.23390	0.7422	0.2574	0	لا	3300	3
4.06082	0.5584	0.4410	1	يملك	4500	4
4.43825	0.3439	0.6552	0	لا	5900	5
15.78045	0.0681	0.9305	0	لا	7700	6
4.49934	0.6656	0.3339	1	يملك	3800	7
4.73480	0.6963	0.3033	0	لا	3600	8
4.02339	0.4665	0.5328	1	يملك	5100	9
11.24680	0.0988	0.8999	1	يملك	7500	10

4.78817	0.2979	0.7011	1	يملك	6200	11
5.04430	0.7269	0.2727	0	لا	3400	12
5.31253	0.252	0.7470	1	يملك	6500	13
6.34217	0.8035	0.1962	0	لا	2900	14
4.32030	0.635	0.3645	1	يملك	4000	15
46.13236	0.0222	0.9764	1	يملك	8000	16
8.46464	0.1371	0.8617	1	يملك	7250	17
5.70394	0.7729	0.2268	0	لا	3100	18
5.23390	0.7422	0.2574	0	لا	3300	19
13.08516	0.0835	0.9152	1	يملك	7600	20
180.56788						

WYX	WY	WX2	XW	X2	
3609.210	1.245	53337687.448	18392.306	8410000	1
113069.016	14.684	935622910.688	121509.469	59290000	2
4446.241	1.347	56997202.125	17271.879	10890000	3
8058.739	1.791	82231649.027	18273.700	20250000	4
17156.150	2.908	154495401.448	26185.661	34810000	5
113069.016	14.684	935622910.688	121509.469	59290000	6
5709.135	1.502	64970494.994	17097.499	14440000	7
5170.185	1.436	61363046.267	17045.291	12960000	8
10932.476	2.144	104648450.691	20519.304	26010000	9
75910.931	10.121	632632738.924	84351.032	56250000	10
20812.353	3.357	184057344.859	29686.669	38440000	11
4677.397	1.376	58312135.728	17150.628	11560000	12
25793.651	3.968	224454406.185	34531.447	42250000	13
3609.210	1.245	53337687.448	18392.306	8410000	14
6299.213	1.575	69124833.109	17281.208	16000000	15
360360.360	45.045	2952471043.497	369058.880	64000000	16
52881.109	7.294	444922686.325	61368.646	52562500	17
4010.868	1.294	54814859.393	17682.213	9610000	18
4446.241	1.347	56997202.125	17271.879	10890000	19
91017.964	11.976	755799027.481	99447.240	57760000	20
931039.465	130.339	7936213718.448	1144026.727	614082500	

باستخدام طريقة المربعات الصغرى الاعتيادية

$$\hat{Y} = -0.24741 + 0.000153\,X$$

وفي هذا المثال يمكننا تقدير $\hat{y}$ لجميع الموظفين فمثلاً $\hat{y}_1$ للموظف الأول هي :

$$\hat{y}_1 = -0.24741 + 0.000153\,(2900) = 0.1962$$

ومن ثم يمكننا إيجاد الوزن W_1 فإن النسبة للموظف الأول

$$\hat{W}_1 = \frac{1}{\hat{y}\,(1-\hat{y})} = \frac{1}{0.1962\,(1-0.1962)} = 6.34216$$

وهكذا لبقية الموظفين

$$\begin{bmatrix} \sum W_i & \sum W_i x_i \\ \sum W_i x_i & \sum W_i x_i^2 \end{bmatrix} \begin{bmatrix} B_0 \\ B_1 \end{bmatrix} = \begin{bmatrix} \sum W_i y_i \\ \sum W_i x_i y_i \end{bmatrix}$$

$$\begin{bmatrix} 180.568 & 1144026.727 \\ 1144026.727 & 7936213718.448 \end{bmatrix} * \begin{bmatrix} B_0 \\ B_1 \end{bmatrix} = \begin{bmatrix} 130.339 \\ 931039.465 \end{bmatrix}$$

$$\begin{bmatrix} \hat{B}_0 \\ \hat{B}_1 \end{bmatrix} = \begin{bmatrix} -7.447 \\ 0.000153 \end{bmatrix}$$

أي: أن علاقة الانحدار

$$\hat{y} = -7.447 + 0.000153\,x$$

وهي تختلف عن العلاقة السابقة التي حصلنا عليها بتطبيق المربعات الصغرى

تباين $\hat{B}_1$ هو : $S^2_{\hat{B}_1} = \dfrac{1}{S_x} = 0.0000007$

الانحراف المعياري لـ $\hat{B}_1$ $S_{\hat{B}_1} = \sqrt{S^2_{\hat{B}_1}} = 0.000027506$

أما تباين $\hat{B}_0$ هو

$$S^2_{\hat{B}_0} = \frac{1}{\sum W_1} + \bar{x}^2 S^2_{\hat{B}_1}$$

$$= 0.035826$$

الانحراف المعياري لـ $\hat{B}_0$

$$S_{\hat{B}_0} = \sqrt{S^2_{\hat{B}_0}} = 0.189277573$$

	قيمة المعامل	$\hat{S}_{B_0}$	قيم t
$\hat{B}_0$	- 7.819258	0.189277573	- 41.31
$\hat{B}_1$	0.001341626	0.000027506	48.78

أي: أن كلاً من $\hat{B}_0$ و $\hat{B}_1$ تختلف معنوياً عن الصفر .

$$\hat{B}_0 \mp S_{\hat{B}_0}(\alpha)$$

$$\hat{B}_1 \mp S_{\hat{B}_1}(\alpha)$$

2- الحالة الثانية : على افتراض أن النموذج غير خطـــي Nonlinear Model

في بعض الحالات عند استخدامنا متغيراً وهمياً للمتغير التابع، نـــرى أن العلاقة بين X و Y غير خطية، وغالباً ما تأخذ دالة الاستجابة الشكل S أي أن الدالة تأخذ الشكل اللوجستي Logistic Function والنموذج له الصـــيغة التالية:

$$E(y/x) = \hat{y} = \frac{\exp(B_0 + B_1 x_1)}{1 + \exp(B_0 + B_1 x_1)}$$

إن ميزة هذا النوع من التوابع هو سهولة تحويلها إلى خطية .

الوسط الحسابي لهذه الدالة هو $E(y/x) = P \Leftarrow \hat{y} = P$

وحيث Y هو متغير وهمي ،وبما أن P هي قيمة احتمالية؛ لذا يمكن جعل الدالة اللوجستية خطية باستخدام التحويل التالي :

$$\hat{P}^* = \ln\left[\frac{E(y/x)}{1 - E(y/x)}\right] = \ln\left(\frac{\hat{y}}{1-\hat{y}}\right) = \frac{P}{1-P}$$

$$\hat{P}^* = \ln\left[\frac{E(y/x)}{1-E(y/x)}\right] = \ln\left(\frac{\hat{y}}{1-\hat{y}}\right) = \frac{P}{1-P}$$

إن هذا التحويل يسمى التحويل اللوجستي للاحتمال P لذا فإن :

$$P^* = B_0 + B_1 X_1$$

ولتوفيق الدالة اللوجستية نفرض بأن هناك تكرارات لكل مستوى من مستويات، فنفرض أن مستويات المتغير المستقل هي : $B_1 , X_2 , \ldots , X_n$

إن Y هو متغير وهمي يأخذ قيمه صفر أو واحد، ويتكرر n_i مرة .

وحيث إن: r_i - هو عدد مرات ظهور الواحد عند كل مستوى

$$P_i = \frac{r_i}{n_i}$$

$$\bar{P}^* = \ln\left(\frac{\bar{P}_i}{1-\bar{P}_i}\right)$$

والآن يمكن أن توفق الدالة اللوجستية بطريقة المربعات الصغرى باستخدام التحويل على النسب المشاهدة P_i أي أن :

$$P_i^* = \ln\left(\frac{P_i}{1-P_i}\right)$$

كمتغير تابع ويلاحظ بأن تباين الخطأ في هذه الحالة غير متجانس، لذلك فإن طريقة المربعات الصغرى الاعتيادية لا تصلح، فلابد من استخدام طريقة المربعات الصغرى الموزونة لتقدير المعلمين B_1 , B_0 ، وعندئذ نختار الوزن W_i .

$$W_i = \frac{1}{VR\left(\bar{P}_i^*\right)}$$

وحيث إن :

$$VER\left(\bar{P}_i^*\right) = \frac{1}{n_i \bar{P}_i (1-P_i)}$$

75.00	66.66	36.75	26.66	4.5	$W_i x_i$
375	266.72	110.25	53.33	4.5	$W_i x_i^2$

$\sum$	10	9	8	7	6
	80	80	80	70	76
	45	42	35	30	25
	0.56	0.53	0.44	0.38	0.33
	0.44	0.47	0.56	0.62	0.67
	1.2857	1.1053	0.7778	0.6000	0.4999
	+0.2513	+0.1001	-0.2513	-0.5108	-0.6933
155.88	19.69	19.95	19.96	18.75	15.56
-84.43	+4.95	+2.0	-4.95	-9.58	-10.78
969.59	196.9	179.55	157.52	131.25	93.36
7137.7	1969	1615.95	126.15	918.75	560.16

5	4	3	2	1	الجرعة
80	120	100	60	50	n_i
20	20	15	8	5	r_i
0.35	0.17	0.15	0.67	0.10	$P_i = \frac{r_i}{n_i}$
0.75	0.83	0.85	0.33	0.90	$(1 - P_i)$
0.3333	0.2000	0.1765	2.0003	0.111	$\frac{P_i}{1 - P_i}$
-1.0987	-1.6094	-1.7344	+0.6933	-2.1973	P_i^*
15.000	16.670	12.25	13.3327	4.5	W_i
-16.48	-26.83	-2.125	+9.24	-9.89	$W_i P_i^*$
75.00	66.66	36.75	26.66	4.5	$W_i x_i$
375	266.72	110.25	53.33	4.5	$W_i x_i^2$

Σ	10	9	8	7	6
	80	80	80	70	76
	45	42	35	30	25
	0.56	0.53	0.44	0.38	0.33
	0.44	0.47	0.56	0.62	0.67
	1.2857	1.1053	0.7778	0.6000	0.4999
	+0.2513	+0.1001	-0.2513	-0.5108	-0.6933
155.88	19.69	19.95	19.96	18.75	15.56
-84.43	+4.95	+2.0	-4.95	-9.58	-10.78
969.59	196.9	179.55	157.52	131.25	93.36
7137.7	1969	1615.95	126.15	918.75	560.16

وباستخدام طريقة المربعات الصغرى الموزونة نجد أن :

$$\hat{\beta}_1 = \frac{\sum W_i \overline{P}^* - \dfrac{\left(\sum W_i \overline{P}^*\right)\left(\sum W_i x_i\right)}{\sum W_i}}{\sum W_i \overline{P}^{*2} - \dfrac{\left(\sum W_i x_i\right)^2}{\sum W_i}}$$

$$= \frac{-351.3338 - \dfrac{(-84.4306)(969.5964)}{155.8818}}{7137.7124 - \dfrac{(969.5964)^2}{155.8818}} = 0.1571$$

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X} =$$

$$\overline{Y} = \overline{P}^{*} = \frac{\sum W \overline{P}}{\sum W_i} = -0.5416$$

$$\overline{X} = \frac{\sum W_i x_i}{\sum W_i} = 6.2201$$

$$\hat{\beta}_0 = -0.5416 - (0.1571)(6.2201) = -1.5188$$

وتصبح معادلة الانحدار على النحو الآتي:

$$\hat{P}^{*} = -1.5188 + 0.1571\, x_i$$

ومن أجل الحصول على القيم المقابلة (الأصلية) نقوم بالخطوات التالية:

I- من أجل الجرعة الأولى :

$$P_1 = -1.5188 + 0.1571\,(1) = -1.3617$$

نوجد القيمة المقابلة لـــ P_1 .

$$P^{*} = \frac{e^{B_0 + B_1 x}}{1 + e^{B_0 + B_1 x}} = \frac{e^{\hat{P}^{*}}}{1 + e^{\hat{P}^{*}}}$$

$$= \frac{e^{-1.3617}}{1 + e^{-1.3617}} = 0.7960$$

أي أن $\hat{P} = 0.796$ وذلك عندما تكون $X_i = 1$

$$S_{\hat{B}_1} = \sqrt{S^2_{\hat{B}_1}} = \sqrt{\frac{1}{\sum W_i x_i^2 - \frac{(\sum W_i x_i)^2}{\sum W_i}}}$$

$$= \sqrt{\frac{1}{1106.7506}} = 0.030059$$

$$S_{\hat{B}_1} = \sqrt{S^2_{\hat{B}_1}} = \sqrt{\frac{1}{\sum W_i} + \overline{x}^2 S^2_{\hat{B}_1}}$$

$$= \sqrt{\frac{1}{155.8818} + (38.6893)(1106.7506)} = 206.9286$$

$\hat{B}_1$	قيمة $\hat{B}_1$	$S_{\hat{B}_1}$	قيمة t
$\hat{B}_0$	- 1.5188	0.030059	50.53
$\hat{B}_1$	0.1571	206.9286	< 1

ومن النتائج أعلاه يتضح بأن معامل الانحدار B_1 معنوي . لاحظ بأنه من الممكن شمول أكثر من متغير واحد للدالة اللوجستية .

فمثلاً لو كانت تحوي على متغيرين X_1 X_2 فإن الدالة اللوجستية المحولة :

$$P^* = B_0 + B_1 x_1 + B_2 x_2$$

تمارين غير محلولة

1- من خلال تحليل عينة مسح الشباب لدراسة 570 مشاهدة لدراسة العلاقة بين الدخل المتحقق بالدولار Y وعدد سنوات الدراسة S وأعلى مستوى تعليمي تم الحصول عليه(لم ينه المدرسة ، دبلوم المدرسة العالي ، درجة المزاملة بالفنون (سنتان من الكلية)، بكالوريوس في الفنون (أربع سنوات في الكلية)) ونرمز لها على التوالي كمتغيرات وهمية *EDUCDO, EDUCHSD, EDUCAA, EDUCBA* صنفت إلى أربعة فئات.

نعمل انحدار لوغاريتم Y على:

1- S فقط

2- على *EDUCDO, EDUCHSD, EDUCAA, EDUCBA*

3- على *S, EDUCDO, EDUCAA,* and *EDUCBA*

الجدول التالي يبين نتائج بالانحدار، وكذلك الخطأ المعياري بين قوسين و مجموع مربع البواقي.

	(1)	(2)	(3)
S	0.079	-	0.040
	(0.008)		(0.019)
EDUCDO	-	-0.173	-0.055
		(0.075)	(0.094)
EDUCAA	-	0.129	0.065
		(0.074)	(0.080)
EDUCBA	-	0.420	0.246
	(0.047)		(0.095)
constant	1.359	2.321	1.824
	(0.113)	(0.027)	(0.236)
R^2	0.141	0.145	0.152
RSS	132.12	131.48	130.44

أ- أعطي تفسيراً منطقياً لمعامل S في النموذج الأول . (ملاحظة: تذكر أن المتغير التابع هو عبارة عن اللوغاريتم الطبيعي للدخل المتحقق بالساعة).

ب- ناقش أي الاثنين يكون منطقياً في تفسير الثابت في النموذج الأول .

ج- أعطي تفسيراً لمعاملات المتغيرات الوهمية في النموذج الثاني.

د- ناقش أي الاثنين ممكن **لإعطاء** تفسير لثابت النموذج الثاني.

هـ - أعطي تفسير لمعامل *EDUCBA* في النموذج الثالث.

و- اختبر قوة العلاقة للمتغيرات الوهمية للنموذج الثالث ، وضح نتائج الاختبار بالتفسير.

ز- أحدهم قال في السيمنار يجب استخدام الانحدار بحذف الفئة؛ لأنه أقل من الفئات التعليمية . كيف سيعيد الباحث هذه العلاقة؟

ح- في السيمنار قال أحد الباحثين: إن المعاملات أقل بين الذكور من الإناث عندما يكون النموذج الثاني من أجل الذكور و الإناث مفصولين.

2- كيف يحدد شكل العلاقة الدالية ؟

1) دراسة العلاقة بين التأمين على الحياة Y وكل من درجة الخطورة X_1 والعمر X_2 لعينة من 20 طبيباً على سبيل المثال .

2) دراسة العلاقة بين درجة تصلب الشرايين Y ونسبته الكولسترول X_1 والوزن X_2

$$y = B_{00} + B_{10}\, x_1 + B_{01}\, x_2 + B_{11} x_1 x_2 + B_{20}\, x_1^2 + B_{02}\, x_2^2 + \ldots\ldots$$

$$y = B_{00}\ {}_{x_1} B_1\ {}_{x_2} B_2\ {}_{x_3} B_3\ e$$

فروض التحليل Assumption

1) Y متغير عشوائي Random Variable وقيمة مستقلة الواحدة عن الأخرى، وتتوزع توزيعاً طبيعياً بوسط حسابي μ_{yx} وتباين $\delta^2 = \delta^2_{y/x}$ (تباين أو تجانس تباين الخطأ) .

2) أن العلاقة بين $\hat{Y}$, X علاقة خطية $\hat{y} = \hat{B}_0 + \hat{B} x_{1i}$

3) القيمة المتوقعة لحد الخطأ، أي: وسطه يساوي الصفر $E(\mu_i) = 0$

وتباين قدرة $E(\mu_i^2) = \delta^2$

4) قيمة حد الخطأ في فترة ما غير مرتبطة أو غير متعلقة بقيمة في أي فترة أخرى .

$$E(\mu_i \mu_j) = 0 \qquad \text{for} \quad i \neq j$$

هذا يكفل أن لقيمة المتوسط للمتغير Y تعتمد فقط على X وليس على u وهذا مطلوب للحصول على تقديرات كفء لمعاملات الانحدار واختبارات غير متحدة لمعنويتها .

5) المتغير المفسر يأخذ قيماً ثابتة، والتي يمكن الحصول في العينات المتكررة ، بحيث إن المتغير المفسر يكون هو الآخر غير مرتبط بعنصر الخطاء أي $E(\mu_i \mu_j) = 0$

Ordinary least Squares (O L S)

3- اكتب معادلة نموذج الانحدار الخطي المتعدد لحالة متغيرين مستقلين أو مفسرين وحالة k متغير مستقل أو مفسر ب) اذكر فروض النموذج الخطي للانحدار المتعدد

أ) 2 متغيران مستقلان $Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + \mu_i$

ب) k متغير مستقل

$y_i = b_0 + \sum b_i x_i + \mu_i$

عرف ما يلي : أ) $\delta \mu_u^2$, S^2 ب) $Vor(\hat{b}_1)$, $Vor(\hat{b}_2)$

جـ) $S_{\hat{b}_1} S_{\hat{b}_2}^2$ د) $S_{\hat{b}_1} S_{\hat{b}_2}$ هـ) ما هي أهمية b_0

أ) $\delta \mu_u^2$ هو تباين حد الخطأ في العلاقة الخطية بين X_{1i} و X_{2i} و Y_i ولكن

$S^2 = \hat{\delta}_u^2$ وهي تباين البواقي، وهي تقدير غير متحد للتباين غير المعلوم δ_u^2 .

k - عدد المعالم المقدرة . درجات الحرية $n - k = d f$

$$Vor(\hat{b}_1) = \delta_u^2 \frac{\sum x_2^2}{\sum x_1^2 \sum x_2^2 - (\sum x_1 x_2)^2}$$

ب)

$$Vor(\hat{b}_2) = \delta_u^2 \frac{\sum x_1^2}{\sum x_1^2 \sum x_2^2 - (\sum x_1 x_2)^2}$$

أن تباين $\hat{b}_1$, $\hat{b}_2$ (أو تقديراتها) مطلوبة لاختبار الفروض وتكوين فترات الثقة لكل من b_1 , b_2 .

$$S_{\hat{b}_1}^2 = S^2 \frac{\sum x_2^2}{\sum x_1^2 \sum x_2^2 - (\sum x_1 x_2)^2} = \frac{\sum e_i^2}{n-k} \cdot \frac{\sum x_2^2}{\sum x_1^2 \sum x_2^2 - (\sum x_1 x_2)^2}$$

$$S_{\hat{b}_2}^2 = S^2 \cdot \frac{\sum x_2^2}{\sum x_1^2 \sum x_2^2 - (\sum x_1 x_2)^2} = \frac{\sum e_i^2}{n-k} \cdot \frac{\sum x_1^2}{\sum x_1^2 \sum x_2^2 - (\sum x_1 x_2)^2}$$

$S_{\hat{b}_1}^2$, $S_{\hat{b}_2}^2$ هما على الترتيب تقديرات غير متحيزين لتباين b_1 وتباين b_2 غير معلومين حيث إن δ_u^2 غير معلومة .

د) $S_{\hat{b}_1}$, $S_{\hat{b}_2}$ هو الانحراف المعياري لكل من $\hat{b}_1$, $\hat{b}_2$ ويسميان بالأخطار المعيارية .

هـ) أحياناً يكون لها أهمية كبيرة في الاستخدام .

$$Vor\left(\hat{b}_0\right)=\delta_u^2 \frac{\sum x_1^2 \sum x_2^2 - \left(\sum x_1 x_2\right)^2}{\begin{array}{l} n\left[\sum x_1^2 x_2^2 - \left(\sum x_1 x_2\right)^2\right] - \\ - \sum x_1\left(\sum x_1 \sum x_2^2 - \sum x_2 \sum x_1 x_2\right) + \\ + \sum x_2\left(\sum x_1 \sum x_1 x_2 - \sum x_2 \sum x_1^2\right) \end{array}}$$

4 - ناقش ما يلي:

أ- مفهوم المتغير الوهمي، ودوره في تقدير النماذج القياسية.

ب- مفهوم جداول تحليل التباين المشترك.

د- مصيدة المتغير الوهمي Dammy Variable Trap

و- التعديل الموسمي باستخدام المتغير الوهمي.

هـ- اختبار Chow

5- من نموذج دالة الاستهلاك التالية :

$$C_1 = \beta_o + \beta_1 Y_t + \beta_2 W_t + \beta_3 W_t Y_t + U_t$$

حيث (C) تشير إلى الاستهلاك ، (Y) تشير إلى الدخل ، (W = 1) تشير إلى زمن السلم ، (u) تشير إلى الخطأ العشوائي .

أ- قارن بين دالة الاستهلاك في زمن الحرب، ودالة الاستهلاك في زمن السلم .

ب- وضح أن تقديرات المربعات الصغرى لمعلمات الانحدار التي تم الحصول عليها تعطي نموذجين منفصلين من (C_t) على (Y_t) ، واحدة مقدرة من بيانات فترة الحرب، والثانية مقدرة من بيانات فترة السلم .

6- نفترض أن النموذج الذي يقدر رواتب أساتذة الجامعة يأخذ الصـــورة الآتية:

$$Y_t = \partial_0 + \partial_1 D_{1i} + \partial_2 D_{2i} + \partial_3 (D_{1i} + D_{2i}) + \beta X_i + U_i$$

حيث إن:

(Y_i) تشير إلى الراتب السنوي لمدرسي الجامعة .

(X_i) تشير إلى مدة الخدمة الجامعية

$(1 = D_1)$ إذا كان المدرس أنثى و $(0 = D_1)$ إذا كان ذكراً .

$(1 = D_2)$ إذا كان المدرس متزوجاً و $(0 = D_2)$ إذا كان عازب .

$(D_2 = D_1)$ $(D_2 = D_1)$ يشير هذا الحد إلى التأثير المتداخل (interaction effect)

المطلوب:

1- ماذا يعني الحد $(D_2 = D_1)$ وما تفسيرك الاقتصادي للمصطلح interaction effect ؟

2- ماذا تعني المعلمة (∂_3) ؟

3- أوجد $E(Y_i / D_1 = 1, D_2 = 1, X_i)$ واشرح معنى ذلك .

7- الجدول التالي يوضح الاستثمار المحلي (Y) ، والناتج القومي الإجمالي (X) ببلايين الدنانير في الاقتصاد العراقي خلال الفترة 1979-1994 قدر أثر المتغير الوهمي (الحرب والسلم) في إجمالي الاستثمار المحلي ؟

Year	Y_i	X_i	(الحرب والسلم)
1979	9.3	90.8	0
1980	13.1	100.0	0
1981	17.9	124.9	0
1982	9.9	158.3	1
1983	5.8	192.0	1
1984	7.2	210.5	1
1985	10.6	212.3	1
1986	30.7	209.3	1
1987	34.0	232.8	0
1988	45.9	259.1	0
1989	35.3	258.0	0
1990	53.8	286.2	0
1991	59.2	330.2	0
1992	52.1	347.2	0
1993	53.3	366.1	0
1994	52.7	366.3	0

8 - الجدول الآتي يعطي الإنفاق الاستهلاكي (C) ، والدخل القابل للتصرف (Yd) ، ونوع جنس (SeX) رب العائلة ، لعينة حجمها (12) . المطلوب :

أ- إيجاد انحدار (C) على $\left(Y^d\right)$ ، (ب) اختبار حد المقطع للعوائل التي يكون رب الأسرة فيها أنثى أو ذكراً ، (ج) اختبار اختلاف الميل (MPC) للعوائل التي يكون رب العائلة فيها أنثى أو ذكر

(د) اختبار اختلاف حد المقطع والميـل . (هـــ) مــا هــي برأيــك أفضــل نتيجة؟

المتغير الوهمي S	الدخل القابل للتصرف (Y_d)	الاستهلاك C	عدد العوائل
M(0)	22.55	18.54	1
M(0)	14.04	11.35	2
F(1)	13.04	12.13	3
M(0)	17.50	15.21	4
F(1)	9.43	8.68	5
M(0)	20.64	16.76	6
M(0)	16.47	13.48	7
F(1)	10.72	9.68	8
M(0)	22.35	17.84	9
F(1)	12.20	11.18	10
F(1)	16.81	14.32	11
M(0)	23.00	19.86	12

حيث يرمز للجنس كما يلي :

ذكر Male(M)=1

أنثى Female(F)=0

9- بافتراض توفر البيانات أدناه عــن اســتهلاك الحنطــة لمــدة ســت سنوات، حيث كانت السنوات الـثلاث الأولــى منهــا ســنوات حــرب فــي السنوات الثلاث الأخيرة هي سنوات سلم.

المطلوب: تقدير معلمات النموذج الخطي البسيط للاستهلاك؟

التطبيق الرابع:

الجــدول الآتــي يعطــي إجمــالي الاســتثمار المحلــي Y_t والنــاتج القومي الإجمالي X_t بملايين الليرات وابتداء من سنة 1939-1954

المطلوب إجراء انحدار Y/X

اختبر التقديرات مستوى 0.05 وأوضح فيما إذا كان استعمال D=1 لسنوات الحرب و D=0 للسنوات الأخرى معنوياً إحصائياً أم لا؟

السنوات	Y_t	X_t	D
1939	9.3	90.8	1
1940	13.1	100.0	1
1941	17.9	124.9	1
1942	9.9	158.3	1
1943	5.8	192.0	1
1944	7.2	210.5	1
1945	10.6	212.3	0
1946	30.7	209.3	0
1947	34.0	232.8	0
1948	45.9	259.1	0
1949	35.3	258.0	0
1950	53.8	296.0	0
1951	59.2	330.2	0
1952	52.1	247.2	0
1953	53.3	347.2	0
1954	52.7	366.3	0

10- من خلال تحليل عينة مسح الشباب لدراسة 570 مشاهدة لدراسة العلاقة بين الدخل المتحقق بالدولار Y وعدد سنوات الدراسة S وأعلى مستوى تعليمي تم الحصول عليه(لم ينهِ المدرسة ، دبلوم المدرسة العالي ، درجة المزاملة بالفنون (سنتان من الكلية)، بكالوريوس في الفنون (أربع سنوات في الكلية) ونرمز لها على التوالي كمتغيرات وهمية *EDUCDO, EDUCHSD, EDUCAA, EDUCBA* صنفت إلى أربعة فئات.

نعمل انحدار لوغاريتم Y على:

1- S فقط

2- على *EDUCDO, EDUCHSD, EDUCAA, EDUCBA*

3- على S, $EDUCDO$, $EDUCAA$, and $EDUCBA$

الجدول التالي يبين نتائج بالانحدار، وكذلك الخطأ المعياري بين قوسين ومجموع مربع البواقي.

	(1)	(2)	(3)
S	0.079	-	0.040
	(0.008)		(0.019)
$EDUCDO$	-	-0.173	-0.055
		(0.075)	(0.094)
$EDUCAA$	-	0.129	0.065
		(0.074)	(0.080)
$EDUCBA$	-	0.420	0.246
	(0.047)		(0.095)
constant	1.359	2.321	1.824
	(0.113)	(0.027)	(0.236)
R^2	0.141	0.145	0.152
RSS	132.12	131.48	130.44

أ- أعطِ تفسيراً منطقياً لمعامل S في النموذج الأول . (ملاحظة: تذكر أن المتغير التابع هو عبارة عن اللوغاريتم الطبيعي للدخل المتحقق بالساعة) .

ب- ناقش أي الاثنين يكون منطقياً في تفسير الثابت في النموذج الأول .

متوسط الكسب، المتحقق يساوي إلى $\$3.89 = 1.359$ لأي فرد ليس عنده سنوات دراسية .

ج- أعطِ تفسيراً لمعاملات المتغيرات الوهمية في النموذج الثاني.

د- ناقش أي الاثنين ممكن لإعطاء تفسير لثابت النموذج الثاني.

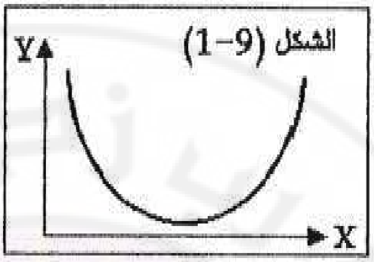
هـ - أعطِ تفسيراً لمعامل $EDUCBA$ في النموذج الثالث.

و- اختبر قوة العلاقة للمتغيرات الوهمية للنموذج الثالث ، و وضح نتائج الاختبار بالتفسير.

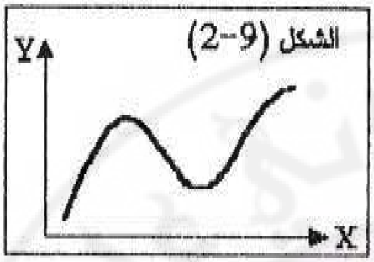
ح- في السيمنار قال أحد الباحثين: أن المعاملات أقل بين الذكور من الإناث؛ عندما يكون النموذج الثاني من أجل الذكور و الإناث مفصولين.

3 -الدالة اللوغاريتمية (Logarithmic Function): الصيغة الرياضية لها:

$$Y = b_o + b_1 \; Log\;(X)$$

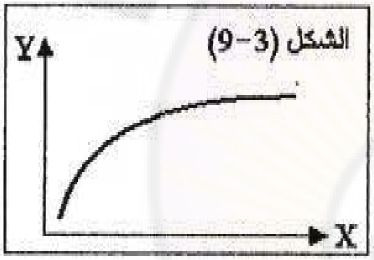


الشكل (1-9)



الشكل (2-9)

أما الشكل البياني لهذه الدالة فهو مبين على الشكل (3-9).

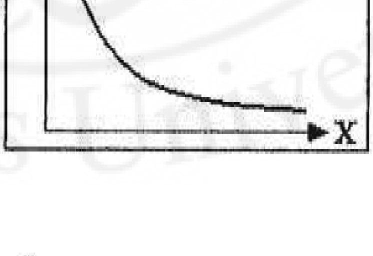


الشكل (9-3)

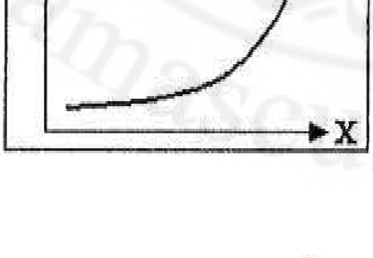
4 - الدالة الأسية (Exponential Function): الصيغة الرياضية لها:

$$Y = a^X$$

والمنحني البياني موضح في الشكل (4-9).



الشكل (9-5)



الشكل (4-9)

5 – الصيغة العكسية (The Inverse Formula): صيغتها الرياضية:

$$Y = b_0 + b_1\left(\frac{1}{X}\right)$$

المنحني البياني لهذه الدالة على شكل قوس، انظر الشكل (5-9).

إذا كيف نميز العلاقات الخطية من اللاخطية؟ لنأخذ النموذج التالي:

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + u$$

إن النموذج المعروض هو نموذج خطي لسببين. لأن القسم الأيمن يتألف من متغيرات خطية، ولأن المتغيرات المتضمنة محددة بقيمها بالذات. وهو أيضاً خطي لأن المعلمات فيه خطية، وإن اختلفت في العوامل المضروبة بها في كل حد.

أما النموذج التالي: $Y = \beta_1 + \beta_2 X_2^2 + \beta_3 \sqrt{X_3} + \beta_4 \log X_4 + u$

فهو لا خطي، وذلك لأنه خطي بالمعلمات β ولكنه غير خطي المتغيرات.

إن النماذج في هذه المعادلات لا يمكن تحويلها إلى خطية بإجراء تحويلات على المتغيرات كما هو مبين: $Z_2 = X_2^2, \quad Z_3 = \sqrt{X_3}, \quad Z_4 = \log X_4$ وبذلك نكون قد حصلنا على علاقة خطية لكل من المتغيرات والمعلمات على النحو الآتي:

$$Y = \beta_1 + \beta_2 Z_2 + \beta_3 Z_3 + \beta_4 Z_4 + u$$

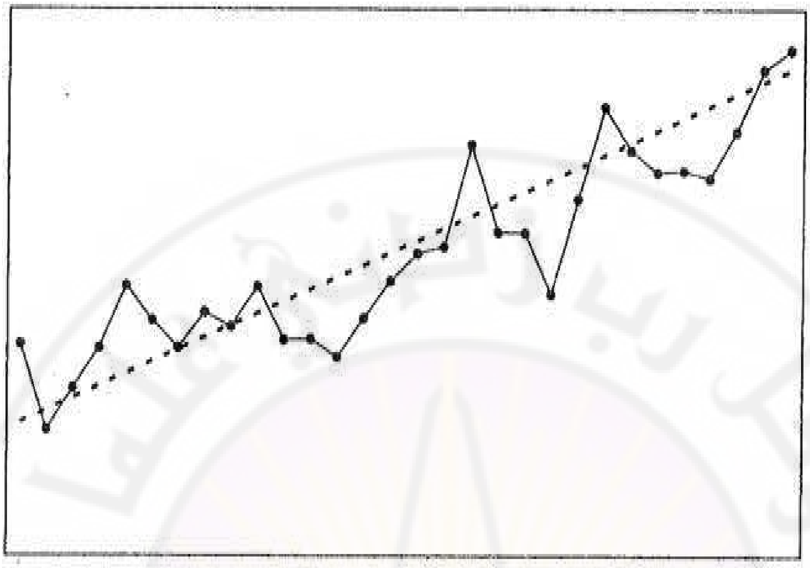
ويعتبر النموذج التالي : $Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_2\beta_3 X_4 + u$ نموذج لا خطي المعلمات؛ لأن معامل X_4 ناتج عن معامل X_2 و X_3 . كما نرى بعض النماذج التي

تكون لا خطية في المعلمات β ولا نستطيع تحويلها إلى خطيـــة ، وكمثـــال على ذلك لندرس العلاقة بين استهلاك الأسرة ودخلها من خـــلال النمـــوذج العكسي $\hat{Y} = b_0 + b_1 X$ ولنفرض أن لدينا البيانات التالية:

household الأسرة	استهلاك الموز Y	الـــــــــدخل income(10,000) X	Z=1/X
1	1.71	1	1.00
2	6.88	2	0.50
3	8.25	3	0.33
4	9.52	4	0.25
5	9.81	5	0.20
6	11.43	6	0.17
7	11.09	7	0.14
8	10.87	8	0.13
9	12.15	9	0.11
10	10.94	10	0.10

```
. reg Y X
  Source |       SS        df       MS      Number of obs =      10
---------+------------------------------          F(  1,       8) =
17.44
   Model |  58.8774834     1  58.8774834    Prob > F       =  0.0031
Residual |   27.003764     8   3.3754705    R-squared      =  0.6856
---------+------------------------------    Adj R-squared  =  0.6463
   Total |  85.8812475     9  9.54236083    Root MSE       =  1.8372
---------------------------------------------------------------------
       Y |   Coef.    Std. Err.     t     P>|t| [95% Conf. Interval]
---------+-----------------------------------------------------------
       X |.8447878    .2022741   4.176  0.003 .378343     1.311233
   _cons |4.618667    1.255078   3.680  0.006 1.724453    7.512881
---------------------------------------------------------------------
```

نرسم المخطط البياني لشكل الانتشار. سوف نتوقع من خـــلال شـــكل الانتشار المشاهد أدناه، أن معامل المتغيرX أكثر معنوية، و أن جودة القياس لـــ R^2 جيدة تماماً، ولكن النتائج كانت غير ذلك.



الشكل (6-9)

ومن خلال رسم خط الانحدار على شكل الانتشار، وتوفيق القيم والبواقي كما هي مبينة على الشكل أعلاه. فإذا كان النموذج صحيحا فالبواقي يجب أن تكون عشوائية. ولكن لدينا هنا قيمة واحدة سالبة من البواقي يتبعها 6 قيم موجبة وثلاث قيم سالبة.

ونعتقد أنه لو كان Y ينتمي إلى $\frac{1}{X}$ سيكون النموذج أكثر منطقية.

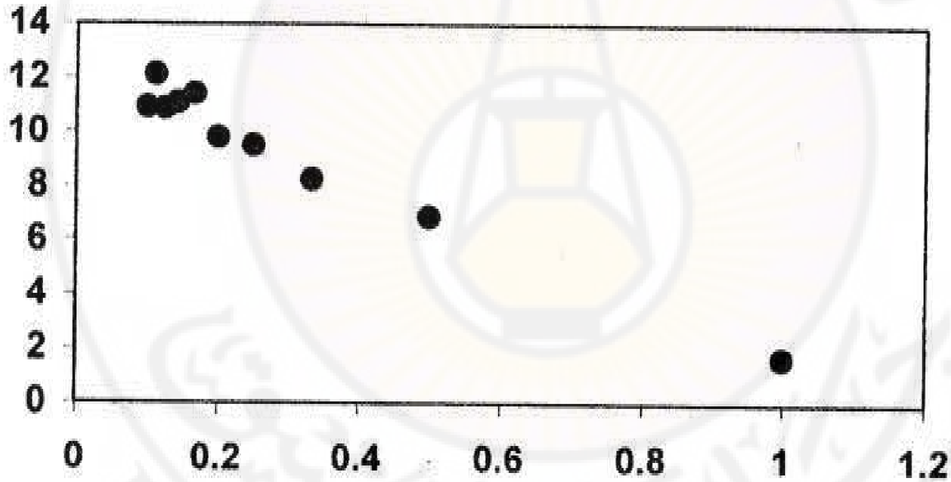
مازال Y ينمو بقيمة X إذا كانت $b_2 < 0$ ، ولكن معدل النمو يهبط، و وكذلك سيكون حد أعلى لـ b_1 . بعد كل ذلك تستطيع استهلاك الكثير من الموز. وسيكون شكل العلاقة: $Y = \beta_1 + \frac{\beta_2}{X} + u$

هذا نموذج غير خطي، ولكن نستطيع تحويله إلى خطي ، بتعريف متغير جديد $Z = \dfrac{1}{X}$ ويكون بدلاً من المتغير X .فنحصل على

$$Y = \beta_1 + \beta_2 Z + u$$

إن الخطوة التالية هي حساب المتغير Z من بيانات المتغير X . كل تطبيقات الانحدار تبين صحة المتغيرات الجديدة الحاصلة، انظر الجدول رقم()أعلاه.

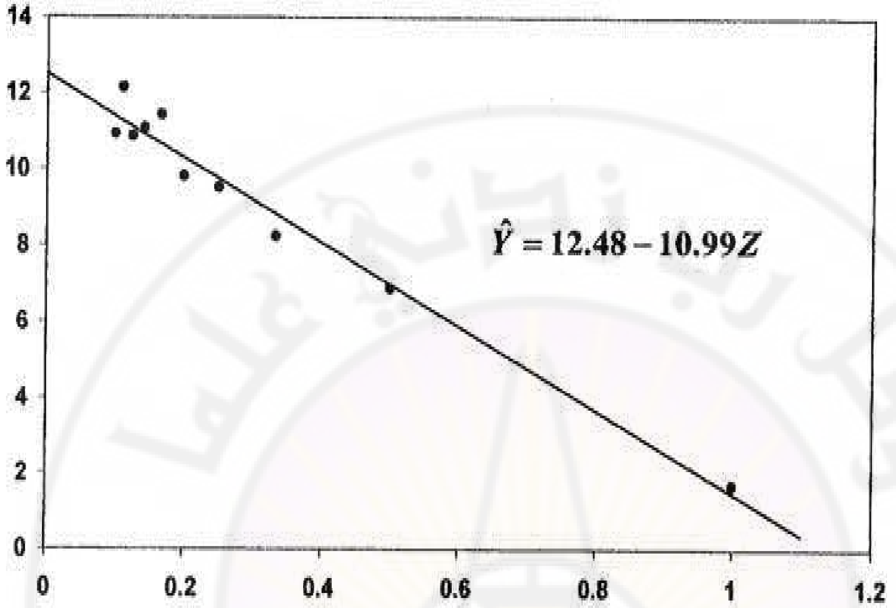
نرسم شكل الانتشار لعلاقة المتغير Y بلمتغير Z ، فنجد:



الشكل (7-9)

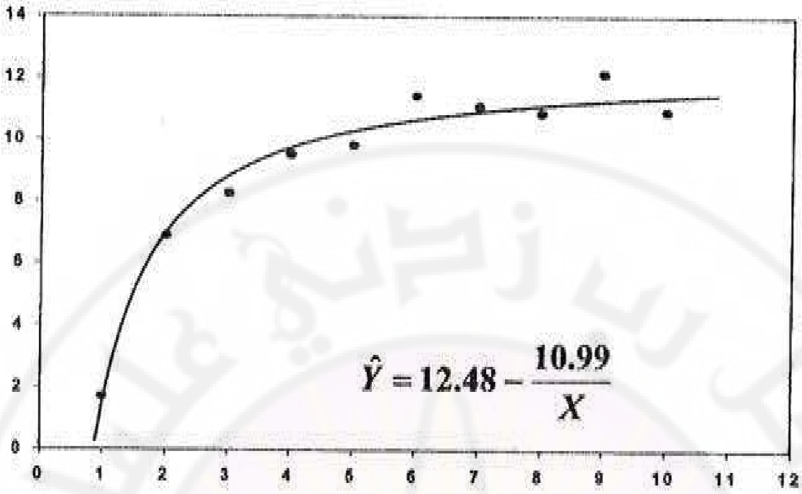
```
. g Z=1/X
. reg Y Z
  Source |       SS         df       MS              Number of obs =
10
---------+------------------------------           F(  1,       8) =
286.10
   Model |  83.5451508      1  83.5451508            Prob > F      =
0.0000
Residual |  2.33609666      8  .292012083            R-squared     =
0.9728
---------+------------------------------           Adj R-squared =
0.9694
   Total |  85.8812475      9  9.54236083            Root MSE      =
.54038
-----------------------------------------------------------------------
------
       Y | Coef.    Std. Err.      t       P>|t|       [95% Conf.
Interval]
---------+-------------------------------------------------------------
------
       Z |-10.98865    .6496573  -16.915    0.000     -12.48677    -
9.490543
   _cons |12.48354     .2557512  48.811     0.000            11.89378
13.07331
-----------------------------------------------------------------------
------
```

لقد حسبنا الجديد Z – (g هذا الحرف اختصاراً لحساب) – ويبين الجدول أعلاه مخرجات الانحدار لهذا النموذج . ومنه نستنتج معادلة الانحدار : $\hat{Y} = 12.48 - 10.99Z$



الشكل (8-9)

نبدل قيمة المتغير X بـ Z ونرسم المنحني الأصلي، ونحصل على توفيق أفضل.



الشكل (9-9)

9-2- المرونة ومضاعفة النماذج اللوغاريتمية

ELASTICITIES AND DOUBLE-LOGARITHMIC MODELS

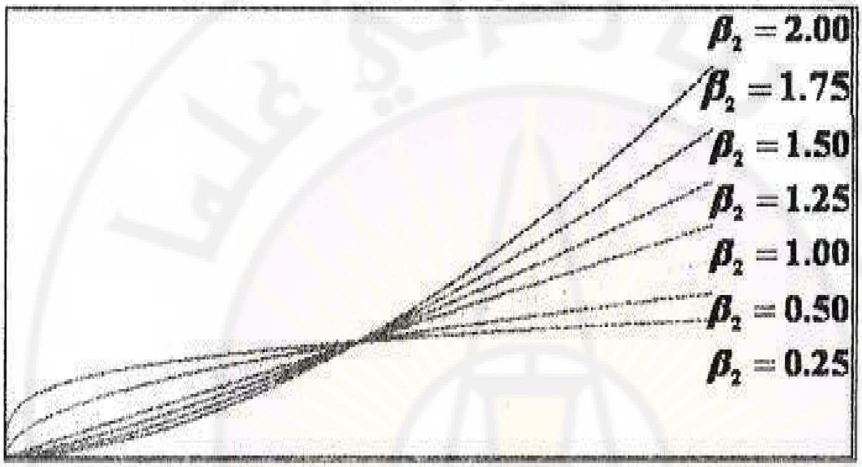
في هذا العرض سنعرف المرونة، وسنعرض كيف نستطيع توفيق واحد من النماذج غير الخطية بثوابت المرونة. أولاً: نحسب المرونة.

المرونة **Elasticity** بين المتغير X و المتغير Y هي عبارة عن التغير النسبي في المتغير Y على التغيير النسبي في المتغير X .

$$\mathbf{E} = \frac{dY/Y}{dX/X} = \frac{dY/dX}{Y/X}$$

إعادة كتابة علاقة المرونة، ونحصل على العلاقة التفسيرية المبينة أعلاه.

وهي قيمة ثابتة لجميع قيم منحني النموذج المذكور . وسنعرض طريقة لتوضيح دور β_1 في المعادلة بالرسم البياني لمدى من قيم β_1. سوف نبدأ بقيمة صغيرة جداً 0.25 . وسوف نزيد بشكل تدريجي قيمة β_1 في خطوات من 0.25 و نرى كيف يكون شكل التغير في المعادلة.



الشكل (12-9)

يكون من السهل توفيق جودة ثابت علاقة المرونة باستخدام عينة المشاهدات. ونستطيع تحويله إلى خطي بأخذ لوغاريتم الطرفين.

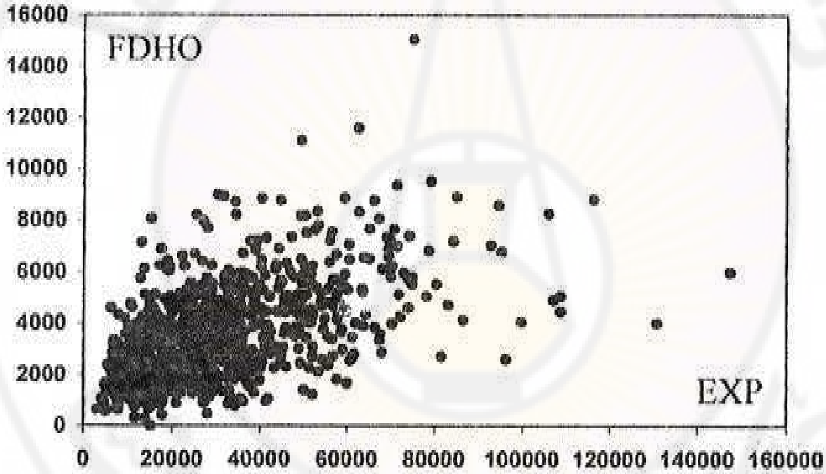
$$\log Y = \log \beta_0 X^{\beta_1} = \log \beta_0 + \log X^{\beta_1}$$
$$= \log \beta_0 + \beta_1 \log X$$

وهكذا تحصل على علاقة خطية بين Y و X كما هو محدد. إن كل تطبيقات الانحدار الجدية تسمح لنا بتوليد المتغيّرات اللوغاريتمية على النحو التالي:

$$Y' = \beta_0' + \beta_1 X'$$
$$Y' = \log Y, X' = \log X$$
$$\text{حيث : } \beta_0' = \log \beta_0 ;$$

المعامل X سوف يكون تقدير مباشر لمرونة β_1 . أما الحد الثابت فسيكون مقدراً من لوغاريتم β_0 . لنحصل على تقدير β_0 يجب حساب $EXP(B_0')$ عندما تكون β_0 تقديراً لـ β_0 : (لهذا الافتراض نستخدم اللوغاريتم الطبيعي لأن أساس اللوغاريتم e لتحويل النموذج) .

ونعرض هنا شكل انتشار للإنفاق الأسري ***FDHO*** على المواد الغذائية مع إجمالي الإنفاق الأسري السنوي، كلاهما مقاس بالدولار من عام 1995 لعينة مكونة من 859 أسرة في الولايات المتحدة (بيانات مسح نفقات المستهلك).



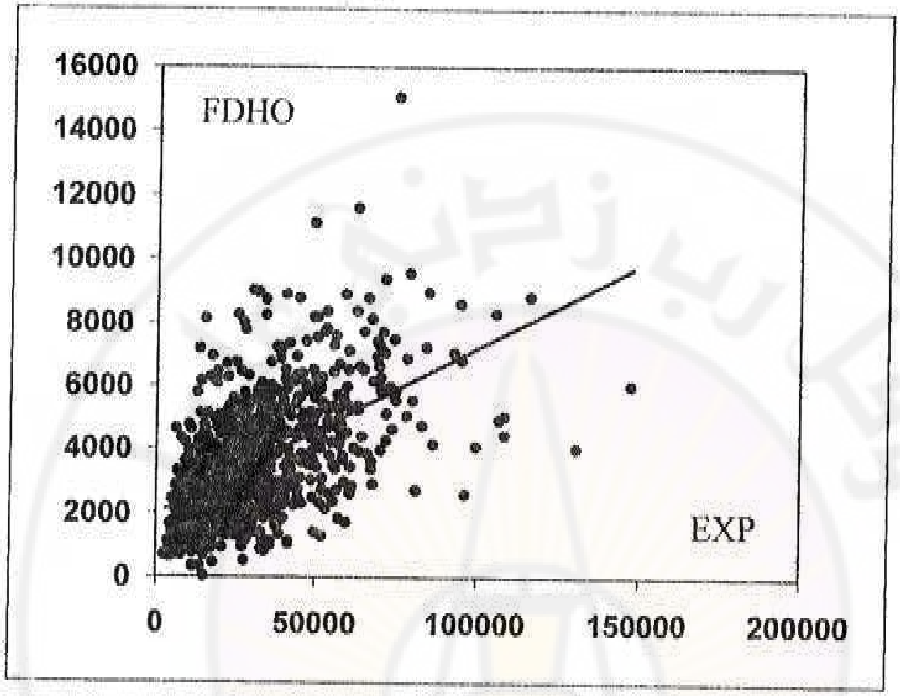
الشكل (13-9)

نبني نموذج انحدار FDHO على EXP . عادة يكون نموذج إنفاق المستهلك على إجمالي الإنفاق ، بالأحرى على الدخل، عندما نستخدم بيانات الأسرة. لا بد من الإشارة إلى أن بيانات الدخل الأسري بشكل نسبي تكون شاذة نوعا ما.

```
. REG FDHO EXP
  SOURCE |       SS          DF        MS             NUMBER OF OBS =
869
---------+------------------------------             F(  1,    867) =
381.47
   MODEL |    915843574       1   915843574           PROB > F       =
0.0000
RESIDUAL |   2.0815E+09     867  2400831.16           R-SQUARED      =
0.3055
---------+------------------------------             ADJ R-SQUARED =
0.3047
   TOTAL |   2.9974E+09     868  3453184.55           ROOT MSE       =
1549.5
------------------------------------------------------------------------
---
    FDHO |    COEF.    STD. ERR.     T      P>|T|     [95% CONF.
INTERVAL]
---------+--------------------------------------------------------------
------------
     EXP |  .0528427    .0027055  19.531    0.000      .0475325
.0581529
   _CONS |  1916.143    96.54591  19.847    0.000      1726.652
2105.634
------------------------------------------------------------------------
---
```

إن الانحدار يتضمن -- من 5 قيم هامشية- جزءاً من الـدولار مـن الإنفاق يكون على الأغلب على الطعام في البيت. هل يعقل ذلك؟ ربما قليل الاحتمال .

من المقترح أن يكون الحد الأدنى من الإنفاق على الطعام في البيت 1916.1 $ إذا كان إجمالي الإنفاق يساوي الصفر. مـن البـديهي أن هـذا مستحيلاً . ربما يكون من الجائز تفسير ذلك مقارنة مع الإنفـاق ، ولكـن سوف نحتاج لأخذ حسابات حجم الأسر و تكوينها. نبين رسم خط الانحدار على شكل الانتشار أنظر الشكل (14-9).



الشكل (14-9)

لنوفق معادلة ثابتة المرونة باستخدام نفس البيانات. فشكل الانتشار في الشكل (15-9) يبين رسم لوغاريتم FDHO مع لوغاريتم EXP .

هنا نبين نتائج انحدار LGFDHO على LGEXP ، حيث أول تعليمتين لحساب لوغاريتم المتغيرات، كما هو مبين في جدول النتائج.

9-3- النماذج نصف اللوغاريتمية SEMI-LOGARITHMIC MODELS :

نعرض في هذه الفقرة مقدمة عن النماذج نصف اللوغاريتمية، وسنبين إمكانية تطبيقها على معادلة الكسب. إن المتغير التابع خطي، ولكن المتغيرات التفسيرية مضاعفة بمعاملاتها، وهي أس للأساس e كما يلي:

$$Y = \beta_0 e^{\beta_1 X}$$

مشتق المتغير Y بالنسبة للمتغير X يساوي العلاقة البسيطة $\beta_1 Y$:

$$\frac{dY}{dX} = \beta_0 \beta_1 e^{\beta_1 X} = \beta_1 Y$$

وهكذا نجد أن المرونات تساوي إلى : $E = \frac{\frac{dY}{dX}}{\frac{Y}{X}} = \frac{\beta_1 Y}{\frac{Y}{X}} = \beta_1 X$

وهكذا نجد أن التغير النسبي في المتغير Y عند التغير بوحدة واحدة من المتغير X يساوي β_1 . أي يكون معتمداً على قيمة X .

$$E = \frac{dY/Y}{dX} = \beta_1$$

وهو القول القاطع في التفسير المنطقي فقط للتغيرات الصغيرة في X. أما عندما لا تكون التغيرات صغيرة فالتفسير يمكن أن يكون معقداً قليلاً .

إذا حسبنا التغير لسنة واحدة مدرسية، فنجد أنه لا يوجد تغير بالقيمة الحدية، و نحصل على قيمة دقيقة زيادة بمقدار 8.2%.

ومن التحليل السابق نحصل على أن:

$$EARNINGS = \beta_0 e^{0.079S}$$

$$EARNINGS' = \beta_0 e^{0.079S} = 3.9e^{0.079S}$$

$$S' = S + 1$$

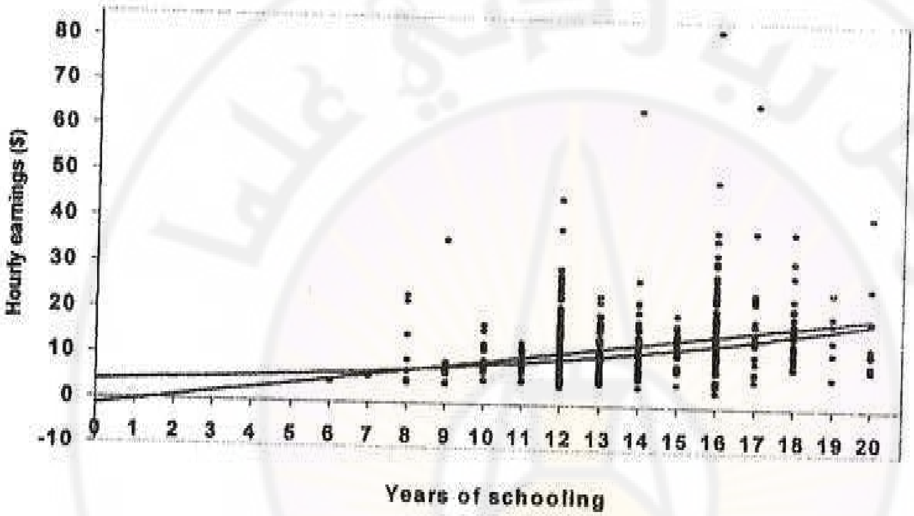
$$\begin{aligned} EARNINGS' &= \beta_0 e^{0.079S} = \beta_0 e^{0.079(S+1)} \\ &= \beta_0 e^{0.079S} e^{0.079} \\ &= EARNINGS e^{0.079} \\ &= EARNINGS(1 + 0.079 + 0.003 + ...) \end{aligned}$$

عندما β_1 تكون أقل من 0.1 يوجد قليلاً من الكسب في مخرجات العمل ذات تأثير دقيق.

إن التقاطع في الانحدار هو مقدر لـ **log** β_0، ومن ذلك نحصل على β_0 حيث إن: $e^{1.36}$ تساوي إلى 3.9 .

وبالضبط إن هذا يعني أن الشخص الذي لم يدخل المدرسة سوف يكسب $3.90 بالساعة. على أي حال يكون ذلك استقراءً محفوفاً بالمخاطر للمدى البعيد في امتلاك البيانات. ويبين الشكل (17–9) شكل انتشار لانحدار نصف اللوغاريتمي مع تحويل البيانات بالمقارنة مع الانحدار الخطي. نلاحظ عدم وجود اختلاف كبير في توفيق خط الانحدار، ولكن نصف اللوغاريتمي يكون مرضياً أكثر في العلاقتين

نجد أن التنبؤ بالعلاقة الخطية سوف يعطينا أن زيادة الدخل بمقدار 1$ في الساعة لكل زيادة سنة دراسية ، وهذا لا يصدق للمستويات العليا من التعليم. النموذج نصف اللوغاريتمي يسمح بالزيادة التدريجية في مستوى التعليم.



الشكل (17-9)

ثانياً: إن النموذج الخطي يعطي نتيجة سلبية لغير الدارسين أثناء التنبؤ ، بينما نصف اللوغاريتمي يساعد على التنبؤ بمقدار النسب للساعة الواحدة 3.9$ و يكون الكسب هو الحد الأدنى ولكن غير مقنع.

9-4- الحد العشوائي في نماذج اللاخطية:

THE DISTURBANCE TERM IN NONLINEAR MODELS

نلاحظ أنه لا احد يملك تصور حول حد الخطأ العشوائي في نماذج الانحدار اللاخطية التي تكتب كما يلي:

$$Y = \beta_1 + \frac{\beta_2}{X} + u$$

$$Z = \frac{1}{X}$$

$$Y = \beta_1 + \beta_2 Z + u$$

من أجل الحصول على نتائج مرضية للنماذج الخطية المحولة؛ يجب أن تحقق شروط غاوس-ماركوف بالإضافة إلى شروط أخرى المساعدة. لا بد من تحقق الاختبارات الاعتيادية، حول اعتيادية توزع الخطأ العشوائي طبيعياً في النموذج المحول. في هذه الحالة بالنسبة للمثال الأول النموذج اللاخطي لا يشكل مشكلة. إذا كان حد الخطأ العشوائي يتطلب خصائص في النموذج الأصلي، هذا يجب أن تتحقق في نموذج الانحدار. هذا يعني أن ليس لها تأثير على عملية التحويل. أثناء مناقشة نموذج اللوغاريتمي المزدوج ، فإن حد الخطأ العشوائي يحذف تماماً.

$$Y = \beta_0 X^{\beta_1} e^{u}$$

$$\log Y = \log \beta_0 + \beta_1 \log X + u$$

على أية حال سنعتبر أن حد الخطأ العشوائي مضافاً بشكل تلقائي. حول هذا الاحتمال فإن العنصر العشوائي في النموذج الأصلي يجب أن يكون مضروباً بالحد e^{u} . سوف نرمز لذلك بالحد v .

$$Y = \beta_1 X^{\beta_2} e^{u} = \beta_1 X^{\beta_2} v$$

عندما u تساوي الصفر لا تعدل قيمة لوغاريتم Y ، لكن v تساوي الواحد وبطريقة مماثلة تعدل قيمة Y .

مثال(4-9): إذا علمت بأن العلاقة ما بين المتغيرين Y و X علاقة اللاخطية من الشكل: $Y = b_0 * b_1^X$.

فاحسب انحدار Y على X. إذا علمت أن البيانات على الشكل التالي:

جدول (1-9)

588	447	366	305	262	228	203	Y
7	6	5	4	3	2	1	X

الحل: يمكن تحويل المعادلة الأسية إلى معادلة خطية وذلك بأخذ لوغاريتم الطرفين، فتصبح العلاقة على الشكل التالي:

$$Log(Y) = Log(b_0) + Log(b_1) * X$$

نكتبها من جديد كعلاقة خطية على الشكل: $Y^* = b_0^* + b_1^* * X$

ونطبق عليها طرق إيجاد معادلة الانحدار الخطية من الدرجة الأولى، بالاعتماد على الجدول (2-9).

جدول (2-9)

المجموع	588	447	366	305	262	228	203	Y_i
28	7	6	5	4	3	2	1	X_i
17.46	2.769	2.850	2.564	2.484	2.418	2.358	2.308	Y^*
72.31	19.38	15.90	12.81	9.937	7.255	4.716	2.308	XY
140	49	36	25	16	9	4	1	X^2

نوجد قيمة المعلمة b_0 من العلاقة:

$$b_0^* = \overline{Y}^* - b_1\overline{X} = 2.465 + (0.0878)*(4) = 2.114$$

من هذه العلاقة نحسب قيمة b_1 , b_0 الحقيقية أي بأخذ مقابل لوغاريتم .

فنجد أن : , $b_0 = 130$

$$\bar{X} = \frac{28}{7} = 4; \qquad \bar{Y}^* = \frac{17.465}{7} = 2.495$$

نوجد قيمة المعلمة $b_1\ b_1^* = \dfrac{n * \sum XY - (\sum X) * (\sum Y)}{n * \sum X^2 - (\sum X)^2}$

$$b_1 = 1.224$$

عندئـــذ نكتـــب معادلـــة الانحـــدار علـــى الشـــكل التـــالي:

$$\hat{Y} = (130) * (1.224)^X$$

6-9- إيجاد معلمات الانحدار للعلاقة اللاخطية البسيطة:

أما إذا تعذر علينا تحويل العلاقة اللاخطية إلى علاقة خطية عند توفيق منحني لوصف تلك العلاقة، لابد من التعامل مع العلاقة كما هي من أجل إيجاد معلمات الانحدار للعلاقة. وسوف نعالج في هذا الفصل المعادلة من الدرجة الثانية (التربيعية) من الشكل:

$$\hat{Y} = b_0 + b_1 X + b_2 X^2$$

باستخدام طريقة المربعات الصغرى وباتباع الخطوات نفسها التي اتبعت من أجل إيجاد معلمات الانحدار للمعادلة من الدرجة الأولى، ولكن مع مراعاة في هذه الحالة وجود ثلاثة مجاهيل وهي المعلمات $(b_2\ ,\ b_1,\ \ b_0)$. ومن أجل تحديد قيم هذه المعلمات نشتق علاقة مربعات الفروق اشتـقاقـات جزئيـة بالنسبة للمعلمات $(b_2\ ,\ b_1,\ \ b_0)$ فنحصل على المعادلات الطبيعية التالية :

$$\sum Y = n * b_0 + b_1 * \sum X + b_2 * \sum X^2$$
$$\sum YX = b_0 * \sum X + b_1 * \sum X^2 + b_2 * \sum X^3$$
$$\sum YX^2 = b_0 * \sum X^2 + b_1 * \sum X^3 + b_2 * \sum X^4$$

وبحل المعادلات الطبيعية حلاً مشتركاً نستطيع تحديد المعلمات (b_2 ، b_1, b_0). وعندئذ تكتب علاقة انحدار Y على X على النحو التالي:

$$\hat{Y} = b_0 + b_1X + b_2X^2$$

مثال (5-9): ليكن لدينا البيانات كما في الجدول (3-9). والمطلوب: إذا علمت أن العلاقة بين X و Y من الدرجة الثانية، أوجد معادلة انحدار Y على X.

جدول (3 - 9)

6	5	4	3	2	1	0	X
21.9	14.6	9.3	5.6	3.2	2.1	2.4	Y

الحل :

جدول(4-11)

Y^2	X^2Y	XY	X^4	X^3	X^2	Y_1	X_1
5.76	0	0	0	0	0	2.4	0
4.41	2.1	2.1	1	1	1	2.1	1
10.24	12.8	6.4	16	8	4	3.2	2
31.36	50.4	16.8	81	27	9	5.6	3
84.89	148.8	37.2	256	64	16	9.3	4
213.16	365	73	625	125	25	14.6	5
479.61	788.4	131.4	1296	216	36	21.9	6
831.03	1367.5	266.9	2275	441	91	59.1	21

نعوض في المعادلات الطبيعية :

$$59.1 = 7 * b_0 + 21 * b_1 + 91 * b_2$$

$$266.9 = 21 * b_0 + 91 * b_1 + 441 * b_2$$

$$1367.5 = 91 * b_0 + 441 * b_1 + 2275 * b_2$$

وبالحل المشترك للمعادلات الطبيعية بطريقة المحددات نجد:

$$b_0 = \frac{\begin{vmatrix} 59.1 & 21 & 91 \\ 266.9 & 91 & 441 \\ 1367.5 & 441 & 2275 \end{vmatrix}}{\begin{vmatrix} 7 & 21 & 91 \\ 21 & 91 & 441 \\ 91 & 441 & 2275 \end{vmatrix}} = 2.51 \;, \quad b_1 = \frac{\begin{vmatrix} 7 & 59.1 & 91 \\ 21 & 266.99 & 441 \\ 91 & 1367.5 & 2275 \end{vmatrix}}{\begin{vmatrix} 7 & 21 & 91 \\ 21 & 91 & 441 \\ 91 & 441 & 2275 \end{vmatrix}} = -1.2$$

$$b_2 = \frac{\begin{vmatrix} 21 & 21 & 59.1 \\ 21 & 91 & 266.9 \\ 91 & 441 & 1367.5 \end{vmatrix}}{\begin{vmatrix} 7 & 21 & 91 \\ 21 & 91 & 441 \\ 91 & 441 & 2275 \end{vmatrix}} = 0.733$$

إذاً معادلة الانحدار لها الشكل التالي:

$$\hat{Y} = 2.51 - 1.2 * X_i + 0.733 * X_i^2$$

9-7- حساب دليل الارتباط في العلاقة اللاخطية:

يمكننا حساب دليل الارتباط كما حسبنا سابقاً معامل الارتباط الخطي بدلالة انحرافات القيم عن مستقيم الانحدار، أو منحني الانحدار؛ لذلك لابد من حساب الخطأ المعياري للتقدير والانحراف المعياري للمتغير Y ، (S_Y).

الخطأ المعياري للتقدير:

$$S_{\hat{Y}}^2 = \frac{\sum \left(Y_i - \hat{Y}\right)^2}{n} = \frac{\sum Y^2 - b_0 * \sum Y - b_1 * \sum YX - b_2 * \sum YX^2}{n}$$

أما الخطأ المعياري لـ Y :
$$S_Y^2 = \frac{\sum Y^2}{n} - \left(\frac{\sum Y}{n}\right)^2$$

ولحساب دليل الارتباط نطبق العلاقة :
$$I = \sqrt{1 - \frac{S_{\hat{Y}}^2}{S_Y^2}}$$

حيث إن: I – تدعى دليل الارتباط (Correlation Index): وهو مقياس للارتباط اللاخطي. و إن دليل الارتباط I يساوي معامل الارتباط r عندما تكون العلاقة ما بين المتغيرين خطية، أو يمكن تحويلها إلى خطية. أي عندما يكون الحد الأخير لحساب الخطأ المعياري للتقدير $S_{\hat{Y}}^2$ مساوياً الصفر (أي عندما $b_2 * \sum YX^2 = 0$).

ولكن لابد من الإشارة إلى أن الخطأ المعياري حول منحني الانحدار أقل من الخطأ المعياري حول مستقيم الانحدار؛ لذلك فإن (I) تكون أكبر من (r). ويمكن القول أيضاً إن معامل الارتباط (r) يكون أقل من (I) دليل الارتباط في حالة الانحدار اللاخطي في مختلف أشكاله.

مثال (9-6): من المثال (9-4) احسب دليل الارتباط .

الحل:

أولاً، نحسب $S_{\hat{Y}}^2$:

$$S_{\hat{Y}}^2 = \frac{831.03 - (2.51)*(59.1) - (-1.2)*(266.9) - (0.733)*(1367.5)}{7}$$
$$= \frac{831.03 - 148.341 + 320.28 - 1002.83}{7} = \frac{0.1357}{7} = 0.0194$$

ثانياً، نحسب: S_Y^2 .

$$S_Y^2 = \frac{831.03}{7} - \left(\frac{59.1}{7}\right)^2 = 118.72 - 71.28 = 47.44$$

ثالثاً ، نطبق علاقة دليل الارتباط :

$$I = \sqrt{1 - \frac{0.0194}{47.44}} = \sqrt{0.999} = 0.999$$

التفسير: إن علاقة الارتباط ما بين المتغيرين Y و X قوية جداً، وهذا يعني أننا وفقنا في اختيار شكل العلاقة بين المتغيرين، هي بالفعل علاقة من الدرجة الثانية (أي علاقة غير خطية). ولكن لو حصلنا على دليل ارتباط ضعيف في هذه الحالة لابد من البحث بشكل تجريبي لاختيار شكل العلاقة الحقيقي، أو القريب من الواقع ما بين المتغيرين، ويساعدنا في ذلك رسم شكل الانتشار أولاً، ومن خلاله نستطيع تحديد أقرب الأشكال لتوزيع نقاط الانتشار، لاختيار الصيغة التي تظهر العلاقة ما بين البيانات المدروسة.

ومن ثم جرب أن يأخذ هذه العلاقة بدلالة النموذج النصف لوغاريتمي فوجد:

Source	SS	df	MS	
Model	15.116284	1	15.116284	Number of obs = 25
Residual	9.04353146	23	.39319702	F(1, 23) = 38.44
				Prob > F = 0.0000
				R-squared = 0.6257
				Adj R-squared = 0.6094
Total	24.1598154	24	1.00665898	Root MSE = .62705

e	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
lgg	1.709096	.2756444	6.200	0.000	1.138883	2.27931
_cons	-.728669	.2830097	-2.575	0.017	-1.314119	-.1432188

والمطلوب:

أ- اكتب معادلة النموذج الأول وفسر الثوابت .

ب- اكتب معادلة النموذج الثاني وفسر الثوابت.

ج- اكتب معادلة النموذج الثالث وفسر الثوابت.

د- برأيك أي النماذج أكثر تمثيلاً للعلاقة وأيد ذلك بالتحليل الاقتصادي أو الإحصائي أو الرياضي أو الاقتصاد القياسي.

هـ- طبق اختبار BOX-COX TESTS لتحديد معنوية معادلة نصف اللوغاريتمي.

5- توصل فيما يلي الباحث إلى معادلة الانحدار لوغاريتم الإنفاق الغذائي المثقل ويمثل الإنفاق العائلي السنوي على الغذاء للأسرة LnX_1 على إجمالي الإنفاق السنوي للأسرة LnX_2 ومكان إقامة الأسرة X_3 وجنس رب الأسرة X_4 مستخدما بيانات مسح دخل ونفقات الأسرة الذي أجره المكتب المركزي للإحصاء عام 2003 لعينة مكونة 29800 على مستوى القطر:

أ- يحاول إيجاد علاقة بين Y التي هي الدخل الساعي وS عدد سنوات الدراسة وPWE، الخبرة العملية، باستخدام صيغة نصف لوغاريتمية:

$$\text{Log } y = \beta_1 + \beta_2 S + \beta_3 PWE + u$$

حيث u هي قيمة المتغير العشوائي من المفترض تتناسب وشروط ماركوف غاوس. PWE تحدد بسنوات الدراسة. بما أن كل الذين استجابوا كانت أعمارهم 33 عام 1991، هذا يصبح:

$$PWE = 33 - S - 5 = 27\text{-}S$$

الباحث وجد أنه من المستحيل تناسب الشكل كما هو محدد. البنود الإحصائية التي توصل إليها موجودة في الجدول التالي:

```
. reg LGY S PWE
      Source |       SS       df       MS              Number of obs =    5660
-------------+------------------------------           F(  1,  5658) = 1232.62
       Model |  237.170265     1  237.170265           Prob > F      =  0.0000
    Residual |  1088.66373  5658  .192411405           R-squared     =  0.1789
-------------+------------------------------           Adj R-squared =  0.1787
       Total |    1325.834  5659  .234287682           Root MSE      =  .43865
------------------------------------------------------------------------------
         LGY |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           S |   .1038011   .0029566    36.11   0.000     .0980051    .1095971
         PWE |  (dropped)
       _cons |   .5000033   .0373785    13.38   0.000     .4267271    .5732795
------------------------------------------------------------------------------
```

أ- اشرح لمَ لمْ يستطع الباحث بناء نموذجه .

ب- اشرح كيف يمكن تفسير معلمة S .

8- حصل الباحث احمد على بيانات الإنفاق السنوي العائلي على الكتب (والتي تمثل بالمتغير B) وعلى بيانات الدخل السنوي العائلي (وتمثل بالرمز Y) لـ (100)عائلة سورية في عام 2005 . ويفترض أن المتغير

(B) يتبع المتغير (Y) ومعدل القدرة الإدراكية للبالغين في العائلة والتي تمثل بالمتغير(IQ) حسب العلاقة التالية :

$$\log B = \beta_1 + \beta_2 \log Y + \beta_3 \log IQ + u \qquad (A)$$

حيث u هي حد الخطأ العشوائي والذي يحقق شروط (ماركوف ـــ قاوس)، ويعتقد الباحث أيضاً إمكانية أن (log B) قد يتحدد من قبل (log Y) لوحده حسب العلاقة التالية:

$$\log B = \beta_1 + \beta_2 \log Y + u \qquad (B)$$

أحمد ليس لديه بيانات حول القدرة الإدراكية للبالغين (معامل الذكاء)، وقرر استخدام معدل سنوات الدراسة للبالغين في العائلة (وتمثل بالمتغير S) كوسيط في النموذج (A). قد يفترض الباحث أن Y و S كلاهما متجانسان. الارتباط في العينة بين (log Y) و (log S) يساوي (0.86) . والباحث جرب نماذج الانحدار التالية :

أولاً: انحدار (log B) على كل من (log Y) و(log S)

ثانياً: انحدار (log B) على (log Y) فقط

والنتائج ظاهرة في الجدول (القيم بين الأقواس هي الأخطاء المعيارية):

	(1)	(2)
$\log Y$	1.10 (0.69)	2.10 (0.35)
$\log S$	0.59 (0.35)	–
constant	–6.89 (2.28)	–3.37 (0.89)
R^2	0.29	0.27

المطلوب مساعدة الباحث احمد في كتابة تقريره بالإجابة على الأسئلة التالية:

أ – على افتراض أن (A) هي البند الصحيح ، اشرح مع البرهان الرياضي فيما إذا تتوقع أن معامل (log Y) سيكون أكبر في الانحدار رقم (2).

ب – افترض أن (A) هي البند الصحيح ، صف المنافع العديدة من استخدام (log S) كوسيط لـ(log IQ) كما في نموذج الانحدار رقم (1) إذا كان (log S) وسيط أو دليل جيد.

ج – اشرح فيما إذا القيمة الصغيرة لـ R^2 لعلاقة الانحدار رقم (1) تشير إلى أن (log S) ليس وسيط جيد.

د – افترض أن (A) هي النموذج الصحيح ، قدم تفسيراً لماذا معاملات المتغيرات (log Y) و(log S) في نموذج الانحدار رقم (1) ليست مختلفة بشكل ملحوظ عن الصفر وذلك باستعمال اختبار(t) من اتجاهين.

أخيراً طبق الباحث إلياس النموذج التالي :

$$\widehat{\log L} = -3.12 + 0.42 \log\frac{Y}{P} - 0.34 \log\frac{W}{P} - 0.03 \log P$$
$$(0.13) \quad (0.09) \qquad (0.10) \qquad (0.04)$$
$$RSS = 1.99 \qquad (3)$$

المطلوب: مساعدة الباحث إلياس في إنجاز تقريره الاقتصادي بالإجابة على الأسئلة التالية:

أ – أعطي تفسير اقتصادي لمعاملات الانحدار في المعادلة رقم (2) .

ب – اشرح لماذا النموذج الثاني هو حالة مقيدة في البداية ، وحدد القيد .

ج – أجر اختبار F المقيد .

د – أجر اختبار t المقيد .

هـ – اشرح فيما إذا اختبار F واختبار t يمكن أن يعطيا نتائج مختلفة.

و – افترض أن القيد كان صحيحاً ، اشرح لماذا نموذج الانحدار رقم (2) مفضل على نموذج الانحدار رقم (1).

ز – نموذج الانحدار رقم (2) أسوأ من نموذج الانحدار رقم (1) ، هل يعني هذا شيء ؟

10– لدى الباحثة ريم بيانات عن المسح الوطني للشباب للعام 2000، والأجر الساعي نرمز له بـ (X) وسنوات التدريس بـ (S) وسنوات خبرة العمل (X)، لدينا عينة مكونة من (1774) ذكر و (1468) أنثى وعرف المتغير الوهمي (MALE) إذا كان ذكر=1 ، وميل متغير وهمي (SMALE) وذلك كناتج للمتغيرين (S) و (MALE) وميل متغير وهمي آخر (XMALE) كناتج لـ (X) و (MALE) ونجري نماذج الانحدار التالية:

(1) ـــ انحدار (log Y)على (S) و(X) لكامل العينة

(2) ـــ انحدار (log Y) على (S) و (X) للذكور فقط

(3) ـــ انحدار (log Y) على (S) و (X) للإناث فقط

(4) ـــ انحدار (log Y) على (S) و (X) و(MALE) لكامل العينة

(5) ـــ انحدار (log Y)على (S) و (X) و(MALE) و(SMALE) و (XMALE) لكامل العينة. النتائج تظهر في الجدول مع الأخطاء المعيارية بين الأقواس ، [RSS] هي مجموع مربع البواقي و(n) هي عدد المشاهدات

	(1)	(2)	(3)	(4)	(5)
S	0.094 (0.003)	0.099 (0.004)	0.094 (0.005)	0.097 (0.003)	0.094 (0.005)
X	0.046 (0.002)	0.042 (0.003)	0.039 (0.002)	0.040 (0.002)	0.039 (0.003)
MALE	–	–	–	0.234 (0.016)	0.117 (0.108)
SMALE	–	–	–	–	0.005 (0.007)
XMALE	–	–	–	–	0.003 (0.004)
constant	5.165 (0.054)	5.283 (0.083)	5.166 (0.068)	5.111 (0.052)	5.166 (0.074)
R^2	0.319	0.277	0.363	0.359	0.359
RSS	714.6	411.0	261.6	672.8	672.5
n	3,242	1,774	1,468	3,242	3,242

قيمـــة الارتبـــاط بـــين MALE و SMALE وبـــين MALE و XMALE كلاهما يساوي 0.96 والارتباط بين SMALE و XMALE يساوي0.93 .

المطلوب: مساعدة الباحثة ريم في كتابة تقريرها بالإجابــة علــى الأسـئلة التالية:

أ – أعطي تفسير لمعاملات (S) و(SMALE) في نموذج الانحدار رقم(5).

ب – أعطي تفسير لمعاملات MALE في نموذجي الانحدار (4) و (5) .

ج – افترضت الباحثة أن وظيفة أو غرض الدخل مختلف بين الذكور والإناث . أجري اختبار بهذا الافتراض مستخدماً نموذج الانحدار رقم (4) وأيضاً باستخدام نموذج الانحدار رقم (1) و رقم (5) علماً F(3,1000) عند مستوى دلالة 1% هو 3.80 .

د – اشرح الفروق في الاختبارات مستعيناً بنموذج الانحدار رقم (4) وبنموذجي الانحدار رقم (1) ورقم(5) .

هـ – في حلقة دراسية اقترح شخص ما أن اختبار Chow يمكن أن يسلط الضوء على فرضية الباحثة ، هل هذا صحيح ؟

و – اشرح أي من (1) أو (4) أو (5) سيكون نموذجك المفضل ؟

الفصل العاشر

مشاكل في تحليل الانحدار

THE PROBLEMS OF THE REGRESSION ANALYSIS

بعد أن استعرضنا بناء نماذج الانحدار، صادفتنا بعض المشاكل. في هذا الفصل سوف نحاول معالجة بعضٍ منها مثل مشكلة التعدد الخطـــي، وانعــدام تجانس الخطأ العشوائي، و الارتباط الذاتي.

10-1- مشكلة التعدد الخطي:

تعدد العلاقات الخطية : عندما يكون في النموذج أكثر من متغيرين مــن المتغيرات من المتغيرات المستقلة (المفسرة) ، فهذا يعطـــي تقـدير النمـــوذج بطريقة المربعات الصغرى OLS قيمة لـــ R^2 تكون عالية، ولكن ليس هنـــاك معنوية إحصائية . وبالطبع مثل هذا ناتج عن وجود علاقة ارتباطية قوية، بـــين المتغيرات المستقلة ولذلك لابد من التخلص من هذه الارتباط، ويكون ذلك على أساس إدخال كل متغير على حدة ومعرفة تأثيره على المتغيـــر التـــابع، ولكـــن النظرية الاقتصادية أحياناً تنص على أنه لابد من أخذ المتغيرين بعين الاعتبار؛ لذلك فإن هذا الأمر يحتاج إلى معالجة.

على فرض أن : $Y = 2 + 3X_2 + X_3$ و علمـــاً أن $X_3 = 2X_2 - 1$. لا يوجد حد الخطأ العشوائي في معادلة Y ، ولكن هذا ليس مهم. وعلى فرض أنه لدينا 6 مشاهدات . $Y = 2 + 3X_2 + X_3$ علماً أن $X_3 = 2X_2 - 1$

التغير في Y	التغير في X_3	التغير في X_2	$X3$	$X2$	Y
5	2	1	19	10	51
5	2	1	21	11	56
5	2	1	23	12	61
5	2	1	25	13	66
5	2	1	27	14	71
5	2	1	29	15	76

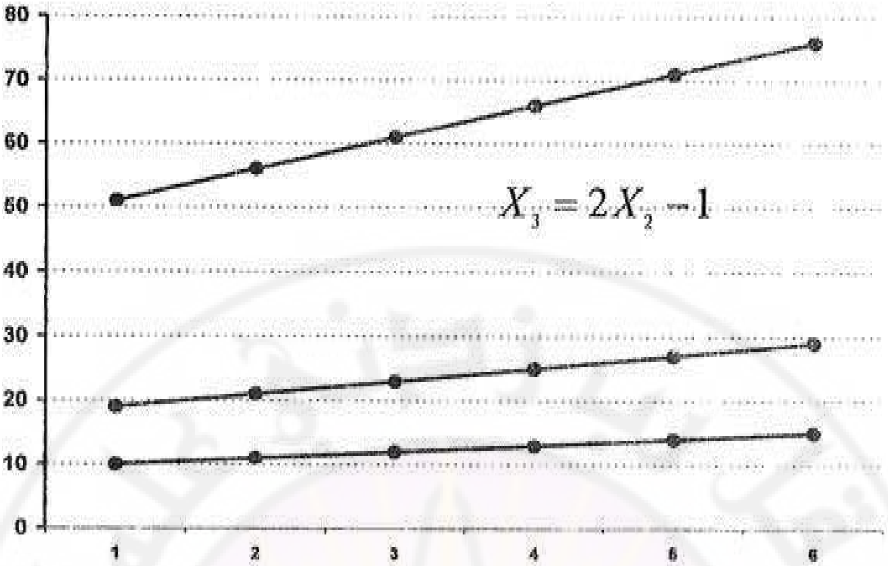
نمثل المتغيرات الثلاث على المحاور الإحداثية. لنتفحص البيانـات نجـد استحالة أن نتبين أيهما يحدث تغيير في Y ، X_1 أو X_2 أو بشكل منفصل لكلا المتغيرين X_2,X_1 .

إن ترجمة المثال بشكل عددي تزداد كل مشاهدة في Y بـ 5 إذا تغيرت X_1 بمقدار 1 انظر الجدول (1).

إذاً العلاقة الحقيقية بين المتغيرات هي من الشكل $Y = 1 + 5X_1$. مهما كانت قيمة Y نجد أن قيمة X_2 زادت بمقدار 2 لكل مشاهدة. أو على فرض أن العلاقة هي من الشكل: $Y = 3.5 +2.5X_2$. ويمكن أن تكون على الشكل إحدى الحالات للعلاقة : $Y = 3.5 -2.5p + 5pX_1 +2.5(1-p)X_2$

أي: سيكون المنحني على علاقة مع قيم p، ليس هناك مـن طريقـة لتحليلِ الانحدار، أو أيِّ تقنية أخرى، يُمْكِنُ أن تُقرِّرَ العلاقةَ الحقيقيةَ مِـنْ هـذه المجموعةِ من الإمكانيات غير المنتهية، لإعطاء بيانات العينة.

ماذا يحدث إذا حاولت تنفيذ الانحدار عندما يكون هناك علاقة خطية تماماً بين المتغيرات المستقلة.



الشكل (1-10)

سوف نبحث ذلك في استخدام نموذج بمتغيرين مستقلين. (**ملاحظة**: الحد العشوائي للخطأ متضمن في النموذج الحقيقي، ولكن الاختلاف في التحليل) .

نحسب معامل الانحدار المتعدد من العلاقة:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u \text{ .}$$

وعلى فرض أن: $X_3 = \lambda + \mu X_2$

$$b_1 = \frac{\mathrm{Cov}(X_1, Y)\mathrm{Var}(X_2) - \mathrm{Cov}(X_2, Y)\mathrm{Cov}(X_1, X_2)}{\mathrm{Var}(X_1)\mathrm{Var}(X_2) - [\mathrm{Cov}(X_1, X_2)]^2}$$

نعوض X_3 بقيمها كما هو مبين في العلاقة أعلاه. و باستخدام قاعدة التباين رقم 4 نستطيع حذف المتغير المضاف μ من حدود التباين.

$$b_1 = \frac{\mathrm{Cov}(X_1, Y)\mathrm{Var}(\lambda + \mu X_1) - \mathrm{Cov}([\lambda + \mu X_1], Y)\mathrm{Cov}(X_1, [\lambda + \mu X_1])}{\mathrm{Var}(X_1)\mathrm{Var}(\lambda + \mu X_1) - [\mathrm{Cov}(X_1, [\lambda + \mu X_1])]^2}$$

بالمماثلة نصلح العلاقة باستخدام قواعد التباين المشترك، ونستطيع عندئذ حذف القيمة المضافة μ من حدود التباين المشترك، و نستطيع إخراج قيمة μ من حدود التباين بعد تربيعها كما تعلمنا سابقاً، ونحصل على:

$$b_1 = \frac{\mathrm{Cov}(X_1,Y)\mathrm{Var}(\mu X_1) - \mathrm{Cov}(\mu X_1,Y)\mathrm{Cov}(X_1,\mu X_1)}{\mathrm{Var}(X_1)\mathrm{Var}(\mu X_1) - [\mathrm{Cov}(X_1,\mu X_1)]^2}$$

التباين المشترك $\mathrm{Cov}(X_1, X_1)$ للمتغير مع نفسه ما هو إلا تباين هذا المتغير $\mathrm{Var}(X_1)$. بناء على جملة الإصلاحات يمكننا أن نقوم بإعادة كتابة البسط والمقام فنجد أن كلاهما يساوي الصفر. وعندئذ نجد أن قيمة معامل التحديد غير معين.

إذاً من غير المفيد وجود علاقة بين المتغيرات المستقلة في الانحدار، وعندما يحدث هذا يكون هناك خطأ منطقي في المواصفات.

$$b_1 = \frac{\mu^2\mathrm{Cov}(X_2,Y)\mathrm{Var}(X_1) - \mu^2\mathrm{Cov}(X_1,Y)\mathrm{Var}(X_1)}{\mu^2\mathrm{Var}(X_1)\mathrm{Var}(X_1) - [\mu\mathrm{Var}(X_1)]^2} = \frac{0}{0}$$

بالتطبيق على مثال انحدار الدخل الساعي *EARNINGS* على المهارة المعرفية*ASVABC* , *و عدد سنوات التحصيل الدراسيS* و مقياس لسرعة الاختبار في المهارة الإنجاز بحساب رياضي بسيط*ASVAB5* .

```
. reg EARNINGS S ASVABC ASVAB5
```

Source	SS	df	MS
Model	4909.11468	3	1636.37156
Residual	33487.9224	566	59.1659406
Total	38397.0371	569	67.4816117

Number of obs	570
F(3, 566)	27.66
Prob > F	0.0000
R-squared	0.1279
Adj R-squared	0.1232
Root MSE	7.6919

EARNINGS	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
S	.7115506	.1612235	4.413	0.000	.3948811	1.02822
ASVABC	.1104595	.0504223	2.191	0.029	.0114219	.2094972
ASVAB5	.0770794	.0463868	1.662	0.097	-.0140319	.1681908
_cons	-5.944977	2.161409	-2.751	0.006	-10.19034	-1.699616

على فرض أن المتغير *ASVABC* في هذا الاختبار مقاسة، و لها متوسط 50 وانحراف معياري 10 .

نتائج الانحدار تنص على أنه إذا زاد التحصيل الدراسي سنة إضافية ستؤدي الزيادة في الدخل بمقدار 0.71 $.أما الوحدة الإضافية في المتغير *ASVABC* سيؤدي إلى زيادة في الكسب بمقدار 0.110 $. نقطة واحدة في الخطأ المعياري حول المتوسط تؤدي إلى 1.10$ لكل ساعة مقارنة مع وسطها الحسابي .

إن زيادة نقطة واحدة حسابياً لسرعة الاختبار يؤدي إلى الزيادة بمقدار 0.08$ في الدخل الساعي.

هل تعتقد أن المتغير *ASVAB5* ينتمي إلى معادلة الدخل؟ إذا طبقنا اختبار t سوف نجد أن هذا المعامل هو غير معنوي في الاختلاف عن الصفر عند مستوى دلالة 5% من اتجاه واحد . (تبرير ذلك: يكون مستبعداً لأن النقاط الجيدة على هذا الاختبار سوف تعكس أثر الدخل).

سوف نلاحظ لذلك في الانحدار معامل المتغير *ASVABC* معنوي عند 5% مستوى اختبار فقط.

لنخرج المتغير من علاقة الانحدار نلاحظ أن t الإحصائية أصبحت 3.60 وهي معنوية حتى عند مستوى دلالة 0.1 بالاختلاف عن الصفر، والسبب في انخفاض قيمة t النسبية عائد إلى ارتفاع معامل الارتباط بين المتغيرين *ASVAB5* و *ASVABC* .

```
. cor ASVABC ASVAB5
(obs=570)
|    ASVABC    ASVAB5
--------+------------------
ASVABC|    1.0000
ASVAB5|    0.6371    1.0000
```

هذا يعني ظهور صعوبة صغيرة فردية التأثير بين المتغيرين *ASVABC* و *ASVAB5* ، نتيجة تقدير الانحدار أدت إلى حدوث شذوذ في النتائج . أي أن الارتباط العالي سبب الأخطاء المعيارية الكبيرة، علماً أن الارتباط الضعيف مؤشر على أن نقاط التقديرات غير موثوقة التقدير.

عندما يكون الارتباط بين المتغيرات المستقلة شديداً يقود إلى فقدان نقاط التقديرات للمعاملات، وحدوث خطأ معياري كبير، وعدم تحقق معنوية t الاحصائية، الانحدار ينص على ضرورة عدم تعدد العلاقات الخطية بين المتغيرات المستقلة لنموذج الانحدار.

```
. reg EARNINGS S ASVABC ASVAB5

------------------------------------------------------------------------------
EARNINGS |   Coef.    Std. Err.     t     P>|t|     [95% Conf. Interval]
---------+--------------------------------------------------------------------
       S |.7115506   .1612235    4.413    0.000     .3948811     1.02822
  ASVABC |.1104595   .0504223    2.191    0.029     .0114219     .2094972
  ASVAB5 |.0770794   .0463868    1.662    0.097    -.0140319     .1681908
   _cons |-5.944977   2.161409   -2.751     0.006      -10.19034        -
1.699616
------------------------------------------------------------------------------

. reg EARNINGS S ASVABC

------------------------------------------------------------------------------
EARNINGS |   Coef.    Std. Err.     t     P>|t|     [95% Conf. Interval]
---------+--------------------------------------------------------------------
       S |.7390366   .1606216    4.601    0.000     .4235506     1.054523
  ASVABC |.1545341   .0429486    3.598    0.000     .0701764     .2388918
   _cons |-4.624749   2.0132     -2.297    0.022    -8.578989    -.6705095
------------------------------------------------------------------------------
```

إن تعدد العلاقة الخطية في هذا الانحدار يجعل من الصعوبة تحديد أي المتغيرات له تأثير في تحديد الدخل *ASVAB5* أو *ASVABC* ، أو لا؛ لأن قيمة t هامشية المعنوية بالاعتماد على الحظ والمصادفة.

10-2- قياس إمكانية تخفيف تعدد العلاقات الخطية

Possible measures for alleviating multicollinearity

ماذا عن هذه المشكلة إذا واجهتنا ؟ سوف نناقش بعض المقاييس الممكنة ، و لنبحث عن ذلك في نموذج متعدد خطي بمتغيرين مستقلين. قبل البدء، لا بد من التأكيد على نقطتين هامتين:

أولاً: إن تعدد العلاقات الخطية ليس سبباً في أن تكون معاملات الانحدار متحيزة. توزيعاتها على الأرجح مازالت متركزة حول القيم الصحيحة، فقد تكون مواصفات الانحدار صحيحة ، ولكنها غير مرضية لكبر تباينها.

ثانياً: إن الخطأ المعياري واختبارات t تبقى صحيحة. الأخطاء المعيارية تكون أكبر مما هي عليه في غياب تعدد العلاقات الخطية ، نحذر أنفسنا من مغبة أن تكون تقديرات الانحدار غير صحيحة.

10-3- طرق التخفيف من مشكلة التعدد العلاقات الخطية :

1- تخفيض قيمة التضمين بالمتغيرات ذات الصلة في النموذج.

إن مقدر تباين المجتمع لحد الخطأ العشوائي هو مجموع مربع البواقي مقسومة على n-k ، حيث n عدد المشاهدات 570 و k عدد المعاملات و هو 3، قيمته تساوي 59.2 حسب مثالنا المدروس.

سوف نضيف متغيرات أخرى مثل منصب العاملين الوظيفي TENURE، جنس العامل المحبب MALE ، ومكان الإقامة في الحضر URBAN لمعادلة الكسب. نجد أن جميع المعاملات معنوية عند 1% مستوى دلالة أو أعلى.

```
. reg EARNINGS S ASVABC ASVAB5 TENURE MALE URBAN

  Source |       SS       df       MS              Number of obs =     570
---------+------------------------------           F(  6,   563) =   23.60
   Model |  7715.87322     6  1285.97887           Prob > F      =  0.0000
Residual |  30681.1638   563  54.4958505           R-squared     =  0.2009
---------+------------------------------           Adj R-squared =  0.1924
   Total |  38397.0371   569  67.4816117           Root MSE      =  7.3821

------------------------------------------------------------------------------
EARNINGS |    Coef.   Std. Err.      t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
       S | .8137184   .1563975    5.203   0.000       .5065245    1.120912
  ASVABC | .0442801    .049716    0.891   0.373      -.0533714    .1419317
  ASVAB5 | .1113769   .0458757    2.428   0.016       .0212685    .2014853
  TENURE |  .287038   .0676471    4.243   0.000       .1541665    .4199095
    MALE |3.123929     .64685    4.829   0.000       1.853395    4.394463
   URBAN |2.061867    .7274286    2.834   0.005       .6330618    3.490672
   _cons |-10.60023   2.195757   -4.828   0.000      -14.91311   -6.287358
------------------------------------------------------------------------------
```

على أي حال هذا الحساب فقط له أهمية نسبية صغيرة في تباين الكسب، و نلاحظ أنه خفض من تقدير تباين المجتمع للخطأ العشوائي بشكل صغير.

```
. reg EARNINGS S ASVABC ASVAB5

  Source |       SS       df       MS              Number of obs =     570
---------+------------------------------           F(  3,   566) =   27.66
   Model |  4909.11468     3  1636.37156           Prob > F      =  0.0000
Residual |  33487.9224   566  59.1659406           R-squared     =  0.1279
---------+------------------------------           Adj R-squared =  0.1232
   Total |  38397.0371   569  67.4816117           Root MSE      =  7.6919

. reg EARNINGS S ASVABC ASVAB5 TENURE MALE URBAN

  Source |       SS       df       MS              Number of obs =     570
---------+------------------------------           F(  6,   563) =   23.60
   Model |  7715.87322     6  1285.97887           Prob > F      =  0.0000
Residual |  30681.1638   563  54.4958505           R-squared     =  0.2009
---------+------------------------------           Adj R-squared =  0.1924
   Total |  38397.0371   569  67.4816117           Root MSE      =  7.3821
```

من هذه النتائج نلاحظ تعاقب أثر الأخطاء المعيارية على المتغيرات *ASVABC* و*ASVAB5* بشكل غير هام.

```
. reg EARNINGS S ASVABC ASVAB5
```

EARNINGS	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
S	.7115506	.1612235	4.413	0.000	.3948811	1.02822
ASVABC	.1104595	.0504223	2.191	0.029	.0114219	.2094972
ASVAB5	.0770794	.0463868	1.662	0.097	-.0140319	.1681908
_cons	-5.944977	2.161409	-2.751	0.006	-10.19034	-1.699616

```
. reg EARNINGS S ASVABC ASVAB5 TENURE MALE URBAN
```

EARNINGS	Coef.	Std. Err.	t	P>|t|	95% Conf.	Interval]
S	.8137184	.1563975	5.203	0.000	.5065245	1.120912
ASVABC	.0442801	.049716	0.891	0.373	-.0533714	.1419317
ASVAB5	.1113769	.0458757	2.428	0.016	.0212685	.2014853
TENURE	.287038	.0676471	4.243	0.000	.1541665	.4199095
MALE	3.123929	.64685	4.829	0.000	1.853395	4.394463
URBAN	2.061867	.7274286	2.834	0.005	.6330618	3.490672
_cons	-10.60023	2.195757	-4.828	0.000	-14.91311	-6.287358

ومن النتائج المعروضة أعلاه يلاحظ عدم استقرار المعاملات، وهذا يشير دائماً لتعدد العلاقات الخطية.

2- زيادة عدد مشاهدات المسح : وهذا حتماً يؤدي إلى زيادة التكاليف. أي نبحث عن n عدد المشاهدات. إذا كنا نعمل على تقاطع قطاعات البيانات (الأفراد ، الأسر ، المشروع ، إلخ...) وكنا نعمل مسحاً ، يجب علينا زيادة حجم العينة بالتفاوض على زيادة الميزانية، أو البحث عن طريقة أخرى غير زيادة الميزانية كأن نستخدم العينة العنقودية في المسح. نستطيع تثبيت الميزانية باستخدام المعلومات ببراعة كالعناقيد. يمكن توزيع المنطقة جغرافياً بالدليل أو ببريد المنطقة. ويمكن تحديد عدد عشوائي ، ربما باستخدام المعاينة العشوائية لتأكيد أبناء العاصمة، وبتصور ملائم للمناطق أخرى حضر و ريف .عندما نحدد المسح، و نحدد المناطق، هذا يعمل على تخفيض عدد مرات السفر، وتكاليف العمل الميداني، يسمح لنا مقابلة عدد كبير من المستجيبين. أما إذا كنا

نعمل مع مواعيد متتالية ، يمكن زيادة العينة بالعمل لفترات قصيرة للبيانــات، على سبيل المثال بيانات ربعية أو شهرية بدلاً من بيانات سنوية، أي: نلجأ إلى استخدام السلاسل الزمنية كائن نستخدم البيانات الفصلية بدلاً عن السنوية.

هنا ننفذ معادلة الانحدار على 2867 مشاهدة على مجموعة بيانات المسح الإحصائي(EAEF).

. reg EARNINGS S ASVABC ASVAB5

Source	SS	df	MS
Model	36689.8765	3	12229.9588
Residual	190928.139	2864	66.664853
Total	227618.016	2867	79.3924017

Number of obs	= 2868
F(3, 2864)	= 183.45
Prob > F	= 0.0000
R-squared	= 0.1612
Adj R-squared	= 0.1603
Root MSE	= 8.1649

EARNINGS	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
S	1.002693	.0787447	12.733	0.000	.8482905	1.157095
ASVABC	.1448345	.0241135	6.006	0.000	.097553	.1921161
ASVAB5	.0483846	.0218352	2.216	0.027	.0055703	.091199
_cons	-9.654593	1.033311	-9.343	0.000	-11.6807	-7.628485

إذا قارنا هذه النتيجة مع استخدام بيانات العينة 570 ، ســوف نــرى أن الخطأ المعياري صغر بشكل كبير كما هو متوقع، كما هو واضح فــي جــدول النتائج أدناه.

. reg EARNINGS S ASVABC ASVAB5

Number of obs = 570

EARNINGS	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
S	.7115506	.1612235	4.413	0.000	.3948811	1.02822
ASVABC	.1104595	.0504223	2.191	0.029	.0114219	.2094972
ASVAB5	.0770794	.0463868	1.662	0.097	-.0140319	.1681908
_cons	-5.944977	2.161409	-2.751	0.006	-10.19034	-1.699616

. reg EARNINGS S ASVABC ASVAB5

Number of obs = 2868

EARNINGS	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
S	1.002693	.0787447	12.733	0.000	.8482905	1.157095
ASVABC	.1448345	.0241135	6.006	0.000	.097553	.1921161
ASVAB5	.0483846	.0218352	2.216	0.027	.0055703	.091199
_cons	-9.654593	1.033311	-9.343	0.000	-11.6807	-7.628485

وبالنتيجة فإن t الإحصائية لبعض المتغيرات المستقلة ارتفعت. في هذه الحالة المتغير ASVABC يكون جزءاً من الحقيقة ؛ لأن نقطة التقدير للمعامل ارتفعت. وكذلك ارتفعت قيمة t الإحصائية للمتغير *ASVAB5* بالرغم من أن المعامل صغير.

3- زيادة $\mathbf{Var}(X_1)$:

الإمكانية الثالثة لتقليل مشكلة تعدد العلاقات الخطية يمكن أن تكون في زيادة تباين المتغيرات المستقلة. هنا يكون من الممكن فقط بمرحلة تصميم المسح. على سبيل المثال إذا كنا نخطط لمسح الأسر بهدف البحث عن نموذج الإنفاق جزء من تنوع الدخل، يجب عليك التأكد من أن العينة تتضمن نسبة من الأسر الغنية و نسبة من الأسر الفقيرة، وكذلك متوسطة الدخل.

4- تخفيض Reduce r_{X_1,X_2} : الإمكانية الأخرى يمكن أن تكون في تخفيض معامل الارتباط بين المتغيرات المستقلة . تكون هذه الإمكانية فقط في مرحلة تصميم عينة المسح، وموازنة ذلك ليس سهلاً.

5- ضم ارتباط المتغيرات(Combine the correlated variables): إذاً كان ارتباط المتغيرات فيما بينها ، يمكن أن نحصل على معقولية جمعهم من خلال بعض الأرقام القياسية العظمى. على سبيل المثال : ماذا نعمل لمعرفة المتغيرات ASVAB . يمكن أن يحسب المتغير ASVABC كوزن متوسط من *ASVAB2* (الاستنتاج المنطقي) *ASVAB3* (عبارة المعرفة) و *ASVAB4* (علامة الفهم أو الإدراك).

```
. reg EARNINGS S ASVAB2 ASVAB3 ASVAB4

  Source |       SS       df       MS              Number of obs =     570
---------+------------------------------           F(  4,   565) =   25.68
   Model |  5906.47726     4  1476.61931           Prob > F      =  0.0000
Residual |  32490.5598   565  57.5054156           R-squared     =  0.1538
---------+------------------------------           Adj R-squared =  0.1478
   Total |  38397.0371   569  67.4816117           Root MSE      =  7.5832

------------------------------------------------------------------------------
EARNINGS |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
       S |  .7362439   .1586812      4.640   0.000      .4245668    1.047921
  ASVAB2 |  .2472668   .0472249      5.236   0.000     .154509     .3400246
  ASVAB3 |  .0137422   .058716       0.234   0.815     -.1015861    .1290705
  ASVAB4 | -.1051868   .0544682     -1.931   0.054     -.2121716    .001798
   _cons | -4.734303   2.06706      -2.290   0.022     -8.794363   -.6742428
------------------------------------------------------------------------------
```

في نموذج انحدار EARNINGS على S ومكونات الثلاثة الأخرى من ASVABC، ASVABC2 له معلمة عالية القيمة ، ولكن المتغير ASVABC3 ليس كذلك والمعامل لـ ASVABC4 إشارته خطأ. هذه ليست مدهشة لأن معاملات الارتباط مع ASVABC شديدة.

```
. cor ASVAB2 ASVAB3 ASVAB4
(obs=570)

        |   ASVAB2   ASVAB3   ASVAB4
--------+---------------------------
  ASVAB2|   1.0000
  ASVAB3|   0.6916   1.0000
  ASVAB4|   0.6536   0.7628   1.0000
```

هنا نقارن بين الانحدار مع ASVABC و مع عناصره (مكوناته). الخطأ المعياري يكون صغيراً مقارنة معهم في معادلة الانحدار الثانية ، كما هو متوقع.

```
. reg EARNINGS S ASVABC

------------------------------------------------------------------------------
EARNINGS |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
       S |  .7390366   .1606216      4.601   0.000      .4235506    1.054523
  ASVABC |  .1545341   .0429486      3.598   0.000      .0701764    .2388918
   _cons | -4.624749   2.0132       -2.297   0.022     -8.578989   -.6705095
------------------------------------------------------------------------------
```

```
. reg EARNINGS S ASVAB2 ASVAB3 ASVAB4

------------------------------------------------------------------------------
EARNINGS |   Coef.   Std. Err.       t     P>|t|      [95% Conf. Interval]
---------+--------------------------------------------------------------------
       S | .7362439   .1586812     4.640   0.000      .4245668    1.047921
  ASVAB2 | .2472668   .0472249     5.236   0.000      .154509     .3400246
  ASVAB3 | .0137422    .058716     0.234   0.815     -.1015861    .1290705
  ASVAB4 |-.1051868   .0544682    -1.931   0.054     -.2121716    .001798
   _cons |-4.734303   2.06706     -2.290   0.022     -8.794363   -.6742428
------------------------------------------------------------------------------
```

قيمة t الإحصائية لـــ ASVABC صغيرة مقارنـــة مـــع ASVAB2 ، ولكن معامل انحدارها أصغر. بالطبع تستطيع المجادلة بأن ASVAB3 لاتنتمي إلى النموذج. لهذا الاقتراح إمكانية حل أخرى لمشكلة تعدد العلاقات الخطية.

6- إسقاط بعض المتغيرات المرتبطة : نلجأ إلى ذلك إذا كانت معاملاتها غير معنوية. في نموذج الانحدار السابق يمكن إسقاط المتغيـــرات *ASVAB3* و *ASVAB4* .

```
. reg EARNINGS S ASVAB2

  Source |       SS       df       MS                  Number of obs =     570
---------+------------------------------               F(  2,   567) =   48.81
   Model |  5639.37111     2  2819.68556               Prob > F      =  0.0000
Residual |   32757.666   567  57.7736613               R-squared     =  0.1469
---------+------------------------------               Adj R-squared =  0.1439
   Total |  38397.0371   569  67.4816117               Root MSE      =  7.6009

------------------------------------------------------------------------------
EARNINGS |   Coef.   Std. Err.       t     P>|t|      [95% Conf. Interval]
---------+--------------------------------------------------------------------
       S | .6449415   .1519755     4.244   0.000      .3464378    .9434452
  ASVAB2 | .2019724   .0376567     5.364   0.000      .1280086    .2759361
   _cons |-5.796398  1.957987    -2.960   0.003     -9.642191   -1.950605
------------------------------------------------------------------------------
```

كما هو متوقع الخطأ المعياري صغير مقارنة مع الانحـــدار المتضـــمن ASVAB3 و ASVAB4 . على أي حال هذا يضاهي تعدد العلاقات الخطية؛ لكونه خطر بسبب إسقاط بعض المتغيرات الأمر؛ الذي يجعل النمـــوذج أكثـــر مصداقية .

7- محدد غير حقيقي **Empirical restriction :**

$$Y = \beta_0 + \beta_1 X + \beta_2 P + u$$

على سبيل المثال: نفترض أن Y معادلة حول الطلب لفئة من المستهلكين على الإنفاق، X حاصل جمع الدخل التصرفي و P الرقم القياسي للأسعار لكل فئة؛ لملائمة النموذج يجب استخدام سلاسل زمنية لبيانات إذا كان X و P على علاقة ارتباط عالية يمكن إزالة ذلك بالطرق التالية:

نحصل على بيانات الدخل والإنفاق والفئات من مسح دخل ونفقات الأسرة، و نبني انحدار Y' على X'. (العلامة ' تعني أن بيانات المسح الأسري غير مجمعة).

$$Y' = \beta_0' + \beta_1' X' + u$$

$$\hat{Y}' = b_0' + b_1' X'$$

هذه علاقة انحدار بسيط؛ لأنها ستكون قليلة نسبياً في السعر المدفوع في الأسرة .

نستبدل b'_1 بقيمتها b_1 في سلسلة النموذج. نطرح b'_1 X من كلا الحدين، ونرجع العلاقة: $Z = Y - b'_1 X$ على السعر. نحصل على علاقة انحدار بسيط، بكذلك أزيل التعدد الخطي العلاقة.

$$Y = \beta_0 + b_1' X + \beta_1 P + u$$

$$Z = Y - b_1' X = \beta_0 + \beta_1 P + u$$

قد نواجه بعض المشاكل التقنية . مثل:

أولاً: المعاملان b_1 قد يكونان مختلفين مفهومياً بمرور الوقت سياقات المقطع العرضي والسلسلة.

ثانياً: سوف نطرح مكون الدخل المقدر $b'_1 X$ وليس $b_1 X$ من الدخل الحقيقي Y عندما نبني Z سوف نقدم عنصراً لقياس الخطأ في المتغير التابع.

8- قيد نظري Theoretical restriction:

حسب المتوسطات الأقل يستخدم قيداً نظرياً ، الذي يحدد كعلاقة فرضية بين المعلمات في نموذج الانحدار. لشرح ذلك نستخدم نموذج مستوى التعليم على فرض أنه أعطينا درجة عالية لمن أكمل S ، بالاعتماد على ASVABC ودرجة عالية لمن أنهى تعليمه من الأب و الأم ، SM و SF على التوالي.

$$S = \beta_0 + \beta_1 ASVABC + \beta_2 SM + \beta_3 SF + u$$

```
. reg S ASVABC SM SF

  Source |       SS       df       MS              Number of obs =     570
---------+------------------------------           F(  3,   566) =  110.83
   Model |  1278.24153     3  426.080508           Prob > F      =  0.0000
Residual |  2176.00584   566  3.84453329           R-squared     =  0.3700
---------+------------------------------           Adj R-squared =  0.3667
   Total |  3454.24737   569  6.07073351           Root MSE      =  1.9607

------------------------------------------------------------------------------
       S |     Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  ASVABC |.1295006   .0099544     13.009   0.000       .1099486     .1490527
      SM |.069403    .0422974      1.641   0.101      -.013676      .152482
      SF |.1102684   .0311948      3.535   0.000       .0489967     .1715401
   _cons |4.914654   .5063527      9.706   0.000       3.920094     5.909214
------------------------------------------------------------------------------
```

التفسير حسب جدول النتائج أعلاه إذا زاد المتغير ASVABC بمستوى واحد سيزداد S بمقدار 0.13 سنة. تزداد S بمقدار 0.07 لكل سنة إضافية في المدرسة بمساعدة الأم و 0.11 سنة بمساعدة الأب لكل سنة إضافية في المدرسة.

نلاحظ أن تعليم الأم ليس أكثر، أهمية من تعليم الأب من أجل بلوغ التعليم، و أتى ذلك على شكل غير متوقع. و الأكثر دهشة معامل SM ليس معنوي من أجل 5% مستوى دلالة والاختبار من اتجاه واحد. مهما كان التصنيف هذا يقود إلى وجود علاقة بين SM و SF ، لذلك الانحدار يعاني من تعدد العلاقة الخطية، دليل ذلك ارتفاع معامل الارتباط بين المتغيرين.

```
. cor SM SF
(obs=570)
   |   SM       SF
---------+------------------
SM|   1.0000
SF|   0.6391   1.0000
```

على فرض أن تعليم الأب والأم متساو بالأهمية . نستطيع فرض قيد $\beta_3 = \beta_2$. هذا يسمح لنا إعادة كتابة المعادلة النحو الآتي:

$$S = \beta_0 + \beta_1 ASVABC + \beta_2 (SM + SF) + u$$

نحدد المتغير SP بحيث يكون عبارة عن مجموع SM و SFكما هو مبين في إعادة كتابة العلاقة:

$$S = \beta_0 + \beta_1 ASVABC + \beta_2 SP + u$$

السبب في المشكلة ارتفاع قيمة معامل الارتباط بين المتغيرين SM و SF تم الحد منها. نجد أن قيمة معامل الانحدار b_2 يساوي 0.094 . وهو معنوي إحصائياً .

```
. g SP=SM+SF

. reg S ASVABC SP

  Source |       SS       df       MS              Number of obs =     570
---------+------------------------------           F(  2,   567) =  166.22
   Model |  1276.73764     2  638.368819           Prob > F      =  0.0000
Residual |  2177.50973   567  3.84040517           R-squared     =  0.3696
---------+------------------------------           Adj R-squared =  0.3674
   Total |  3454.24737   569  6.07073351           Root MSE      =  1.9597

------------------------------------------------------------------------------
       S |   Coef.   Std. Err.      t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  ASVABC |.1295653   .0099405    13.024   0.000       .1100249    .1491057
      SP |.093741    .0165688     5.658   0.000       .0611973    .1262847
   _cons |4.823123   .4844829     9.955   0.000       3.871523    5.774724
------------------------------------------------------------------------------
```

الخطأ المعياري لـ SP صغير جداً مقارنة مع المتغيرين SM و SF .
تستخدم هذه العلاقة للحد من المعيقات الكبيرة في عدم الكفاية و من أجل إزالة

مشكلة تعدد العلاقات الخطية . ونجد أن قيمة t الإحصائية عالية المعنوية. وهذا سيؤدي إلى معنوية نتائج الانحدار. على كل حال القيد لن يكون قانونياً. سوف نختبر ذلك فيما بعد.

10-4- مواصفات المتغيرات (إغفال علاقة المتغير OMISSION OF A RELEVANT VARIABLE):

في هذه الفقرة سنحاول البحث عن نتيجة محددة في نموذج الانحدار في حدود المتغيرات المستقلة.

للتحليل المبسط سوف نفترض وجود إمكانيتين. إدخال Y يعتمد على X_1 أو X_1 و X_2 . إذا كانت Y تعتمد على X_1 و سوف نوفق نموذج انحدار بسيط، لن نصطدم بأي مشكلة مع شروط غاوس-ماركوف.

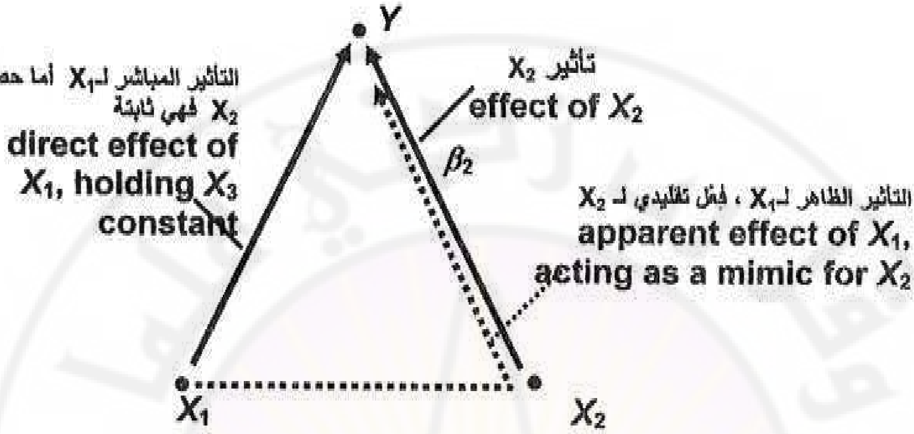
بطريقة مماثلة لن نصطدم بأي مشكلة إذا كانت Y تعتمد على كل من X_1 و X_2 و سوف نوفق انحدار متعدد.

سنحاول فحص سلسلة التوفيقات لنموذج الانحدار البسيط عندما يكون النموذج الحقيقي متعدداً. وفي الحالة اللاحقة سنأخذ العكس، ونفحص سلسلة التوفيقات النموذج الانحدار المتعدد عندما يكون نموذج الحقيقي بسيطاً.

إن إغفال علاقة المتغير المستقل يؤدي إلى معاملات انحدار مبنية على أساس متحيز و الخطأ المعياري يكون معتلاً . هنا إغفال X_2 يؤدي إلى b_1 مبني على قيمة $\beta_2 \, \mathrm{Cov}(X_1, X_2)/\mathrm{Var}(X_1)$. سوف نعرض هذه الحالة أولاً بشكل استنتاجي ومن ثم رياضياً.

$$E(b_1) = \beta_1 + \beta_2 \frac{\mathrm{Cov}(X_1, X_2)}{\mathrm{Var}(X_1)}$$

السبب البديهي يبين أن الإضافات تؤدي إلى التأثير في b_1 حيث أن الظهور يكون غير مباشر على X_1 نتيجة فعل فقدان المتغير X_2.



الشكل(10-2)

قوة تأثير بالنيابة تعتمد على عاملين : قوة تأثير X_2 على Y ، التي تعطى بالمعلمة b_2 ، و قدرة X_1 على محكاة X_2 .

قدرة المتغير X_1 على محكاة X_2 يتحدد بمعامل الميل الحاصل عندما X_3 يكون منحدر على X_1 التي خطؤها هو $\mathrm{Cov}(X_1, X_2)/\mathrm{Var}(X_1)$.

لنشتق هذه العلاقة رياضياً، قد نقع بخطأ أثناء توفيق نموذج الانحدار البسيط ، نحسب معامل الميل من العلاقة:

$\mathrm{Cov}(X_1, Y)/\mathrm{Var}(X_1)$. الخطوة الأولى نستبدل قيمة Y من النموذج الحقيقي.

$$b_1 = \frac{\mathrm{Cov}(X_1, Y)}{\mathrm{Var}(X_1)} = \frac{\mathrm{Cov}(X_1, [\beta_0 + \beta_1 X_1 + \beta_2 X_2 + u])}{\mathrm{Var}(X_1)}$$

سوف نستخدم علاقة التباين المشترك الأولى في البسط.

$$b_1 = \frac{\text{Cov}(X_1, \beta_0) + \text{Cov}(X_1, \beta_1 X_1) + \text{Cov}(X_1, \beta_2 X_2) + \text{Cov}(X_1, u)}{\text{Var}(X_1)}$$

الحد الأول يساوي الصفر لأن b_0 ثابتة .يخرج b_1 و b_2 خارج قوس في الحد الثاني والثالث. وبإجراء الإصلاحات الرياضية والاختصارات نحصل على:

$$b_1 = \frac{0 + \beta_1 \text{Cov}(X_1, X_1) + \beta_2 \text{Cov}(X_1, X_2) + \text{Cov}(X_1, u)}{\text{Var}(X_1)}$$

$$b_1 = \beta_1 + \beta_2 \frac{\text{Cov}(X_1, X_2)}{\text{Var}(X_1)} + \frac{\text{Cov}(X_1, u)}{\text{Var}(X_1)}$$

للتأكد من تحيز أو عدم تحيز هذه العلاقة نأخذ توقع قيمة b_2 .

$$b_1 = \frac{\text{Cov}(X_1, Y)}{\text{Var}(X_1)} = \beta_0 + \beta_1 \frac{\text{Cov}(X_1, X_2)}{\text{Var}(X_1)} + \frac{\text{Cov}(X_1, u)}{\text{Var}(X_1)}$$

$$E(b_1) = E\left\{\beta_1 + \beta_2 \frac{\text{Cov}(X_1, X_2)}{\text{Var}(X_1)} + \frac{\text{Cov}(X_1, u)}{\text{Var}(X_1)}\right\}$$

باستخدام قاعدة التوقع الأولى، نستطيع تحليل قيمة التوقع كمجموع توقع مركبات عدة قيم.

$$E(b_1) = E(\beta_1) + E\left\{\beta_2 \frac{\text{Cov}(X_1, X_2)}{\text{Var}(X_1)}\right\} + E\left\{\frac{\text{Cov}(X_1, u)}{\text{Var}(X_1)}\right\}$$

إن المركبتان الأوليتان مثبتتان عند b_1 و b_2 وهما ثوابت و X_2, X_1 من المفروض متجانستي التباين. نلاحظ أن القيمة المتوقعة للحد الثالث تساوي الصفر . (لبرهان ذلك انظر الفصل الثالث).

$$E(b_1) = \beta_1 + \beta_2 \frac{\text{Cov}(X_1, X_2)}{\text{Var}(X_1)}$$

وهكذا تكون b_3 تقديراً غير متحيز لقيمة $\beta_2 \,\mathrm{Cov}(X_1, X_2)/\mathrm{Var}(X_1)$.
وبالنتيجة فإن مواصفات الأخطاء المعيارية اختبارات t و اختبار F تكون باطلة.

سوف نوضح ذلك بمثال لتحيز باستخدام نموذج مستوى التعليم المتحقق ، بالاحتفاظ بالتحليل البسيط ، سوف نستخدم لهذا الافتراض S على ASVABC و SM . المخرجات تبين تتطابق الانحدار باستخدام EAEF بيانات المجموع 21 .

```
. reg S ASVABC SM

  Source |       SS       df       MS              Number of obs =     570
---------+------------------------------           F(  2,   567) =  156.81
   Model |   1230.2039     2  615.101949           Prob > F      =  0.0000
Residual |  2224.04347   567  3.92247526           R-squared     =  0.3561
---------+------------------------------           Adj R-squared =  0.3539
   Total |  3454.24737   569  6.07073351           Root MSE      =  1.9805

------------------------------------------------------------------------------
       S |      Coef.   Std. Err.      t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  ASVABC |   .1381062   .0097494   14.166    0.000      .1189567    .1572556
      SM |    .154783   .0350728    4.413    0.000      .0858946    .2236715
   _cons |   4.791277   .5102431    9.390    0.000       3.78908    5.793475
------------------------------------------------------------------------------
```

سوف ننفذ الانحدار مرة ثانية بحذف المتغير SM . قبل عمل ذلك سوف نحاول التنبؤ باتجاه التحيز للمعامل ASVABC .

$$S = \beta_0 + \beta_1 ASVABC + \beta_2 SM + u$$

$$E(b_1) = \beta_1 + \beta_2 \frac{\mathrm{Cov}(ASVABC, SM)}{\mathrm{Var}(ASVABC)}$$

يكون ذلك الافتراض معقولا لأهمية الحساسية الشائعة لأن β_2 حقيقية .
هذا الافتراض يدعم بقوة حقيقة تقديره في معاملات الانحدار الحقيقية والمعنوية.

الارتباط بين ASVABC و SM حقيقي، وكذلك تباينهم المشترك يجب أن يكون واقعياً . المتغير $\mathrm{Var}(ASVABC)$ يكون بشكل أتوماتيكي موجباً.
وهكذا التحيز يجب أن يكون حقيقياً.

```
. cor SM ASVABC
(obs=570)

         |       SM   ASVABC
---------+------------------
      SM|  1.0000
  ASVABC|  0.3819   1.0000
```

كما نرى معامل ASVABC يكون فعلاً مرتفعاً عند استبعاد SM . إن جزءاً من الاختلاف يكون عائداً لقوى الحظ والمصادفة، ولكن الجزء الآخر عائد للتحيز .

```
. reg S ASVABC SM
------------------------------------------------------------------------------
       S |    Coef.   Std. Err.      t     P>|t|      [95% Conf. Interval]
---------+--------------------------------------------------------------------
  ASVABC |.1381062    .0097494   14.166    0.000      .1189567     .1572556
      SM | .154783    .0350728    4.413    0.000      .0858946     .2236715
   _cons |4.791277    .5102431    9.390    0.000      3.78908      5.793475
------------------------------------------------------------------------------
. reg S ASVABC
------------------------------------------------------------------------------
       S |    Coef.   Std. Err.      t     P>|t|      [95% Conf. Interval]
---------+--------------------------------------------------------------------
  ASVABC |.1545378    .0091559   16.879    0.000      .1365543     .1725213
   _cons |5.770845    .4668473   12.361    0.000      4.853888     6.687803
------------------------------------------------------------------------------
```

سوف نتوقع أن يكون تحيز b_2 أعلى في نموذج الانحدار بعد استبعاد ASABC بدلاً من SM ، . سوف نفترض أن β_1 ستكون حقيقية، و أن حدود التباين المشترك والتباين في علاقة التحيز هي حدود حقيقية.

في هذه الحالة يكون التحيز ملحوظاً ، لأن معامل SM قيمته أكثر من الضعف. (السبب في ذلك الأثر الكبير للمتغير Var(SM) يكون أكثر من الأثر الصغير للمتغير Var(ASVABC), بينما b_1 و b_2 يكونان صغيرين في الحجم لذلك نحكم بالتقدير .

وأخيراً سوف نبحث عن سلوك R^2 عندما نستبعد المتغير . في الانحدار البسيط لـــ S على AVABC R^2 بلغت 0.33 و في الانحدار البسيط لـــ S على SM بلغ 0.13 .

```
. reg S ASVABC SM
  Source |       SS       df       MS              Number of obs =     570
---------+------------------------------           F(  2,   567) =  156.81
   Model |  1230.2039      2  615.101949           Prob > F      =  0.0000
Residual |  2224.04347   567  3.92247526           R-squared     =  0.3561
---------+------------------------------           Adj R-squared =  0.3539
   Total |  3454.24737   569  6.07073351           Root MSE      =  1.9805

. reg S ASVABC
  Source |       SS       df       MS              Number of obs =     570
---------+------------------------------           F(  1,   568) =  284.89
   Model |  1153.80864     1  1153.80864           Prob > F      =  0.0000
Residual |  2300.43873   568  4.05006818           R-squared     =  0.3340
---------+------------------------------           Adj R-squared =  0.3329
   Total |  3454.24737   569  6.07073351           Root MSE      =  2.0125

. reg S SM
  Source |       SS       df       MS              Number of obs =     570
---------+------------------------------           F(  1,   568) =   83.59
   Model |  443.110436     1  443.110436           Prob > F      =  0.0000
Residual |  3011.13693   568  5.30129742           R-squared     =  0.1283
---------+------------------------------           Adj R-squared =  0.1267
   Total |  3454.24737   569  6.07073351           Root MSE      =  2.3025
```

وهل هذا يعني أن ASVABC تفسر 33% من التباين S و 13% من SM ؟ الجواب بالطبع لا، لأن الانحدار المتعددَ يَكْشفُ بأنّ قوّتَهم التوضيحيةَ المشتركةَ 0.36، ولَيسَ 0.46 .

في الانحدار الثاني ASVABC عملياً جزء من نيابة هذا المتغير عائدة لـ SM ، والذي بدوره يضخم قوة المتغير التفسيرية . وبالمثل في الحالة الثالثة SM يكون جزء من نيابة المتغير ASVABC ومرة ثانية هذا يضخم ظاهر القوة التفسيرية.

على أية حال، من المحتمل أن يكون لتحيز المتغير المَحْذوف أثر على تخفيض في القوّةِ التوضيحيةِ الظاهرةِ في متغيّر. في هذه الحالة سوف نستخدم نموذج معادلة الكسب البسيط؛ مفترضين لوغاريتم الكسب يعتمد على S و MALE .

```
. reg LGEARN S MALE

  Source |       SS       df       MS              Number of obs =     570
---------+------------------------------           F(  2,   567) =   65.74
   Model |  28.951332      2  14.475666            Prob > F      =  0.0000
Residual | 124.850561    567  .220194992           R-squared     =  0.1882
---------+------------------------------           Adj R-squared =  0.1854
   Total | 153.801893    569  .270302096           Root MSE      =  .46925

------------------------------------------------------------------------------
  LGEARN |     Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
       S |.0818944    .0079976    10.240    0.000       .0661858     .097603
    MALE |.2285156    .0397695     5.746    0.000       .1504021    .3066291
   _cons | 1.19254    .1134845    10.508    0.000       .9696386    1.415441
------------------------------------------------------------------------------
```

وإذا استبعدنا MALE من الانحدار فمعامل S يجب أن يكون اقل تحيزاً. و β_3 يكون مشابهاً للحقيقي، و نستطيع أن نعرف أن المتغير $\mathrm{Var}(S)$ إيجابي ولكن التباين المشترك يكون سالباً لأن الارتباط بين S و MALE سالب.

$$LGEARN = \beta_0 + \beta_1 S + \beta_2 MALE + u$$

$$E(b_1) = \beta_1 + \beta_2 \frac{\mathrm{Cov}(S, MALE)}{\mathrm{Var}(S)}$$

```
. cor S MALE
(obs=570)
        |        S      MALE
--------+------------------
       S|   1.0000
    MALE|  -0.0581    1.0000
```

من أجل السبب نفسه معامل MALE في الانحدار البسيط لـ LGEARN على MALE يجب أن يكون التحيز منخفضاً. نرى أن المعاملين لـ S و MALE منخفضين في الانحدار البسيط.

```
. reg LGEARN S MALE
------------------------------------------------------------------------------
  LGEARN |  Coef.   Std. Err.    t      P>|t|      [95% Conf. Interval]
---------+--------------------------------------------------------------------
       S | .0818944   .0079976  10.240   0.000    .0661858        .097603
    MALE |.2285156    .0397695   5.746   0.000    .1504021       .3066291
   _cons | 1.19254    .1134845  10.508   0.000    .9696386       1.415441

. reg LGEARN S
------------------------------------------------------------------------------
  LGEARN |       Coef.   Std. Err.        t      P>|t|        [95% Conf.
Interval]
---------+--------------------------------------------------------------------
       S |.0792256    .0082061    9.655   0.000    .0631077       .0953435
   _cons | 1.358919    .1127785   12.049   0.000  1.137406        1.580433

. reg LGEARN MALE
------------------------------------------------------------------------------
  LGEARN |    Coef.    Std. Err.   t       P>|t|      [95% Conf. Interval]
---------+--------------------------------------------------------------------
    MALE | .2048652    .0431797  4.744   0.000     .1200538       .2896767
   _cons | 2.313324     .032605 70.950   0.000     2.249282       2.377365
```

من مقارنة R^2 في العلاقات الانحدار الثلاثة المعروضة نجد أن R^2 في الانحدار البسيط فعلياً اقل من R^2 في الانحدار المتعدد . على كل حال في هذه الحالة إن الاختلاف صغير بسبب الارتباط الضعيف والسالب بين S و MALE صغير حيث يساوي ‎-0.05 .

```
. reg LGEARN S MALE
  Source |       SS         df        MS            Number of obs =     570
---------+------------------------------            F(  2,    567) =   65.74
   Model |   28.951332       2   14.475666          Prob > F       =  0.0000
Residual |  124.850561     567  .220194992          R-squared      =  0.1882
---------+------------------------------            Adj R-squared  =  0.1854
   Total |  153.801893     569  .270302096          Root MSE       =  .46925

. reg LGEARN S
  Source |       SS         df        MS            Number of obs =     570
---------+------------------------------            F(  1,    568) =   93.21
   Model |   21.681253       1   21.681253          Prob > F       =  0.0000
Residual |   132.12064     568   .23260676          R-squared      =  0.1410
---------+------------------------------            Adj R-squared  =  0.1395
   Total |  153.801893     569  .270302096          Root MSE       =  .48229

. reg LGEARN MALE
  Source |       SS         df        MS          Number of obs =     570
---------+------------------------------          F(  1,    568) =   22.51
   Model |  5.86288165       1  5.86288165        Prob > F       =  0.0000
Residual |  147.939011     568  .260456005        R-squared      =  0.0381
---------+------------------------------          Adj R-squared  =  0.0364
   Total |  153.801893     569  .270302096        Root MSE       =  .51035
```

لنعود إلى تفنيد أهمية مواصفات اسم المتغير وسوف نبحث عن نتيجة إدخال متغير غير متضمن في علاقة الانحدار البسيط.

أهمية مواصفات اسم المتغير

Consequences of Variable Misspecification

		$Y = \beta_0 + \beta_1 X_1 + u$	$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$
Fitted Model النموذج المقدر	$\hat{Y} = b_0 + b_1 X_1$	Correct specification ,no problems مواصفات صحيحة ولا يوجد مشاكل	Coefficients are biased (in general). Standard errors are invalid. المعاملات متحيز (عموما)، معيار الأخطاء غير صحيحة.
	$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2$	Coefficients are unbiased (in general), but inefficient. Standard errors are valid (in general) المعلمات غير متحيز (عموما)، لكن غير كفء، الأخطاء المعيارية صحيحة (عموما)	Correct specification, no problems مواصفات صحيحة ولا يوجد مشاكل

هذه التقنية يمكن عرضها بشكل سريع .

$$Y = \beta_0 + \beta_1 X_1 + u$$

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2$$

يمكن إعادة كتابة النموذج الصحيح مضاف إليه X_2 كمتغير تفسيري، بمعامل يساوي الصفر. نلاحظ أنه قد تطابق النموذج الحقيقي مع النموذج المقدر. وهكذا يعتبر b_1 تقديراً غير متحيز لـــ β_1 و b_2 ستكون تقديراً غير متحيز عن الصفر .

وعلى كل حال فإن تباين المجتمع لـــ b_1 سيكون أكبر منه إذا تم تصحيح الانحدار البسيط الذي تم تنفيذه؛ لأنه يتضمن الحد $1/\sqrt{(1 - r^2_{X_1 X_2})}$

$$Y = \beta_0 + \beta_1 X_1 + 0X_2 + u$$

$$\textbf{population variance of } b_1 = \sigma^2_{b_1} = \frac{\sigma^2_u}{n\text{Var}(X_1)} \times \frac{1}{1 - r^2_{X_1,X_2}}$$

وتعتبر كفاءة التقدير b_1 باستخدام نموذج الانحدار المتعدد أفضل من استخدام نموذج الانحدار البسيط.

إن السبب البديهي لأن نموذج الانحدار البسيط يستفيد من المعلومات ولأن X_2 يجب ألا يكون في الانحدار ، بينما في نموذج الانحدار المتعدد نجد هذا خارج نتائج الانحدار . و تَبقى الأخطاء المعيارية صحيحةً، لأن النموذجَ يُحدَدُ رسمياً بشكل صحيح، لَكنْها ستَميل إلى أن تكون أكبر منْها في المحصلة في الانحدار البسيط، الذي يَعْكسُ خسارةَ الكفاءة.

هذه النَتائج بشكل عام. مع مُلاحظة أنه إذا صادف، وكان X_1 وX_2 غيرَ مرتبطين فيما بينها، لَنْ يكون هناك خسارة في الكفاءة تذكر.

من أجل التحليل سنستخدم أساس معادلة نصف اللوغاريتمي للكسب، أي انحدار LGEARN على S و ASVABC كما هو مبين أدناه:

```
. reg LGEARN S ASVABC

  Source |       SS       df       MS              Number of obs =     570
---------+------------------------------           F(  2,   567) =   57.45
   Model |  25.9166749     2  12.9583374           Prob > F      =  0.0000
Residual |  127.885218   567  .225547121           R-squared     =  0.1685
---------+------------------------------           Adj R-squared =  0.1656
   Total |  153.801893   569  .270302096           Root MSE      =  .47492

------------------------------------------------------------------------------
  LGEARN |     Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
       S |.0544266     .0099018     5.497    0.000       .034978     .0738753
  ASVABC |.0114733     .0026476     4.333    0.000       .0062729    .0166736
   _cons |1.118832      .124107     9.015    0.000       .8750665    1.362598
------------------------------------------------------------------------------
```

وسنقوم الآن بإضافة متغيرات تعليم الأهل SM,SF . وهذان المتغيران يحددان اثر التعليم على الدخل بشكل غير مباشر ، ولكن لن يكون هناك اثر دليل؛ مباشر على الدخل ، بسبب وجود الإضافات.

```
. reg LGEARN S ASVABC SM SF

  Source |       SS       df       MS              Number of obs =     570
---------+------------------------------           F(  4,   565) =   29.22
   Model |  26.3617806     4  6.59044515           Prob > F      =  0.0000
Residual |  127.440112   565  .22555772            R-squared     =  0.1714
---------+------------------------------           Adj R-squared =  0.1655
   Total |  153.801893   569  .270302096           Root MSE      =  .47493

------------------------------------------------------------------------------
  LGEARN |      Coef.   Std. Err.  t       P>|t|     [95% Conf. Interval]
---------+--------------------------------------------------------------------
       S | .0511811    .0101812    5.027    0.000     .0311835     .0711788
  ASVABC | .010444     .0027481    3.800    0.000     .0050463     .0158417
      SM | .0071835    .0102695    0.699    0.485    -.0129876     .0273547
      SF | .004794     .0076389    0.628    0.531    -.0102101     .0197981
   _cons |1.073972     .1324621    8.108    0.000     .8137933     1.33415
------------------------------------------------------------------------------
```

إن الاختبار الإحصائي يبين قيمة t الإحصائية لكلا المتغيرين وستكون دليلاً ضعيفاً؛ لأنه سيكون احتمالاً غير موضوعي.

وفي هذه المسألة ليس هناك دليل على أن مضمون تعليم الأهل سيجعل معاملات أخرى متحيزة . ستتغير المعاملات الأخرى ، ولكن التغيرات ستكون صغيرة في علاقة الأخطاء المعيارية بسبب قوى الحظ والمصادفة .

```
. reg LGEARN S ASVABC

------------------------------------------------------------------------------
  LGEARN |    Coef.    Std. Err.      t      P>|t|     [95% Conf. Interval]
---------+--------------------------------------------------------------------
       S | .0544266    .0099018     5.497    0.000     .034978     .0738753
  ASVABC | .0114733    .0026476     4.333    0.000     .0062729    .0166736
   _cons | 1.118832    .124107      9.015    0.000     .8750665    1.362598
------------------------------------------------------------------------------

. reg LGEARN S ASVABC SM SF

------------------------------------------------------------------------------
  LGEARN |    Coef.    Std. Err.      t      P>|t|     [95% Conf. Interval]
---------+--------------------------------------------------------------------
       S | .0511811    .0101812     5.027    0.000     .0311835     .0711788
  ASVABC | .010444     .0027481     3.800    0.000     .0050463     .0158417
      SM | .0071835    .0102695     0.699    0.485    -.0129876     .0273547
      SF | .004794     .0076389     0.628    0.531    -.0102101     .0197981
   _cons | 1.073972    .1324621     8.108    0.000     .8137933     1.33415
------------------------------------------------------------------------------
```

إنّ الأخطاء المعيارية أكبر في مواصفات النموذج ، وهذا يَعكسُ خسارة الكفاءةِ. على أية حال، خسارة الكفاءةِ لَيستْ كبيرة جداً. بما أنه يوجد علاقة ارتباط قوية بين متغيّر تعليم الأهل من S وAsvabc وهذا سيجعل العلاقة أقل الكفاءة، و تظهر مشكلة خطيرة ألا وهي مشكلة التعدد الخطي.

```
. cor S ASVABC SM SF
(obs=570)

        |        S   ASVABC       SM       SF
--------+------------------------------------
       S|   1.0000
  ASVABC|   0.5779   1.0000
      SM|   0.3582   0.3819   1.0000
      SF|   0.4066   0.4179   0.6391   1.0000
```

5-10- متغيرات النيابة (PROXY VARIABLES) :

على فرض أن المتغير Y يكون افتراضياً يعتمد على مجموعة متغيرات تفسيرية $X_1, \ldots, X_k$ كما هو مبين و نفترض أن لبعض الأسباب لا يوجد بيانات عن المتغير X_1 .

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + u$$

سنشاهد علاقة انحدار Y على $X_1, \ldots, X_k$ خاضع لتحيز تقديرات المعلمات و يعاني من الأخطاء المعيارية للاختبارات.

إن هذه المشاكل في بعض الأحيان تنخفض أو تزول باستخدام المتغير النائب في مكان X_1 . لأن أثر المتغير النائب يكون واحداً على العلاقة الخطية للمتغيرات المفقودة . في هذه الحالة على سبيل المثال نطلق عليه Z ، يعمل عمل المتغير الأصلي X_1 .

$$X_1 = \lambda + \mu Z$$

لتكون علاقة النائب علاقة قانونية يجب أن تكون مبررة على أساس النظرية ، أو الإحساس، أو التجربة . ولا يمكن التأكد بشكل مباشر، وذلك بسبب عدم توفر البيانات عن X_1 .

إذا كان من المناسب أن يحدد النائب نموذج الانحدار نستطيع إعادة كتابته كما هو مبين.

$$Y = \beta_0 + \beta_1(\lambda + \mu Z) + \beta_2 X_2 + ... + \beta_k X_k + u$$
$$= (\beta_0 + \beta_1\lambda) + \beta_1\mu Z + \beta_2 X_2 + ... + \beta_k X_k + u$$

وهكذا نحصل على نموذج مع كل المتغيرات الجديرة بالملاحظة . إذا كانت علاقة النيابة واحدة مضبوطة، سنحصل على نتائج علاقة انحدار جيدة لأن:

1- تقديرات المعاملات لـ $X_2, ..., X_k$ ستكون نفسها إذا حصلنا على انحدار Y على $X_2, ..., X_k$.

2- الأخطاء المعيارية و t الإحصائية لمعلمات لـ $X_2, ..., X_k$ سوف تكون نفسها إذا حصلنا على انحدار Y على $X_2, ..., X_k$.

3- R^2 سيكون نفسه إذا حصلنا على انحدار Y على $X_2, ..., X_k$.

4- معامل المتغير Z سيكون تقديراً لـ $b_1\mu$ و كذلك لا يكون ممكن الحصول على تقدير b_1 ما لم نحدد قيمة μ .

5- على كل حال إن قيمة t الإحصائية لـ Z ستكون نفسها التي نحصل عليها من أجل X_1 إذا استطعنا أن نوفق انحدار Y على $X_1, ..., X_k$، و كذلك إذا استطعنا تخمين معنوية المتغير X_1 وحتى إذا لم نقدر معنوية ذلك.

عادة (b_0) تكون نسبيا قليلة الأهمية ' لذلك من الممكن أن لا نحصل على تقدير الميل b_0 في النموذج الجديد والذي قيمته تساوي إلى $(b_0+b_1\lambda)$.

بشكل عام كثير من الافتراضات ستكون حقيقية لأن العلاقة بين المتغيرين X_1 و Z تقريباً أكثر دقة. وفي هذه الحالة تكون النتائج المعروضة أعلاه تقريبية.

على أي حال إذا كان Z نائباً ضعيفاً لـ X_1 فإن النتائج ستكون سارية المفعول لموضوع قياس الخطأ (انظر الفصل التاسع). وما يعزز ذلك إمكانية بعض المتغيرات لـ X تحاول ان تنوب عن X_2 وستكون هناك مشكلة استبعاد المتغير المتحيز.

إن استخدام المتغير النائب سنوضحه من خلال مثال نموذج التعليم. سوف نفترض أن بلوغ التعليم يعتمد على مفصل مهارة الإدراك و الخلفية العائلية.

$$S = \beta_1 + \beta_2 ASVABC + \beta_3 INDEX + u$$

عادة ASVABC نستخدمه لقياس مهارة الإدراك. على أي حال لا يوجد متغير مستقل يعكس خلفية الأسرة في مجموعة البيانات. بالحقيقة من الصعب تخيل كيف يحدث مثل هذا المتغير؟ وكيف يمكن تحديده؟

لذلك سوف نحاول إيجاد نائب عنه. احد هذه البديهيات أنه من الممكن أن يكون المتغير المستوى التعليمي للأم SM أو المستوى التعليمي للأب SF. على فرض أن الرقم القياسي للخلفية العائلية يعتمد على كلاهما:

$$INDEX = \lambda + \mu_1 SM + \mu_2 SF$$

هكذا نحصل على علاقة معبرة S كعلاقة من ASVABC و SM و SF.

$$S = \beta_0 + \beta_1 ASVABC + \beta_2(\lambda + \mu_1 SM + \mu_2 SF) + u$$
$$= (\beta_0 + \beta_2\lambda) + \beta_1 ASVABC + \beta_2\mu_1 SM + \beta_2\mu_2 SF + u$$

لنستخدم البيانات المتوفرة لدينا في بناء نموذج لانحدار S على ASVABC SM SF سنحصل على:

```
. reg S ASVABC SM SF

  Source |       SS       df       MS              Number of obs =     570
---------+------------------------------           F(  3,   566) =  110.83
   Model |  1278.24153     3  426.080508           Prob > F      =  0.0000
Residual |  2176.00584   566  3.84453329           R-squared     =  0.3700
---------+------------------------------           Adj R-squared =  0.3667
   Total |  3454.24737   569  6.07073351           Root MSE      =  1.9607

------------------------------------------------------------------------------
       S |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
---------
  ASVABC |  .1295006    .0099544   13.009    0.000      .1099486    .1490527
      SM |  .069403     .0422974    1.641    0.101      -.013676     .152482
      SF | .1102684     .0311948    3.535    0.000      .0489967    .1715401
   _cons | 4.914654     .5063527    9.706    0.000      3.920094    5.909214
------------------------------------------------------------------------------
```

سنبدأ بالاستغناء عنها تبعاً واحداً تلو الآخر لمقارنة متغير النيابة :

```
. reg S ASVABC

  Source |       SS       df       MS              Number of obs =     570
---------+------------------------------           F(  1,   568) =  284.89
   Model |  1153.80864     1  1153.80864           Prob > F      =  0.0000
Residual |  2300.43873   568  4.05006818           R-squared     =  0.3340
---------+------------------------------           Adj R-squared =  0.3329
   Total |  3454.24737   569  6.07073351           Root MSE      =  2.0125

------------------------------------------------------------------------------
       S |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  ASVABC | .1545378     .0091559   16.879    0.000      .1365543    .1725213
   _cons |5.770845      .4668473   12.361    0.000      4.853888    6.687803
------------------------------------------------------------------------------
```

بمقارنة نتائج الانحدار أعلاه فإن معامل ASVABC يكون متحيزاً نحو الأعلى، إذا أغفلنا محاولة السيطرة على متغير خلفية العائلة. ماذا يعني هذا بالضبط؟ إن كل من المتغيرين SM و SF متشابهان في التأثير الايجابي على بلوغ التعليم، وكلاهما على ارتباط موجب مع المتغير ASVABC ، كما هو موضح أدناه:

```
. cor ASVABC SM SF
(obs=570)

         |  ASVABC       SM       SF
---------+---------------------------
  ASVABC |  1.0000
      SM |  0.3819   1.0000
      SF |  0.4179   0.6391   1.0000
```

لنأخذ عدة متغيرات أخرى لتكوين المتغير النائب لخلفية العائلة، وليكن متغير المكتبة *LIBRARY* (وهو متغير وهمي يعطي القيمة واحد إذا أحد أفراد الأسرة ملك بطاقة مكتبة عندما بلغ 14) و *SIBLINGS*(عدد الأخوة والأخوات للمستجوب).

```
. reg S ASVABC SM SF LIBRARY SIBLINGS

  Source |       SS       df       MS                 Number of obs =     570
---------+------------------------------             F(  5,   564) =   66.87
   Model |  1285.58208     5  257.116416             Prob > F      =  0.0000
Residual |  2168.66529   564  3.84515122             R-squared     =  0.3722
---------+------------------------------             Adj R-squared =  0.3666
   Total |  3454.24737   569  6.07073351             Root MSE      =  1.9609

------------------------------------------------------------------------------
       S |   Coef.    Std. Err.      t     P>|t|      [95% Conf. Interval]
---------+--------------------------------------------------------------------
  ASVABC |.1277852    .010054    12.710   0.000      .1080373     .147533
      SM |.0619975    .0427558    1.450   0.148     -.0219826    .1459775
      SF |.1045035    .0314928    3.318   0.001      .042646      .166361
 LIBRARY |.1151269    .1969844    0.584   0.559     -.2717856    .5020394
SIBLINGS |-.0509486   .039956    -1.275   0.203     -.1294293    .027532
   _cons |5.236995    .5665539    9.244   0.000      4.124181     6.349808
------------------------------------------------------------------------------
```

إن متغيّر *LIBRARY* كَان أحد المتغيرات الثلاثة التـــي تَضـــمنت فـــي المسحِ الطوليِ الوطنيِ للشابِ للمُسَاعَدَة على التقاط تأثيرِ الخلفيةِ العائليةِ علـــى التعليم. لَهُ معامل إيجابي مُتَوَقع، لَكنّه لَيس معنوياً.

ما الهدف والطموح عنـــد العائلـــة لتنشـــئة الأولاد، ولتحديـــد عـــددهم (SIBLINGS متغير عدد الأولاد) يجب أن يكون توقع الإشارة السالبة لمعامل هذا المتغير، و بالفعل تحقق ذلك بالإضافة إلى أنه متغير غير معنـــوي؛ لـــذلك

سوف نبحث عن متغيّرات أخرى لتكون خلفيةِ تعبر عن المستوى التربوي مثل: الإيمان، الانتماء العرق، ومنطقة السكنِ. هذه المتغيّرات مجهّزة في مجموعـــة المعلومات، لكن سيكون أمراً ميسراً للتّجْريب فيها.

6-10- اختبارات الحد من الخطية TESTING A LINEAR :RESTRICTION

في الفقرة السابقة كان هناك جدل حول علاقة بلوغ التعليم والمهارة والخلفية الأسرية للشخص، أي: حول نيابة متغير تعليم الأب أو الأم من خلال استخدام العلاقة التالي:

$$S = \beta_0 + \beta_1 ASVABC + \beta_2 SM + \beta_3 SF + u$$

حيث وجدنا في نتائج الانحدار على مثالنا السابق عدم معنوية متغير تعليم الأم.

```
. reg S ASVABC SM SF

  Source |       SS       df       MS              Number of obs =     570
---------+------------------------------           F(  3,   566) =  110.83
   Model |  1278.24153     3  426.080508           Prob > F      =  0.0000
Residual |  2176.00584   566  3.84453329           R-squared     =  0.3700
---------+------------------------------           Adj R-squared =  0.3667
   Total |  3454.24737   569  6.07073351           Root MSE      =  1.9607

------------------------------------------------------------------------------
       S |  Coef.   Std. Err.      t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  ASVABC |.1295006   .0099544   13.009   0.000       .1099486    .1490527
      SM | .069403   .0422974    1.641   0.101       -.013676     .152482
      SF |.1102684   .0311948    3.535   0.000       .0489967    .1715401
   _cons |4.914654   .5063527    9.706   0.000       3.920094    5.909214
------------------------------------------------------------------------------
```

كملاحظة على ما تم استعراضه سابقاً ، يمكن أن يكون ذلك عائداً للارتباط الخطي بسبب الارتباط العالي بين تعليم الأب وتعليم الأم :

```
. cor SM SF
(obs=570)
        | SM     SF
--------+------------------
     SM|  1.0000
     SF|  0.6391   1.0000
```

في مناقشة التعدد الخطي، يوجد عدة إجراءات لتَخفيف المشــكلة نقتــرح استخدام، تقييد نظري ملائم. بشكل خاص، في حالة النموذج الحالي، نفتــرض بأن تأثير تعليم الأهل قد يكون نفسه للأبويين، هذا يعني أن b_2 و b_3 قَد يَكونان متساويين.

$$S = \beta_0 + \beta_1 ASVABC + \beta_2 SM + \beta_3 SF + u$$

أي : $\beta_3 = \beta_2$

يمكننا إعادة كتابة النموذج في هذه الحالة كما هو مبــين، بعــد اســتبدال متغير تعليم الأهل الإجمالي SP بدلاً من التمييز بين تعليم الأب والأم، وبــذلك نتخلص من الارتباط الخطي المتعدد الذي سبب التعدد الخطي قد زال.

$$S = \beta_0 + \beta_1 ASVABC + \beta_2 (SM + SF) + u$$
$$= \beta_0 + \beta_1 ASVABC + \beta_2 SP + u$$
$$SP = SM + SF$$

إن نتائج الانحدار مع المتغير SP بدلاً من SM و SF كانت كما يلي:

```
. g SP = SM + SF

. reg S ASVABC SP
  Source |       SS       df       MS              Number of obs =     570
---------+------------------------------           F(  2,   567) =  166.22
   Model |  1276.73764     2  638.368819           Prob > F      =  0.0000
Residual |  2177.50973   567  3.84040517           R-squared     =  0.3696
---------+------------------------------           Adj R-squared =  0.3674
   Total |  3454.24737   569  6.07073351           Root MSE      =  1.9597

------------------------------------------------------------------------------
       S |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  ASVABC |.1295653      .0099485    13.024    0.000       .1100249    .1491057
      SP | .093741      .0165688     5.658    0.000       .0611973    .1262847
   _cons |4.823123      .4844829     9.955    0.000       3.871523    5.774724
------------------------------------------------------------------------------
```

إن نتائج الانحدار تبين أن الخطأ المعياري لمعامل لـ SP صغير جداً بالمقارنة مع SM و SF ، وبالتالي قيمة t الإحصائية عالية المعنوية. إذاً يعتبر هذا المعامل عبارة عن تسوية بين SM و SF . على أي حال إن استعمال التقييد سيؤدي إلى زيادة في الكفاءة فقط إذا كان التقييد صحيحاً. إذا كان استخدامه ليس صحيحاً، سيؤدي إلى معلمات متحيزة وستكون الأخطاء المعيارية والاختبارات الإحصائية غير المعنوية.

هل معلمات SM و SF غير مقيدة في الانحدار حتى لو أزلنا التقيد؟

في هذه الحالة ليس صحيحاً، لأن المعامل SM صغير جداً أكثر من SF ولكن كما نلاحظ أن الأخطاء المعيارية كبيرة جداً.

سوف نعمل اختباراً ملائماً. إن عبء التقييد يجعل ذلك صعب في توفيق نموذج الانحدار، بسبب وجود معلمات متكفية. وسوف يكون هناك زيادة في RSS و (انخفاض في R^2) عندما نفرض ذلك.

إذا كان التقييد صحيحاً ، فإن التوفيق يتدهور ويصبح صغير القيمة. وعلى أي حال إذا كان التقييد غير صحيح سيسبب انحرافاً، وبذلك يقودنا إلى تدهور المعنوية في التوفيق.

```
. reg S ASVABC SM SF

------------------------------------------------------------------------------
       S |     Coef.   Std. Err.      t     P>|t|      [95% Conf. Interval]
---------+--------------------------------------------------------------------
  ASVABC |.1295006     .0099544   13.009   0.000      .1099486      .1490527
      SM | .069403     .0422974    1.641   0.101     -.013676       .152482
      SF | .1102684    .0311948    3.535   0.000      .0489967      .1715401
   _cons | 4.914654    .5063527    9.706   0.000      3.920094      5.909214
------------------------------------------------------------------------------
```

```
. reg S ASVABC SP

------------------------------------------------------------------------------
       S |     Coef.   Std. Err.     t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  ASVABC | .1295653   .0099485   13.024   0.000       .1100249    .1491057
      SP |  .093741   .0165688    5.658   0.000       .0611973    .1262847
   _cons | 4.823123   .4844829    9.955   0.000       3.871523    5.774724
------------------------------------------------------------------------------
```

في هذه الفقرة سنحاول لحظ الزيادة في RSS الصغيرة و الأمر الذي يؤدي إلى استبعاد رفض التقييد. لنختبر هذه الزيادة من خلال فرضية العدم صحيحة التقييد، و أحد البدائل على الأقل غير صحيح.

$$H_0 : \beta_3 = \beta_2 , \quad H_1 : \beta_3 \neq \beta_2$$

إن الاختبارات الإحصائية تكون من عائلة اختبارات F حيث يكون البسط يعبر عن التحسن بالتوفيق على أريحية التقييد ، مقسوماً (على تكلفة درجات الحرية ، وذلك بسبب إضافة معلمة واحدة للتقدير).

$$F(1,n-k)=\frac{(RSS_R-RSS_U)/1}{RSS_U/(n-k)}=\frac{2177.51-2176.01}{2176.01/566}=0.39$$

أما المقام في الاختبار الإحصائي هو RSS بعد عمل التحسين (لأن RSS في حالة عدم تقييد النموذج)، والتقسيم على n-k عدد درجات الحرية الباقية. حيث K عدد المعلمات في النموذج غير المقيد. وفي هذه الحالة كانت قيمة F الإحصائية تساوي إلى 0.39 . وبما أن F الإحصائية اقل من الواحد غير معنوية (انظر جداول F)، هذا يعني سوف لن نرفض فرضية العدم H_0 ، أي: التقييد صحيح.

وعلى العموم إن عملية التقييد الخطي يمكن اختبارها باستخدام اختبار t . وهذا يتضمن كتابة النموذج في الشكل المقيد، و إضافة حد لضمان حالة العودة لعدم التقييد . يقيم الاختبار فيما إذا كان التقييد الإضافي مطلوب. لإيجاد حد التحويل نعيد كتابة التقييد في النموذج تحت نسخة عدم التقييد وبالطرح. نجد

الحد الذي يعيد تحول نموذج المقيد إلى الخلف، أي: نموذج غير مقيد ويكون : $(b_3 - b_2)SF$.

$$\begin{aligned}0 &= \beta_2 SM + \beta_3 SF - \beta_2 SP\\ &= \beta_2 SM + \beta_3 SF - \beta_2(SM + SF)\\ &= (\beta_3 - \beta_2)SF\end{aligned}$$

$$S = \beta_0 + \beta_1 ASVABC + \beta_2 SP + u$$

إن الحد الذي يعيد نموذج المقيد إلى الخلف، أي: غير نموذج مقيد هو: $(b_3 - b_2)SF$.

نضيف هذا الحد إلى النموذج المقيد، ونبحث عما إذا كنا بحاجته. فرضية العدم ضرورية؛ لأن معامل التحويل يساوي الصفر، و الفرض البديل يكون مختلفاً عن الصفر.

بالطبع فرضية العدم تكون صحيحةً للنموذج المقيد. إذا كانت صحيحة فإن حد التحويل لا نحتاجه. وعملية التقييد غير ملائمة لتصوير واقع البيانات.

$$S = \beta_0 + \beta_1 ASVABC + \beta_2 SP + (\beta_3 - \beta_2)SF + u$$

$$H_0 : \beta_3 - \beta_2 = 0, \quad H_1 : \beta_3 - \beta_2 \neq 0$$

ونجد هنا تطابقاً في الانحدار، حيث نشاهد أن معامل SF غير معنوي الاختلاف عن الصفر؛ لأنه يتضمن حداً غير ضروري، و أن حالة التقييد غير ملائمة لتصوير واقع حال البيانات.

و يبن الشكل الرياضي إن اختبار F و اختبار t متكافئان؛ لأن F الإحصائية هي مربع t الإحصائية، و القيمة الحرجة لـــ F الإحصائية هي مربع القيمة الحرجة لـــ t .

```
. reg S ASVABC SP SF

  Source |       SS       df       MS              Number of obs =     570
---------+------------------------------           F(  3,   566) =  110.83
   Model |  1278.24153     3  426.080508           Prob > F      =  0.0000
Residual |  2176.00584   566  3.84453329           R-squared     =  0.3700
---------+------------------------------           Adj R-squared =  0.3667
   Total |  3454.24737   569  6.07073351           Root MSE      =  1.9607

------------------------------------------------------------------------------
       S |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
---------
  ASVABC |.1295006    .0099544    13.009    0.000      .1099486      .1490527
      SP | .069403    .0422974     1.641    0.101      -.013676       .152482
      SF |.0408654    .0653386     0.625    0.532     -.0874704      .1692012
   _cons |4.914654    .5063527     9.706    0.000      3.920094      5.909214
------------------------------------------------------------------------------
```

مثال(1-10) : لدراسة العلاقة بين مستوى الواردات Y والناتج القومي X_1 ، بملايين الليرات، والرقم القياسي العام لأسعار المستهلكين X_2 للولايات المتحدة الأمريكية من 1964 إلى 1979 . ومن المتوقع أن يرتفع مستوى الواردات مع زيادة GNP وارتفاع مستوى الأسعار المحلية، بإجراء انحدار Y على X_1 و X_2 نحصل على :

$$\hat{Y} = -101.49 + 0.08X_1 + 0.76X_2 \qquad R^2 = 0.97$$
$$(1.40) \quad (1.00) \qquad \overline{R^2} = 0.85$$
$$r_{12} = 0.997$$

وهنا نلاحظ أنه $\hat{b}_2 , \hat{b}_1$ ليس معنوياً إحصائياً عند مستوى معنوية % 5 وبغض النظر عن أن $0.985 = \overline{R^2}$ ، فإن هناك إشارة واضحة لوجود ازدواج خطي بين X_1 و X_2 . ويؤكد هذا معامل الارتباط البسيط الجزئي المرتفع جداً بين X_1 و X_2 ($r_{12} = 1.0$) ، وبإعادة تقدير الانحدار دون X_2 نحصل على :

$$\hat{Y} = -69.03 + 0.03X_1 \qquad R^2 = 0.986$$
$$(-12.00)(31.87)$$

أما إعادة تقدير الانحدار بدون X_1 نحصل على :

$$\hat{Y} = -146.52 + 1.82X_2 \qquad R^2 = 0.985$$
$$(-17.58)(30.79)$$

نجد أن الانحدار البسيط يبين معنوية كل مـــن X_1 و X_2 عنـــد مســـتوى أفضل منهما .

ولكن استبعاد أي منها من علاقة الانحدار يؤدي إلـــى تقـــديرات OLS متحيزة؛ لأن النظرية الاقتصادية تشير إلى وجوب وجود دخول كل من GNP ومستوى الأسعار في دالة الواردات.

10-7- انعدام تجانس تباين الخطأ Heteroscedasticity:

سنتعرف في هذا الفصل على موضوع عدم التجانس ؛ الذي قد يتصف به توزيع حد الخطأ العشوائي في نموذج الانحدار.قبل الدخول في تفصـــيل هـــذا الموضوع نستعرض بعض المفاهيم ذات الصلة به، وهي:

1- عشوائية الانحدار STOCHASTIC REGRESSORS

حتى الآن نفترض أن المتغيرات المستقلة في نموذج الانحدار عشـــوائية؛ لأنه لا يوجد لها أجزاء عشوائية. الشيء المريح في هذا الافتراض أنه لا يمكن تطبيق طريقة المربعات الصغرى للانحدار.

$$y = \alpha + \beta X + u$$

$$b = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)} = \frac{\mathrm{Cov}(X,[\alpha + \beta X + u])}{\mathrm{Var}(X)} = \beta + \frac{\mathrm{Cov}(X,u)}{\mathrm{Var}(X)}$$

إن التقدير بطريقة المربعات الصغرى يبقى غير متحيز وكفوأً ومتناســـقاً كشرط لأن مكونات المتغيرات المستقلة تخضع لتوزيع مستقل عن حد الخطـــأ العشوائي كشرط لصحة خصائص النموذج.

سوف يكون كافياً أن نوضح أنه غير متحيز و متناسق. وبعد تحليل معامل الميل بأي طريقة سوف نركز على حد الخطأ، ونبين أن توقعه يساوي الصفر.

للبرهان على ذلك نعيد هيكلة حد الخطأ ولكنه كمجموع جداءات من سلسلة الحدود، حيث كل ناتج من المعادلة لـ X و معادلة لـ u . سنبدأ بكتابة حد التباين المشترك كما يلي:

$$\frac{\mathrm{Cov}(X,u)}{\mathrm{Var}(X)}=\frac{\frac{1}{n}\sum(X_i-\overline{X})(u_i-\overline{u})}{\mathrm{Var}(X)}$$

سوف نضيف الحد $\mathrm{Var}(X)$ إلى داخل علاقة المجموع؛ فنحصل على الشكل التالي:

$$\frac{\mathrm{Cov}(X,u)}{\mathrm{Var}(X)}=\frac{\frac{1}{n}\sum(X_i-\overline{X})(u_i-\overline{u})}{\mathrm{Var}(X)}=\frac{1}{n}\sum\left\{\frac{X_i-\overline{X}}{\mathrm{Var}(X)}\right\}(u_i-\overline{u})$$

هكذا يكون كل حد داخل المجموع عبارة عن ناتج ضرب معادلة X و حد الخطأ u . معادلة X و معادلة u هي (u_i-u) كما هي مبينة.

$$\frac{\mathbf{Cov}(X,u)}{\mathbf{Var}(X)}=\frac{1}{n}\sum f(X_i)(u_i-\overline{u})\ f(x_i)=\frac{X_i-\overline{X}}{\mathrm{Var}(X)}$$

إن القيمة المتوقعة لحد الخطأ تساوي القيمة المتوقعة للعلاقة. سوف نخرج كعامل مشترك، وسوف نعكس عامل التقسيم للمجموع والعلاقة . (القيمة المتوقعة للمجموع هي عبارة عن مجموع التوقعات) لذلك نكتب:

$$E\left\{\frac{\text{Cov}(X,u)}{\text{Var}(X)}\right\} = E\left\{\frac{1}{n}\sum f(X_i)(u_i - \overline{u})\right\} =$$

$$= \frac{1}{n}\sum E\{f(X_i)(u_i - \overline{u})\} = 0$$

بما أن استقلالية المتغيرات محقق بين X و u (هذا شرط هام – يبقى هنا ثابتاً) سوف ندخل القيمة المتوقعة لكل حد داخل علاقة التوقع الناتجة فنحصل على المطلوب.

تذكرة: المتغيران العشوائيان X و Y يكونان مستقلين إذا كان :

$$E[f(x)g(y)] = E[f(x)\}\, E\{g(y)]$$

من أجل أي معادلة : $f(x)$ و $g(y)$.

$$E[f(X_i)(u_i - \overline{u})] = E[f(X_i)]\, E(u_i - \overline{u})$$

$$= E[f(X_i)][E(u_i) - E(\overline{u})] = E[f(X_i)] \times [0 - 0]$$

$$E\left\{\frac{\text{Cov}(X,u)}{\text{Var}(X)}\right\} = E\left\{\frac{1}{n}\sum f(X_i)(u_i - \overline{u})\right\} =$$

$$= \frac{1}{n}\sum E\{f(X_i)(u_i - \overline{u})\} = 0$$

إن القيمة المتوقعة لـ u جزئياً تساوي الصفر، وكذلك القيمة المتوقعة لكل حد في المجموع يساوي الصفر، و كذلك القيمة المتوقعة لحد الخطأ يساوي الصفر ، وهكذا b تكون تقديراً غير متحيز لـ b.

التناسق يمكن توضيحه بسهولة. نهاية القيم للبسط والمقام لحد الخطأ يكون تباين المجتمع المشترك، والتباين على التوالي:

$$\text{plim}\, b = \beta + \frac{\text{plim Cov}(X,u)}{\text{plim Var}(X)}$$

$$= \beta + \frac{\sigma_{x,u}}{\sigma_x^2} =$$

إذا كان X و u مستقلين يمكننا تحليل التباين المشترك إلى علاقتين للتوقع، نجد:

$$\sigma_{x,u} = E\{(X-\mu_x)(u-\mu_u)\}$$

$$= E(X-\mu_x)E(u-\mu_u)$$

إن كلا هاتين العلاقتين يساوي الصفر، لهذا السبب plim b يساوي β .

$$\text{plim}\, b = \beta + \frac{\text{plim Cov}(X,u)}{\text{plim Var}(X)}$$

$$= \beta + \frac{\sigma_{x,u}}{\sigma_x^2} = \beta + \frac{0}{\sigma_x^2} = \beta$$

2- قياس الخطأ MEASUREMENT ERROR

في هذه الفقرة نبحث أهمية قياس الأخطاء في المتغيرات لنموذج الانحدار. بالتحليل البسيط سوف نحدد نموذج الانحدار البسيط.

$$y = \alpha + \beta z + v$$

$$x = z + w$$

سوف نبدأ بقياس الاخطأ في المتغيرات التفسيرية. نفترض أن Y يتحدد بالمتغير z ولكن z يكون موضوع قياس الخطأ w . سوف يكون علامة القياس المتغير التفسيري X .

نستبدل قيمة Z من المعادلة الثانية، ونعيد كتابة النموذج كما هو مبين.

$$y = \alpha + \beta(X - w) + v$$

هكذا نعبر عن y كعلاقة خطية لمشاهدات المتغير x مع حد الخطأ العشوائي يبدأ بشكل مختلط حد الخطأ في النموذج الأصلي وقياس الخطأ

$$\begin{aligned} Y &= \alpha + \beta(X - w) + v \\ &= \alpha + \beta X + v - \beta w \\ &= \alpha + \beta X + u \end{aligned}$$

$$u = v - \beta w$$

على أي حال لتوفيق هذا النموذج باستخدام طريقة المربعات الصغرى، يجب أن نتأكد من تحقق شروط غاوس – ماركوف سوف لن يتحقق . تعتبر X عشوائية المركبات لقياس الخطأ w . و w أيضاً واحد من مكونات خليط الخطأ العشوائي. هكذا u لا تتوزع بشكل مستقل عن X .

$$\begin{aligned} Y &= \alpha + \beta(X - w) + v \\ &= \alpha + \beta X + v - \beta w \\ &= \alpha + \beta X + u \end{aligned}$$

سوف نوضح أن طريقة تقدير ميل المربعات الصغرى غير متعارضة في العينات الكبيرة غير المتحيزة هبوطاً إذا كانت b موجبة، و صعوداً إذا كانت b سالبة.

$$b = \frac{\text{Cov}(X,Y)}{\text{Var}(X)} = \frac{\text{Cov}(X,[\alpha + \beta X + u])}{\text{Var}(X)}$$

سوف نبدأ بتقدير المربعات الصغرى، و نبدل قيمة Y بما يساويها في النموذج الحقيقي. في هذه الحالة نرى التطور البديل للنموذج. وإن التحليل البسيط يكون باستخدام مساواة تنتمي Y إلى X .

نستخدم القاعدة الأولى للتباين المشترك بتحليل البسط . $\text{Cov}(X, a)$ يساوي الصفر لأن a قيمة ثابتة. $\text{Cov}(x, bX)$ يساوي $b\text{Var}(X)$ وخصائص الحد لـ b بعد التقسيم على $\text{Var}(X)$:

$$b = \frac{\text{Cov}(X, y)}{\text{Var}(X)} = \frac{\text{Cov}(X, [\alpha + \beta X + u])}{\text{Var}(X)}$$

$$= \frac{\text{Cov}(X, \alpha) + \text{Cov}(X, \beta X) + \text{Cov}(X, u)}{\text{Var}(x)}$$

$$= \beta + \frac{\text{Cov}(X, u)}{\text{Var}(X)}$$

$$b = \frac{\text{Cov}(X, y)}{\text{Var}(X)} = \frac{\text{Cov}(X, [\alpha + \beta X + u])}{\text{Var}(X)}$$

$$= \frac{\text{Cov}(X, \alpha) + \text{Cov}(X, \beta X) + \text{Cov}(X, u)}{\text{Var}(x)}$$

$$= \beta + \frac{\text{Cov}(X, u)}{\text{Var}(X)} = \beta + \frac{\text{Cov}([z + w], [v - \beta w])}{\text{Var}(z + w)}$$

$$= \beta + \frac{\text{Cov}(z, v) + \text{Cov}(z, -\beta w) + \text{Cov}(w, v) + \text{Cov}(w, -\beta w)}{\text{Var}(z) + \text{Var}(w) + 2\text{Cov}(z, w)}$$

نستطيع تحليل البسط كما هو مبين باستخدام القواعد تحليل التباين المشترك، وكذلك المقام نستطيع تحليله باستخدام قاعدة التباين الأولى.

$\text{Cov}(z, -bw)$ يساوي $-b\,\text{Cov}(z, w)$ و $\text{Cov}(w, -b\,w)$ يساوي $-b\,\text{Var}(w)$.

$$b = \beta + \frac{\text{Cov}(z, v) - \beta\text{Cov}(z, w) + \text{Cov}(w, v) - \beta\text{Var}(w)}{\text{Var}(z) + \text{Var}(w) + 2\text{Cov}(z, w)}$$

$$b = \beta + \frac{\mathrm{Cov}(z,v) + \mathrm{Cov}(z,-\beta w) + \mathrm{Cov}(w,v) + \mathrm{Cov}(w,-\beta w)}{\mathrm{Var}(z) + \mathrm{Var}(w) + 2\mathrm{Cov}(z,w)}$$

يمكننا إجراء تحليل أعمق من ذلك' لذلك سوف نبحث الآن عن خصائص b. والشيء الطبيعي البدء بالبحث فيما إذا كانت b متحيزاً أو غير متحيز.

لسوء الحظ إذا أخذنا التوقع لا يمكننا توضيح توقع حد الخطأ العشوائي . السبب في ذلك نسبته في كلا المعادلتين المتغير العشوائي نفسه و w و توقع القيمة القاعدة لا تقبل مثل هذه العبارة.

$$b = \beta + \frac{\mathrm{Cov}(z,v) - \beta\mathrm{Cov}(z,w) + \mathrm{Cov}(w,v) - \beta\mathrm{Var}(w)}{\mathrm{Var}(z) + \mathrm{Var}(w) + 2\mathrm{Cov}(z,w)}$$

$$E(b) = \beta + E\left\{\frac{\mathrm{Cov}(z,v) - \beta\mathrm{Cov}(z,w) + \mathrm{Cov}(w,v) - \beta\mathrm{Var}(w)}{\mathrm{Var}(z) + \mathrm{Var}(w) + 2\mathrm{Cov}(z,w)}\right\}$$

على أي حال نستطيع أن نبحث فيما سيحدث في العينات الكبيـــرة ؛ لأن المكونات سوف ترشدنا إلى نهاية القيم إذا كانت القيم النهائية موجودة.

$$\mathrm{plim}\, b = \beta + \frac{\begin{array}{c}\mathrm{plim}\,\mathrm{Cov}(z,v) - \beta\mathrm{plim}\,\mathrm{Cov}(z,w) \\ + \mathrm{plim}\,\mathrm{Cov}(w,v) - \beta\mathrm{plim}\,\mathrm{Var}(w)\end{array}}{\mathrm{plim}\,\mathrm{Var}(z) + \mathrm{plim}\,\mathrm{Var}(w) + 2\mathrm{plim}\,\mathrm{Cov}(z,w)}$$

نحتاج إلى بعض الافتراضات ذات الاهتمام لــ z و v و w . سوف نفترض في التقارب الأول أنها جميعاً تتوزع بشكل مستقل فيما بينها. في هذه الحالة نهاية المقادير التالية يساوي الصفر:

$$\mathrm{Plim}\,\mathrm{Cov}(z, v),\ \mathrm{Plim}\,\mathrm{Cov}(z, w),\ \mathrm{Plim}\,\mathrm{Cov}(w, v)$$

$$\operatorname{plim} b = \beta + \frac{\operatorname{plim}\operatorname{Cov}(z,v) - \beta\operatorname{plim}\operatorname{Cov}(z,w) + \operatorname{plim}\operatorname{Cov}(w,v) - \beta\operatorname{plim}\operatorname{Var}(w)}{\operatorname{plim}\operatorname{Var}(z) + \operatorname{plim}\operatorname{Var}(w) + 2\operatorname{plim}\operatorname{Cov}(z,w)}$$

$$= \beta + \frac{0 + 0 + 0 - \beta\sigma_w^2}{\sigma_z^2 + \sigma_w^2}$$

على أي حال نهاية قيمة تباين العينة هو تباين المجتمع نفسه.

وهكذا في العينات الكبيرة b تتحيز نحو الصفر، و حجم التحيز يعتمد على نسبية الحجم لتباين كل من w و z .

$$\operatorname{plim} b = \beta + \frac{\operatorname{plim}\operatorname{Cov}(z,v) - \beta\operatorname{plim}\operatorname{Cov}(z,w) + \operatorname{plim}\operatorname{Cov}(w,v) - \beta\operatorname{plim}\operatorname{Var}(w)}{\operatorname{plim}\operatorname{Var}(z) + \operatorname{plim}\operatorname{Var}(w) + 2\operatorname{plim}\operatorname{Cov}(z,w)}$$

$$= \beta + \frac{0 + 0 + 0 - \beta\sigma_w^2}{\sigma_z^2 + \sigma_w^2} = \beta - \beta\frac{\sigma_w^2}{\sigma_z^2 + \sigma_w^2}$$

نظراً لأن تقدير b غير ثابت يكون في مأمن من افتراض ، الوقوع في التحيز لأن العينات الصغيرة .

إذا كانت افتراضاتنا غير صحيحة حول z,v,w سيكون تصحيح b دائماً ساكناً أي: تقديراً متحيزاً، ولكن العلاقة للتحيز ستكون أكثر تعقيداً.

فيما بعد سيكون انتهاكاً لشروط غاوس – ماركوف لوجود الأخطاء المعيارية لأن مقاييس اختبارات t و F غير مرضية. مع الأخذ بعين الاعتبار إن قياس الخطأ في المتغير التابع قليل الأهمية. لنفترض صحة المتغير التابع يكون q ومقياس المتغير التابع y و مقياس الخطأ r .

نستطيع إعادة كتابة النموذج في حدود المتغيرات المشاهدة وباستبدال q من المعادلة الثانية.

$$q = \alpha + \beta X + v \qquad Y = q + r$$

$$Y - r = \alpha + \beta X + v$$

في هذه الحالة إن وجود مقياس خطأ لا يقود إلى انتهاك الشروط الأربعة لغاوس –ماركوف . إذا v تبدد الشك، لأن الشروط الأساسية للنموذج u تبدد الشكوك في تعديل واحد ما لم يوجد سبب قوي لـ r لأنه بتوزيع مستقل لـ x .

$$Y = \alpha + \beta X + v + r$$
$$= \alpha + \beta X + u$$

$$u = v + r$$

إن الأخطاء المعيارية و الاختبارات سوف تبقى قانونية. على أي حال إن الأخطاء المعيارية سوف تكون أكبر من أن لا نملك قياس الخطأ يعكس حقيقة تباين المشترك للمجتمع للمعاملات أكبر.

$$\sigma_b^2 = \frac{\sigma_u^2}{n\sigma_x^2} = \frac{\sigma_v^2 + \sigma_r^2}{n\sigma_x^2}$$

3- البحث النقدي لفريدمان تقديرات المربعات الصغرى لمعادلة الاستهلاك(FRIEDMAN'S CRITIQUE OF OLS ESTIMATION OF THE CONSUMPTION FUNCTION)

نظرية ميلتون فريدمان حول معاملات الدخل كمثال كلاسيكي لقياس نظرية الخطأ. فكرة التحيز هنا ان معامل الاستهلاك C^P هو قيمة نسبية من معامل الدخل Y^P.

فمعاملا الاستهلاك والدخل يكونان فكرة قابلة للتطبيق للتعبير عن الحد المتوسط للاستهلاك والدخل على التوالي، ولكن لا يمكن قياسهما مباشرة.

الدخل الحقيقي Y (مقاس) له مركبتان : معامل الدخل و الدخل المؤقت Y^T(العشوائية ، الجزئية القصيرة) . بطريقة مماثلة الاستهلاك الحقيقي C (مقاس) ويملك مكونتين : معامل الاستهلاك و الاستهلاك المؤقت C^T.

النموذج الفعلي	$q = a + b z + v$	$C^P = b Y^P$
قياس الأخطاء	$x = z + w$ $y = q + r$	$Y = Y^P + Y^T$ $C = C^P + C^T$

وفقاً لفرضية معامل الدخل كنموذج للانحدار بطريقة المربعات الصغرى، أي: الاستهلاك الحقيقي على الدخل الحقيقي سيكون موضوعاً لقياس الخطأ في كلا المتغيرين المستقل والتفسيري مقياس الأخطاء يبدأ من الاستهلاك المؤقت C^T و الدخل المؤقت Y^T . من المعادلات الثلاثة الأولى نستطيع أن نشتق المساواة؛ التي تبين نسبة الاستهلاك الحقيقي إلى الدخل الحقيقي.

لشرح التحليل سوف نفترض أن المركبات المؤقتة للاستهلاك و الدخل تعتمد على معاملات المركبات لكل واحدة مع الأخرى.

النموذج الفعلي	$q = a + b z + v$	$C^P = b Y^P$
قياس الأخطاء	$x = z + w$ $y = q + r$	$Y = Y^P + Y^T$ $C = C^P + C^T$
النموذج الفعلي لقياس المتغيرات	$y = a + b x + v + r - b w$ $= a + b x + u$	$C = b Y + C^T - b Y^T$ $= b Y + u$
الفرضيات	كل من $v, w,$ و r تتوزع بشكل مستقل كل واحدة عن الأخرى وكذلك (z,q)	كل من Y^T and C^T يتوزع بشكل مستقل الواحدة عن الأخرى وكذلك (Y^T , C^T)

حتى مع هذه الافتراضات لطريقة المربعات الصغرى سوف لن نصل لتقدير ميل الاستهلاك حق قدره بسبب تحيز العينات الكبيرة، كما هو مبين أدناه:

النموذج الحقيقي	$q = a + b z + v$	$C^P = b Y^P$
قياس الخطأ	$x = z + w$ $y = q + r$	$Y = Y^P + Y^T$ $C = C^P + C^T$

النموذج الفعلي لقياس المتغيرات	$y= a + b x + v+ r - b w = a + b x + u$	$C= b Y + C^T - b Y^T = b Y + u$
نهاية المقدار b_{OLS} plim b_{OLS}	$\beta - \beta \frac{\sigma_w^2}{\sigma_z^2 + \sigma_w^2}$	$\beta - \beta \frac{\sigma_{Y^T}^2}{\sigma_{Y^P}^2 + \sigma_{Y^T}^2}$

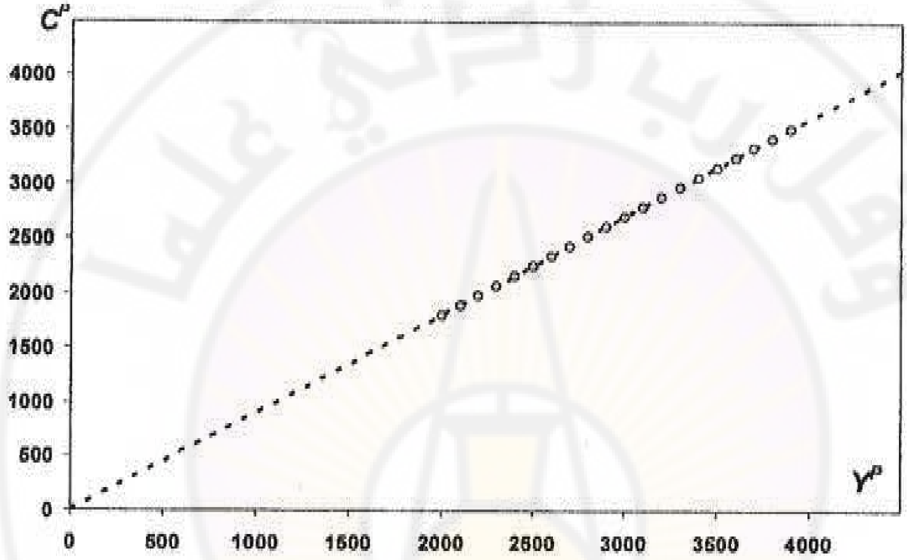
إذا كان هذا الافتراض المقلق للتباين المشترك للمركبات المؤقتة في التحليل لا بد من تعديله . سوف نوضح ذلك بمثال للتحليل مع تجربة مونتي كارلو . سوف نختار القيمة 0.9 لمعامل الدخل بعلاقة الاستهلاك، لعينة مكونة من 20 و البيانات الدخل الدائم من 2000 و 3900 يزداد في كل خطوة بمقدار 100 .

النموذج الحقيقي	$q= a + b z + v$	$C^P = b Y^P$
بيانات الانحدار للنموذج الفعلي لقياس المتغيرات		$YP = 2,000, 2,100..., 3,900$
النموذج الفعلي لقياس المتغيرات	$y= a + b x + v+ r - b w = a + b x + u$	$Y^T = 400N(0,1)$ $C^T = 0$
نهاية المقدار b_{OLS} plim b_{OLS}	$\beta - \beta \frac{\sigma_{Y^T}^2}{\sigma_{Y^P}^2 + \sigma_{Y^T}^2}$	$0.9 - 0.9 \frac{160,000}{325,000 + 160,00}$ $= 0.9 - 0.29 = 0.61$

قياس الخطأ للدخل (المركبة المؤقتة) سيكون نتاجاً من التوزيع الطبيعي مع متوسط صفر وتباين يساوي الواحد ، و متناسباً يساوي 400.

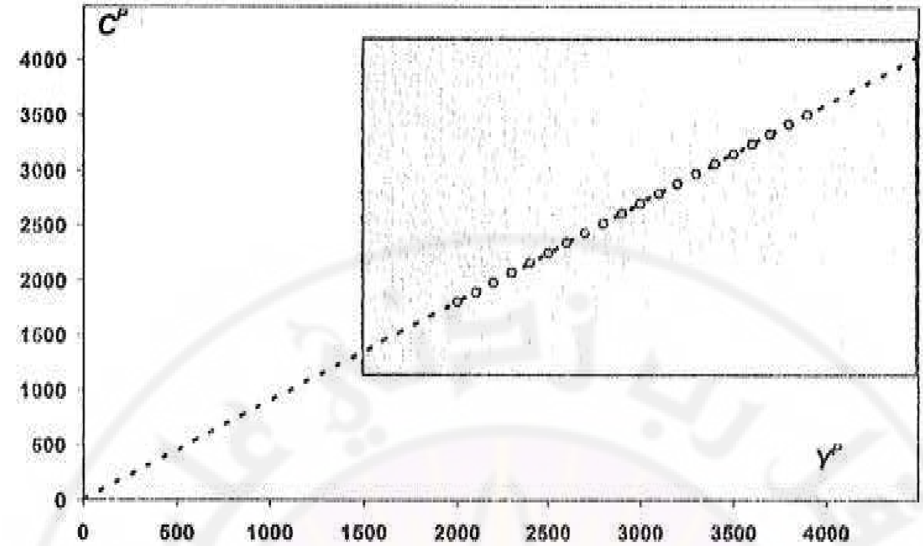
لماذا مقياس العامل 400؟ العينة الكبيرة لها تحيز في الجزء على المجتمع ستكون مدركة.

تم اختيار مقياس العامل 400 على مبدأ التجربة و الخطأ؛ لأنه ينتج تحيزاً منظوراً. الاختيار يعطي معاملات وبيانات لتقدير b_{OLS} سيكون نحو 0.61 في العينات الكبيرة. انظر الشكل البياني التالي كجزء لعلاقة غير عشوائية :



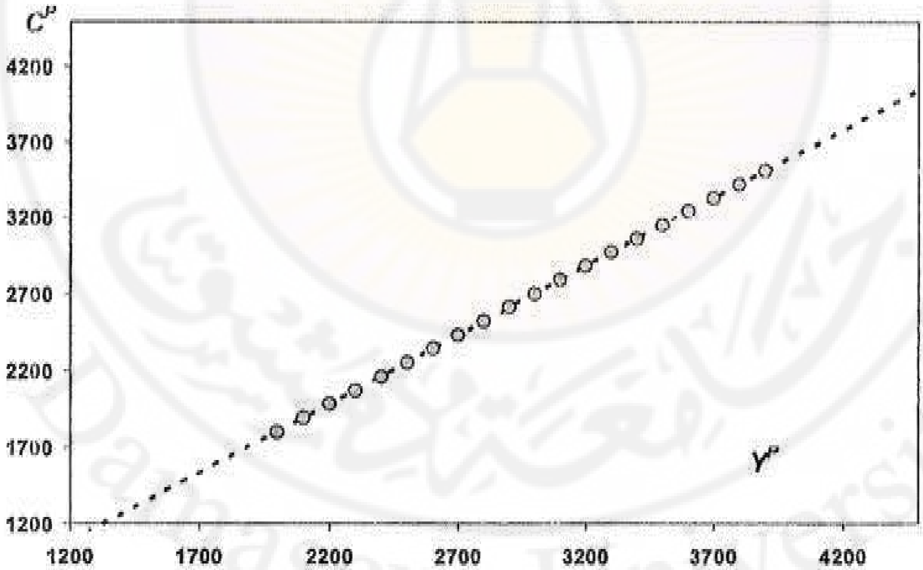
الشكل (3-10)

نكبر الجزء موضع الاهتمام :



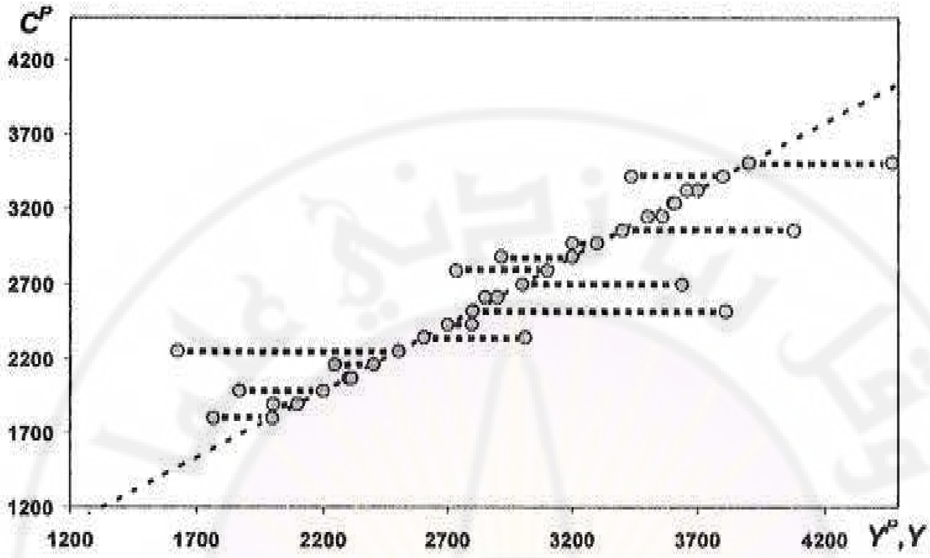
الشكل (4-10)

نتابع قياس الخطأ في النموذج للدخل المؤقت.



الشكل (5-10)

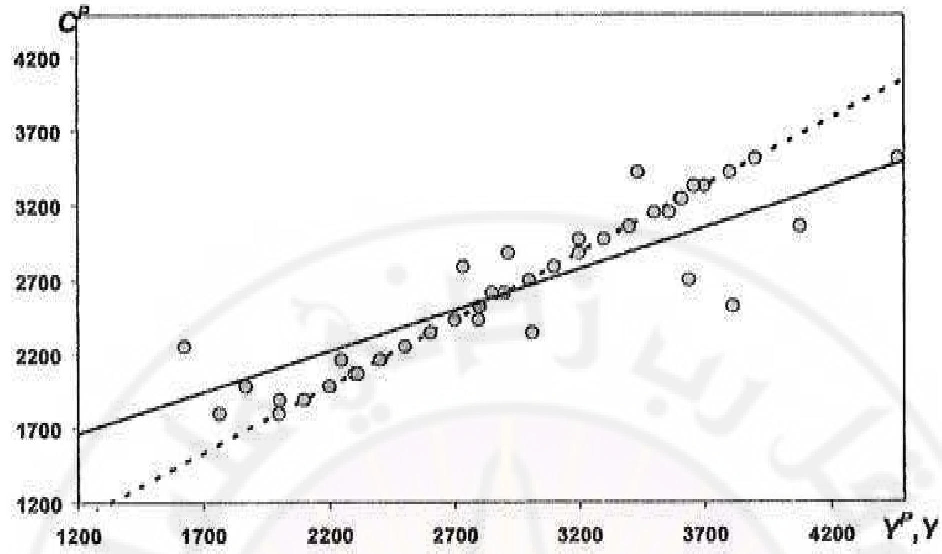
نلاحظ أن الاستهلاك غير متكافئ مع الـدخل المؤقـت، وكـذلك مـن المشاهدات نلاحظ التغيير في مقياس Y فقط كما في الشكل التالي:



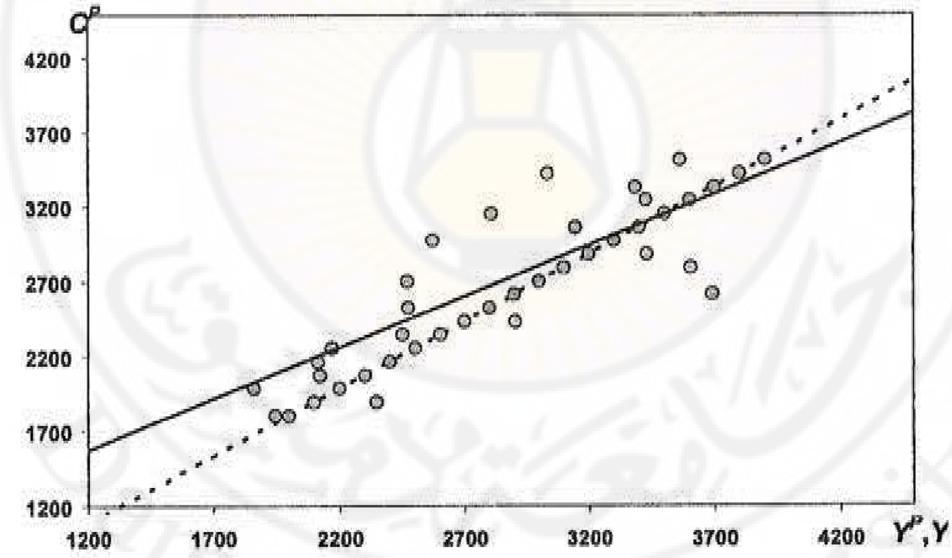
الشكل (6-10)

وبالتالي خط الانحدار لم يعكس حقيقة الميل.

سوف نعيد التجربة بمجموعة أعداد عشوائية ينتج استهلاكاً مؤقتاً من الدخل في 20 مشاهدة . مرة ثانية خط الانحدار لم يعكس حقيقة الميـل. نعيـد التجربة من جديد ، بإحداث الأرقام العشوائية، ينتج استهلاك مؤقت من الدخل، مرة أخرى خط الانحدار لم يعكس حقيقة الميـل، انظـر الشـكلين (7-10) و (8-10).



الشكل (7–10)



الشكل (8–10)

يبين الجدول التالي 10 مجموعات من النتائج للتجربة. القيم الحقيقية لـــ a و b تساوي الصفر و 0.9 .

Sample	a	s.e.(a)	b	s.e.(b)
1	1,001	251	0.56	0.08
2	755	357	0.62	0.11
3	756	376	0.68	0.13
4	668	290	0.66	0.09
5	675	179	0.64	0.06
6	982	289	0.57	0.10
7	918	229	0.56	0.07
8	625	504	0.66	0.16
9	918	181	0.58	0.06
10	679	243	0.65	0.08

نلاحظ وجود انحدار في التحيز للمعلمة b و هذه القيم تظهر حول توزيع نهاية القيمة 0.61. بعد ذلك نجد أن a تساوي الصفر و يجب أن تكون موجبة أو سالبة العشوائية . ومن الواضح ارتفاع التحيز المتعاقب للميل لم يقدر حـق قدره. الخطأ المعياري يكون ضعيفاً لقياس خطأ التحيز، و يجب أن نحاول إنجاز أي من الاختبارات.

10-8- عدم التجانسHetero Scedasticity

إن مفهوم عدم التجانس Hetero Scedasticity مكون من عبارتين، الأولى Hetero وتعني مختلفاً غير متساو، والثانية Scedasticity وتعني التباعد أو الانتشار أو التشتت أي عدم التجانس . ومجمل العبارة نقصد به عدم ثبات التباين أو عدم تساوي تباين حد الخطأ العشوائي. وهذا يعني الخروج عن احد الفروض الأساسية لغاوس-ماركوف. يلاحظ في الكثير من الدراسات التي تعتمد على البيانات المقطعية (العرضية) فإن ثبات تباين حد الخطأ العشوائي تصبح غير واقعية ، فمثلاً عند دراسة ميزانية الأسرة فإن تباين البواقي في

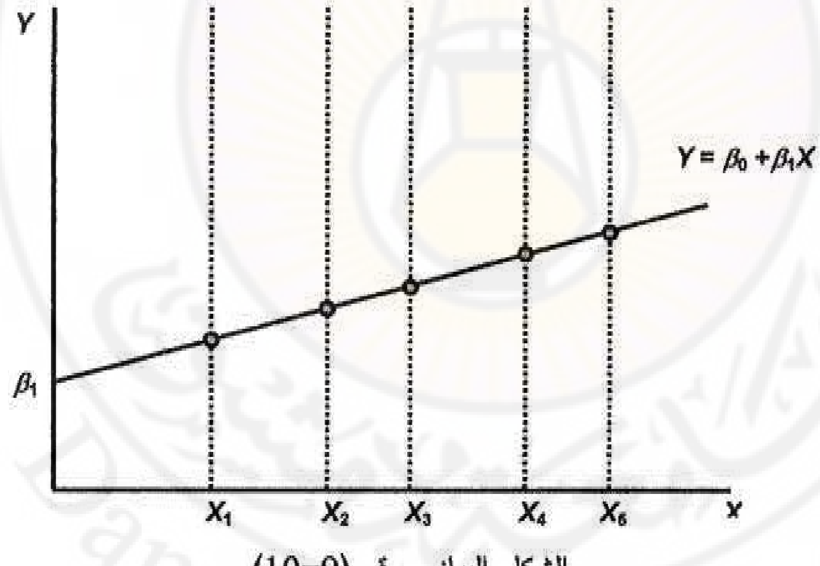
نموذج الانحدار من النادر أن يثبت مع تزايد الدخل وكذلك فإن بيانات المقطع العرضي لدراسة سلوك الشركة. وإن تباين البواقي من المحتمل أن يتزايد مع حجم الشركة، وعليه فإن عند تزايد أو تناقص تباينات حد الخطأ العشوائي مع تزايد قيم المتغيرات المستقلة؛ فعندئذ نحصل على ما يسمى بعدم التجانس.

لنناقش المسألة ببعض التفصيل:

سوف نناقش تركيبية نموذج الانحدار من الشكل :

$$Y = b_{01} + b_1 X + u$$

لنعالج هذا الموضوع من خلال الرسم البياني لنستفيد من ترتيب التقسيمات في الشكل البياني، سوف نفترض أنه لدينا عينة من خمس مشاهدات، من قيم X التي نبينها على الرسم.



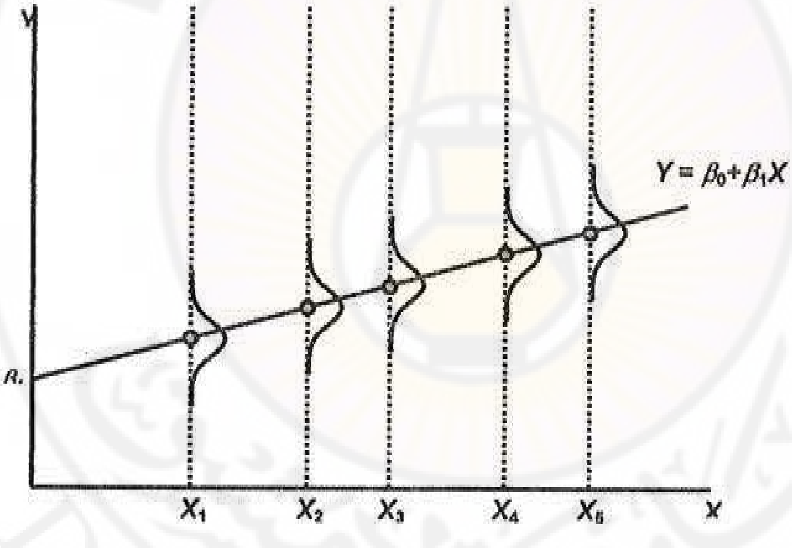
الشكل البياني رقم (9-10)

إذا كان لا يوجد حداً للخطأ العشوائي في النموذج، سوف تقع جميع النقاط المشاهدات على خط واحد؛ كما هو مبين في الشكل (9-10).

الآن لنأخذ قيماً لتأثير حد الخطأ، و نستبدل كل مشاهدة بشكل عــامودي على المقياس ، سنعدل قيم Y خارج نطاق X .

إن حد الخطأ العشوائي في كل مشاهدة يفترض له توزيع عشوائي . مــن الرسم البياني أعلاه يمكن أن نستنتج ثلاثة فرضيات:

الأولى: إن القيم لــ u في كل مشاهدة يساوي الصــفر، و الثانيــة: لأن التوزيع في كل مشاهدة يخضع للتوزيع الطبيعي. و سوف نفترض أن جميعهــا صحيح. أما الثالثة: لأن التوزيع نفسه لكل مشاهدة. في هذه الحالة المتوســطات لها توزيع طبيعي؛ كما هو مبين في الشكل البياني (10–10) ولها التباين نفسه.

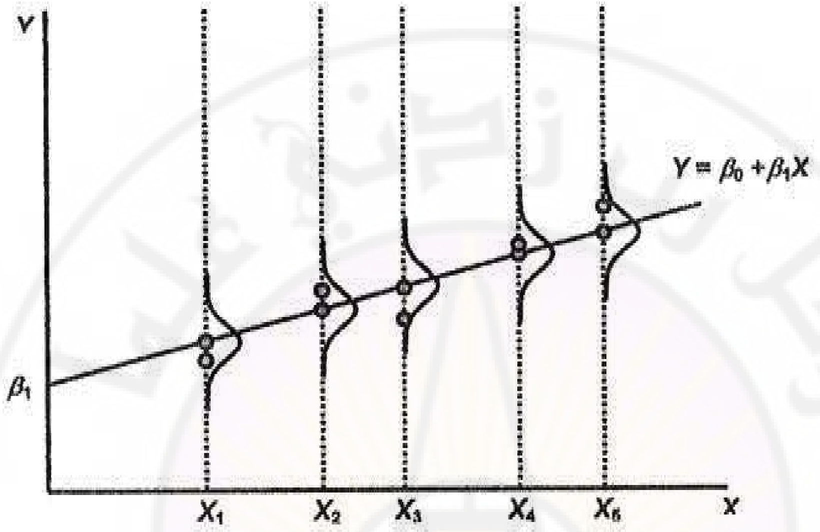


الشكل (10–10)

إذا كانت هذه الشروط محققة فإن حد الخطأ العشوائي يشير إلى أنه متجانس (لها الانتشار نفسه). كل مشاهدة لها الاحتمال نفسه (قبل سحب العينة) يمكن الاعتماد عليه في الوصول إلى موضع كل مشاهدة على خط الانحدار:

$$Y = b_0 + b_1 X$$

طالما إن العينة تبدو واضحة فإن بعض المشاهدات سوف تقع خلف الخط؛ وليس لدينا من طريقة لتحسين الوضع على مكان مفترض ان يكون عليه.

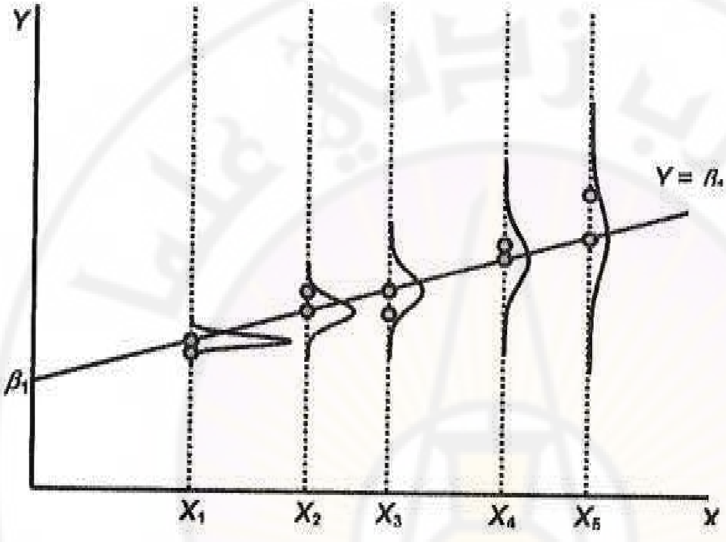


الشكل (11-10)

الآن لنمعن النظر في الشكل البياني رقم (11-10) . نجد أن توزيع حد الخطأ العشوائي مترافق مع كل مشاهدة توقعه صفر و طبيعي . على أي حال إن التباين ليس ثابتاً لفترة طويلة.

من البديهي المشاهدات لها أفضل توزيع عندما u لها اقل تباين ، كما نلاحظ عند النقطة X_1 سوف تتجه ليكون لها أفضل دليلاً للتوزيع من النقطة X_5 التي لها تباين مرتفع نسبياً. عندما التوزيع لا يكون نفسه لكل المشاهدات، يمكننا القول أن حد الخطأ العشوائي هو الذي يشكل موضوع عدم تجانس الخطأ.

يوجد موضوعان من الموضوعات الجامعة لعدم التجانس. واحد منها هو الأخطاء المعيارية لمعاملات الانحدار عندما تكون خاطئة، و اختبار t (و F) عندما تكون غير مرضية .

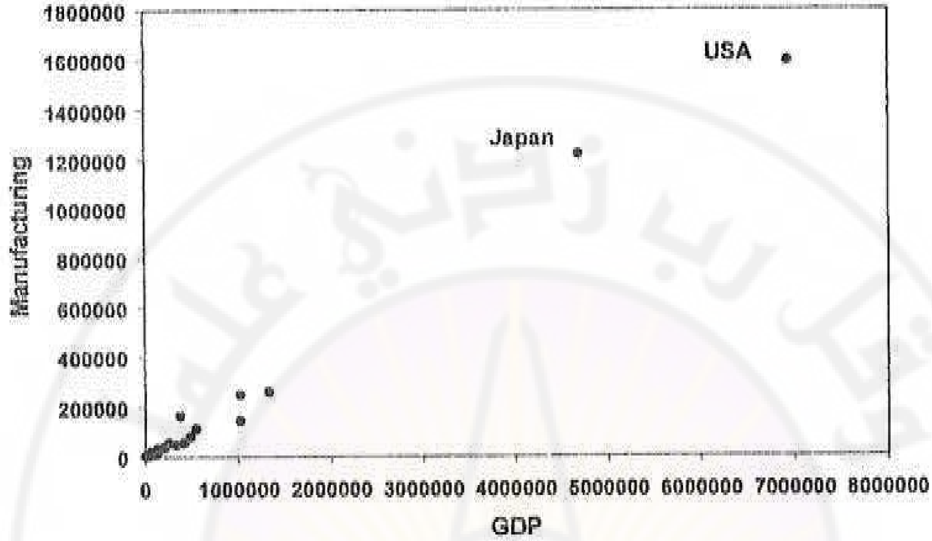


الشكل (12-10)

والأخر طريقة المربعات الصغرى غير مجدية كتقنية تقدير . أما التقنية البديلة التي تعطي وزناً مرتفعاً نسبياً ، لأن تباين المشاهدات قليل يجب أن يخدم ناتج التقديرات أكثر.

مثال(2-10): يبين شكل الانتشار للمخرجات المتوضعة في الرسم مقابل GDP مقاسة بملايين الدولارات لثلاثين بلد من عام 1997. (البيانات من المجموعة الإحصائية لـ UNIDO . العينة محصورة في البلدان التي فيها GDP اقل من 10 بليون و رأسمال اقل من 2000 دولاراً) .

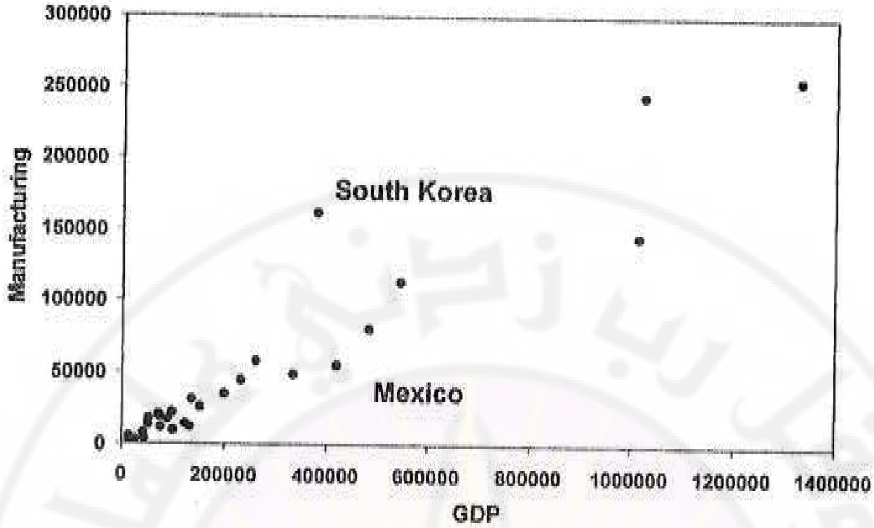
إن شكل الانتشار مسيطر عليه مشاهدات اليابان، و أمريكا شاذة عن بيانات النموذج، ومن الصعب الكشف عنها في أي نوع من النماذج.



الشكل (13-10)

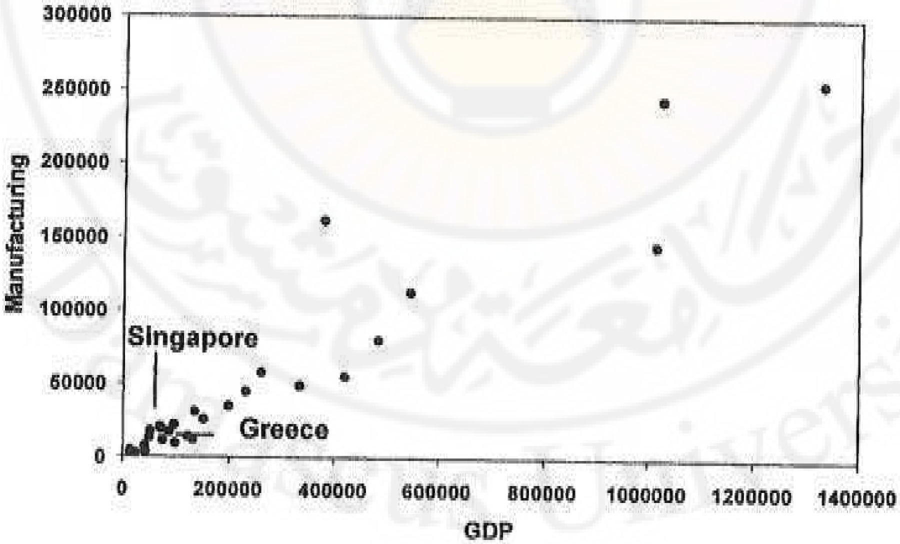
على أي حال لنسقط هذين البلدين، و ونعيد رسم شكل الانتشار. واضح من الشكل(13-10) ظهور عدم التجانس من جديد. السبب في عدم التجانس هو الاختلاف في حجم قطاع الصناعة تقريباً علاقة الاتجاه المتزايد مع حجم GDP.

إن GDP في كل من كوريا الجنوبية و المكسيك كبير نسبياً . القطاع الصناعي يكون نسبياً أكثر أهمية في كوريا الجنوبية؛ لذلك المشاهدة تبتعد عن اتجاه الخط . وعكس الحالة في المكسيك تنخفض في عام 1997.



الشكل (10-14)

سنغافورة واليونان زوج مختلف من البلدان، إن نسبية اخـتلاف العلاقـة كبيرة وصغيرة من علاقة الاتجاه العام.



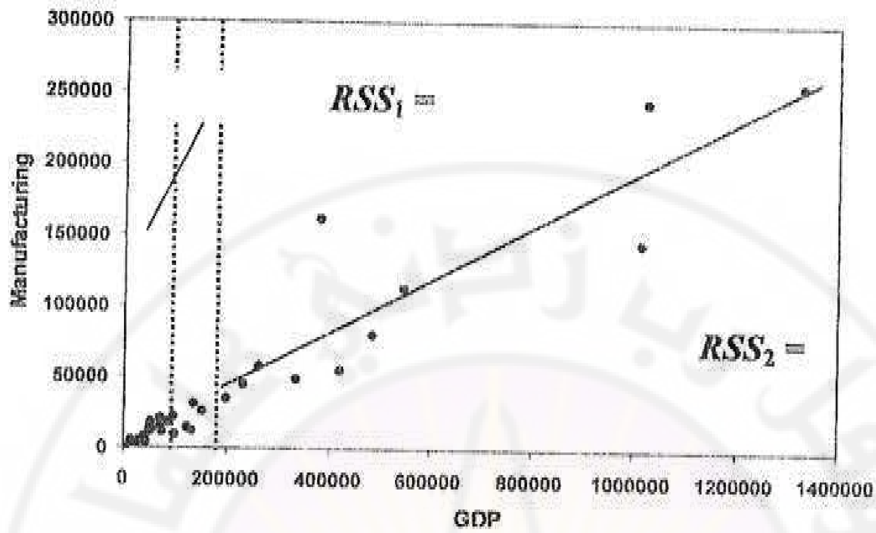
الشكل (10-15)

10-9- اختبار جولدفيلد – كوانت GOLDFELD-QUANDT TEST

إن حد الخطأ العشوائي في نموذج الانحدار يفترض تجانسه إذا كان يملك نفس قوة التوزيع لكل المشاهدات . إذا كان هذا الشرط غير محقق يمكن القول: ان هناك عدم تجانس التباين ، واضح إن إمكانية تحديد نوع انعدام تجانس الخطأ اللانهائي. على أي حال شائع جزئياً نوع الخطأ المعياري للتوزيع يكون نسبياً إلى حجم واحد من المتغيرات المستقلة. هذا النوع من انعدام التجانس يمكن توضيحه بالمثال من خلال الرسم البياني. الخطأ المعياري للتوزيع يكون جزئياً من X ، انظر الشكل (15-10) .

إن اختبار جولد فيلد -كوانت يمكن اختبار هذا النوع من انعدام التجانس. من خلال العينة تنقسم البيانات إلى ثلاثة مجالات اشتملت على 8/3 من المشاهدات مع القيم الصغيرة للمتغير X ، الـ 8/3 من المشاهدات مع القيم العظمى و 1/4 من قيم المتوسطة. في هذه الحالة من أجل 28 تتوزع المجالات الأقل و المتوسط، و الأعلى لها , 11 , 6 , 11 على التوالي .

إن خط الانحدار في المجالات الأقل والأعلى من المشاهدات كما هي مبينة على الرسم البياني رقم (16-10).



الشكل (16-10)

إن خط الانحدار للمجال الأقل مطمور تحت المشاهدات . الجزء المبين على الشكل البياني أعلاه باللون الغامق .

يمكننا مقارنة البواقي مجموع المربعات لنموذجي الانحدار. سوف يدل RSS_1 و RSS_2 للمجال الأدنى و الأعلى على التوالي.

إذا كان حد الخطأ المتجانس يجب أن يكون منتظم الاختلاف بين RSS_1 و RSS_2 . على أي حال إذا كان الانحراف المعياري للتوزيع لحد الخطأ المعياري يكون جزئيا إلى X ، RSS_2 يكون أكبر من RSS_1 . إذا كان ذلك كبيراً والسؤال هل يكون ذو معنوية. نستخدم الاختبار الإحصائي F ، من أجل عدد درجات حرية في الانحدار الأقل n_1 و n_2 الأعلى . (طبيعيا n_1 و n_2 سوف يكون نفسه) .

$$F(n_2,n_1)=\frac{RSS_2/n_2}{RSS_1/n_1}=\frac{13,518,000,000/9}{157,000,000/9}=86.1$$

$$F(9,9)_{crit,0.1\%} = 10.1$$

في هذه الحالة سوف نرفض فرضية العدم لتباين التجانس عند مستوى دلالة 0.1% . من أجل ذلك نحتاج لإيجاد الخيارات لمستقيم الانحدار، حسب طريقة المربعات الصغرى. إذاً لماذا تم تقسيم العينة إلى ثلاثة مجالات؟ لماذا لا نقسم ذلك بالتساوي و نقارن RSS ، ونستخدم فقط النصفين؟! السبب لأن حذف المجال الأوسط يزيد من التفاوت بين التباين والبواقي، وسيكون الحظ الأوفر لقبول فرضية العدم لتجانس التباين. على أي حال نحذف القطاع الأكبر والأصغر سيكون عدد درجات الحرية في العينة الجزئية للانحدار أقل وهذا ويجعل من الصعوبة قبول فرضية العدم . وهكذا تكون التسوية بين حذفنا المجال الكبير والصغير، و على هذا الأساس التجريبي ينصح جولد فيلد و كوانت بحذف الربع من المشاهدات فقط.

10-10- انعدام التجانس : التوزين و الانحدار اللوغاريتمي

HETEROSCEDASTICITY:(WEIGHTED AND LOGARITHMIC REGRESSIONS)

في هذه الفقرة نتبع طريقتين للتعامل مع مشكلة عدم التجانس. سنبدأ من الحالة العامة، حيث التباين لتوزيع حد الخطأ العشوائي في المشاهدة i هو s_{i2}.

$$Y = \beta_0 + \beta_1 X + u$$

و تباين المجتمع هو : $u_i = \sigma_i^2$

إذا كان s_i معلوم في كل مشاهدة ، نستطيع أن نشتق تجانس النموذج بالتقسيم إلى المساواة التالية: $\frac{Y_i}{\sigma_i} = \beta_0 \frac{1}{\sigma_i} + \beta_1 \frac{X_i}{\sigma_i} + \frac{u_i}{\sigma_i}$

تباين المجتمع لحد الخطأ العشوائي في النموذج المعدل يساوي الواحد في كل المشاهدات، و كذلك حد الخطأ العشوائي يكون متجانساً.

في النموذج المعدل ، نجد انحدار Y' على X' و H كما هـو محـدد.
ملاحظة: لا يوجد جزء محدد في النموذج المعدل. b_1 أصبحت معامل الميل من المتغير الصناعي $1/s_i$.

$$\text{population variance of } \left\{\frac{u_i}{\sigma_i}\right\} = \frac{1}{\sigma_i^2} \text{population variance of } u_i$$

$$= \frac{\sigma_i^2}{\sigma_i^2} = 1$$

في النموذج المعدل ، نجد انحدار Y' على X' و H كما هـو محـدد.
ملاحظة: إنه لا يوجد جزء محدد في النموذج المعدل. أصبحت β_0 معامل الميل للمتغير الصناعي $1/s_i$.

$$Y' = \beta_0 H + \beta_1 X' + u'$$

$$Y' = \frac{Y_i}{\sigma_i}, \quad H = \frac{1}{\sigma_i}, \quad X' = \frac{X_i}{\sigma_i}, \quad u' = \frac{u_i}{\sigma_i}$$

النموذج المعدل يمكن وصفه بنموذج الانحـدار المثقـل ؛ لأنـه سـنثقل المشاهدة i بالعامل $1/s_i$.

ملاحظة: سوف نعطي الوزن العالي لمعظم المشاهدات المعتمدة (هكـذا مع القيمة الدنيا لـ s_i).

بالطبع في التجربة لا نعرف قيمة s_i لكل مشاهدة، لكن يمكـن أن تكـون معقولة، ويفترض أن تكون متناسب القياس مع المتغير Z .

$$\sigma_i = \lambda Z_i$$

إذاً في هذه الحالة ، نستطيع عمل نموذج متجانس بالتقسيم على Z .

$$\frac{Y_i}{Z_i} = \beta_0 \frac{1}{Z_i} + \beta_1 \frac{X_i}{Z_i} + \frac{u_i}{Z_i}$$

$$\textbf{Population variance of } \left\{\frac{u_i}{Z_i}\right\} = \frac{1}{Z_i^2}\sigma_i^2 = \frac{\sigma_i^2}{\sigma_i^2 / \lambda^2} = \lambda^2$$

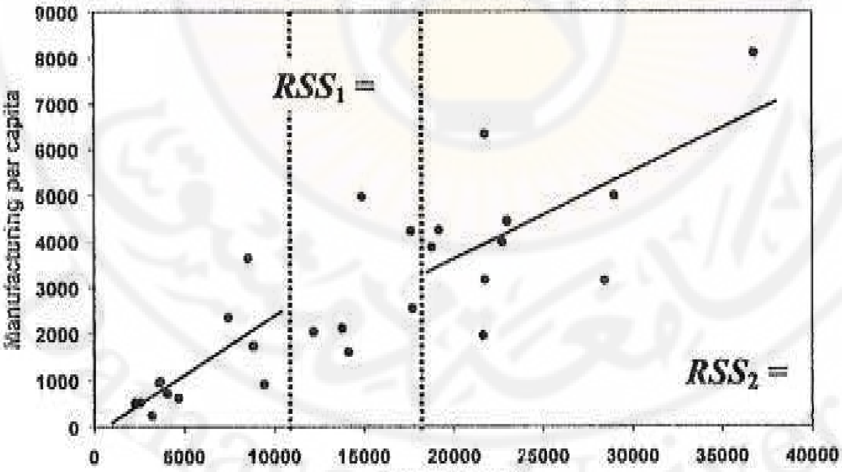
$$Y' = \beta_0 H + \beta_1 X' + u'$$

$$Y' = \frac{Y_i}{Z_i},\ H = \frac{1}{Z_i},\ X' = \frac{X_i}{Z_i},\ u' = \frac{u_i}{Z_i}$$

إن حد الخطأ العشوائي في النموذج المعدل له تباين ثابت λ^2، ونحن لا نحتاج لمعرفة قيمة λ^2 ، ولكن النقطة الحرجة افتراضية يمكن أن تكون ثابتة.

سوف نستخدم مثل هذا الإجراء على بيانات UNIDO على مخرجات قطاع الصناعة و GDP . مخرجات انحدار قطاع الصناعة بالنسبة لرأسمال على GDP يكون أقل انعداماً للتجانس.

نعرض أدناه شكل الانتشار المعدل . هل هو متجانس ؟ في الواقع لا . هذا سيكون جزءاً من انعدام التجانس RSS_2 أكبر بكثير من RSS_1.



الشكل (17-10)

على أي حال العينة الجزئية أصغر ونسبة عالية تحصل على أساس الحظ. تنص فرضية العدم لانعدام التجانس على الرفض فقط عند 0.05 مستوى دلالة.

$$F(n_2, n_1) = \frac{RSS_2 / n_2}{RSS_1 / n_1} = \frac{17,362,000/9}{5,378,000/9} = 3.23$$

$$F(9,9)_{crit,5\%} = 3.18$$

حتى المتغير X نفسه سيكون مستقراً عند توزين المتغير . بعد إجراء اختبار جولد فيلد – كوانت يفترض أن الخطأ المعياري لحد الخطأ العشوائي يكون متناسبا معه.

$$\sigma_i = \lambda X_i$$

ملاحظة: إنه عندما نوزن ذلك نجد أن، الحد b_1 يصبح تقاطع في النموذج المعدل.

$$\frac{Y_i}{X_i} = \beta_0 \frac{1}{X_i} + \beta_1 + \frac{u_i}{X_i}$$

نتابع عندما نفسر نتائج الانحدار، فمعامل الميل المقدر لـ b_1 في النموذج الأصلي، والتفسير يكون كتقدير b_2 .

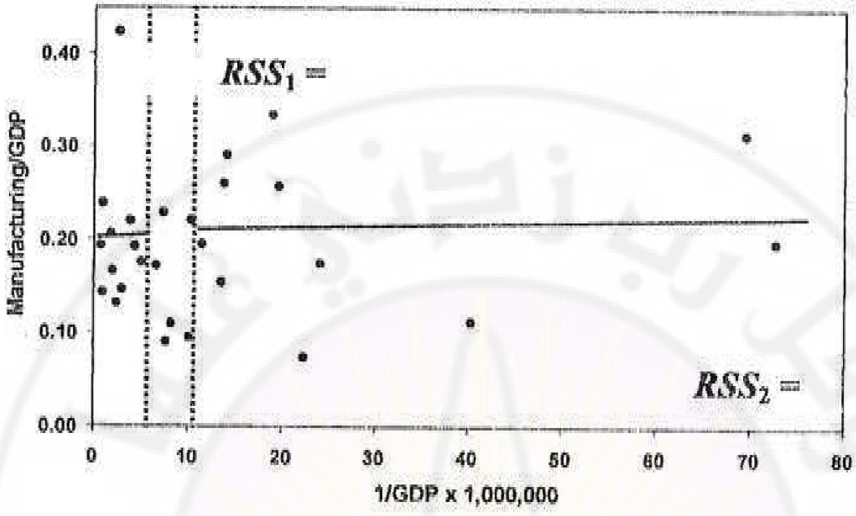
$$\text{population variance of} \left\{\frac{u_i}{X_i}\right\} = \frac{1}{X_i^2}\sigma_i^2 = \frac{\sigma_i^2}{\sigma_i^2 / \lambda^2} = \lambda^2$$

$$Y' = \frac{Y_i}{X_i},\ H = \frac{1}{X_i},\ u' = \frac{u_i}{X_i}$$

$$Y' = \beta_1 H + \beta_2 + u'$$

انظر نجد تتطابق شكل الانتشار الرسم البياني() . هل نحتاج لأي دليل على انعدام التجانس؟

يجب أن نتوقع واحدة على أساس الحظ لنفس فرضية العدم لمجموع مربع البواقي للعينتين الجزئيتين المتماثلتين المغلقتين.



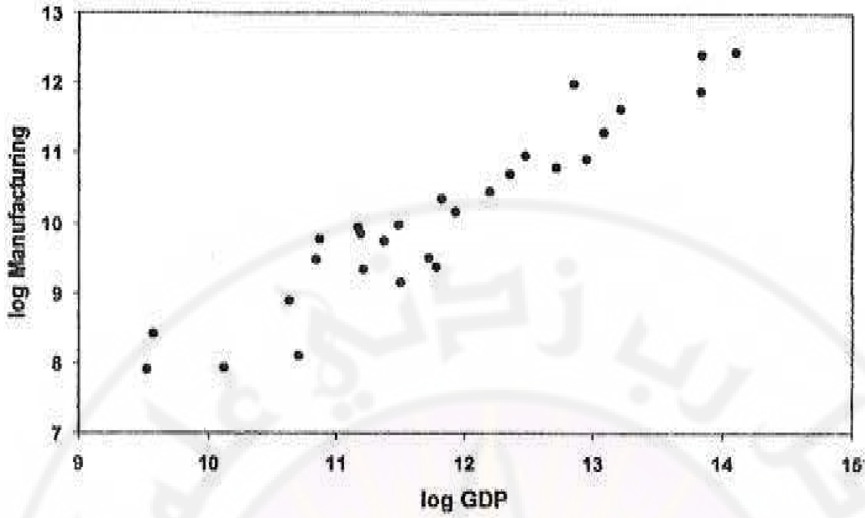
الشكل (18-10)

نتابع ونلاحظ أن F الإحصائية ليست معنوية ، أي لم نتخلص من انعدام التجانس.

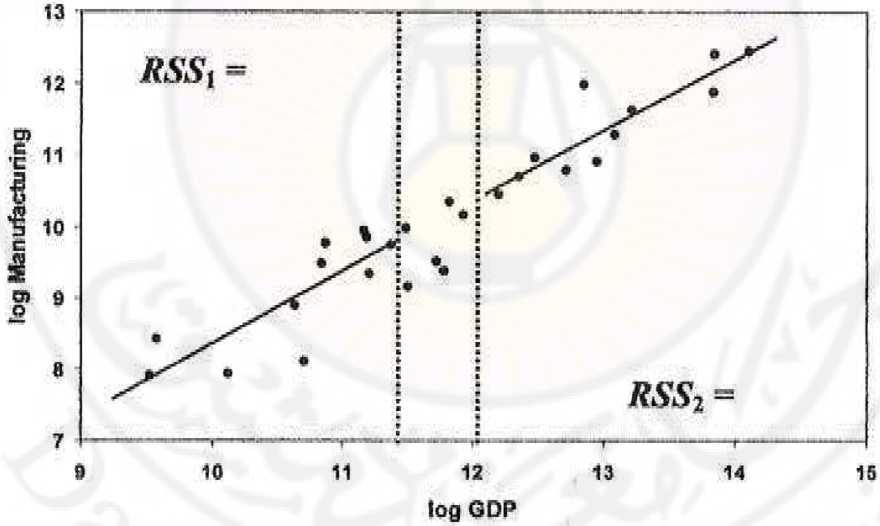
$$F(n_2,n_1)=\frac{RSS_2/n_2}{RSS_1/n_1}=\frac{0.070/9}{0.065/9}=1.08$$

$$F(9,9)_{crit,5\%}=3.18$$

سوف نبحث عن الحل البديل للمشكلة . من الممكــن أن يكــون انعــدام التجانس بسبب تغير الخصائص الرياضية غير المناسبة. نفترض أنــه بشــكل خاص إن العلاقة الحقيقية لها شكل لوغاريتمي. نلاحظ من خلال شكل الانتشار لا يوجد إشارة لانعدام التجانس.



الشكل (10-19)



الشكل (10-20)

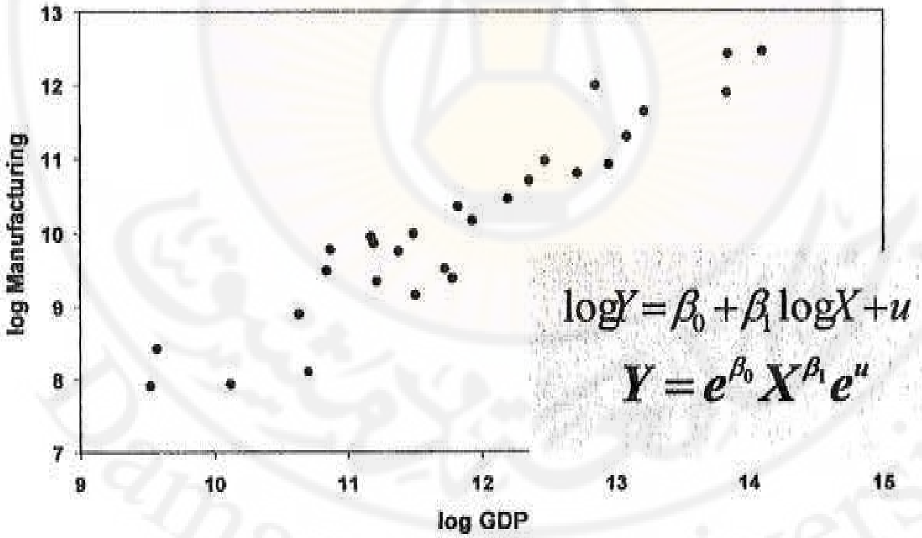
نتأكد من ذلك بإجراء هذا مع اختبار جولد فيلد – كوانت . في هذه الحالة لا يوجد نقطة محسوبة باختبار إحصائي تقديري، ونجد RSS_2 أصغر من RSS_1 وكذلك هذا لا يعني أن RSS_2 أكثر معنوية من RSS_1 .

في هذه الحالة يجب أن نختبر فيما إذا كان هناك دليل على الخطأ المعياري لحد الخطأ العشوائي و إذا كان متناسب عكسياً مع المتغير X . لهذا الفرض F الإحصائية تكون عكس التقليدية.

$$F(n_1,n_2)=\frac{RSS_1/n_1}{RSS_2/n_2}=\frac{2.140/9}{1.037/9}=2.06$$

$$F(9,9)_{crit,5\%}=3.18$$

لن نرفض فرضية العدم لانعدام التجانس. لأن إضافة حد الخطأ العشوائي في نموذج اللوغاريتمي يجب أن يساوي ميل النموذج الأصلي.



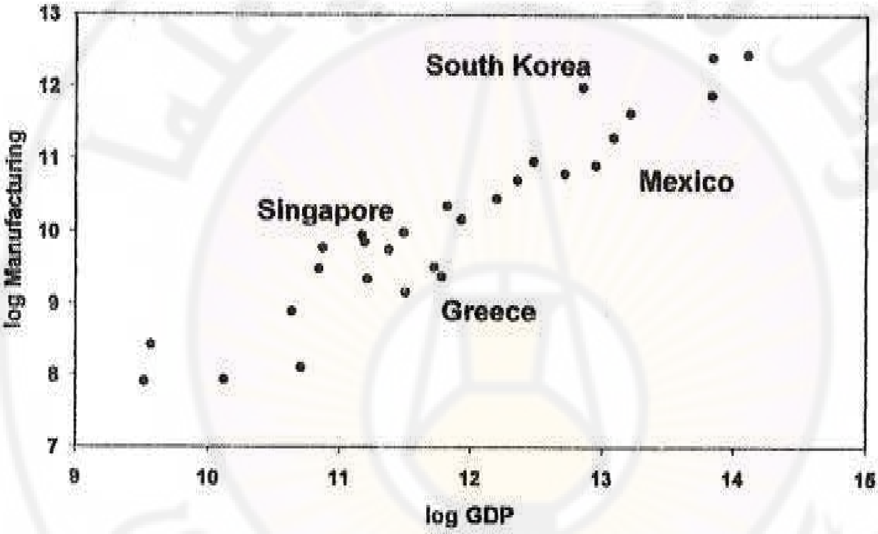
الشكل (10-21)

$$\log Y=\beta_0+\beta_1\log X+u$$

$$Y = e^{\beta_0} X^{\beta_1} e^{u}$$

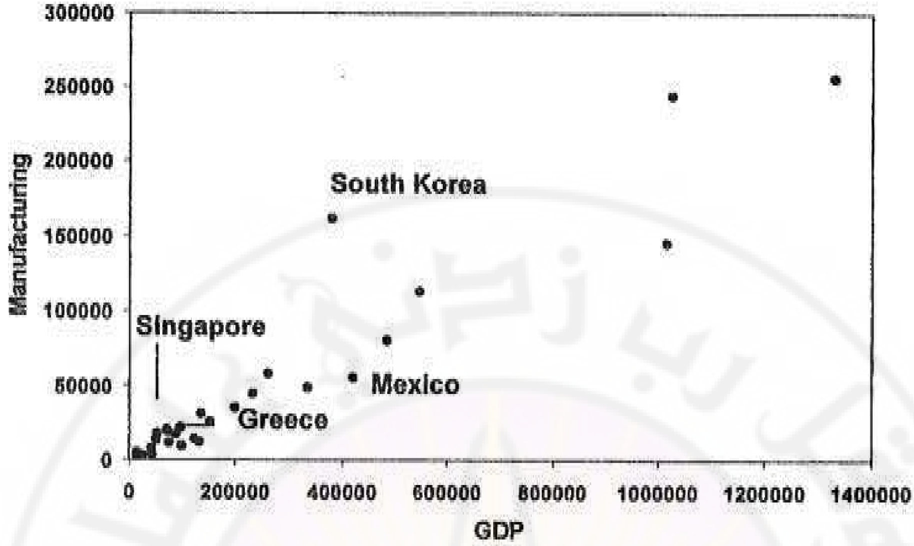
في هذه المتوسطات إن الحجم المطلق للحد الخطأ العشوائي يكون كبير القيمة بالنسبة للمتغير X و صغير لأحدها عندما يسحب شكل الانتشار مع المتغير في الشكل الأصلي .

على سبيل المثال : سنغافورة وكوريا الجنوبية تملك نسبياً قطاعات صناعية كبيرة واليونان و المكسيك نسبياً أصغر منها .



الشكل (10-22)

إن الاختلاف بين كوريا الجنوبية والمكسيك متشابهة عندما يكون الرسم بالشكل اللوغاريتمي، ولكن بين هذه البلدان كبيرة جداً عندما تكون المتغيرات في الرسم بالوحدات الطبيعية.



الشكل (23-10)

سنعرض ملخصاً للبدائل الأربعة للنموذج الانحدار:

$$MANU = 604 + 0.194GDP$$
$$(5700) \quad (0.013) \qquad R^2 = 0.89$$

$$\frac{M\hat{A}NU}{POP} = 612\frac{1}{POP} + 0.182\frac{GDP}{POP}$$
$$(1371) \qquad (0.016) \qquad R^2 = 0.70$$

$$\frac{M\hat{A}NU}{GDP} = 0.189 + 533\frac{1}{GDP}$$
$$(0.019) \quad (841) \qquad R^2 = 0.02$$

$$\log M\hat{A}NU = -1.694 + 0.999\log GDP$$
$$(0.785) \quad (0.066) \qquad R^2 = 0.92$$

في الانحدار المقترح كلما ازداد 1$ مليون في GDP فــإن مخرجــات الصناعة تزداد بمقدار 194000$. وهذا يعني أن الزيادة الهامشية في قطــاع الصناع هو 0.19 من GDP ، ولا بد من الإشارة إلى أن التقاطع لــيس لــه معنى يذكر في هذه العلاقة.

على أي حال إن هذا الانحدار كان موضوع هيمنة انعدام التجانس. بالرغم من تقدير معامل GDP ليكون غير متحيز، ولكنه ليكون نســبياً غيــر دقيــق. وكذلك يكون تأثير التباعد لانعدام التجانس على الأخطاء المعيارية واختبــار F غير مرضي.

أما تقدير معامل الميل في الانحدار الثاني فهو صغيراً. على أي حال هذا الانحدار حسب فرضية العدم لانعدام التجانس يعتبر مقبولاً، ولكن فقط عند 5% مستوى دلالة.

أما في نموذج الانحدار الثالث الموزن بـــ GDP ، يصبح تقدير معامل الميل الأصلي غير مرضي.

لن نرفض فرضية العدم في هذا النموذج. مبدئيا سنسلم للتقديرين السابقين من المعلمات وسننجز الاختبارات لهما.

أما النموذج اللوغاريتمي أيضا لن نرفض فرضية العدم، و كذلك نملــك نموذجين حيث اختبارات جولد فيلد – كوانت محققة. ولكن ما هــو الأفضــل؟ فكر بذلك قليلاً!

من المحتمل نختار النموذج اللوغاريتمي لأن R^2 مرتفعة بشــكل ملفــت، على أي حال في هذا المثال يوجد اختيار بين النموذجين الثالث والرابع ، لأنــه لهما التفسير نفسه.

في النموذج الثالث نجد اختبار t الإحصائية للحد GDP صغيرة، ويظهر ليكون المتغير ذا صلة. يخبرنا النموذج بأن مخرجات قطاع الصناعة بالنسبة لـGDP ثابتة، لذلك R^2 تأثيره صفر.

ويخبرنا النموذج الرابع أن مرونة مخرجات قطاع الصناعة مع GDP يساوي الواحد. بكلمات أخرى مخرجات قطاع الصناعة يزداد نسبياً مع GDP وبقاء الحد الثابت نسبياً كما هو.

نعيد المسارات اللوغاريتمية إلى شكلها بالوحدات الطبيعية نحصل على الشكل المبين. وهو مشابه للنموذج الثالث تدلنا ضمنياً على أن حسابات القطاع الصناعي تشكل حوالي 0.18 من القيمة الهامشية لـ GDP .

$$MANU = e^{-1.694} GDP^{0.999} = 0.184 GDP^{0.999}$$

لقد ذكرنا سابقاً إن تحليل الانحدار يعتمد على عدة فروض، وعندها ذكرنا أن أهم هذه الفروض التي يقوم عليها النموذج الخطي البسيط والعام هو:

$$\text{Var}(U) = \sigma^2 I$$

أي أن تباين المتغير العشوائي يساوي إلى التباين الثابت للمجتمع الإحصائي مضروباً بالمصفوفة المتطابقة I . لتوضيح ذلك نكتب مصفوفة التباين والتغاير للمتغير العشوائي U .

إذا كان K = 4 متغيراً مستقلاً

عندئذ U :

$$\mathbf{U} = \begin{bmatrix} \mathbf{U}_1 \\ \mathbf{U}_2 \\ \mathbf{U}_3 \end{bmatrix}$$

أما التباين للمتغير العشوائي :

$$\text{Var}(U) = E(U U')$$

$$\text{Var}(U) = E\begin{bmatrix} U_1 \\ U_2 \\ U_3 \end{bmatrix}\begin{bmatrix} U_1 & U_2 & U_3 \end{bmatrix}$$

$$\mathbf{Var}(\mathbf{u}) = \begin{bmatrix} \mathbf{Var}(\mathbf{U}_1) & \mathbf{Cov}(\mathbf{U}_1, \mathbf{U}_2) & \mathbf{Cov}(\mathbf{U}_1, \mathbf{U}_3) \\ \mathbf{Cov}(\mathbf{U}_2, \mathbf{U}_1) & \mathbf{Var}(\mathbf{U}_2) & \mathbf{Cov}(\mathbf{U}_2, \mathbf{U}_3) \\ \mathbf{Cov}(\mathbf{U}_3, \mathbf{U}_1) & \mathbf{Cov}(\mathbf{U}_3, \mathbf{U}_2) & \mathbf{Var}(\mathbf{U}_3) \end{bmatrix}$$

نظراً لأن تحليل الانحدار يقوم أساساً على افتراض ثبات تباين المتغير العشوائي، وعلى افتراض انعدام العلاقة بين القيم المتتالية للمتغير العشوائي :

$$\text{Var}(U_1) = \text{Var}(U_2) = \text{Var}(U_3) = \sigma^2$$

$$\text{Cov}(U_j, U_i) = 0 \qquad j \neq i$$

فإن المصفوفة أعلاه تصبح كالآتي :

$$\mathbf{Var}(\mathbf{U}) \begin{bmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \sigma^2 \mathbf{I}$$

أي أن تباين المتغير العشوائي يساوي حاصل ضرب التباين الثابت للمتغير التابع في المجتمع الإحصائي بالمصفوفة الأحادية المتطابقة .

الجدير بالذكر، أنه كثيراً ما يصادف الباحث في أبحاث الاقتصاد القياسي التطبيقية أن يكون : $\text{Var}(U) \neq \sigma^2 I$

أي أن الباحث كثيراً ما يخالف الفروض الأساسية التي يعتمدها من أجل إدخال المتغير العشوائي في تحليل الانحدار، فيطرح السؤال الهام عندئذ نفسه للإجابة، وماذا عن الخصائص الإحصائية لمعاملات الانحدار ؟ يمكن القول أن قيمة تباين المتغير العشوائي قد تختلف في المصفوفة $\sigma^2 I$ وذلك في حالتين معروفتين بالاقتصاد القياسي :

في المصفوفة $\sigma^2 I$ ، لكن تكون القيمة خارج القطر الرئيسي غير مساوية للصفر، بسبب وجود التغاير بين القيم المتتالية للمتغير العشوائي، وتسمى حالة انعدام التجانس (Heteroscedasticity)

لابد من الإشارة إلى أنه إذا كان النموذج يعاني من مخالفة الحالة الأولى والحالة الثانية، فإن النموذج سوف يعطينا تقديرات خطية غير متميزة (Unbiased Linear Estimates) لمعاملات الانحدار، ولذلك هذه التقديرات غير كفوءة (Inefficient) .

سواء استخدمنا توفيق خط الانحدار بطريقة المربعات الصغرى، أو بطريقة أخرى وبالتالي فإن الخطأ المعياري لتوزيع المعاينة لمعاملات الانحدار، سيكون كبيراً، مما يؤدي إلى إعطاء نتائج مضللة لكل من فترة الثقة، واختيار الفروض . فيزيد احتمال ارتكاب الخطأ من النوع الأول (Type I error) أو من النوع الثاني (Type I error) في اتخاذ القرار الإحصائي .

لذلك يتطلب منا كباحثين، وقبل استخدام أسلوب المربعات الصغرى من إجراء التحويلات (Transformation) المناسبة لمتغيرات النموذج الاقتصادي، حتى نتمكن من الحصول على معاملات الانحدار الكفوءة .

تمارين محلولة

1- يقدم الجدول التالي نتائج الانحدار المتعدد و البسيط للمتحول LGFDHO (أي: لوغاريتم المعدل السنوي للإنفاق الأسري على الطعام المُستهلك ضمن المنزل) على LGEXP (أي: لوغاريتم المعدل السنوي للإنفاق الأسري)، و على LGSIZE (أي: لوغاريتم عدد أفراد الأسرة) و ذلك باستخدام نموذج من 868 أسرة حسب مسح الإنفاق الاستهلاكي لعام 1995، مع العلم أن معامل الارتباط من أجل LGEXP و LGSIZE هو: 0.45.

في ضوء ما سبق، اشرح التباين في معاملات الانحدار.

البيان	(1)	(2)	(3)
LGEXP	0.29	0.48	–
	(0.02) (0.02)	-	-
LGSIZE	0.49	–	0.63
	(0.03)	(0.02)	-
constant	4.72	3.17	7.50
	(0.22)	(0.24)	(0.02)
R2	0.52 0.31	0.42	-

2- يعتقد باحث اجتماعي أن مستوى النشاط في اقتصاد الخفي(الظل)، Y يعتمد على أمرين:

- مستوى العبء الضريبي X، الذي يؤثر بشكل إيجابي .
- الإنفاق الحكومي Z ، الذي يؤثر بشكل سلبي ويؤدي لتثبيط نشاط اقتصاد الظل(الخفي).

ويجب على Y أن تعتمد على X و Z معاً، و جميع البيانات العالمية في القطاعات تكون حول Y,X,Z وكلها مقدرة بالمليون دولار أمريكي وتم تطبيقها على عينة من 30 دولة صناعية، و على عينة أخرى من 30 دولة نامية، و قام باحثنا ببناء علاقات الانحدارات التالية:

1. Log Y على كل من Log X و Log Z.
2. Log Y على Log X فقط. (لكل عينة)
3. Log Y على Log Z فقط. (لكل عينة)

إن النتائج مدرجة في الجدول التالي: (تم وضع الأخطاء المعيارية بين قوسين)

	Industrialized Countries			Developing Countries		
	(1)	(2)	(3)	(1)	(2)	(3)
log X	0.699 (0.154)	0.201 (0.112)	–	0.806 (0.137)	0.727 (0.090)	–
log Z	–0.646 (0.162)	–	–0.053 (0.124)	–0.091 (0.117)	–	0.427 (0.116)
constant	–1.137 (0.863)	–1.065 (1.069)	1.230 (0.896)	–1.122 (0.873)	–1.024 (0.858)	2.824 (0.835)
R2	0.44	0.10	0.01	0.71	0.70	0.33

إذا علمت أن X مرتبطة بشكل إيجابي مع Z في كلتا العينتين ، نتيجة تنفيذ الاختبار الإحصائي المناسب . اكتب تقريرا قصيرا تنصح فيه الباحث الاجتماعي عن كيفية تفسير النتائج

3- لدى الباحثة بيانات عن إنتاج كل عامل من العمال Y ورأس المال لكل عامل من العمال K وكلاهما يقاس بآلاف الدولارات من اجل 50 شركة في صناعة النسيج عام 2001. تفترض الباحثة إن إنتاج كل عامل يعتمد على

رأس المال لكل عامل وربما أيضا على المستوى التكنولوجي للشركة (TECH):

$$Y=\beta_1+\beta_2K+\beta_3TECH+U$$

بما أن حد التوزيع U. الباحثة غير قادرة على قياس المستوى التكنولوجيTECH فقررت استخدام النفقات لكل عامل في البحث والتطوير عام 2001 في شركة R&D كبديل عنها وقد وضعت علاقة الانحدار التالية (تم وضع الأخطاء المعيارية ضمن أقواس).

$$\hat{Y}=1.02+0.32K \qquad R^2=0.749$$
$$(0.45)\ (0.04)$$

$$\hat{Y}=0.34+0.29K+0.05\ R\&D \qquad R^2=0.750$$
$$(0.61)\ (0.22)\ (0.15)$$

أن ارتباط معلمات K مع R&D هو(0.92).ناقش نتائج الانحدار :

-افترض أن Y تعتمد على K و TECH كليهما .

-افترض أن Y تعتمد على K فقط.

4- تظهر نتائج الانحدار للمتغير LGFDHO (أي: لوغاريتم المعدل السنوي للإنفاق الأسري على الطعام المُستهلك ضمن المنزل) على LGEXP (أي: لوغاريتم المعدل السنوي للإنفاق الأسري)، و على LGSIZE (أي: لوغاريتم عدد أفراد الأسرة) و ذلك باستخدام نموذج من 868 أسرة حسب مسح الإنفاق الاستهلاكي لعام 1995 .

ثم تم اخذ انحدار LGFDHOPC (لوغاريتم الإنفاق على الطعام للشخص الواحد) (FDHO/SIZE) علىLGEXPPC (لوغاريتم الإنفاق الكلي للشخص)(EXP/SIZE). في التوصيف الثالث تم اخذ انحدار LGFDHOPC على LGEXPPC و LGSIZE .

```
. reg LGFDHO LGEXP LGSIZE

      Source |       SS       df       MS              Number of obs =     868
-------------+------------------------------           F(  2,   865) =  460.92
       Model |  138.776549     2  69.3882747           Prob > F      =  0.0000
    Residual |  130.219231   865  .150542464           R-squared     =  0.5159
-------------+------------------------------           Adj R-squared =  0.5148
       Total |  268.995781   867  .310260416           Root MSE      =    .388

------------------------------------------------------------------------------
      LGFDHO |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       LGEXP |   .2866813   .0226824    12.639   0.000      .2421622    .3312003
      LGSIZE |   .4854698   .0255476    19.003   0.000      .4353272    .5356124
       _cons |   4.720269   .2209996    21.359   0.000      4.286511    5.154027
------------------------------------------------------------------------------

. reg LGFDHOPC LGEXPPC

      Source |       SS       df       MS              Number of obs =     868
-------------+------------------------------           F(  1,   866) =  313.04
       Model |  51.4364364     1  51.4364364           Prob > F      =  0.0000
    Residual |  142.293973   866  .164311747           R-squared     =  0.2655
-------------+------------------------------           Adj R-squared =  0.2647
       Total |   193.73041   867  .223449146           Root MSE      =  .40535

------------------------------------------------------------------------------
    LGFDHOPC |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     LGEXPPC |    .376283   .0212674    17.693   0.000      .3345414    .4180246
       _cons |   3.700667   .1978925    18.700   0.000      3.312262    4.089072
------------------------------------------------------------------------------

. reg LGFDHOPC LGEXPPC LGSIZE

      Source |       SS       df       MS              Number of obs =     868
-------------+------------------------------           F(  2,   865) =  210.94
       Model |  63.5111811     2  31.7555905           Prob > F      =  0.0000
    Residual |  130.219229   865  .150542461           R-squared     =  0.3278
-------------+------------------------------           Adj R-squared =  0.3263
       Total |   193.73041   867  .223449146           Root MSE      =    .388

------------------------------------------------------------------------------
    LGFDHOPC |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     LGEXPPC |   .2866813   .0226824    12.639   0.000      .2421622    .3312004
      LGSIZE |  -.2278489   .0254412    -8.956   0.000     -.2777826   -.1779152
       _cons |   4.720269   .2209996    21.359   0.000      4.286511    5.154027
------------------------------------------------------------------------------
```

المطلوب:

أ- تفسير نتائج النموذج الثاني.

ب- إنجاز اختبار F عند التقييد.

ج- إنجاز اختبار t عند التقييد.

د- لخص نتائجك من تحليلك لنتائج الانحدار.

الفصل الحادي عشر

الارتباط الذاتي

Autocorrelation

11-1- مقدمة:

إن مخالفة فرضية انعدام التغاير بين القيم المتتالية للمتغير العشوائي ، تؤدي إلى معاناة النموذج الاقتصادي من وجود الارتباط الذاتي (Autocorrelation)، علماً أن أكثر البيانات الاقتصادية للسلاسل الزمنية تعاني من مشكلة وجود ارتباط قيم المتغير العشوائي في الفترة t ، وبين قيم المتغير العشوائي في الفترات السابقة للفترة t مثل t-1 ، t-2 ... إلخ . دعنا للتبسيط نفترض بإمكاننا أن نقدر إنتاج الأسمنت (Y_t) في الفترة t بالاعتماد على عدد المستفيدين من الجمعيات السكنية(X) من خلال علاقة الانحدار التالية:

$$Y_t = \beta_0 + \beta_1 X_t + U_t$$

و يقيس المتغير العشوائي U_t أثر المتغيرات الأخرى(غير عدد الجمعيات السكنية) على إنتاج الأسمنت، والتي لم نتمكن من قياسها وإدخالها بشكل صريح في النموذج ، ومثال ذلك الوضع الاقتصادي السائد في الفترة t ... إلخ .

[1]نظراً لأن الوضع الاقتصادي السائد في الفترة t ، يتأثر إلى حد بعيد بالوضع الاقتصادي الذي كان سائداً في الفترات t-1 ، t-2 ، t-3 ...... إلخ .

---

[1] (سلبيات المترتبة على تقدير المربعات الصغرى : مما تقدم يتبين أن في حال وجود ارتباط ذاتي لا تتحقق العلاقة : $E(U_j , U_i) = \sigma^2 I$ والتي كانت الأساس في جعل تقدير المربعات الصغرى أفضل تقدير خطي غير متحيز (Blue). ويترتب على ذلك عدد من السلبيات:

ويتأثر إلى حد بعيد بالوضع الاقتصادي الذي سوف يسود في الفترة t+1 t+2 ، لذلك لم يعد باستطاعتنا افتراض انعدام العلاقة بين القيم U_t ، U_{t-1} ، U_{t-2} ، U_{t+1} ..... الخ . وبما أن استخدام طريقة المربعات الصغرى في تحليل الانحدار يفترض انعدام وجود مثل هذه العلاقة ، لذلك يتوجب على الباحث، وقبل تطبيق طريقة المربعات الصغرى على البيانات الزمنية ، أن يختبر وجود الارتباط المتسلسل الذاتي في بياناته. فإن وجد علاقة بين القيم المتتالية للمتغير العشوائي توجب عليه عندئذ تحويل المتغيرات إلى صيغة أخرى، تكون نتيجتها انعدام مثل هذه العلاقة.

إذاً الارتباط الذاتي هو حالة خاصة من قانون الارتباط ، ويعزى الارتباط الذاتي بين U_j , U_i إلى أن العلاقة ليست بين متغيرين مختلفين أو أكثر ولكن بين القيم المتتابعة لنفس المتغير.

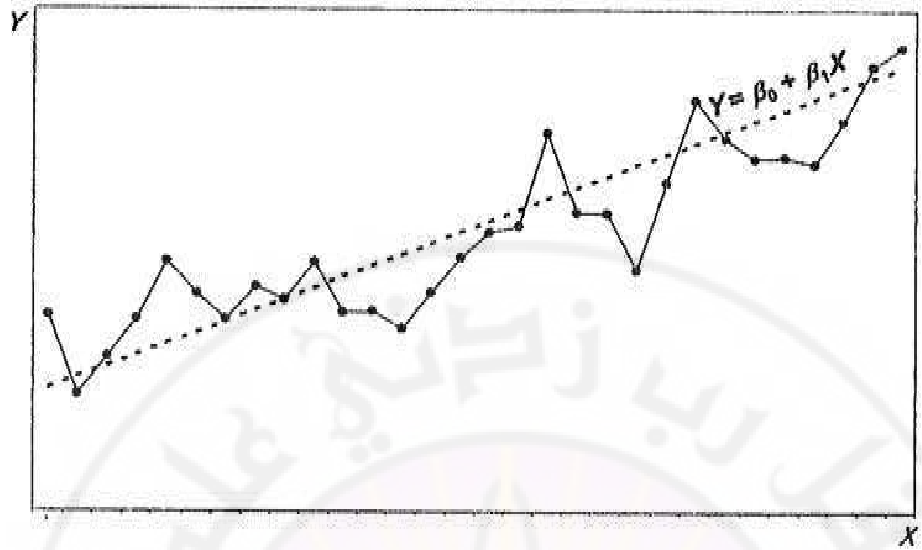
من الشكل البياني (1-11) ، نلاحظ أن القيم المتعاقبة متشابهة بالإشارة سواء بالإيجاب أوالسلب ، وهذا ما يسمى بالارتباط الذاتيautocorrelation الموجب.

1) لا يؤثر الارتباط الذاتي في خاصية عدم التحيز، وذلك لاعتمادها الفرضية التالية: $E(U_t) = 0$

2) لا يؤثر الارتباط الذاتي في الاتساق (Consistency) .

3) يقلل الارتباط الذاتي من الكفاءة (Efficiency) .

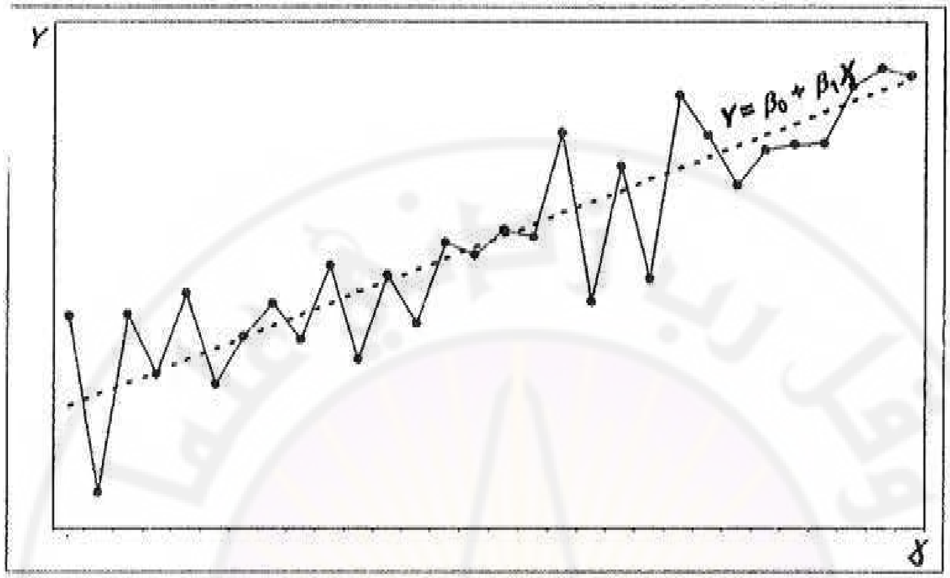
4) في حالة وجود ارتباط ذاتي تكون صيغة تقدير التباين وفق طريقة المربعات الصغرى الاعتيادية (متحيزة سلباً (Negative Bias) في مجال التحليل الاقتصادي بسبب الارتباط الذاتي الموجب Positive Autocorrelation). الأمر الذي يسبب رفض فرضية العدم في بعض الحالات بصورة غير صحيحة).



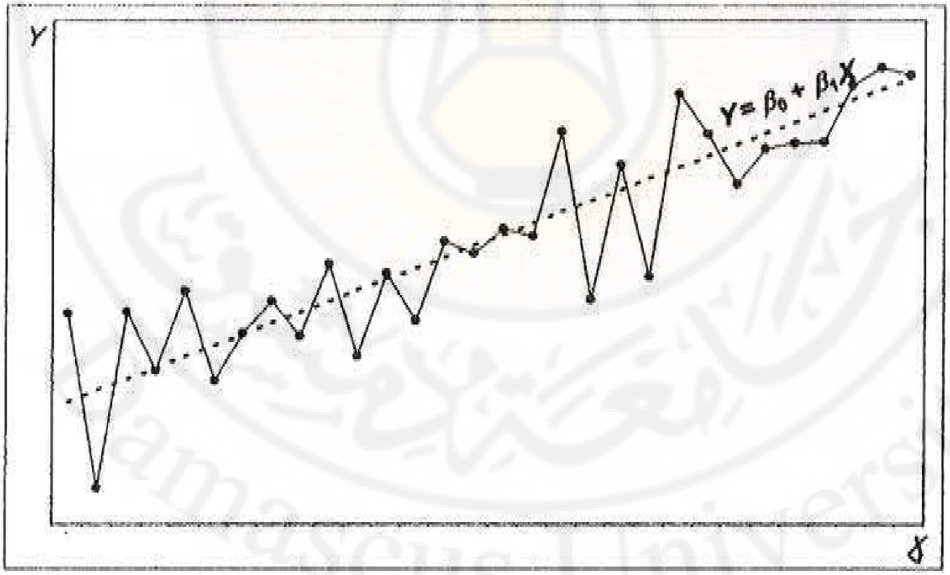
الشكل (1-11)

أما الشكل البياني (1-11) ، فإننا نلاحظ أن القيم متذبذبة ، فالقيمة الموجبة يتلوها قيمة سالبة ، والقيمة السالبة يتلوها قيمة موجبة وهذا يعتبر مثال على واضح على الارتباط الذاتي السالب.

أما الشكل البياني (2-11) ، فإننا نلاحظ أن القيم متذبذبة ، فالقيمة الموجبة يتلوها قيمة سالبة ، والقيمة السالبة يتلوها قيمة موجبة وهذا يعتبر مثال على واضح على الارتباط الذاتي السالب.



الشكل (2-11)



الشكل (3-11)

أما الشكل البياني (3–11) ، فإننا نلاحظ أن القـيم متذبذبـة ، فالقيمـة الموجبة يتلوها قيمة سالبة ، والقيمة السالبة يتلوها قيمة موجبة وهذا يعتبر مثال على واضح على الارتباط الذاتي السالب.

يعتمد الارتباط ذاتي لــ u_t على قيم التباطؤ نفسها وهو الحالة الأول لأنه يعتمد على القيم السابقة. وتعتمد u_t أيضاً على e_t البواقي حسـب الـزمن t وحتى يتم تحديد زمن t.

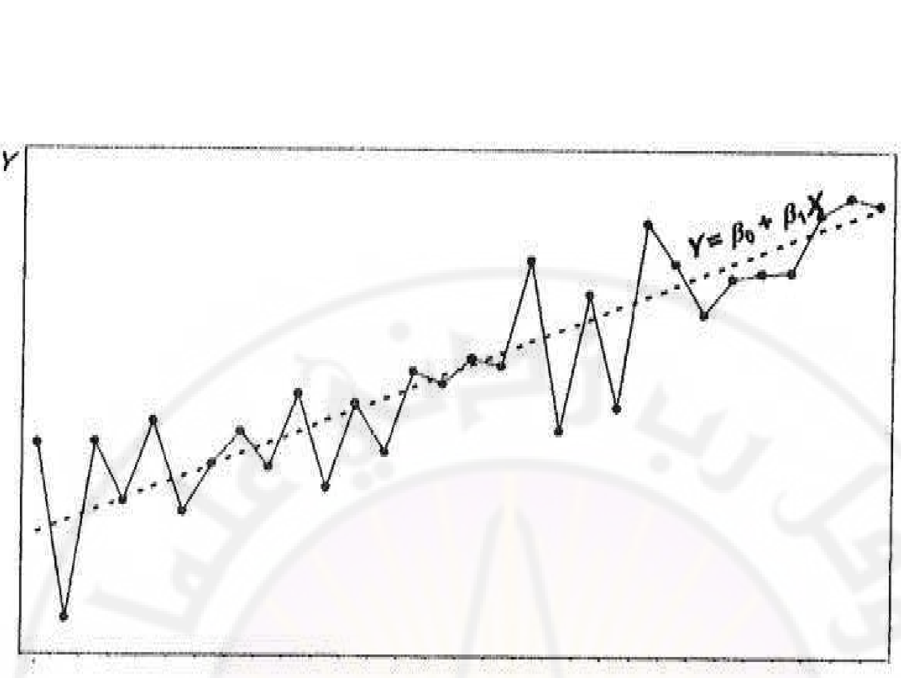
الحالة الأولى: الارتباط الذاتي من الدرجة الأولى AR(1) :

$$u_t = \rho u_{t-1} + \varepsilon_t$$

سنعرض مثالاً معقداً نوع ما حول الارتباط الذاتي المتبـاطئ ، والـذي يصف حالة التباطؤ من المرتبة الخامسة ، لأنه يعتمد على القيم المتباطئ لــ u_t لخمس سنوات.

الحالة الخامسة: الارتباط الذاتي من الدرجة الخامسة AR(5) :

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \rho_3 u_{t-3} + \rho_4 u_{t-4} + \rho_5 u_{t-5} + \varepsilon_t$$



الشكل البياني رقم (4-11)

إنّ النوع الرئيسي الآخر للارتباط الذاتي هو المتوسط المتحرك ، حيث أنّ تعبير الخطأ العشوائي بشكل علاقة خطيّة من المشاهدات و عدد محدود من الوحدات السابقة.

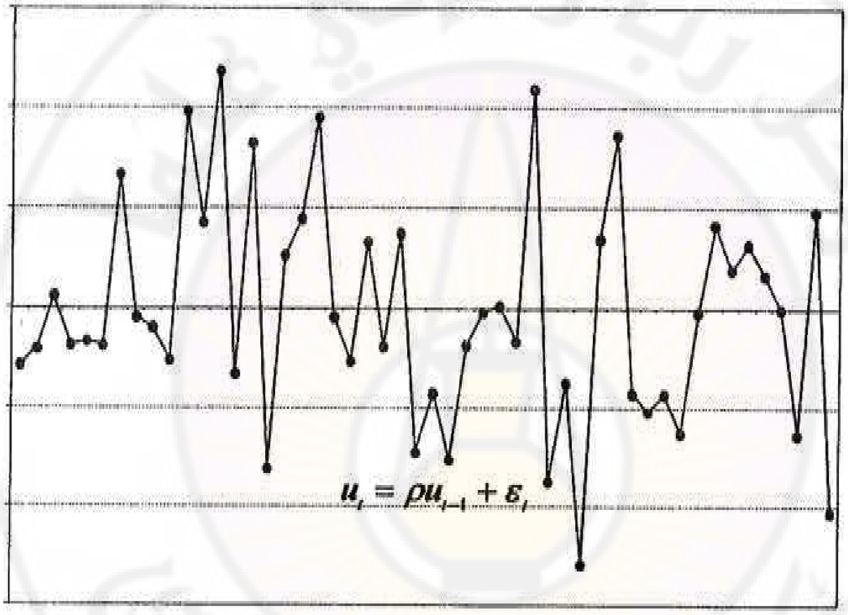
الحالة الثالثة: المتوسط المتحرك الارتباط الذاتي من الدرجة الثالثة(3) MA:

$$u_t = \lambda_0\varepsilon_t + \lambda_1\varepsilon_{t-1} + \lambda_2\varepsilon_{t-2} + \lambda_3\varepsilon_{t-3}$$

هذا المثال يصف الحالة الثالثة كمتوسطات متحركة للارتباط الذاتي(3)MA، يدلّ على لأنه يعتمد على الحالات الثلاثة السابقة بالإضافة إلى الحالية.

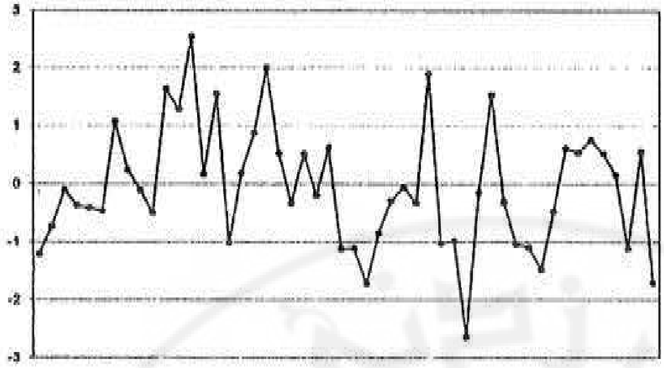
يمكن إعطاء أمثلة عن هذه السلسلة، بتوليد الأنماط عندما يكون حد الخطأ العشوائي خاضع للارتباط الذاتي(1)AR ، ويمكن برهان هذا من خلال تمثيل البواقي بالرسوم البيانية في انحدار السلاسل الزمنية.

سنَستعمل 50 قيمة مستقلة من e، أخذت مِن توزيع طبيعي بمتوسط 0 ، ونولد سلسلة لـــ u مستخدمين قِيَمَ مختلفة ρ . لنبدأ مَع ρ تَساوي إلى 0، لذا ليس هناك ارتباط ذاتي. وسوف نَزِيد ρ بشكل تدريجي في عـــدد مـــن الخطـــوات بالقيمة 0.1 ، ونلاحظ ماذا سيحدث.



الشكل البياني رقم(5-11)

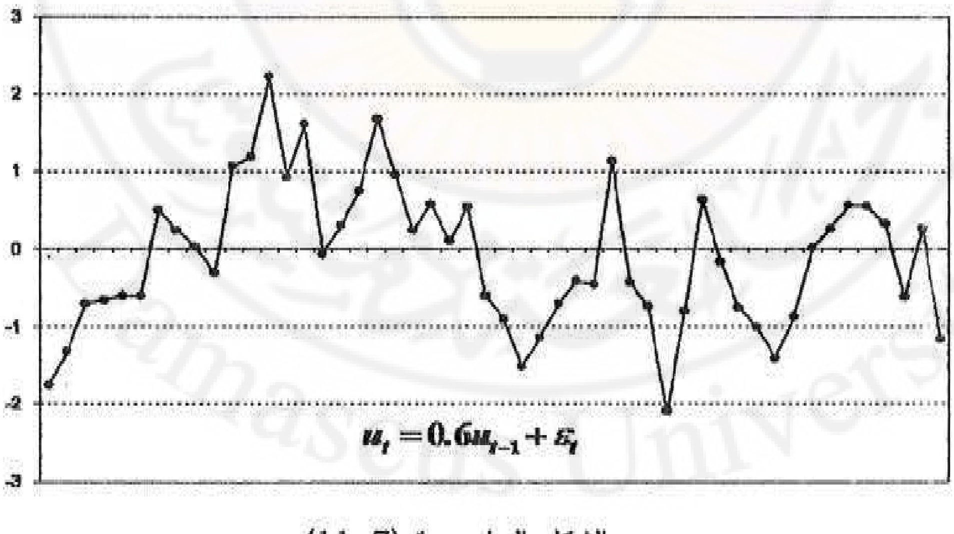
نجد انه عندما ρ تَساوي إلى 0.3, أن نمط الارتباط الذاتي إيجابي كما هو ظاهر في الشكل البياني رقم(5-11).



$$u_t = 0.3u_{t-1} + \varepsilon_t$$

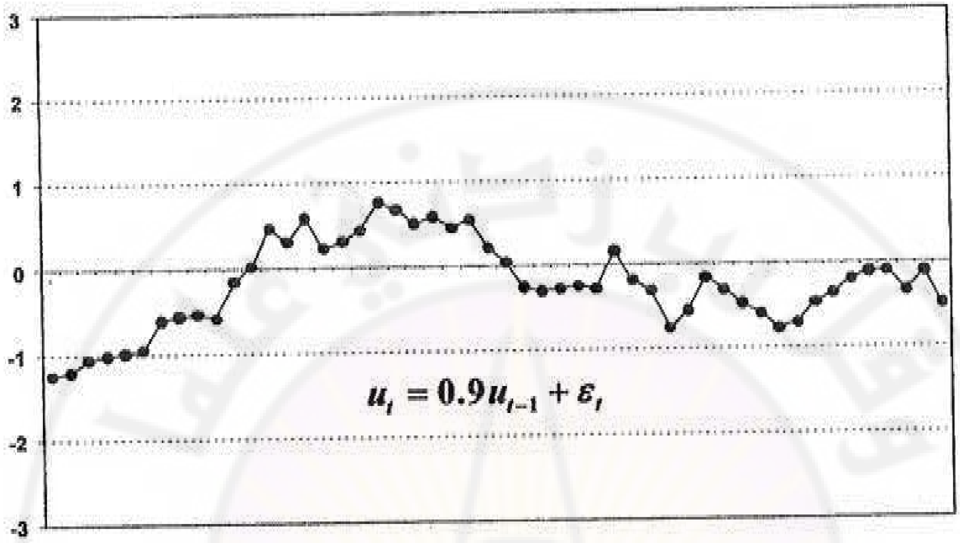
الشكل البياني رقم(6-11)

كذلك من الواضح أنه عندما ρ تساوي القيمة 0.6، يكون الحد u خاضع للارتباط الذاتي الإيجابي،حيث تميل القيم الإيجابية إلى أن تكون متبوعة بواحدة إيجابية والقيم السلبية بواحدة سلبية ، انظر الشكل البياني رقم(6-11).



الشكل البياني رقم(7-11)

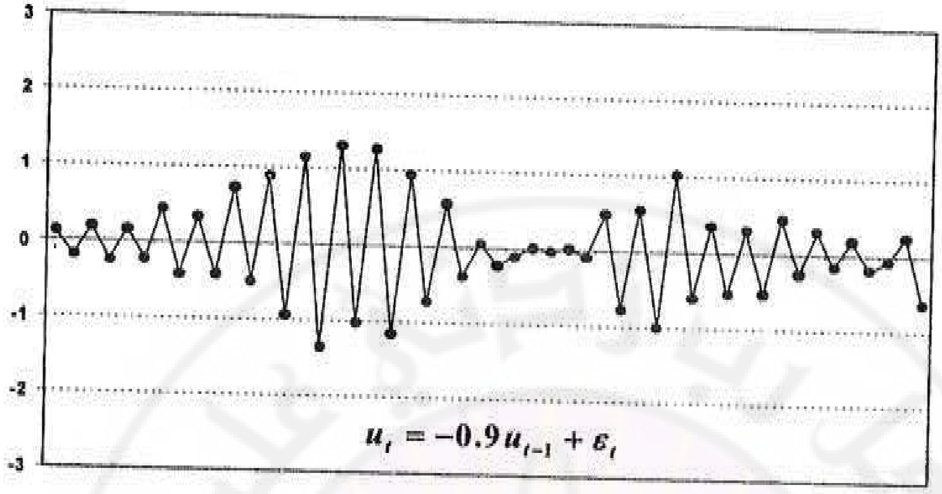
وعندما تصبح ρ مساوية إلى 0.9، فإن القيم في السلسلة تصبح لها الإشارة نفسها على طول والعودة الـ0 أصبح ضعيفاً.



الشكل البياني رقم(8-11)

إنّ هذه العملية أصبحت تقترب لما يعرف بالخطوة العشوائية، حيث أنّ ρ مساوي إلى 1 والعملية تصبح غير مستقرة. تحديد الخطوات العشوائية وحالة عدم الاستقرار سنتعرض لها في الفصل القادم . في الوقت الحاضر سنفترض أن: $|\rho| < 1$. بالطريقة نفسها نبحث مشكلة الارتباط الذاتي السالب، مبتدئين بالمجموعة نفسها المكونة من 50 قيمة لـ e_t موزّعة بشكل مستقل.

نجد أن نمط الارتباط الذاتي سالب كما هو واضح بشكل جلي عندما قيمة $\rho = -0.9$ في الشكل البياني رقم(8-11).



الشكل البياني رقم(9-11)

لنأخذ مثال تطبيقي على ذلك :

مثال (1-11): ليكن لدينا البيانات التالية عن إجمالي الإنفاق DPI والإنفاق على الغذاء FOOD إلى الدخل وكذلك الأسعار النسبية PTPE ، في بلد ما خلال الفترة من 1974-2008 .

Year	DPI	FOOD	PTPE	RESID
1973	1629.1	319.5	22.8	0.0088
1974	1664.8	322.7	23.2	0.0045
1975	1719.8	326.9	23.4	-0.0021
1976	1793.8	329.6	23.7	-0.0206
1977	1852.5	332.6	24.0	-0.0295
1978	1980.7	346.3	24.3	-0.0282
1979	2096.5	366.6	24.7	-0.0042
1980	2198.9	381.2	25.3	0.0135
1981	2296.7	388.0	26.0	0.0002
1982	2392.3	406.0	27.0	0.0200
1983	2469.1	414.7	28.2	0.0246
1984	2567.9	425.8	29.5	0.0316
1985	2662.5	427.4	30.8	0.0095
1986	2771.9	438.7	31.9	0.0143
1987	2959.4	436.3	33.6	-0.0106
1988	2937.7	426.9	37.0	-0.0145
1989	2986.4	437.8	40.0	0.0005
1990	3094.8	462.1	42.3	0.0254
1991	3185.8	470.2	45.1	0.0251
1992	3345.3	472.8	48.4	0.0075
1993	3436.2	479.1	52.8	0.0082
1994	3466.0	480.8	58.5	0.0020
1995	3544.8	479.5	63.7	-0.0171
1996	3576.0	486.4	67.4	-0.0145
1997	3668.8	499.9	70.5	-0.0096
1998	3905.3	510.6	73.1	-0.0259
1999	4009.3	521.4	75.8	-0.0252
2000	4135.8	532.2	78.0	-0.0234
2001	4170.8	539.8	81.0	-0.0147
2002	4316.3	560.1	84.3	0.0008
2003	4393.2	568.2	88.4	0.0064
2004	4468.7	578.2	92.9	0.0136
2005	4486.5	580.4	96.8	0.0106
2006	4613.7	583.9	100.0	-0.0065
2007	4666.2	597.1	102.6	0.0067
2008	4775.6	609.5	105.1	0.0127

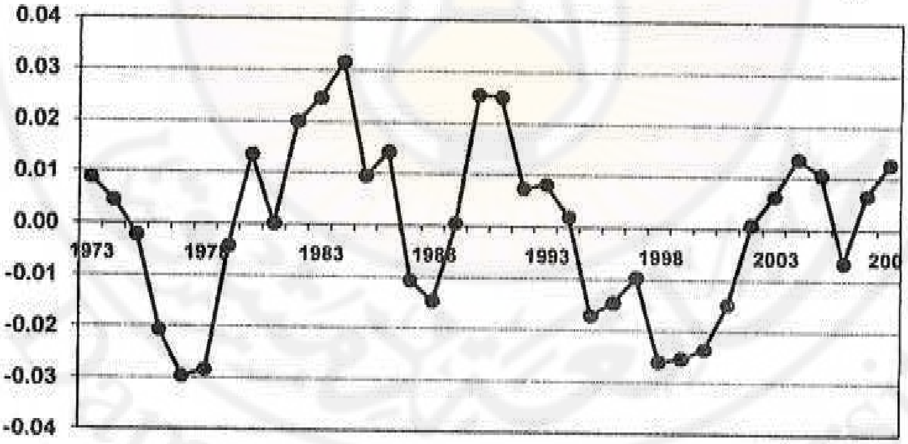
لنقوم بناء انحدار البواقيRESID اللوغاريتمي لإجمالي الإنفاق على الغذاء منسوبا إلى الدخل والسعر النسبي، لهذه البيانات. فنحصل:

Dependent Variable: LGFOOD
Method: Least Squares
Sample: 1959 1994
Included observations: 36

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.658875	0.278220	9.556745	0.0000
LGDPI	0.605607	0.010432	58.05072	0.0000
LGPRFOOD	-0.302282	0.068086	-4.439712	0.0001

R-squared	0.992619	Mean dependent var	6.112169
Adjusted R-squared	0.992172	S.D. dependent var	0.193428
S.E. of regression	0.017114	Akaike info criter	-5.218197
Sum squared resid	0.009665	Schwarz criterion	-5.086238
Log likelihood	96.92755	F-statistic	2219.014
Durbin-Watson stat	0.613491	Prob(F-statistic)	0.000000

بالطبع، فإن الرسم البياني للبواقي ليس حداً عشوائياً. لكن إذا كان الحد العشوائي خاضع للارتباط الذاتي فإن البواقي ستكون خاضعة أيضاً للارتباط الذاتي .



الشكل البياني رقم(10-11)

يمكننا أن نرى بأنّ هناك دليل قوي على وجود الارتباط الذاتي الإيجابي. مقارنة مع الرسم البياني للأنماط المولّدة عشوائياً، وجدنا أن ρ كانت حوالي 0.6 أو 0.7.

11-2- مصادر الارتباط الذاتي:

يمكننا إرجاع الارتباط الذاتي، بسبب القيم المرتبطة ذاتياً بالخطأ العشوائي إلى عدة أسباب:

1- حذف بعض المتغيرات المستقلة من النموذج ، والتي لها تأثير على المتغيرات المستقلة الأخرى، مما ينعكس ذلك على المتغير العشوائي Ui.

2- نقص في تحديد الشكل الرياضي، كأن نجعل المتغير التابع متغير مستقل والعكس صحيح وعندئذٍ لحدود الخطأ العشوائي U's تظهر وكأن بعضها تابع لبعض.

3- بسبب العمليات التي تجري على المشاهدات الإحصائية ، أي ما يسمى بتهيئة البيانات أو تنعيمها أو تنظيفها Cleaning Data.

4- عدم تحديد المتغير العشوائي بشكل صحيح أو نقص بياناته. قد نتوقع أن تكون القيم المتتالية لـ U مرتبطة، وذلك على الرغم من العوامل العشوائية، لأنها قد تمثل حروب أو كوارث ، عندئذٍ يظهر تأثيرها على أكثر من فترة زمنية.

11-3- اختبار وجود الارتباط الذاتي:

ويمكننا اختبار وجود الارتباط الذاتي في بيانات النموذج باستخدام اختبار داربين- وستون (Durban and Watson test) .

$$E(U_t , U_{t-s}) = 0$$

$$Cov(U_t , U_{t-s}) = E[U_t - E(U_t)] [U_{t-s} - E(U_{t-s})]$$

$$E(U_t) = 0$$

$$Cov(U_t , U_{t-2}) = E[U_t , U_{t-s}] = 0$$

أما حالة الارتباط الذاتي فيعبر عنها بالعلاقة: $E(U_t, U_{t-s}) \neq 0$

هذا يعني بأن خطأ الفترة (t) مرتبطة بخطأ الفترة السابقة (t-s)

يرمز عادة إلى معامل الارتباط بين المتغيرات العشوائية ρ حيث أن :

$-1 < \rho < 1$.

نفترض وجود علاقة خطية لتوليد الخطأ على الشكل التالي، ونسميها معادلة (نظام الارتباط الانحدار من الدرجة الأولى **AR(1)**) .

$$U_t = \rho U_{t-1} + \varepsilon_t \quad \textbf{(1)}$$

ε_t – عبارة عن متغير عشوائي يتوزع توزعاً طبيعياً

$E\varepsilon_t \sim N(0, \sigma^2_{\varepsilon t})$ ولجميع المشاهدات (t) $\quad E(\varepsilon_t, \varepsilon_{t-1}) = 0$

$$\mathbf{E}(\varepsilon t^2) = \sigma^2_{\varepsilon t})$$

من المعادلة (1) نجد أنه عندما نبدل (U_{t-s}) بدلاً من (U_{t-1}) نجد أن :

$$U_t = \rho U_{t-1} + \varepsilon_t$$

$$U_t = \rho(\rho U_{t-2} + \varepsilon_{t-1}) + \varepsilon_t$$

$$U_t = \varepsilon_t + \rho \varepsilon_{t-1} + \rho^2 \varepsilon_{t-2} + \ldots..$$

$$\mathbf{U}_t = \sum_{\rho=0}^{\infty} \rho^r \varepsilon_{t-r}$$

$$\mathbf{E}(\mathbf{U}_t^2) = \mathbf{E}(\varepsilon_t^2) + \rho^2 \mathbf{E}(\varepsilon_{t-1}^2) + \rho^4 \mathbf{E}(\varepsilon_{t-2}) + \cdots$$

$$= (\mathbf{1} + \rho^2 + \rho^4 + \rho^6 + \cdots)\sigma_\varepsilon^2$$

نرمز لتباين الخطأ بالرمز (σ_u^2) وذلك لتمييزه عن تباين المتغير العشوائي (σ_ε^2) أي أن:

$$\mathbf{E}(\mathbf{U}_t^2) = \sigma_u^2$$

وبما أن : $\quad \mathbf{1} + \rho^2 + \rho^4 + \rho^6 + \cdots = \dfrac{\mathbf{1}}{\mathbf{1} - \rho^2}$

فإن $\quad \sigma_u^2 = \dfrac{\sigma_\varepsilon^2}{\mathbf{1} - \rho^2} \qquad \textbf{(2)}$

والآن نشتق صيغة التباين المشترك $\text{Cov}(U_t, U_{t-1})$

بما أن: $U_t = \varepsilon_t + \rho\,\varepsilon_{t-1} + \rho^2\varepsilon_{t-2} + \cdots$

وكذلك $U_{t-1} = \varepsilon_{t-1} + \rho\,\varepsilon_{t-1} + \rho^2\,\varepsilon_{t-2} + \ldots$

فإن : $E(U_t, U_{t-1}) = E\,[\varepsilon_t + \rho\,\varepsilon_{t-1} + \rho^2\,\varepsilon_{t-2} + \ldots] \times$

$$\times\ [\varepsilon_{t-1} + \rho\,\varepsilon_{t-2} + \rho^2\,\varepsilon_{t-3} + \ldots]$$

$$= \rho\, E\,[\,\rho\,\varepsilon_{t-1} + \rho^2\,\varepsilon_{t-2} + \ldots\,]^2$$

و بناء على ذلك:

$$E(U_t U_{t-1}) = \rho\,\frac{\sigma_\varepsilon^2}{1-\rho^2} \qquad (3)$$

وبالأخذ بالمعادلة (2) بعين الاعتبار يمكن إعادة صياغة المعادلة (3)على الوجه التالي :

$$E(U_t U_{t-1}) = \rho^2\,\sigma_u^2 \qquad (4)$$

وبالإمكان تقيم المعادلة(4) للتباين المشترك بين (U_{t-1}) , (U_t)

$$E(U_t U_{t-2}) = \rho^2\,\sigma_u^2 \qquad (5)$$

$$E(U_t U_{t-S}) = \rho^S\,\sigma_u^2 \qquad S = \overline{1,n}$$

نعود إلى القيمة $E(UU')$ ولكن في حالة الارتباط الذاتي

$$E(UU') = \begin{bmatrix} E(U_1^2) & E(U_1U_2) & \cdots & E(U_1U_n) \\ E(U_2U_1) & E(U_2^2) & \cdots & E(U_2U_n) \\ \vdots & \vdots & \cdots & \vdots \\ E(U_nU_1) & E(U_nU_2) & \cdots & E(U_n^2) \end{bmatrix}$$

نعوض عن $E(UU')$ بـــ $E(U_t U_{t-S})$ بقيمها من المعادلة (*)

$$E\left(u\,u'\right)=\begin{bmatrix} \sigma_u^2 & \rho\,\sigma_u^2 & \cdots & \rho^{n-1}\sigma_u^2 \\ \rho\,\sigma_u^2 & \sigma_u^2 & \cdots & \rho^{n-2}\sigma_u^2 \\ \vdots & \vdots & \cdots & \vdots \\ \rho^{n-1}\sigma_u^2 & \rho^{n-2}\sigma_u^2 & \cdots & \sigma_u^2 \end{bmatrix}$$

$$E\left(U\,U'\right)=\sigma_u^2\begin{bmatrix} 1 & \rho & \cdots & \rho^{n-1} \\ r & 1 & \cdots & \rho^{n-2} \\ \vdots & \vdots & \cdots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \cdots & 1 \end{bmatrix}$$

$$E\left(U\,U'\right)=\sigma_u^2\,\Omega$$

الآن نعيد كتابة علاقات المربعات الصغرى المهمة على الشكل التالي:

$$E\left(\hat{\beta}\right)=\left(X'\Omega^{-1}X\right)^{-1}X'\Omega^{-1}Y$$

$$Var\left(\hat{\beta}\right)=S^2\left(X'\Omega^{-1}X\right)^{-1}$$

$$S^2=e'\Omega^{-1}e\,/\,n-k$$

$$e=Y-X\beta$$

أما Ω^{-1} فهي على الشكل التالي :

$$\Omega^{-1}=\frac{1}{1-\rho^2}\begin{bmatrix} 1 & -\rho & 0 & \cdots & 0 & \cdots & 0 \\ -\rho & \left(1+\rho^2\right) & -\rho & \cdots & 0 & \cdots & 0 \\ 0 & -\rho & \left(1+\rho^2\right) & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & 0 & & -\rho & \left(1+\rho^2\right) & -\rho \\ 0 & 0 & 0 & & 0 & -\rho & 1 \end{bmatrix}$$

ومن أجل تطبيق طريقة المربعات الصغرى، ولتكن (T) من أجل تحويل النموذج الاعتيادي: $Y = X\beta + U$

إلى الصيغة : $TY = TXB + TU$

التي من شأنها تحقق الآتي $\mathbf{E}\left(\mathbf{T\,u\,u'\,T}\right) = \mathbf{E}\left(\mathbf{U_{*} U'_{*}}\right) \sigma_{\varepsilon}^{2} I_{n}$

T – ما هي إلا عبارة عن مصفوفة الانتقال التي تساعد على الانتقال إلى العلاقة الممكنة حسب الصيغة المصفوفية $T_{(nxn)}$ تعطي على الشكل التالي :

$$T_1 = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & 0 & \cdots & 0 & 0 & 0 \\ -r & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -\rho & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots\vdots\vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\rho & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & -\rho & 1 \end{bmatrix}$$

وبالإمكان التحقق بسهولة حول إمكان تحقق الآتي :

$$\mathbf{E}\left(\mathbf{T_1\, U\, U'\, T'_1}\right) = \sigma_{\varepsilon}^{2}\, \mathbf{I_n}$$

ومن ثم يمكن استخدام المربعات الصغرى بعد تحويل البيانات (T_1Y) و (T_1X) كالآتي :

$$T_1Y = \begin{bmatrix} \sqrt{1-r^2}\; Y_1 \\ Y_2 - r\,Y_1 \\ Y_3 - r\,Y_2 \\ \cdots\cdots \\ Y_n - r\,Y_{n-1} \end{bmatrix}$$

$$\mathbf{T_1X}=\begin{bmatrix}\left(\sqrt{1-\rho^2}\right) & \left(\sqrt{1-\rho^2}\,\mathbf{X}_{21}\right) & \cdots & \left(\sqrt{1-\rho^2}\,\mathbf{X}_{k1}\right)\\ (1-\rho) & (X_{12}-\rho\mathbf{X}_{11}) & \cdots & (\mathbf{X}_{k2}-\rho\mathbf{X}_{k1})\\ \vdots & & & \\ (1-\rho) & (\mathbf{X}_{1n-\rho}\mathbf{X}_{1n-1}) & \cdots & (\mathbf{X}_{kn-1})\end{bmatrix}$$

ولتسهيل الاحتساب يترك عادة الصف الأول من المصفوفة (T_1) ، حيث نحصل على مصفوفة بأبعاد $[n-1\,X_n]$ نرمز لها بـ T_2

$$\mathbf{T_2}=\begin{bmatrix}-\rho & 1 & 0 & 0 & \cdots & 0 & 0 & 0\\ 0 & -\rho & 1 & 0 & \cdots & 0 & 0 & 0\\ \vdots & \vdots & \vdots & \vdots & & \vdots & & \\ 0 & 0 & 0 & 0 & \cdots & -\rho & 1 & 0\\ 0 & 0 & 0 & 0 & \cdots & 0 & -\rho & 1\end{bmatrix}$$

إن استخدام المصفوفة (T_2) في التحويل بدلاً من (T_1) يجعل التقدير قريب من الكمال، وتؤدي إلى المواصفات نفسها للمربعات الصغرى الاعتيادية تقريباً .

11-4- أساليب تطبيق أسلوب المربعات الصغرى المعممة في حالة الارتباط الذاتي :

لابد من تقدير الحدود التالية :ρ و α و β

أ) طريقة التكرار : تبدأ العملية وفق الشكل الرياضي التالي:

$$Y_t=\alpha+\beta X_t+U_t$$

1- نقوم بتقدير $\hat{\alpha}$ و $\hat{\beta}$

2- ثم نحسب البواقي e_t ونقدر من خلالها ρ

حيث إن: $\hat{\rho}=\dfrac{\sum e_t e_{t-1}}{\sum e_{t-1}^2}$

3- نحسب قيمة المشـــاهدات للمتغيـــرين الجديـــدين $\left(Y_t - \hat{\rho}\, Y_{t-1}\right)$ و $\left(X_t - \hat{\rho}\, X_{t-1}\right)$

ويتم تقدير معلمات النموذج الجديد بطريقة المربعات الصغرى من جديـــد فنحصل على قيم المعلمات $\left(\hat{\alpha}\right)$ و $\left(\hat{\hat{\beta}}\right)$ و $\hat{\hat{\rho}}$ حيث:

$$\hat{\hat{\rho}} = \frac{\sum \hat{\hat{e}}_t \hat{\hat{e}}_{t-1}}{\sum \hat{\hat{e}}_{t-1}^2}$$

4- نحسب قيمة المشاهدات للمتغيرين الجديدين

$$\left(X_t - \hat{\hat{\rho}}\, X_{t-1}\right) \text{ و } \left(Y_t - \hat{\hat{\rho}}\, Y_{t-1}\right)$$

ويعاد تقدير النموذج على ضوء المعطيات الجديدة :

$$\left(Y_t - \hat{\hat{\rho}}\, Y_{t-1}\right) = \alpha\left(1 - \hat{\hat{\rho}}\right) + \beta\left(X_t\ \hat{\hat{\rho}}\, X_{t-1}\right) + \varepsilon_t$$

5- نكرر هذا العمل إلى أن نصل إلى مرحلة تتطابق فيها القيم التقديرية لـــ α و β مع المرحلة السابقة للعملية الأخيرة.

ب) طريقة التقدير على مرحلتين (Two Stage Estimation)

تختزل طريقة التكرار ويكتفي عند المرحلة الثانية ، أي عند $\left(\hat{\hat{\alpha}}\right)$ و $\left(\hat{\hat{\beta}}\right)$

ج) طريقة داربن (Durbins Method)

$$(Y_t - \rho\, Y_{t-1}) = \alpha\,(1 - \rho) + \beta\,(X_t - \rho\, X_{t-1}) + \varepsilon_t$$

$$Y_t = \alpha\,(1 - \rho) + \rho\, Y_{t-1} + \beta\, X_t - \beta\, \rho\, X_{t-1} + \varepsilon_t$$

$$Y_t = \alpha^* + \rho\, Y_{t-1} + \beta\, X_t + \gamma\, X_{t-1} + \varepsilon_t$$

يتبين من المعادلة أعلاه أن النموذج يحتوي على ثلاثة متغيرات مســتقلة هي (X_t) و (X_{t-1}) و Y_{t-1} يمكن تطبيق طريقة المربعات الصغرىفي تقــدير معلمات النموذج التالي :

$$\left(Y_t\ \hat{r}\ Y_{t-1}\right) = \alpha^* + \rho\, Y_{t-1} + \beta\, X_t + \gamma\, X_{t-1} + \varepsilon_t$$

$$\alpha^* = (1 - \rho)$$

اختبار فرضية انعدام الارتباط الذاتي باستخدام (داربرن- واتسون ((Durbin-Watson)

- فرضية العدم : $H_0 : \rho = 0$

انعدام الارتباط الذاتي H_0 = (No Autocorrelation)

- الفرضية البديلة : $H_1 : \rho > 0$

وجود الارتباط الذاتي الموجب H_1=(Positive Autocorrelation)

ملاحظة : إن الارتباط الذاتي عادة في الاقتصاد يكون موجباً ، كمثال زيادة استهلاك يكون عادة مترتب على زيادة سابقة في الدخل، وليس العكس (السبب والنتيجة) . ويحسب قيمة مؤشر الارتباط الذاتي من العلاقة:

$$d = \frac{\sum_{t=2}^{n} (e_t - e_{t-2})^2}{\sum_{t=1}^{n} e_t^2}$$

هناك جدول إحصائي لـ (d) من أجل الاختبار، وهو يتضمن قمتين القيمة الأولى تمثل الحد الأدنى (Lower Limit) (d_L) والثانية تمثل الحد الأعلى (Upper Limit) ويرمز لها بـ d_U .

إن تحديد القيمتين مرتبط بدرجات الحرية، وعدد المتغيرات المستقلة في النموذج ، ويتم الاختبار على الشكل التالي: $d > 4 - d_L$.

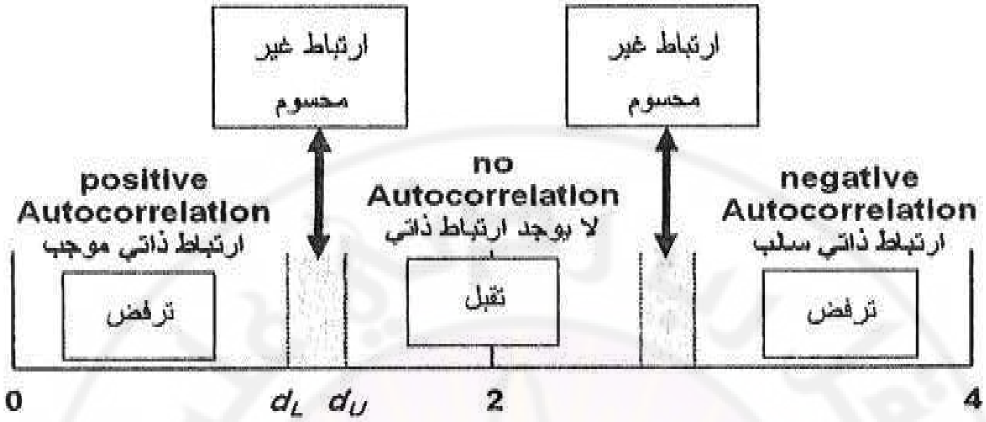
1) ترفض فرضية العدم في حالة $d < d_L$ أي أن هناك ارتباط ذاتي موجب .

2) تقبل فرضية العدم في الحالة : $d_u < d < 4 - d_U$

3) يعتبر الاختبار (غير محسوم INCONLUSIVE) في الحالة :

$d_L \leq d \leq d_U$ أو $4 - d_U \leq d \leq 4 - d_L$

والشكل البياني أدناه يوضح كيفية إجراء اختبار داربن – واتسون



الشكل (11-11)

التفسير الرياضي:

نحسب قيمة المؤشر:

$$d = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}$$

أو من العلاقة:

$$d = \frac{\sum_{t=2}^{n} e_t^2 \sum_{t=1}^{n} e_{t-1}^2 - 2 \sum_{t=2}^{n} e_t \, e_{t-1}}{\sum_{t=1}^{n} e_t^2}$$

$$\sum_{t=2}^{n} e_t^2 \cong \sum_{t=2}^{n} e_{t-L} \cong \sum_{t=1}^{n} e_t$$

$$d = 2 - \frac{2\sum e_t e_{t-1}}{\sum e_t^2} = 2\left[1 - \frac{\sum_{t=2}^{n}(e_t e_{t-1})}{\sum_{t=1}^{n} e_t^2}\right]$$

بما أن e_t هي القيمة التقديرية للخطاء العشوائي U_t فإن القيمة $\frac{\sum e_t e_{t-1}}{\sum e_t^2}$ تمثل القيمة التقديرية (لمعامل الارتباط الذاتي الخطي من الدرجة الأولى) ونرمز له بـ ρ

$$\rho = \frac{\sum e_t e_{t-1}}{\sum e_t^2}$$

$$d \cong 2[1-\rho] \quad \Rightarrow \quad -1 \le \rho \le 1$$

فإن قيمة (d) تتراوح بصورة تقريبية بين الصفر في حالة ($\rho = 1$) و (4) في حالة ($\rho=-1$) وقيمتها الوسطية هي (2) وتعني انعدام الارتباط الذاتي كلياً .

ارتباط موجب	غير محسوم	لا يوجد ارتباط ذاتي		غير محسوم	ارتباط سالب
0	d_L	d_U	2	$4-d_U$	$4-d_L$	4

ونخلص إلى أن التحليل واختبار وجود الارتباط الذاتي للنموذج:

$$\hat{Y}_t = \hat{\alpha} + \hat{\beta} X_t$$

نحسب :

$$d = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}$$

علماً:

$$e_t = \left(\hat{Y}_t - Y_t\right)$$

فإذا كانت $d_U < d$ فإننا نأخذ بفرضية العدم، والتي تنص على عدم وجود ارتباط ذاتي، أما إذا كانت $d_L > d$ فإننا نرفض فرضية العدم، ونأخذ

الفرضية البديلة التي تنص بأن هناك ارتباط ذاتي وعندئذ نقدر معامل الارتباط الذاتي ρ تمهيداً لتحويل البيانات على النحو الآتي:

$(Y_t - \rho Y_{t-1})$, $(X_t - \rho Y_{t-1})$ إلى (Y_t) , (X_t)

$$\hat{\rho} = \frac{\sum e_t e_{t-1}}{\sum e_{t-1}^2}$$

مثال(2-11) : إن الجدول التالي يبين العلاقة بين الإنفـــاق الاســـتهلاكي (Y_t) والدخل الشخصي المتاح (X_t) لإحدى الدول من عام 1999 إلـــى 2008 والمطلوب إجراء انحدار Y_t على X_t واختبار وجود الارتباط الذاتي من الدرجة الأولى عند مستوى معنوية 5 % .

السنة	Y_t	X_t	$\hat{Y}_t$	$Y_t - \hat{Y}_t$	$e_t - e_{t-1}$	$(e_t - e_{t-1})^2$	e_t^2
1999	199	212	198.38	0.62	....	....	0.3844
2000	204	214	200.19	3.81	3.19	10.1761	14.5161
2001	216	231	215.53	0.47	-3.34	11.1556	0.2209
2002	218	237	220.94	-2.94	-33.41	11.6281	8.6436
2003	224	244	227.26	-3.26	-0.32	0.1024	10.6276
2004	235	255	237.19	-2.16	1.07	1.1449	4.7961
2005	238	257	238.99	-0.99	1.20	1.4400	0.9801
2006	256	273	253.43	2.17	3.56	12.6736	6.6049
2007	264	284	263.36	0.64	-1.93	3.7249	0.4096
2008	270	290	268.78	1.22	0.58	0.3364	1.4884
						52.382	48.672

$$D_W = \frac{\sum (e_t - e_{t-1})^2}{\sum e_t^2} = \frac{52.382}{48.672} = 1.0762$$

ومن الجداول نوجد قيمة d_U , d_L عند مستوى معنوية (دلالة) 5 % و $n = 10$ و $k' = 1$ نجد أن : $d_U = 1.320$ و $d_L = 0.879$ الآن نقارن القيمة المحسوبة مع القيمة الجدولية فنجد أن :

$$d_L = 0.879 \quad < D_W = 1.0762 \quad < d_U = 1.320$$

فإن هذا الاختبار لا يبين لنا فيما إذا كان هناك ارتباط ذاتي ، ويتركنا في حالة غير محسومة.

ولكن بفرض أن قيمة الاختبار كانت تساوي 0.69

$$D_W = 1.69 \ < \ d_L = 0.879$$

عند مستوى دلالة 5 % و $k' = 1$ و $n = 10$

فنجد أن هناك ارتباط ذاتي موجب من الدرجة الأولى. أما إذا كانت كما يلي:

$$d_U = 1.320 \ < \ D_W = 2 \ < \ 4 - 1.320$$
$$1.320 \qquad < 2 \ < \qquad 2.68$$

وبالتالي نقول: ليس هناك ارتباط ذاتي .

5-11- إزالة الارتباط الذاتي AR(1) (ELIMINATING AUTOCORRELATION)

سوف نعرض في هذه الفقرة كيفية التخلص من الارتباط الذاتي من علاقة الانحدار من المرتبة الأولى AR(1). سنبدأ بنموذج الانحدار البسيط:

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$
$$u_t = \rho u_{t-1} + \varepsilon_t$$

إذا كان نموذج الانحدار صحيحاً في الزمن t سيكون صحيحاً أيضاً في الزمن t-1. لذلك يمكننا مضاعفة المعادلة في الزمن t-1 من بالمقدار ρ .

$$\rho Y_{t-1} = \beta_1 \rho + \beta_2 \rho X_{t-1} + \rho u_{t-1}$$

لنطرح معادلة الزمن t-1 من معادلة الزمن t من الشكل :

$$Y_t - \rho Y_{t-1} = \beta_1(1-\rho) + \beta_2 X_t - \beta_2 \rho X_{t-1} + u_t - \rho u_{t-1}$$

يتم تحويل الحد العشوائي لـ ε_t في الزمن t في عملية AR(1)، حسب الفرضيات، بحيث يتوزّع الحد العشوائي بشكل مستقل، عندئذٍ فقط تكون مشكلة الارتباط الذاتي قد أزيلت.

$$Y_t = \beta_1(1-\rho) + \rho Y_{t-1} + \beta_2 X_t - \beta_2 \rho X_{t-1} + \varepsilon_t \quad (1)$$

بالتدقيق في العلاقة نجد أن هناك مشكلة بسيطة واحدة، هي أن المواصفات الملاحظة تتضمّن تقييد لاخطّي. لأنّ معامل X_{t-1} يكون سالب نتيجة الضرب بمعاملات X_t و Y_{t-1} .

هذا يعني أنّه يجب أن لا نحاول توفيق المعادلة التي تستعمل المربعات الصغرى العادية OLS ، لأن هذه الطريقة لا تأخذ بالحسبان حساب التقييد ولذا ننتهي بالتقديرات المتعارضة للمعلمات.

على فرض حصلنا على المعادلة على النحو الآتي:

$$\hat{Y}_t = 100 + 0.5Y_{t-1} + 0.8X_t - 0.6X_{t-1}$$

كما هو مبين يمكن أن نستنتج التقديرات من 0.5 لـ ρ و 0.8 لـ β_2 .
لكن هذه الأعداد ستكون غير متوافقة مع التقدير 0.6 لـ $\beta_2 \rho$. لذا نحتاج لاستعمال تقنية التقدير اللاخطي. قبل عمل ذلك، لا بد من تحويل النموذج إلى الانحدار المتعدد بمتغيّرين تفسيريين.

لنعيد كتابة النموذج ثانية مع فترة إبطاء ومضروبا بـ ρ.

$$\rho Y_{t-1} = \beta_1 \rho + \beta_2 \rho X_{2t-1} + \beta_3 \rho X_{3t-1} + \rho u_{t-1} \quad (2)$$

نقوم بطرح المعادلة الثانية من الأولى،فنجد:

$$Y_t - \rho Y_{t-1} = \beta_1(1-\rho) + \beta_2 X_{2t} - \beta_2 \rho X_{2t-1} + \beta_3 X_{3t} - \beta_3 \rho X_{3t-1} + u_t - \rho u_{t-1}$$

نحصل مرة ثانية على نموذج خالي من الارتباط الذاتي:

$$Y_t = \beta_1(1-\rho) + \rho Y_{t-1} + \beta_2 X_{2t} - \beta_2 \rho X_{2t-1} + \beta_3 X_{3t} - \beta_3 \rho X_{3t-1} + \varepsilon_t$$

نلاحظ وجود قيدين، واحد يتضمّن معاملات Y_{t-1}، X_{2t} و X_{2t-1} . وكذلك نلاحظ وجود قيدين، واحد يتضمّن معاملات Y_{t-1}، X_{3t} و X_{3t-1}

يبن الجدول التالي مخرجات الانحدار اللوغاريتمي للإنفاق على الخدمات السكنية إلى الدخل والسعر، يفترض أنها عملية AR(1)، مستخدمين البرنامج Eviews.

```
Dependent Variable: LGHOUS
Method: Least Squares      Sample(adjusted): 1960 1994
LGHOUS=C(1)*(1-C(2))+C(2)*LGHOUS(-1)+C(3)*LGDPI-C(2)*C(3)
     *LGDPI(-1)+C(4)*LGPRHOUS-C(2)*C(4)*LGPRHOUS(-1)
```

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	6.131576	0.727244	8.431247	0.0000
C(2)	0.972488	0.004167	233.3565	0.0000
C(3)	0.275879	0.078318	3.522532	0.0013
C(4)	-0.303387	0.085802	-3.535896	0.0013

R-squared	0.999695	Mean dependent var	6.017555
Adjusted R-squared	0.999665	S.D. dependent var	0.362063
S.E. of regression	0.006622	Akaike info criter	-7.089483
Sum squared resid	0.001360	Schwarz criterion	-6.911729
Log likelihood	128.0660	F-statistic	33865.14
Durbin-Watson stat	1.423030	Prob(F-statistic)	0.000000

يسمح برنامج Eviews بطرقتين لتحديد معادلة الانحدار. واحدة أن يدرج المتغيّرات، بدء من المتغير التابع، بالاستمرار مع التقاطع C ، وينتهي بقائمة المتغيّرات التفسيرية. هذا رائع بالنسبة للانحدارات الخطية.

ولكن بطريقة أخرى تكتب النموذج كمعادلة، بالإشارة إلى المعلمات C(1) و C(2) إلخ... . هذا ما يجب أن نعمله عندما نوفق نموذج لاخطّي.

أما هنا فنجد أن : C(1) تدل على β_1 . و تدل C(2) على معامل المتغير التابع المتخلف ρ . هو أيضا معامل التقاطع في هذا النموذج. الملاحظة التي تبين أن تقدير ρ عالي جدا ويساوي إلى 0.97.

تدل C(3)، معامل الدخل β_2 . والشيء الملفت يبدو التقدير منخفض . معامل الدخل المتباطئ يجب يكون محدّدا بالعلاقة C(2)*C(3)- . أما C(4) فتدل على معامل السعر β_3 . إنّ قيمة التقدير منخفضة، ولكن ليس لدرجة غير قابل للتصديق .

معامل السعر المتباطئ يجب أن يكون محدّداً بالعلاقة C(2)*C(4)-.

إنّ الوظيفة الوحيدة لهذه طريقة AR(1) هي لازالت الارتباط الذاتي، وتكمن بتحديد النموذج في شكل المعادلة، فهي عملية معقدة ومن السّهل ارتكاب الأخطاء.

تعتبر مواصفات AR(1) عامة، تزوّد أكثر تطبيقات الانحدار الجدّية بعض الطرق المختصرة لتحديده بسهولة. في حالة Eviews، AR(1) تقدير يحدث بسبب إضافة AR(1) إلى قائمة المتغيّرات التفسيرية.

```
=================================================================
Dependent Variable: LGHOUS
Method: Least Squares     Sample(adjusted): 1960 1994
Included observations: 35 after adjusting endpoints
Convergence achieved after 24 iterations
=================================================================
     Variable     Coefficient Std. Error t-Statistic   Prob.
=================================================================
        C           6.131573   0.727241    8.431276   0.0000
      LGDPI         0.275879   0.078318    3.522534   0.0013
     LGPRHOUS      -0.303387   0.085802   -3.535896   0.0013
      AR(1)         0.972488   0.004167    233.3540   0.0000
=================================================================
R-squared            0.999695   Mean dependent var    6.017555
Adjusted R-squared   0.999665   S.D. dependent var    0.362063
S.E. of regression   0.006622   Akaike info criter   -9.927360
Sum squared resid    0.001360   Schwarz criterion    -9.749606
Log likelihood       128.0660   F-statistic           33865.14
Durbin-Watson stat   1.423031   Prob(F-statistic)     0.000000
=================================================================
```

الثابت هو عبارة عن تقدير β_1.

```
=================================================================
Dependent Variable: LGHOUS
LGHOUS=C(1)*(1-C(2))+C(2)*LGHOUS(-1)+C(3)*LGDPI-C(2)*C(3)
      *LGDPI(-1)+C(4)*LGPRHOUS-C(2)*C(4)*LGPRHOUS(-1)
=================================================================
                  Coefficient Std. Error t-Statistic   Prob.
=================================================================
      C(1)          6.131576   0.727244    8.431247   0.0000
      C(2)          0.972488   0.004167    233.3565   0.0000
      C(3)          0.275879   0.078318    3.522532   0.0013
      C(4)         -0.303387   0.085802   -3.535896   0.0013
=================================================================
=================================================================
     Variable     Coefficient Std. Error t-Statistic   Prob.
=================================================================
        C           6.131573   0.727241    8.431276   0.0000
      LGDPI         0.275879   0.078318    3.522534   0.0013
     LGPRHOUS      -0.303387   0.085802   -3.535896   0.0013
      AR(1)         0.972488   0.004167    233.3540   0.0000
=================================================================
```

إنّ معامل الدخل يكون تقديراً للمرونة ، فيما يتعلق بالدخل الحالي. وإنّ معامل السعر هو تقدير للمرونة فيما يتعلق بالسعر الحالي. وكذلك فإنّ معامل AR(1) هو تقدير لـ ρ .

معاملات الدخل المتباطئ وتخلّفت عن تبطأ السعر، لأنها مضمنة في تقديرات ρ ، β_2 ، β_3 .

الطريقة الأولى: الارتباط الذاتي في أي نموذج مع أي متغير تابع متباطئ:

رأينا في الفقرة السابقة بأنّ المستوى الأول AR(1) من الارتباط الذاتي، يمكن أن يزال من خلال التلاعب البسيط في النموذج. إنّ نموذج الانحدار يمكن أن يكون انحدار لاخطّي المعلمات، لكن لا يوجد أي مشكلة لتوفيقها.

$$\rho Y_{t-1} = \beta_1 \rho + \beta_2 \rho X_{t-1} + \rho u_{t-1}$$

$$Y_t - \rho Y_{t-1} = \beta_1 (1-\rho) + \beta_2 X_t - \beta_2 \rho X_{t-1} + u_t - \rho u_{t-1}$$

$$Y_t = \beta_1 (1-\rho) + \rho Y_{t-1} + \beta_2 X_t - \beta_2 \rho X_{t-1} + \varepsilon_t$$

على أية حال، في بداية أيام استخدام الحاسبات، فان حساب التقدير اللاخطي ما كان بسيط ، وكان يتم الابتعاد عنه قدر الإمكان. اختبار Cochrane Orcutt التكراري كان طريقة جيدة لاستعمال تحليل الانحدار الخطي المناسب لتوفيق نموذج اللاخطي.

وهو ليس اهتمامنا العملي الآن، لكنّ قد نورد إشارات إليه من حين لآخر أثناء المناقشة في هذه الفقرة.

نعود للعلاقة:

$$Y_t - \rho Y_{t-1} = \beta_1 (1-\rho) + \beta_2 X_t - \beta_2 \rho X_{t-1} + u_t - \rho u_{t-1}$$

ونلاحظ بأنّ النموذج يمكن أن يعاد كتابته بالشكل المناسب، عندما يكون نموذج انحدار بسيط خالي من الارتباط الذاتي ، على النحو الآتي:

$$\tilde{Y}_t = \beta_1' + \beta_2 \tilde{X}_t + \varepsilon_t$$

حيث أن:

$$\tilde{Y}_t = Y_t - \rho Y_{t-1}$$

$$\tilde{X}_t = X_t - \rho X_{t-1}$$

$$\beta_1' = \beta_1(1-\rho)$$

على أية حال، لبناء المتغيّرات الاصطناعية نحتاج لتقدير ρ نحصل على استعمال واحد من البواقي. إذا كان حد الخطأ العشوائي مولّد من خلال عملية AR(1)، و e_t سيتعلّق الأمر بـــ e_{t-1} من خلال عملية مماثلة.

الخطوة الأولى: نحسب الانحدار الأصلي حسب طريقة المربعات الصغرىOLS ، أي نحسب انحدار Y_t على X_t.

وفي الخطوة الثانية: نحسب البواقي و ومن ثم نحسب انحدار العلاقة e_t على e_{t-1} . فإن قيمة معامل الانحدار سيكون عبارة مقدر ρ .

وفي الخطوة الثالثة: سوف نحسب المتغيّرات الاصطناعية Y_tو X_t وانحدار Y_tعلى X_t ، فإن معامل الانحدار سيكون عبارة عن تقدير β_2 وتقدير β_1 والذي يمكن أن يشتقّ منه التقاطع.

نعود للخطوة الثانية، ونحسب البواقي لـــ e_t على e_{t-1} مرة ثانية للحصول على تقدير أفضل لـــρ .

عندما نعود للخطوة الثالثة، والاختيار بين خطوة الثانية والخطوة الثالثة حتى نرى مدى التقارب بين النتائج. هكذا النموذج اللاخطي يمكن أن يلاءم استعمال تحليل الانحدار الخطي.

ذكرنا سابقاً أنّ المتغير التابع يكون متباطئ مع أحد المتغيّرات التفسيرية، عند استخدام تقديرات المربعات الصغرى OLS وقد يكون ذلك فيه بعض

الشيء من التحيّز بالنسبة للعينات الصغيرة، حتى إذا كانت قيمة u محققة لشروط غاوس – ماركوف.

هذا التحيّز لا يعتبر على العموم مشكلة خطيرة وعادة لا يأخذ به عملياً. أما في العينات الكبيرة فهذه المشكلة غير موجودة. على أية حال، فإن حد الخطأ العشوائي خاضع للارتباط الذاتي، وهذه حالة مختلفة تماماً. أما طريقة المربعات الصغرى ستعطي تقديرات متناقضة. على سبيل المثال، يفترض بأنّ النموذج يعاني الارتباط الذاتي من الدرجة الأولى AR(1) هو مبين أدناه :

$$u_t = \rho u_{t-1} + \varepsilon_t$$

إذا النموذج صحيح في الزمن t ، وكذلك يجب أن يكون صحيح في الزمن t-1 .

$$Y_{t-1} = \beta_1 + \beta_2 X_{t-1} + \beta_3 Y_{t-2} + u_{t-1}$$

هكذا Y_{t-1} يحتوي على الحد العشوائي u_{t-1}، وكذلك u_t ؛ لذا شرط غاوس– ماركوف الرابع منتهك، وهذا يعني بأنّ طريقة المربعات الصغرى سوف تعطي تقديرات متناقضة وكذلك فإنّ اختبارات الأخطاء المعيارية و t و F غير صحيحة.

عندما نستخدم المتغير التابع المتباطئ كمتغير تفسيري، فإن اختبار داربن– وتسون للارتباط الذاتي غير ملائم لأن d الإحصائية تتحيّز نحو القيمة 2. لذلك سوف نستخدم اختبار h إحصائية لداربن بدلا من ذلك، للكشف عن الارتباط الذاتي.

$$h = \hat{\rho}\sqrt{\frac{n}{1 - n s^2_{b_{Y(-1)}}}}$$

ثلاث أمور مطلوبة للاختبار: تقدير r المعامل في عملية AR(1) و n، عدد المشاهدات في الانحدار، وتقدير تباين معامل المتغير التابع المتباطئ

$s^2_{b_{Y(-1)}}$. على أية حال، يوجد عدّة طرق يمكن من خلالها الحصول على تقدير ρ، على وجه التحديد صحيح فقط في العينات الكبيرة،علماً إنه يجب أن تعطي مختلف الطرق وفي العينات الكبيرة نتائج مماثلة. هو أسهل لتوضيح العلاقة بين d و ρ.

$$d \approx 2 - 2\rho$$

$$\hat{\rho} = 1 - 0.5d$$

هنا مخرجات لنتائج الانحدار اللوغاريتمي للإنفاق على الخدمات السكنية منسوبة إلى الدخل والسعر النسبي، مع المتغير التابع المتباطئ.

```
LS // Dependent Variable is LGHOUS
Sample(adjusted): 1960 1994
Included observations: 35 after adjusting endpoints
```

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.390249	0.152989	-2.550839	0.0159
LGDPI	0.313919	0.052510	5.978243	0.0000
LGPRHOUS	-0.067547	0.024689	-2.735882	0.0102
LGHOUS(-1)	0.701432	0.045082	15.55895	0.0000

R-squared	0.999773	Mean dependent var	6.017555
Adjusted R-squared	0.999751	S.D. dependent var	0.362063
S.E. of regression	0.005718	Akaike info criter	-10.22102
Sum squared resid	0.001014	Schwarz criterion	-10.04327
Log likelihood	133.2051	F-statistic	45427.98
Durbin-Watson stat	1.718168	Prob(F-statistic)	0.000000

$$h = \hat{\rho}\sqrt{\frac{n}{1 - ns^2_{b_{Y(-1)}}}} \quad h = 0.14 \times \sqrt{\frac{35}{1 - 35 \times 0.0020}} = 0.86$$

قيمة اختبار ديربن -وتسون يساوي إلى 1.72، لذلك نحصل على تقدير ρ من العلاقة $\hat{\rho} = 1 - 0.5d = 1 - 0.5 * 1.72 = 0.14$.

عدد مشاهدات 35 لمتغيرات علاقة الانحدار . (كان من المفروض أن تكون 36 مشاهدة حسب حجم العينة، لكن المشاهدة الأولى لا تدخل لأنه يوجد فترة تبطأ للمتغير Lghous (-1) و لم يحدد لها قيمة).

إنّ تقدير تباين معامل المتغير التابع المتباطئ $s^2_{b_{Y(-1)}}$ يحسب بتربيع الخطأ المعياري له.

وبالتطبيق علاقة h نجد :

$$h = 0.14 \times \sqrt{\frac{35}{1 - 35 \times 0.0020}} = 0.86$$

فرضية العدم: تنص على أنه لا يوجد ارتباط ذاتي لان h الإحصائية تخضع للتوزيع الطبيعي بمتوسط 0 وتباين 1. هذا يعني بأنّ القيمة الحرجة عند 5 % مستوى دلالة 1.96، لذلك سوف لن نرفض فرضية العدم في هذه الحالة.

للمقارنة بين طريقة تأثيرات الارتباط الذاتي من المرتبة الأولى AR(1) وطريقة المربعات العادية مع متغير تابع متباطئ ، تم الاستعانة بتجربة مونتي كارلو، لتوليف 10 نتائج للمقارنة :

$$Y_t = 10 + 0.8Y_{t-1} + u_t \qquad u_t = 0.7u_{t-1} + \varepsilon_t$$

	OLS					AR(1)				
Sample	b_1	b_2	s.e.(b_2)	d	h	b_1	b_2	s.e.(b_2)	d	h
1	4.7	0.95	0.07	1.14	2.50	17.3	0.79	0.10	2.05	-0.16
2	5.0	0.92	0.07	0.95	3.09	19.1	0.70	0.14	1.94	0.26
3	0.9	0.94	0.05	0.40	4.47	4.0	0.84	0.11	2.06	-0.20
4	6.1	0.84	0.11	0.89	3.68	14.2	0.65	0.15	1.50	2.32
5	5.1	0.83	0.07	1.11	2.60	8.5	0.75	0.13	2.09	-0.34
6	-1.7	1.01	0.05	1.04	2.70	5.0	0.91	0.09	2.01	0.03
7	5.3	0.90	0.08	0.78	3.65	17.4	0.71	0.13	1.93	0.27
8	-1.3	0.96	0.04	0.83	3.22	2.4	0.80	0.11	1.62	1.26
9	-0.6	0.98	0.04	0.55	4.01	3.8	0.83	0.10	1.83	0.54
10	-0.9	1.00	0.07	1.03	2.81	11.8	0.80	0.12	1.70	1.08

لإبقاء التحليل بسيط بقدر الإمكان، النموذج بدائي جدا. المتغير Y محدّد فقط بواسطة قيمته المتباطئ، بتقاطع 10 ومعامل انحدار 0.8 . إنّ حد الخطأ العشوائي لـــ AR(1) خاضع للارتباط الذاتي مع ρ و يساوي إلى 0.7 و إنّ حجم العيّنة 30.

في كلّ العينات 10 نجد أن تقدير معامل الانحدار b_2 حسب طريقة المربعات الصغرى OLS أعلى من القيمة الحقيقية، وهذا دليل التحيّز الإيجابي. بينما تقديرات معامل التقاطع b_1 يتحيّز بشكل تنازلي، انظر الجدول أعلاه.

على الرغم من حقيقة مقياس ديربن – وتسون بأنه يتحيّز نحو القيمة 2، النتائج تخبرنا بأنه يوجد ارتباط ذاتي موجب حيادي، انظر الجدول أعلاه.

أما اختبار h الذي يستخدم في العيّنات الكبيرة، وهو مطبق في هذا المثال على 29 مشاهدة فقط حسب طرية المربعات الصغرى، فهو يخبرنا على وجود ارتباط ذاتي، انظر الجدول أعلاه .

في العيّنة الأولى التي يتبين لنا أنه يجب أن نرفض فرضية عدم بأنه لا يوجد ارتباط ذاتي عند 5 % مستوى معنوية $h> 1.96$ ، أما بالنسبة للقيم التسعة الأخرى فإننا سوف نرفض فرضية العدم عند 1 % مستوى معنوية $h> 2.58$.

بينما بالمقارنة طريقة AR(1)، للتقدير معلمات النموذج β_1 و β_2 ، فنجدها متفرّقة بشكل عشوائي حول قيمها الحقيقية. في تسعة من العينات العشرة، نجد أن قيمة h الإحصائية منخفضة بما فيه الكفاية، لدرجة لا يمكننا أن نرفض فرضية عدم لعدم وجود ارتباط ذاتي. ملاحظة: (في العيّنة الرابعة، على سبيل المثال، نرتكب خطأ، إذا استعملنا 5 % مستوى معنوية).

إنّ الأخطاء المعيارية لـــ b_2 أكبر من تلك في الانحدار بطريقة المربعات الصغرى، لكن الحالة الأخيرة كانت غير ذلك، وهذا لا يدعو للقلق، انظر الجدول أعلاه.

من هذه المناقشة نستنتج أن الارتباط الذاتي يتضمن تعديل جزئي وتكييف للنماذج المتوقعة.

في نموذج الجزئي المعدل، حد الخطأ العشوائي في النموذج الموفق هو تماما كما في علاقة الهدف، ماعدا بأنّه تضاعف بمقدار الثابت λ .

$$Y_t^* = \beta_1 + \beta_2 X_t + u_t$$

$$Y_t = \lambda Y_t^* + (1-\lambda) Y_{t-1}$$

$$Y_t - Y_{t-1} = \lambda (Y_t^* - Y_{t-1})$$

$$Y_t = \lambda(\beta_1 + \beta_2 X_t + u_t) + (1-\lambda) Y_{t-1}$$

$$= \beta_1 \lambda + \beta_2 \lambda X_t + (1-\lambda) Y_{t-1} + \lambda u_t$$

هكذا، تحقق شروط غاوس-ماركوف في علاقة الهدف، سيكون محقق أيضاً في العلاقة الموفقة. المشكلة الوحيدة هي تحيّز العيّنة الصغير، وهذا نتجاهله في الممارسة العملية على أية حال.

بالطبع، إن الحد العشوائي في علاقة الهدف مرتبط ذاتياً، و سيكون مرتبط ذاتياً في العلاقة الموفقة. فطريقة المربعات الصغرى تعطي تقديرات متناقضة لذلك يتوجب علينا استخدام في بعض الأحيان AR(1) كطريقة للتقدير بدلا من ذلك.

في هذه الحالة نموذج التوقّعات التكيفية تشكّل، مع نموذج الانحدار خيار بديل. نموذج واحد سريع لـــY كمعادلة للقيم الحالية و المتباطئ لـــ X، يأخذ بعين الاعتبار بشكل كافي إهمالنا لمعامل المتغيّر X^e_{t-s+1} غير الجدير بالملاحظة، انظر العلاقات أدناه:

$$Y_t = \beta_1 + \beta_2 X^e_{t+1} + u_t$$

$$X^e_{t+1} - X^e_t = \lambda(X_t - X^e_t)$$

$$Y_t = \beta_1 + \beta_2\lambda X_t + \beta_2\lambda(1-\lambda)X_{t-1} + \beta_2\lambda(1-\lambda)^2 X_{t-2} + \dots$$
$$+ \beta_2\lambda(1-\lambda)^{s-1} X_{t-s+1} + \beta_2(1-\lambda)^s X^e_{t-s+1} + u_t$$

إنّ حد الخطأ العشوائي في نموذج الانحدار، تماما هو كما في النموذج الأصلي. إذا لم تتحقق شروط غاوس- ماركوف في النموذج الأصلي ، يمكن عمل ذلك في نموذج الانحدار، الموفق باستخدام نموذج التقدير اللاخطي المعياري.

إذا كان النموذج الأصلي مرتبط ذاتياً، سيكون نموذج الانحدار مرتبط ذاتياً أيضاً ، عندئذٍ لا بد من استخدام طريقة تقدير AR(1).

إنّ النسخة الأخرى لنموذج الانحدار تظهر Y كنموذج على X والمتغير المتباطئ Y . أما حد الخطأ العشوائي فهو مركّب من: u_t و u_{t-1} .

$$Y_t = \beta_1\lambda + (1-\lambda)Y_{t-1} + \beta_2\lambda X_t + u_t - (1-\lambda)u_{t-1}$$

هكذا إذا حد الخطأ العشوائي في النموذج الأصلي محقق شروط غاوس- ماركوف، و حد الخطأ العشوائي في نموذج الانحدار سيكون خاضع MA(1) (المتوسط المتحرك للارتباط الذاتي من المرتبة الأولى).

إذا تقارن شروط الخطأ العشوائي المركّبة للمشاهدات t و t-1، سنجد بأنّه يوجد u_{t-1} كفرق مشترك، قارن المعادلة التالية مع السابقة.

$$Y_{t-1} = \beta_1\lambda + (1-\lambda)Y_{t-2} + \beta_2\lambda X_{t-1} + u_{t-1} - (1-\lambda)u_{t-2}$$

مجموعة المتوسط المتحرك للارتباط الذاتي و وجود المتغير التابع المتباطئ في نموذج الانحدار يشكلان سبباً لانتهاك شرط غاوس -ماركوف الرابع.

u_{t-1} مكوّن من Y_{t-1} و حد الخطأ العشوائي المركّب. وهكذا فالمطلوب هو أن حد الخطأ العشوائي تتوزع بشكل مستقل عن المتغيّرات التفسيرية غير المرضية. تحت هذه الشروط، سوف نستخدم نموذج الانحدار الأخر بدلاً من ذلك.

على أية حال، يفترض بأنَ حد الخطأ العشوائي في النموذج الأصلي يعاني من ارتباط ذاتي من المرتبة الأولى AR(1) .

$$u_t = \rho u_{t-1} + \varepsilon_t$$

أما حد الخطأ العشوائي المركّب في الزمن t سيكون على النحو الآتي :

$$u_t - (1-\lambda)u_{t-1} = \rho u_{t-1} + \varepsilon_t - (1-\lambda)u_{t-1}$$
$$= \varepsilon_t + (\rho + \lambda - 1)u_{t-1}$$

الآن، تحت الفرضيات المعنوية، نجد كل من ρ و λ يجب أن تقع بين 0 و 1. لذلك من الممكن يكون أن معامل u_{t-1} صغيراً بما فيه الكفاية لإهمال الارتباط الذاتي.

في هذه الحالة يجب تطبيق طريقة المربعات الصغرى OLS لتوفيق نموذج الانحدار مع ذلك. يجب ، أن يؤدّي بالطبع إلى أن اختبار h للتدقيق بأن الارتباط الذاتي (غير معنوي) .

11-6- طرق أخرى لتصحيح الارتباط الذاتي :

الطريقة الثانية(GLS): في حال مخالفة فرضية ثبات التباين المتغير العشوائي في المجتمعات الفرعية نلاحظ في هذه الحالة، أن القطر في المصفوفة σ_1^2 لا يحتوي قيم متساوية ، وذلك بالرغم من أن القيم خارج القطر تساوي صفراً . ويعود السب في ذلك إلى مخالفة فرضية ثبات التباين المتغير العشوائي في المجتمع الإحصائي لتقديرات معاملات الانحدار . علماً أنه باستطاعتنا في

مثل هذه الحالة ، استخدام طريقة (Aikken) أيكين لتحويل المتغيرات إلى صيغة تتناسب مع فرضية ثبات التباين للمتغير العشوائي في العينات .

أوجد أيكين طريقة عام 1930 تسمى GLS =General least square وتعتبر هذه الطريقة تعميماً وتوسيعاً لطريقة المربعات الصغرى الاعتيادية، بحيث تمكن الإحصائي من الحصول على تقديرات خطية غير متحيزة وكفوءة لمعاملات الانحدار الجزئية . فإذا فرضنا أن النموذج انحدار Y على X يأخذ الشكل التالي:

$$Y = \beta_0 + B_1X_1 + e$$

وكانت تعاني من مشكلة مخالفة لفرضية الخاصة بثبات تباين المتغير العشوائي في العينات، فإنه يتوجب على الباحث عندئذ تحويل متغيرات النموذج إلى صيغة أخرى ينتج عنها قيم متساوية في قطر مصفوفة σ_1^2 . علماً أنه يوجد العديد من الطرق التي تمكن الباحث من تحويل المتغيرات إلى صيغة تحقق ثبات التباين في العينات، من أهمها :

1) إذا كان تباين المتغير التابع Y يزيد بشكل تناسبي (Proportional) مع الزيادة في X أي إذا كان :

$$K = \frac{\sigma y}{\sqrt{x}} \qquad \Leftarrow \sigma_y^2 = K^2 \cdot X$$

فباستطاعة الباحث عندئذ أن يلجأ إلى قسمة طرفي معادلة الانحدار على $\sqrt{X}$ وهذا بدوره يحقق ثبات المتغير العشوائي في المجتمع الإحصائي المدروس .

عندئذ تصبح العلاقة :

$$\frac{Y}{\sqrt{X}} = \frac{\beta_0}{\sqrt{X}} + \frac{\beta_1 X_1}{\sqrt{X}} + \frac{e}{\sqrt{X}}$$

أما إذا كانت إحدى قيم X تساوي الصفر، فعندئذ يمكن استخدام العلاقة $\sqrt{X+0.5}$ أي أنه إذا كان تباين المتغير Y يتزايد بشكل تناسبي على X أو $\overline{X}$ فحينئذ يتوجب علينا أن نأخذ انحدار $\frac{Y}{\sqrt{X}}$ دالة للمتغيرات المستقلة $1/\sqrt{X}$ و $\sqrt{X}$. بمعنى أدق أن علاقة الانحدار أصبحت على الشكل التالي:

$$Y/\sqrt{X} \quad ، \quad 1/\sqrt{X} \quad ، \quad \sqrt{X}$$

$$\frac{Y}{\sqrt{X}} = \beta_0 \cdot \frac{1}{\sqrt{X}} + \beta_1 \sqrt{X} + e/\sqrt{X}$$

عندئذ فإن قيمة الثابت بعد التحويل تساوي صفراً ، وهذا بدوره يعطي قيمة مرافقة لمعامل التحديد R^2 ، ولكن هناك صعوبة في تفسير النتائج، وخاصة موقع قيم المتغير التابع .

2) إذا كان تباين المتغير التابع Y يزيد تناسباً مع القيمة المربعة للوسط الحسابي ، أي إذا كان :

$$\sigma_Y^2 = k^2\,\overline{X}^2$$

$$\sigma_Y^2 = k^2 X^2 \qquad \Rightarrow k = \sqrt{\frac{\sigma_Y^2}{X^2}} = \frac{\sigma_Y}{X}$$

$$\frac{Y}{X} = \frac{\beta_0}{X} + \beta_1 \frac{X}{x} + \frac{e}{X}$$

$$\frac{Y}{X} = \beta_0 \frac{1}{X} + \beta_1 + e \cdot 1/X$$

نلاحظ أن معامل الانحدار β_1 أصبح يمثل القيمة الثابتة في المعادلة بدلاً من β_0 .

	(1)	(2)	(3)	(4)
LGDPI	-0.96 (0.35)	1.68 (1.08)	1.13 (1.14)	-0.05 (0.18)
LGPRTAXI	2.74 (1.19)	-1.67 (0.83)	-1.58 (0.89)	0.03 (0.61)
LGTXI (-1)	-	-	0.90 (0.07)	0.91 (0.08)
LGDPI (-1)	-	-	-1.32 (1.08)	-
LGPRTACI(-1)	-	-	2.04 (0.89)	-
Constant	-3.41 (2.87)	-16.83 (112.04)	-0.50 (1.35)	0.39 (1.43)
$\hat{\rho}$	-	1.00 (0.04)	-	-
R^2	0.1960	0.8676	0.8770	0.8448
Rss	1.8334	0.3018	0.2803	0.3538
d	0.63	0.99	0.94	0.77

6 – لدى الباحث بيانات سنوية من 1973 إلى 2002 عن الاستهلاك الكلي (C) والدخل الكلي (Y) لبلد معين . الباحث مهتم باكتشاف العلاقة بين (C_t) و (Y_t) سامحاً لحركة ديناميكية قصيرة الأجل ، وهو يهيئ نماذج الانحدار التالية :

(1)- نموذج انحدار ADL (1,1) لـ Ct على (Y_t , C_{t-1} , Y_{t-1}) مستخدماً طريقة المربعات الصغرى الاعتيادية (OLS)

(2)- نموذج انحدار طريقة المربعات الصغرى الاعتيادية (OLS) لـ (C_t) على (Y_t) ،

(3)- نموذج انحدار (C_t) على (Y_t) باستخدام طريقة تقدير والتي تسمح لـــ AR(1) بالارتباط التلقائي . يظهر الجدول أدناه نتائج نموذج الانحدار . تم وضع الأخطاء المعيارية بين قوسين و (RSS) هي مجموع مربع البواقي . d هي إحصائية (دربن ـــ واتسون) . ρ هي تقدير معامل الارتباط التلقائي في العملية AR(1) .

	(1)	(2)	(3)
Y_t	0.23 (0.03)	0.72 (0.07)	0.67 (0.12)
C_{t-1}	0.48 (0.12)	–	–
Y_{t-1}	0.19 (0.04)	–	–
constant	−15.21 (18.27)	130.02 (42.15)	80.61 (30.33)
$\hat{\rho}$	–	–	0.34 (0.13)
d	1.79	1.08	1.53
RSS	2332.0	4195.2	3582.0

أ – بين فيما إذا كان النموذج (1) ، يمكن أن يكون ممثلا كافياً للبيانات

ب– برهن أن النموذج (2)، هو تقييد للنموذج (1) وحدد فيما إذا كان مقبولاً.

ج – برهن أن النموذج (3) هو تقييد للنموذج (1) وحدد فيما إذا كان مقبولا.

د – اشرح فيما إذا كان النموذج (1) يمكن أن يُفسَّر اشتقاقه من قبل نموذج جزئي معدل .

هـــ – قدر الميل الحدي للاستهلاك ضمن كل نموذج من النماذج الثلاثة مميزاً بين المعلمات الحدية، فيما إذاً كانت طويلة الأجل وقصيرة الأجل إذاً أمكن ذلك.

المصطلحات العلمية

A

Acceptance region	منطقة القبول
Additive constant	ثابت تجميعي
Adjusted R^2	معامل التحديد المعدل
Almon Lag	ابطاء المون
Alternative hypothesis	الفرضية البديلة
Assumption	افتراض
Autocorrelation	الارتباط الذاتي
Autoregressive (AR) process	نموذج الانحدار الذاتي

B

Bias	تحيز
Binary variables	المتغيرات الثنائية
Bounded	محدودة

C

Categorical variables	المتغيرات الترتيبية
Central Tendency	نزعة مركزية
Coefficient	معامل
Coefficient of Determination	معامل التحديد
Combination	مكونات
Consistency	اتساق
Continuous	مستمر

Correlation	ارتباط
Covariance	تباين(تغاير)
Critical region	المنطقة الحرجة
Cross-sectional	مقطعي

D

Deflated	المخفض
Degenerated	منحل
Density	كثافة
Dependence	تابع
Determination coefficient	معامل التحديد
Discrepancy	تناقض
Discrete	منقطع
Distributed lag	فترة إبطاء
Disturbance term	خطأ عشوائي (حد الخطأ)
Dummy variable	متغير وهمي(صوري)

E

Econometrics	اقتصاد قياسي(قياس اقتصادي)
Endogenous variable	متغير داخلي
Endpoint	طرفي
Error sum of squares(ESS)	مجموع مربع الأخطاء
Estimate	تقدير
Estimators	مقدرات
Events	أحداث

Exact	مؤكد(يقني)
Exogenous variable	متغير خارجي
Expectation	توقعات
Explanatory variable	متغير تفسيري
Explicit	صريح

F

Fallacy of Composition	مكونات خاطئة
Finite	محدودة
First difference equations	معادلات فروق من الدرجة الأولى
Fit	توفيق
Forecast	تنبؤ
Formalization	تأطير
Formally	رياضيا، اصطلاحياً
Function	دالة (معادلة)

G

Goldfeld-Quandt test (G-Q)	اختبار غولدفيلد-كوانت(**G-Q**)
Goodness of fit	جودة التوفيق

H

Heteroscedasticity	اختلاف التباين
Homoscedasticity	ثبات التباين
Hypotheses testing	اختبار الفرضيات

I

Identification	تمييز

Independence variable	المتغير المستقل
Inflection points	نقاط الانعطاف
Instrumental	مساعد
Interaction term	حد مركب

J

Joint probability	احتمال مشترك

K

Koyck lag	ابطاء كويك

L

Lag	ابطاء
Limited-information	معلومات محدودة
Linear	خطي
Lower bound	الحد الأدنى

M

Marginal propensity to consume(MPC)	الميل الحدي للاستهلاك
Mean	الوسط الحسابي
Mean Square error	متوسط مربع الخطأ
Monotonic	مضطرد
Multicolliearity	تعدد العلاقات الخطية

N

Nonlinear Simple Regression	الانحدار اللاخطي البسيط
Normal distribution	توزيع طبيعي

Null hypothesis	فرضية العدم

O

Offsetting	المعادل(المقابل)
One tailed test	اختبار من اتجاه واحد
Open ended	ذو نهاية مفتوحة
Operator	معامل
Ordinary least squares(OLS)	طريقة المربعات الصغرى
Overall mean	متوسط إجمالي

P

Parameter	معلمة
Point estimation	تقدير النقطة
Polynomial	متعدد الحدود
Population	مجتمع
Predetermined	محدد مسبقاً
Prediction	تنبؤ
Probability	احتمال
Proxy	تقريبي

Q

Qualitative variables	المتغيرات النوعية

R

Random	عشوائي
Reasonable	معنوية
Reciprocal transformation	تحويل عكسي

Reduced form	شكل مختزل
Regression	انحدار
Regression sum of squares(RSS)	مجموع مربعات الانحدار
Rejection region	منطقة الرفض
Relabeling	اعادة صياغة
Residuals	بواقي
Restricted model	نموذج مقيد

S

Sampling distribution	توزيع المعاينة
Scale factors	عوامل ترجيح
Scatter diagram	شكل انتشار
Semilog transformation	تحويل شبه لوغاريتمي
Significance level	مستوى معنوية
Sorting	تصنيف (عزل أو فصل)
Specification of model	تحديد النموذج
Spurious	زائف
Spurious Regression	الانحدار الزائف
Standard deviation	انحراف معياري
Standard error	خطأ معياري
Structural equation	معادلات هيكلية
Subscripts	دليل سفلي
Subset	مجموعة جزئية

T

Time series	سلسلة زمنية
Total sum of squares (TSS)	المجموع الكلي للمربعات
Trend	اتجاه عام
Trend-off	تبادل
Two Stage least squares (2STS)	المربعات الصغرى على مرحلتين
Two tailed test	اختبار من اتجاهين
Type I error	الخطأ من النوع الأول
Type II error	الخطأ من النوع الثاني
Typical	نمطي(نموذجي)

U

Unbiased	غير متحيز
Under-identified equation	معادلة ناقصة التميز
Univariate case	حال المتغير الواحد
Unrestricted model	نموذج غير مقيد

V

Variable	متغير
Variance	تباين

الجدول(ا): المساحات تحت المنحني الطبيعي المعياري Z:N~(0,1)

P(0 < Z < z)   The Standard Normal Distribution

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.000000	0.003989	0.007978	0.011967	0.015953	0.019939	0.023922	0.027903	0.031881	0.035856
0.1	0.039828	0.043795	0.047758	0.051717	0.055670	0.059618	0.063559	0.067495	0.071424	0.075345
0.2	0.079260	0.083166	0.087064	0.090954	0.094835	0.098706	0.102568	0.106420	0.110261	0.114092
0.3	0.117911	0.121719	0.125516	0.129300	0.133072	0.136831	0.140576	0.144309	0.148027	0.151732
0.4	0.155422	0.159097	0.162757	0.166402	0.170031	0.173645	0.177242	0.180822	0.184386	0.187933
0.5	0.191462	0.194974	0.198468	0.201944	0.205402	0.208840	0.212260	0.215661	0.219043	0.222405
0.6	0.225747	0.229069	0.232371	0.235653	0.238914	0.242154	0.245373	0.248571	0.251748	0.254903
0.7	0.258036	0.261148	0.264238	0.267305	0.270350	0.273373	0.276373	0.279350	0.282305	0.285236
0.8	0.288145	0.291030	0.293892	0.296731	0.299546	0.302338	0.305106	0.307850	0.310570	0.313267
0.9	0.315940	0.318589	0.321214	0.323814	0.326391	0.328944	0.331472	0.333977	0.336457	0.338913
1.0	0.341345	0.343752	0.346136	0.348495	0.350830	0.353141	0.355428	0.357690	0.359929	0.362143
1.1	0.364334	0.366500	0.368643	0.370762	0.372857	0.374928	0.376976	0.378999	0.381000	0.382977
1.2	0.384930	0.386860	0.388767	0.390651	0.392512	0.394350	0.396165	0.397958	0.399727	0.401475
1.3	0.403199	0.404902	0.406582	0.408241	0.409877	0.411492	0.413085	0.414656	0.416207	0.417736
1.4	0.419243	0.420730	0.422196	0.423641	0.425066	0.426471	0.427855	0.429219	0.430563	0.431888
1.5	0.433193	0.434478	0.435744	0.436992	0.438220	0.439429	0.440620	0.441792	0.442947	0.444083
1.6	0.445201	0.446301	0.447384	0.448449	0.449497	0.450529	0.451543	0.452540	0.453521	0.454486
1.7	0.455435	0.456367	0.457284	0.458185	0.459071	0.459941	0.460796	0.461636	0.462462	0.463273
1.8	0.464070	0.464852	0.465621	0.466375	0.467116	0.467843	0.468557	0.469258	0.469946	0.470621
1.9	0.471284	0.471933	0.472571	0.473197	0.473810	0.474412	0.475002	0.475581	0.476148	0.476705
2.0	0.477250	0.477784	0.478308	0.478822	0.479325	0.479818	0.480301	0.480774	0.481237	0.481691
2.1	0.482136	0.482571	0.482997	0.483414	0.483823	0.484222	0.484614	0.484997	0.485371	0.485738
2.2	0.486097	0.486447	0.486791	0.487126	0.487455	0.487776	0.488089	0.488396	0.488696	0.488989
2.3	0.489276	0.489556	0.489830	0.490097	0.490358	0.490613	0.490863	0.491106	0.491344	0.491576
2.4	0.491802	0.492024	0.492240	0.492451	0.492656	0.492857	0.493053	0.493244	0.493431	0.493613
2.5	0.493790	0.493963	0.494132	0.494297	0.494457	0.494614	0.494766	0.494915	0.495060	0.495201
2.6	0.495339	0.495473	0.495603	0.495731	0.495855	0.495975	0.496093	0.496207	0.496319	0.496427
2.7	0.496533	0.496636	0.496736	0.496833	0.496928	0.497020	0.497110	0.497197	0.497282	0.497365
2.8	0.497445	0.497523	0.497599	0.497673	0.497744	0.497814	0.497882	0.497948	0.498012	0.498074
2.9	0.498134	0.498193	0.498250	0.498306	0.498359	0.498411	0.498462	0.498511	0.498559	0.498605
3.0	0.498650	0.498694	0.498736	0.498777	0.498817	0.498856	0.498893	0.498930	0.498965	0.498999
3.1	0.499032	0.499064	0.499096	0.499126	0.499155	0.499184	0.499211	0.499238	0.499264	0.499289
3.2	0.499313	0.499336	0.499359	0.499381	0.499402	0.499423	0.499443	0.499462	0.499481	0.499499
3.3	0.499517	0.499533	0.499550	0.499566	0.499581	0.499596	0.499610	0.499624	0.499638	0.499650
3.4	0.499663	0.499675	0.499687	0.499698	0.499709	0.499720	0.499730	0.499740	0.499749	0.499758
3.5	0.499767	0.499776	0.499784	0.499792	0.499800	0.499807	0.499815	0.499821	0.499828	0.499835
3.6	0.499841	0.499847	0.499853	0.499858	0.499864	0.499869	0.499874	0.499879	0.499883	0.499888
3.7	0.499892	0.499896	0.499900	0.499904	0.499908	0.499912	0.499915	0.499918	0.499922	0.499925
3.8	0.499928	0.499930	0.499933	0.499936	0.499938	0.499941	0.499943	0.499946	0.499948	0.499950
3.9	0.499952	0.499954	0.499956	0.499958	0.499959	0.499961	0.499963	0.499964	0.499966	0.499967

الجدول (II): t توزيع ستودنت

α	0.2	0.1	0.05	0.02	0.01	0.005	0.002
1	3.078	6.314	12.706	31.821	63.656	127.321	318.289
2	1.886	2.920	4.303	6.965	9.925	14.089	22.328
3	1.638	2.353	3.182	4.541	5.841	7.453	10.214
4	1.533	2.132	2.776	3.747	4.604	5.598	7.173
5	1.476	2.015	2.571	3.365	4.032	4.773	5.894
6	1.440	1.943	2.447	3.143	3.707	4.317	5.208
7	1.415	1.895	2.365	2.998	3.499	4.029	4.785
8	1.397	1.860	2.306	2.896	3.355	3.833	4.501
9	1.383	1.833	2.262	2.821	3.250	3.690	4.297
10	1.372	1.812	2.228	2.764	3.169	3.581	4.144
11	1.363	1.796	2.201	2.718	3.106	3.497	4.025
12	1.356	1.782	2.179	2.681	3.055	3.428	3.930
13	1.350	1.771	2.160	2.650	3.012	3.372	3.852
14	1.345	1.761	2.145	2.624	2.977	3.326	3.787
15	1.341	1.753	2.131	2.602	2.947	3.286	3.733
16	1.337	1.746	2.120	2.583	2.921	3.252	3.686
17	1.333	1.740	2.110	2.567	2.898	3.222	3.646
18	1.330	1.734	2.101	2.552	2.878	3.197	3.610
19	1.328	1.729	2.093	2.539	2.861	3.174	3.579
20	1.325	1.725	2.086	2.528	2.845	3.153	3.552
21	1.323	1.721	2.080	2.518	2.831	3.135	3.527
22	1.321	1.717	2.074	2.508	2.819	3.119	3.505
23	1.319	1.714	2.069	2.500	2.807	3.104	3.485
24	1.318	1.711	2.064	2.492	2.797	3.091	3.467
25	1.316	1.708	2.060	2.485	2.787	3.078	3.450
26	1.315	1.706	2.056	2.479	2.779	3.067	3.435
27	1.314	1.703	2.052	2.473	2.771	3.057	3.421
28	1.313	1.701	2.048	2.467	2.763	3.047	3.408
29	1.311	1.699	2.045	2.462	2.756	3.038	3.396
30	1.310	1.697	2.042	2.457	2.750	3.030	3.385
40	1.303	1.684	2.021	2.423	2.704	2.971	3.307
50	1.299	1.676	2.009	2.403	2.678	2.937	3.261
60	1.296	1.671	2.000	2.390	2.660	2.915	3.232
70	1.294	1.667	1.994	2.381	2.648	2.899	3.211
80	1.292	1.664	1.990	2.374	2.639	2.887	3.195
90	1.291	1.662	1.987	2.368	2.632	2.878	3.183
100	1.290	1.660	1.984	2.364	2.626	2.871	3.174
200	1.286	1.653	1.972	2.345	2.601	2.838	3.131
500	1.283	1.648	1.965	2.334	2.586	2.820	3.107

جدول(III): توزيع كاي مربع χ^2

λ	0.05	0.025	0.01	0.005	0.995	0.99	0.975	0.95
1	3.841	5.024	6.635	7.879	0.00004	0.0002	0.001	0.004
2	5.991	7.378	9.210	10.597	0.010	0.020	0.051	0.103
3	7.815	9.348	11.345	12.838	0.072	0.115	0.216	0.352
4	9.488	11.143	13.277	14.860	0.207	0.297	0.484	0.711
5	11.070	12.832	15.086	16.750	0.412	0.554	0.831	1.145
6	12.592	14.449	16.812	18.548	0.676	0.872	1.237	1.635
7	14.067	16.013	18.475	20.278	0.989	1.239	1.690	2.167
8	15.507	17.535	20.090	21.955	1.344	1.646	2.180	2.733
9	16.919	19.023	21.666	23.589	1.735	2.088	2.700	3.325
10	18.307	20.483	23.209	25.188	2.156	2.558	3.247	3.940
11	19.675	21.920	24.725	26.757	2.603	3.053	3.816	4.575
12	21.026	23.337	26.217	28.300	3.074	3.571	4.404	5.226
13	22.362	24.736	27.688	29.819	3.565	4.107	5.009	5.892
14	23.685	26.119	29.141	31.319	4.075	4.660	5.629	6.571
15	24.996	27.488	30.578	32.801	4.601	5.229	6.262	7.261
16	26.296	28.845	32.000	34.267	5.142	5.812	6.908	7.962
17	27.587	30.191	33.409	35.718	5.697	6.408	7.564	8.672
18	28.869	31.526	34.805	37.156	6.265	7.015	8.231	9.390
19	30.144	32.852	36.191	38.582	6.844	7.633	8.907	10.117
20	31.410	34.170	37.566	39.997	7.434	8.260	9.591	10.851
21	32.671	35.479	38.932	41.401	8.034	8.897	10.283	11.591
22	33.924	36.781	40.289	42.796	8.643	9.542	10.982	12.338
23	35.172	38.076	41.638	44.181	9.260	10.196	11.689	13.091
24	36.415	39.364	42.980	45.558	9.886	10.856	12.401	13.848
25	37.652	40.646	44.314	46.928	10.520	11.524	13.120	14.611
26	38.885	41.923	45.642	48.290	11.160	12.198	13.844	15.379
27	40.113	43.195	46.963	49.645	11.808	12.878	14.573	16.151
28	41.337	44.461	48.278	50.994	12.461	13.565	15.308	16.928
29	42.557	45.722	49.588	52.335	13.121	14.256	16.047	17.708
30	43.773	46.979	50.892	53.672	13.787	14.953	16.791	18.493
40	55.758	59.342	63.691	66.766	20.707	22.164	24.433	26.509
50	67.505	71.420	76.154	79.490	27.991	29.707	32.357	34.764
60	79.082	83.298	88.379	91.952	35.534	37.485	40.482	43.188
70	90.531	95.023	100.425	104.215	43.275	45.442	48.758	51.739
80	101.879	106.629	112.329	116.321	51.172	53.540	57.153	60.391
90	113.145	118.136	124.116	128.299	59.196	61.754	65.647	69.126
100	124.342	129.561	135.807	140.170	67.328	70.065	74.222	77.929
150	179.581	185.800	193.207	198.360	109.142	112.668	117.985	122.692
200	233.994	241.058	249.445	255.264	152.241	156.432	162.728	168.279
500	553.127	563.851	576.493	585.206	422.303	429.387	439.936	449.147

1% Critical Values of the *F* Distribution

Denominator Degrees of Freedom	Numerator Degrees of Freedom: 1	2	3	4	5	6	7	8	9	10
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63
90	6.93	4.85	4.01	3.54	3.23	3.01	2.84	2.72	2.61	2.52
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32

Example: The 1% critical value for numerator $df = 3$ and denominator $df = 60$ is 4.13.
Source: This table was generated using the Stata® function invfprob.

Table A.5 Critical Values for the Durbin–Watson Test: 5% Significance Level[a]

n	$k = 3$ L	$k = 3$ U	$k = 6$ L	$k = 6$ U	$k = 9$ L	$k = 9$ U	$k = 12$ L	$k = 12$ U	$k = 15$ L	$k = 15$ U	$k = 21$ L	$k = 21$ U
10	0.70	1.64	0.24	2.82								
15	0.95	1.54	0.56	2.22	0.25	2.98						
20	1.10	1.54	0.79	1.99	0.50	2.52	0.26	3.06	0.10	3.54		
25	1.21	1.55	0.95	1.89	0.70	2.28	0.47	2.70	0.27	3.12	0.04	3.79
30	1.28	1.57	1.07	1.83	0.85	2.14	0.64	2.48	0.45	2.82	0.16	3.47
35	1.34	1.58	1.16	1.80	0.97	2.05	0.78	2.33	0.60	2.62	0.30	3.19
40	1.39	1.60	1.23	1.79	1.06	2.00	0.90	2.23	0.73	2.47	0.43	2.97
45	1.43	1.62	1.29	1.78	1.14	1.96	0.99	2.16	0.84	2.37	0.55	2.81
50	1.46	1.63	1.34	1.77	1.20	1.93	1.06	2.10	0.93	2.29	0.66	2.68
55	1.49	1.64	1.37	1.77	1.25	1.91	1.13	2.06	1.00	2.23	0.75	2.57
60	1.51	1.65	1.41	1.77	1.30	1.89	1.19	2.03	1.07	2.18	0.84	2.49
65	1.54	1.66	1.44	1.77	1.34	1.88	1.23	2.01	1.12	2.14	0.91	2.42
70	1.55	1.67	1.46	1.77	1.37	1.87	1.27	1.99	1.17	2.11	0.97	2.36
75	1.57	1.68	1.49	1.77	1.40	1.87	1.31	1.97	1.22	2.08	1.03	2.32
80	1.59	1.69	1.51	1.77	1.43	1.86	1.34	1.96	1.25	2.06	1.08	2.28
85	1.60	1.70	1.53	1.77	1.45	1.86	1.37	1.95	1.29	2.04	1.12	2.24
90	1.61	1.70	1.54	1.78	1.47	1.85	1.40	1.94	1.32	2.03	1.16	2.21
95	1.62	1.71	1.56	1.78	1.49	1.85	1.42	1.93	1.35	2.01	1.20	2.19
100	1.63	1.72	1.57	1.78	1.51	1.85	1.44	1.92	1.37	2.00	1.23	2.16
150	1.71	1.76	1.67	1.80	1.62	1.85	1.58	1.90	1.54	1.94	1.44	2.04
200	1.75	1.79	1.72	1.82	1.69	1.85	1.65	1.89	1.62	1.92	1.55	1.99

[a] k is the number of explanatory variables + 1 (constant term).

Source: This table is adapted from N. E. Savin and K. J. White, "The Durbin–Watson Test for Serial Correlation with Extreme Sample Sizes or Many Parameters," *Econometrica*, Vol. 45, 1977, p. 1989–1996. We have given the table only for some sample sizes and number of variables. For intermediate sample sizes or number of variables, interpolation can be used. Given the limitations of the DW test, we find

المراجع باللغة العربية

- د.أبو سدرة، فتحي صالح ؛ أ. الكيخا، نجاة رشيد ؛ - الإحصاء والاقتصاد القياسي؛ - الطبعة الأولى - جامعة فارس يونس 1999
- اف والس - ترجمة د.عادل عبد الغني محبوب؛ مقدمة في الاقتصاد القياسي؛ جامعة المستنصرية 1981 .
- د.البشير، زين العابدين عبد الرحيم؛ د. عودة، أحمد عبد المجيد ؛ - الاستدلال الإحصائي؛ - جامعة الملك سعود 1997 .
- جون نتر - ويليام وازرمان - ميخائيل كتنر - ترجمة أ.د. كنجو، أنيس إسماعيل وآخرون؛ - نماذج إحصائية خطية تطبيقية- انحدار،تحليل تباين وتصاميم تجريبية- الجزء الأول ؛ النشر العلمي والمطابع- جامعة الملك سعود- الرياض- 1995
- جون نتر - ويليام وازرمان - ميخائيل كتنر- ترجمة أ.د.أنيس إسماعيل كنجو وآخرون؛ - نماذج إحصائية خطية تطبيقية- انحدار،تحليل تباين وتصاميم تجريبية - الجزء الثاني ؛ النشر العلمي والمطابع- جامعة الملك سعود- الرياض- 1995
- د.حليمي، عبدالقادر ؛ - مدخل إلى الإحصاء؛ - منشورات عويدات- بيروت-1985
- د.حميدان، عدنان عباس ؛ د. النعيمي، قاسم محمد ؛ د. آغا، عمار ناصر ؛ - مبادئ الإحصاء؛ جامعة دمشق - مركز التعلم المفتوح- 2005

- دومينيك سالفاتور -ترجمة د. منتصر، سعدية حافظ ؛ مراجعة د. أنيس، عبدالعظيم ؛ سلسلة ملخصات شوم نظريات ومسائل في الاحصاء والاقتصاد القياسي؛ دار ماكجروهيل للنشر -1982
- د.الزعبي، محمد بلال – أ.عباس الطلافحة؛ – النظام الإحصائي Spss – فهم وتحليل البيانات الإحصائية؛ – الجامعة الأردنية – الطبعة الأولى 2000 .
- د.السيد ،عباس ؛ – الاقتصاد القياسي؛ – دار الجامعات المصرية 1990.
- د.السيفو، وليد إسماعيل؛ د. أحمد محمد مشعل؛ – الاقتصاد القياسي بين النظرية والتطبيق؛ – عمان – دار المجدلاوي – 2003 .
- د.شربجي ،عبد الرزاق ؛ – الاقتصاد القياسي التطبيقي – نماذج قياسية تطبيقية لاقتصاديات الدول العربية؛ – الطبعة الثانية – دار العلم للملايين 1993 .
- د.الشربجي، مجدي؛ التنبؤ الكمي للمشروعات الحكومية والأساليب والنماذج والتطبيقات؛ الدار المصرية اللبنانية 1994
- د.الشربجي، مجدي ؛ – الاقتصاد القياسي النظرية والتطبيق؛ – الدار المصرية اللبنانية – طبقة أولى 1994 .
- د.شريف، عصام عزيز ؛ – مقدمة في الاقتصاد القياس؛ – الطبعة الثالثة – دار الطليعة 1983.
- د..عبد الرحمن، عبد المحمود محمد ؛ – مقدمة في الاقتصاد القياسي؛ – جامعة الملك سعود 1995
- د.عبد الرحمن، نصر ؛ عبد المحمود محمد؛ – مقدمة في الاقتصاد والقياس؛ – جامعة الملك سعود 1995 .

- د.عطية، عبد القادر محمد عبد القادر ؛ – الاقتصاد القياسي بين النظرية والتطبيق؛ – الدار الجامعية – 1998 .
- د.عكاشة، محمود خالد ؛ – استخدام نظام SPSS في تحليل البيانات الإحصائية ؛ – فلسطين – غزة –2002
- د.العلي، إبراهيم محمد ؛ – الرياضيات الاقتصادية (المالية والإدارية)؛ – جامعة تشرين–1992
- د.العلي، إبراهيم محمد ؛ – نظرية الارتباط ؛ – جامعة حلب–
- العلي، إبراهيم محمد – د. كابوس، أمل ؛ – الإحصاء الرياضي؛ – جامعة حلب – 1986
- د.العلي، إبراهيم محمد؛ د. كابوس، أمل ؛د.حلاق، عمر ؛ نظرية الاحتمالات؛ جامعة حلب – 1985
- د.عوض، طالب محمد ؛ – مقدمة في الاقتصاد القياسي؛ الجامعة الأردنية 2000 .
- د.العيسوي، إبراهيم ؛ – القياس والتنبؤ في الاقتصاد؛ – دار النهضة .
- د.فرحات، محمد لطفي ؛ – مبادئ الاقتصاد القياسي (قياس العلاقات الاقتصادية)؛ – الدار الجماهيرية للتوزيع والإعلان 1986 .
- د.كابوس، أمل ؛ – مبادئ الاقتصاد القياسي؛ – جامعة حلب– 1981
- موراي ر. شبيجل–ترجمة د. شعبان، شعبان عبدالحميد؛ مراجعة د. الموازيني، أحمد ؛ سلسلة ملخصات شوم نظريات ومسائل في الإحصاء؛ دار ماكجروهيل للنشر– الطبعة الثالثة –1993
- د.النعيمي محمد عبد العال ؛ الحمداني، رفاه شهاب ؛ عبد الرزاق، كنعان عبد اللطيف ؛ نظرية الاقتصاد القياسي؛ العراق الجامعة المستنصرية – 1991

- د.النعيمي، قاسم محمد؛ أساسيات الإحصاء؛ الطبعة الثالثة – دار الشوكاني 2001 .
- هادي كلجيان والاس أوتس ترجمة د. المرسي سيد حجازي – د. عبد القادر محمد عطية؛ مقدمة في الاقتصاد القياسي المبادئ والتطبيقات؛ جامعة الملك سعود 2001 .
- هندرسون وكواندت – ترجمة د . متوكل عباس مهلهل؛ نظرية اقتصاديات الوحدة – أسلوب رياضي؛ دار ماكجروهيل للنشر 1980 .
- والتر فاندل– تعريب د. عزام، عبد المرضي حامد ود. هارون، أحمد حسين ؛ السلاسل الزمنية – من الوجهة التطبيقية ونماذج بوكس–جنكنز ؛ الرياض – دار المريخ – 1992

المراجع الأجنبية باللغة الانكليزية

- Introduction to Econometrics – James H . Stock – Mark W . Watson – International Edition – 2003
- Introductory Econometrics A Modern Approach – 2 E Jeffrey M . Wooldridge .
- Introduction to econometrics G . S Medal second Edition – 1992 .
- Copyright Christopher Dougherty 1992 – 2002
- Econometrics and analysis for Developing countries, Chandan mukkerjee, Howard and marc Wuyts – 1998.
- G.S. Maddala_Introduction to Econometrics_second edition. University of Florida and OHIO State University
- Introductory Econometrics-A Modern Approach_ Wooldridge (2004)
- CONSUMER EXPENDITURE SURVEY - STATA COMMANDS- c.dougherty@lse.ac.uk-January-2002
- SCHOOL COSTS DATA SET REGRESSION EXERCISES c.dougherty@lse.ac.uk April 2002
- Statistical Case Studies;Wolfgang H ardle, Yuichi Mori, Philippe Vieu May 6, 2004 >
- EDUCATIONAL ATTAINMENTAND EARNINGS FUNCTIONS -REGRESSION EXERCISES c.dougherty@lse.ac.uk September 2003 –
- Computer-Aided Introduction to Econometrics
- Juan M. Rodriguez Poo;In cooperation with;Ignacio Moral, M. Teresa Aparicio, Inmaculada Villanua,
- Pavel _C_zek, Yingcun Xia, Pilar Gonzalez, M. Paz Moral, Rong Chen,Rainer Schulz, Sabine

Stephan, Pilar Olave,J. Tomas Alcala and Lenka Cizkova

- January 17, 2003
- Handbook of statistics 12- Enviaronmental Statistics –Edited by G.P.Patil ;C.R.Rao-1994
- ***Introduction to Econometrics* (Oxford: Oxford University Press, 2002), second edition [ISBN 0 19 877643 8].**

المراجع الأجنبية باللغة الروسية

- Вороновицкий М.М., Мейталь Ш. **Модель социального влияния на цены.** – Экономика и математические методы. Т.39, № 4. – Москва: "Наука", 2003.
- Замков О.О., Толстопятенко А.В., Черемных Ю.Н. **Математические методы в экономике.** – Москва: "ДИС", 1998.
- Казанцев С.В. **Теоретические модели цен.** – Новосибирск: "Наука", 1987.
- Козырь Ю.В. **Модель прогноза *Formod* : принципы, структура и механизм функционирования.** – Экономика и математические методы. Т.39, № 3. – Москва: "Наука", 2003.
- Кубонива М., Табата М., Табата С., Хасэбэ Ю. **Математическая экономика на персональном компьютере**: Пер. с япон./ Под ред. и с предисл. Е.З. Демиденко. – Москва: "Финансы и статистика", 1991.
- Магнус Я.Р., Катышев П.К., Пересецкий А.А. **Эконометрика. Начальный курс.** – Москва: "Дело", 1997.
- Маленво Э. **Лекции по микро-экономическому анализу.** – Москва: "Наука", 1985.
- **Моделирование глобальных экономических процессов**: Под редакцией В.С. Дадаяна. – Москва: "Экономика", 1984.
- Соложенцев Е.Д., Карасев В.В. **Логико-вероятностные модели риска в бизнесе с группами несовместных событий.** – Экономика и

математические методы. Т.39, № 1. – Москва: "Наука", 2003.

- Фишер Ф. **Проблема идентификации в эконометрии.** – Москва: "Статистика", 1978.

اللجنة العلمية

الأستاذ الدكتور إبراهيم محمد العلي

الأستاذ الدكتور خلف مطر الجراد

الأستاذ المساعد الدكتور زياد زنبوعة

المدقق اللغوي

الدكتور خالد الحلبوني

كلية الآداب / قسم اللغة العربية