



مبادئ الإحصاء





منشورات جامعة دمشق
كلية الآداب والعلوم الإنسانية
قسم الإعلام

مبادئ الإحصاء

الدكتور

عدنان غاتم

أستاذ مساعد في قسم الإحصاء التطبيقي
كلية الاقتصاد

الدكتور

رمضان درويش

أستاذ مساعد في قسم القياس والتقويم
كلية التربية

1431 - 1430

2010 - 2009

جامعة دمشق



الفهرس

رقم الصفحة	الموضوع
7	المقدمة
11	الفصل الأول : أهمية علم الإحصاء
13	تعريف علم الإحصاء
25	الفصل الثاني : جمع وتبويب وعرض البيانات
27	مصادر البيانات الإحصائية
81	الفصل الثالث : مقاييس النزعة المركزية
84	المونوال Mode
101	الفصل الرابع : مقاييس التشتت
103	المدى المطلق (R)
131	الفصل الخامس : الارتباط والاتحاد الخطي البسيط
134	العلاقة الارتباطية
163	الفصل السادس : مبادئ الاحتمالات
165	تفسير الاحتمال بأنه تكراراً نسبياً
181	الفصل السابع : الاستدلال الإحصائي للعينات الكبيرة الحجم
185	قانون الأعداد الكبيرة
229	الفصل الثامن : الاستدلال الإحصائي للعينات الصغيرة الحجم
232	خصائص التوزيع t
261	الفصل التاسع : الإحصاء اللامعلمي
264	اختبار حسن المطابقة

271	اختبار الاستقلال لجداول التوافق
281	اختبار الإشارة Sign test
282	حالة وسيط عينة
327	الفصل العاشر : تحليل التباين
329	تحليل التباين أحادي التصنيف
341	الملاحق : الرموز والأحرف الأبجدية
347	قائمة المراجع
351	الجداول الإحصائية
373	المصطلحات الإحصائية

مقدمة:

إن كلمة " الإحصاء " تعني لغوياً كماً من المعطيات العددية، لها معانٍ كثيرة ومتباينة، فعند سماع هذه الكلمة يتبادر إلى أذهان البعض من الناس مفهوم الأرقام الخاصة بعدد السكان أو عدد الأسر أو عدد صناديق الاقتراع، والبعض الآخر وبعضهم يبني مفهوم الإحصاء لديهم بناءً على ما تنشره الصحف والمجلات من بيانات مختلفة عن استطلاعات الرأي العام أو تلك البيانات التي توضح الصادرات والواردات لدولة ما.. إلى غير ذلك من الأمثلة المتداولة يومياً.

عُرف الإحصاء منذ أقدم العصور، فقد عرفه المصريون القدماء والرومان والصينيون، وكان الهدف منه معرفة المقدار العددي للقوة البشرية لمعرفة مقدرة الدولة على الدفاع، وأصبح فيما بعد يتصل بفكرة الأرقام والاعداد ويسعى لجمع البيانات والمعلومات عن الظواهر المختلفة.

برزت أهمية الإحصاء في الستينيات من القرن الماضي كأساس ضروري لعملية التنمية الاقتصادية والاجتماعية وفي تطوير العلوم ونموها باعتباره العلم المساعد لأغلب العلوم.

في الوقت الحاضر، أصبح الإحصاء بمفهومه العلمي، كطريقة للحصول على البيانات الإحصائية وكأداة للتحليل، يحتل مكاناً مرموقاً في مضمار العلم والمعرفة ويحظى باهتمام جميع الدول في رسم سياساتها ووضع خططها التنموية وذلك لوضع اقتصادياتها على المسار الذي يحقق الأهداف المرجوة من هذه الخطط.

لقد أصبح التفكير الإحصائي ضرورة ملحة لكل فرد ومؤسسة، ومع تطور العلم الحديث تغيرت وظيفة الإحصائي واتخذ علم الإحصاء معنى آخر هو علم اتخاذ القرارات الذي يمهّد الطريق لكشف الحقائق.

لقد بات معروفاً، أن الإحصاء لم يعد حكراً على الرياضيين، بل اعتمد العديد من الباحثين تطبيقه في المجالات التي يعملون بها، فالباحث الاقتصادي أو الاجتماعي اليوم يستخدم العديد من الموضوعات التي يمكن أن تخضع للقياس والتعبير عنها كمياً مثل: العلاقة بين مستوى الدخل ومستوى المشاركة السياسية والعلاقة بين مستوى الدخل والانتماء السياسي؟ وأثر وسائل الإعلام في السلوك السياسي؟ وغيرها من العلاقات..

نأمل في فصول هذا الكتاب، أن نكون قدّمنا مفهوماً واضحاً لعلم الإحصاء والتحليل الإحصائي ومدى فائدتهما لتخصص الإعلام، إذ تجلّى هدفنا في تزويد طلاب قسم الإعلام بالمفاهيم الإحصائية الأكثر استخداماً والتي نفيدهم في مجال تخصصهم.

لقد توخينا البساطة والسهولة في العرض الرياضي، وابتعدنا عن البراهين المعقدة ليكون هذا الكتاب في متناول الدارسين والمهتمين بعلم الإحصاء.

توزعت محتويات هذا الكتاب إلى عشرة فصول، كالآتي:

الفصل الأول: أهمية علم الإحصاء

الفصل الثاني: جمع وتبويب وعرض البيانات الإحصائية

الفصل الثالث: مقاييس النزعة المركزية

الفصل الرابع: مقاييس التشتت.

الفصل الخامس: الارتباط والاتحاد الخطي البسيط.

الفصل السادس: مبادئ الاحتمالات.

الفصل السابع: الاستدلال الإحصائي للعينات الكبيرة الحجم.

الفصل الثامن: الاستدلال الإحصائي للعينات الصغيرة الحجم.

الفصل التاسع: الإحصاء اللامعلمي.

الفصل العاشر: تحليل التباين.

وفي نهاية الكتاب، قمنا بوضع الجداول الإحصائية التي تساعد الباحث في فهم أصول التحليل الإحصائي والمتعلقة بفصول الكتاب، بالإضافة إلى قائمة بالمصطلحات الإحصائية والمراجع العربية والأجنبية ومواقع الانترنت. إن إعداد هذا الكتاب هو إعداد جماعي ومسؤولية مشتركة، لكل من الدكتور عدنان غانم والدكتور رمضان درويش.

لقد بذلنا جهدنا في تقديم الأفكار التي تثير اهتمام طلاب قسم الإعلام، وإننا نضع هذا الكتاب بين أيديهم، ونأمل منهم جميعاً أن يتفضلوا بتقديم أي نقد يساهم في إغناء موضوعاته، وأخيراً نقدم شكرنا وتقديرنا لكل من ساهم في إخراج هذا الكتاب، والله ولي التوفيق.

المؤلفان



الفصل الأول

أهمية علم الإحصاء

- تعريف علم الإحصاء
- تعاريف ومصطلحات هامة في الإحصاء
- خطوات البحث الإحصائي
- مثال تطبيقي عن مراحل البحث الإحصائي.
- مجالات استخدام الإحصاء (التخطيط، العلوم الاجتماعية، الإعلام، الاقتصاد، السكان، السياسة، علم النفس والتربية).
- مزايا علم الإحصاء



أهمية علم الإحصاء

1-1- مقدمة:

تتبع أهمية الإحصاء، باعتباره العلم المساعد لأغلب العلوم وأحد أدوات البحث العلمي الذي يفسر الغموض حول ظاهرة معينة، وتتبع أهميته أيضاً من خلال استخدام أدواته في مختلف مجالات الحياة للتأكد من ظاهرة معينة أو نفيها أو قياسها وعرض النتائج في مؤشرات وصفية أو استدلالية.

1-2- تعريف علم الإحصاء:

هناك الكثير من التعاريف الشائعة والمألوفة للإحصاء، على سبيل

المثال:

- هو أداة علمية تتناول الأسس والطرائق العلمية في جمع، وتبويب، وعرض، وتحليل البيانات العددية وتفسيرها، ومن ثم اتخاذ القرارات العقلانية في ضوء النتائج المستخلصة.

هو الأسلوب أو الطريقة التي يستخدمها الباحث من أجل الوصول إلى الحقيقة عن طريق افتراضات معينة يحاول الباحث إثباتها.

هو علم يقدم النماذج الرياضية والتحليلية لاتخاذ القرارات العقلانية.

هو فرع من العلوم الرياضية التطبيقية يقوم بجمع واختصار وتحليل البيانات العددية لاتخاذ القرارات العقلانية.

هو جزء من الرياضيات التطبيقية يعالج سلوك الأقسام ويستخلص القوانين الرياضية.

هو العلم الذي يدرس الأفراد في المجتمع كونهم وحدات رياضية ويقوم على تمثيل الوقائع الحيوية بجدول وخطوط بيانية بالاعتماد على الأساس الرقمي.

نلاحظ أن معظم هذه التعاريف تركز على ما يهدف إليه الإحصاء وهو تجميع البيانات العددية وتنظيمها وتحليلها وتفسيرها.

3-1- تعريف ومصطلحات هامة مستخدمة في الإحصاء:

— الرقم الإحصائي : عبارة عن مؤشر يعكس حالة الظاهرة المدروسة (حجم، عدد...)

— المجتمع الإحصائي: هو مجموعة المفردات (الوحدات الإحصائية) المطلوب جمع البيانات منها حول ظاهرة معينة ما، فإذا كان البحث يتعلق بإجراء بحث عن واقع التعليم المفتوح في سورية، فإن مجتمع البحث سيكون هنا جميع طلبة التعليم المفتوح في سورية، وهنا نميز:

— المجتمع الإحصائي المحدود: عبارة عن المجتمع الذي يمكن حصر مفرداته مثل عدد طلاب جامعة دمشق، عدد طلاب التعليم المفتوح في سورية.

— المجتمع الإحصائي غير المحدود: عبارة عن المجتمع الذي لا يمكن حصر مفرداته مثل عدد أسماك البحر.

ونذكر أيضاً أن هناك نوعين من المجتمعات الإحصائية:

- 1 — مجتمع الهدف Target population وهو المجتمع المستهدف بالدراسة.
- 2 — مجتمع العينة Sampled population وهو المجتمع الذي سيتم اختيار العينة منه.

وتجدر الإشارة، إلى أن مفهوم المجتمع الإحصائي هو مفهوماً مرناً، أي يمكن زيادة وتقليص حجمه حسب موضوع الدراسة في ضوء عدة اعتبارات، منها:

أ — هدف الدراسة.

ب — الإمكانيات المادية.

ج — الموارد المتاحة.

— الوحدة الإحصائية: هي أصغر جزء مستقل تجري عليه الدراسة الإحصائية، أي يتم جمع البيانات عنها وهي ثابتة ومستقرة باستثناء (الوحدة النقدية)، فعندما يتعلق البحث بالطلبة، فإن وحدة البحث هي الطالب، وإذا كان البحث يتعلق

بالعمال، فإن وحدة البحث هي العامل وهكذا ...، مع العلم أن هناك أنواعاً للوحدات الإحصائية هي:

أ - وحدات بسيطة: وهي وحدات العد ووحدات القياس.

ب - وحدات مركبة: وهي عبارة عن الوحدات المركبة من وحدتين مثل: كم/سا.

- العينة الإحصائية: عبارة عن جزء من المجتمع الإحصائي، ويجب أن يمثل هذا الجزء المجتمع الذي سحب منه أصدق تمثيل، ويتم ذلك من خلال الاختيار العشوائي للمفردات لتحقيق إمكانية تعميم النتائج على المجتمع الذي سحبت منه العينة.

المعايينة: هي الإجراء الذي يمكن بواسطته أن نستقرأ خصائص المجتمع من خلال عدد من مفرداته (العينة) وحتى نتمكن من تعميم نتائج العينة على المجتمع يجب:

أ - تحديد حجم العينة بدقة.

ب - الاختيار العشوائي.

ج - إمكانية تقدير الخطأ الناتج من استخدام العينة.

د - تحديد درجة الثقة في تقديرات المجتمع في نتائج العينة.

هـ - أن تكون العينة بسيطة بشكل كافٍ لتنفيذها في الواقع.

- إطار البحث: هو أي وسيلة يمكن بها حصر مفردات المجتمع مثل: سجلات الطلاب في الكليات أو سجلات الموظفين في الوزارات.

- المتغير: عبارة عن قيم الظاهرة المدروسة مثل: حجم الصادرات، أجر العامل، درجات الطالب في مقرر ما.

- متغير كمي: نقول عن متغير ما أنه كمي، إذا كان بالإمكان التعبير عنه رقمياً، مثل: درجات الحرارة، أرباح شركة ما، درجات طالب في مقرر ما.

متغير نوعي: يقصد به المتغير الذي لا يمكن قياسه مثل: اللون، الجنسية.

— المتغير العشوائي: عبارة عن متغير لا يمكن التنبؤ بقيمته مسبقاً، مثل: كمية الأمطار الساقطة في منطقة ما، إنفاق أسرة ما على الثقافة، وينقسم المتغير العشوائي إلى:

— المتغير العشوائي المنفصل (المتقطع): عبارة عن التغير الذي يأخذ قيماً صحيحة ولا يمكن التعبير عنها بأجزاء مثل: عدد الطلاب، عدد العمال، عدد أفراد الأسرة.

— المتغير العشوائي المتصل (المستمر): عبارة عن المتغير الذي يأخذ أي قيمة ضمن مجال مثل: الطول، الوزن، العمر.

الإحصاء الوصفي: عبارة عن الإحصاء الذي يهتم بجمع البيانات وعرضها، أي تقديم البيانات في صورة يسهل فهمها دون الوصول إلى النتائج.

الإحصاء الاستدلالي: عبارة عن الإحصاء الذي يأتي دوره بعد إنجاز الإحصاء الوصفي، حيث يستخدم المعلومات التي توصل إليها الإحصاء الوصفي للتوصل إلى نتائج معينة من خلال اختبار الفرضيات والتقدير الإحصائي.

— البحث الإحصائي: عبارة عن وسيلة لدراسة ظاهرة معينة أو مشكلة ما في محاولة للتعرف على العوامل التي أدت إلى ظهورها من خلال جمع وتبويب وعرض البيانات.

— البيانات: هي مؤشرات أولية تجمع من المجتمع مباشرة أو بأسلوب غير مباشرة وليس لها معنى.

— المعلومات: هي مؤشرات طبقت عليها الطرائق الإحصائية لكي تجعلها أكثر وضوحاً للباحث.

4-1- خطوات البحث الإحصائي:

المرحلة الأولى (الأعمال المكتبية) وتتضمن:

- اختيار الموضوع (الاهتمام الشخصي، توفر البيانات، التمويل).
- تحديد إطار البحث (الوحدة الإحصائية، المجتمع الإحصائي).
- تحديد المشكلة والهدف من البحث
- تحديد فروض البحث، بحيث تكون الفرضيات دقيقة وواضحة وقابلة للاختبار.
- خطة البحث
- الكلفة

— تصميم استمارة البحث (الاستبيان).

المرحلة الثانية (الأعمال الميدانية) وتتضمن:

- جمع البيانات (المصادر الميدانية، المصادر الوثائقية)

المرحلة الثالثة (مرحلة تجهيز وتحليل ونشر النتائج) وتتضمن:

- اختيار الطرائق الإحصائية اللازمة.
- تحليل البيانات واستخلاص النتائج
- اختبار صحة الفروض
- دراسة احتمالات تعميم النتائج التي يحصل عليها الباحث عند اختبار الفروض
- اتخاذ القرار والتوصيات اللازمة.
- كما ينبغي على الباحث عند إجراء البحث مراعاة، عدم المغالاة في جمع البيانات وبالإضافة إلى إلمامه إلى حد كبير بالموضوع المرغوب دراسته.
- من خلال ذلك نجد أن الإحصاء يخدم هدفين أساسيين:
- توفير الأسس العلمية لجمع البيانات وتبويبها وحساب المقاييس التلخيصية والوصفية منها.

— تحليل البيانات التي تم جمعها، بهدف بناء تقديرات إحصائية لبعض الظواهر المدروسة وتفسيرها واستخلاص الاستنتاجات، وبناءً على جزء من المجتمع (عينة) للتوصل إلى قرارات يمكن تعميمها على كامل أجزاء المجتمع، وهو ما يعرف بالإحصاء الاستدلالي.

* مثال تطبيقي عن مراحل البحث الإحصائي:

لنفرض أننا أمام ظاهرة تزايد عدد حوادث المرور في مدينة ما، ونريد اتخاذ قرار بشأن الحلول الممكنة للحد من هذه الظواهر نظراً لآثارها السلبية على المجتمع، وتحليل أسباب الحوادث وملابساتها. مثل: (عدم التقيد بإرشادات المرور، السرعة، التجاوز غير القانوني، السائق تحت تأثير مخدر.. إلخ)، تتبع الخطوات التالية:

1 — تحديد المشكلة وهدفها: الظاهرة أو المشكلة هنا تزايد عدد حوادث

المرور بالمقارنة مع السنوات الماضية مثلاً، والهدف هو الحد من تزايد حوادث المرور.

2 — الوحدة الإحصائية والمجتمع الإحصائي: إن الوحدة الإحصائية هنا فسي مثالنا سائق السيارة، والمجتمع هو مجتمع السائقين (مالكي السيارات).

3 — تحديد نوع البيانات المطلوب جمعها: ويكون نوع البيانات المطلوب جمعها في هذه الحالة عدد الحوادث بالنسبة لكل نوع من أنواع السيارات (خاصة، أجرة، شاحنات..) وكذلك لكل نوع من أنواع الحوادث (تصادم سيارتين، تصادم مع جسم ثابت، تصادم سيارة بأحد المارة) ويمكن التقسيم أيضاً بالنسبة لآثار الحادث (إصابة سطحية، إصابة خطيرة، إصابة قاتلة) ويمكن أيضاً (اصطدام، دهس، انقلاب).

4 — تحديد مصادر جمع البيانات: يمكن جمع البيانات في هذه الحالة بأسلوبين:

الأسلوب الأول : من سجلات إدارة المرور بالمدينة (المصادر الوثائقية أو التاريخية).

الأسلوب الثاني: عن طريق عمل استقصاء ميداني للسائقين لمختلف أنواع السيارات للسؤال عن الرأي في نظام المرور، والإشارات الضوئية، ومدى سلامة الطرق، (المصادر الميدانية).

5 - تحديد كيفية جمع البيانات: بالنسبة للأسلوب الأول يتم الاتصال المباشر بإدارة المرور بالمدينة والحصول على البيانات المطلوبة كافة، أما بالنسبة للأسلوب الثاني فيتم عمل مجموعة مدربة من العدادين وتوزيعهم في مناطق مختلفة من المدينة لجمع البيانات ميدانياً بواسطة استمارات وجدول إحصائية أعدت خصيصاً لهذا الغرض.

6 - تحديد كلفة البحث: من خلال مكافأة القائمين على جمع البيانات من الميدان وتكلفة المطبوعات.

7 - تحديد الفترة الزمنية للدراسة: حيث يتم تحديد هذه الفترة بحد أقصى ثلاثة شهور مثلاً، نظراً للأهمية في سرعة الإنجاز، وللوصول لجميع الجوانب الخاصة بالمشكلة، واتخاذ القرارات المناسبة في ضوء نتائج الدراسة.

8 - استخدام المقاييس الإحصائية المناسبة: من أجل الحصول على المؤشرات التلخيصية والوصفية العامة، لجميع عناصر المشكلة (مثلاً: متوسط عدد حوادث المرور في منطقة معينة، علاقة الارتباط بين عمر السائق وعدد الحوادث، كثافة المرور على طريق معينة ... إلخ).

9 - اتخاذ القرار: بعد الحصول على البيانات وتحليلها إحصائياً يتم اتخاذ القرار والتوصيات المناسبة الكفيلة بالحد من هذه الظاهرة.

شكل توضيحي لأهم مراحل البحث الإحصائي



5-1- مجالات استخدام الإحصاء:

جميع العلوم تستخدم الإحصاء لجمع البيانات وتفسيرها للوصول إلى النتائج التي تفسر الغموض الذي يدور حول الظاهرة:

- في التخطيط: يعتبر الإحصاء أساساً يعتمد عليه، في رسم أية خطة مستقبلية تحدد بها السياسة العليا لدولة لتحقيق أهدافاً معينة في مختلف المجالات والأنشطة، ومن المعروف أنه لا يمكن أن توضع خطة دون توفر البيانات الإحصائية عن الاقتصاد المدروس.
- في الاقتصاد: يمثل الإحصاء مكانة هامة في الاقتصاد وبخاصة في تطويره ونموه لأن توافر البيانات الإحصائية عن اقتصاد دولة ما يشكل الأساس في وضع السياسات الاقتصادية الملائمة للنظريات الاقتصادية.
- في العلوم الاجتماعية: يعتبر الإحصاء وسيلة مهمة لقياس مدى رفاهية الناس وتقدم المجتمع ورفقي مستوى الأفراد ثقافياً أو اجتماعياً من حيث إمكانية جميع البيانات المتعلقة بالحالة التعليمية.
- في الإعلام: يمكن استخدام الإحصاء في دراسة استطلاعات الرأي العام للحصول على معلومات وبيانات ضرورية للمخططين والقادة في مجالات السياسة والاقتصاد والاجتماع، والقرارات ناجحة بقدر ما تكون دراسة الرأي العام ناجحة في انتقائها عينات الدراسة، مثال: التعرف على عدد المؤيدين لأحد المرشحين في الانتخابات في منطقة معينة أو دراسة العلاقة بين المستوى التعليمي والمشاركة السياسية وغيرها.
- في إدارة الأعمال: يؤدي علم الإحصاء دوراً مميزاً في مختلف الدراسات الإدارية لما يقدمه من طرائق وأساليب إحصائية تعد منهجاً علمياً وأداة للبحث في مجال إدارة الأعمال، فالإحصاء يساهم في :

- دراسة اتجاه المبيعات وبيان الآثار الموسمية والدورية لها بهدف إعداد الخطط المستقبلية للمؤسسات والمنشآت المدروسة.
- تتبع المخزون من المواد الخام ومستلزمات الإنتاج وغيرها.
- دراسة احتياجات المستهلكين ورغباتهم وأنواقهم.
- توقع المبيعات في ضوء أساليب الإنتاج والخطط الإنتاجية.
- تحديد درجة الجودة والرقابة عليها.
- في الدراسات السكانية: يطبق علم الإحصاء في مجال الدراسات السكانية بغية التعرف على أهم المؤشرات الديمغرافية المتعلقة بحالة السكان (ولادات، وفيات، هجرة، زواج، طلاق وغيرها).
- في المحاسبة: يعتبر الإحصاء وسيلة مهمة للمحاسب في اتخاذ القرارات المبنية على البيانات المتعلقة بالحالة الاقتصادية لشركة ما لإجراء التنبؤات والتقديرات والاستنتاجات عن مجموعة من المتغيرات.
- في العلوم المصرفية والمالية: يمكن استخدام الإحصاء في دراسة النقد المتداول، والودائع، القروض، أسعار الفائدة، أسعار الصرف، الأوراق المالية، الأسهم، باعتبار العلوم المالية المصرفية تتميز بالتغير الدائم والتنبؤ بها صعب ويتطلب أساليب إحصائية متقدمة.
- في السياسة: تتبع أهمية الإحصاء من خلال استخدام أدواته في العديد من الموضوعات التي يمكن أن تخضع للقياس للتعبير عنها كمياً مثل: قياس استطلاع الرأي العام.
- في التربية وعلم النفس: يستخدم الإحصاء بشكل كبير في دراسة:
 - القدرات العقلية للطلاب (مستوى الذكاء).
 - دراسة النتائج التربوية لنظم وطرق التعليم المختلفة.

في الصناعة والعلوم الهندسية: يطبق الإحصاء في مجال الصناعة للتعرف على:

— تحديد مستويات التخزين.

— مراقبة وضبط جودة الإنتاج.

في الطب: يمكن استخدام الإحصاء في تحديد دقة القياسات ومقارنة النتائج وتقويم الاختبارات التشخيصية وفي حساب المعدلات في مجال الخدمات الطبية مثل: معدل أشغال الأسرة في المشافي ومعدل انتشار الأمراض المعدية وغيرها.

6-1- مزايا علم الإحصاء:

يتميز يمكن القول أن الإحصاء بالمزايا التالية:

— يدرس الناحية الكمية للظواهر.

— يدرس تطور الظواهر عبر الزمن.

— يدرس التأثير المتبادل بين الظواهر الاقتصادية والاجتماعية.



الفصل الثاني

جمع وتبويب وعرض البيانات الإحصائية

- جمع البيانات الإحصائية
- مصادر البيانات الإحصائية (المصادر المباشرة، المصادر التاريخية)
- أساليب جمع البيانات الإحصائية (أسلوب الحصر الشامل، أسلوب العينات)
- طرق جمع البيانات الإحصائية (الطريقة المباشرة، الطريقة غير المباشرة)
- تفريغ وتبويب البيانات الإحصائية (التبويب اليدوي، التبويب الآلي).
- إعداد وتجهيز البيانات باستخدام (SPSS)
- عرض البيانات الإحصائية (العرض الجدولي، العرض البياني)
- عرض البيانات باستخدام (SPSS, Excel)
- التوزيعات التكرارية
- مفهوم الالتواء ومفهوم التفرطح
- أسئلة وتمارين غير محلولة



جمع وتبويب وعرض البيانات الإحصائية

2-1- جمع البيانات الإحصائية :

تعتبر عملية جمع البيانات الإحصائية أهم مرحلة من مراحل البحث الإحصائي، لأنها بمثابة الأساس الذي تبنى عليه بقية المراحل، باعتبار أي خلل في عملية جمع البيانات سيؤدي إلى نتائج مضللة عن الظاهرة المدروسة، وتستخدم البيانات التي نجم عنها من المفردات لنوعين من الأهداف:

- 1 - لوصف المجموعة أي وصف شكل التوزيع الذي تخضع له ولحساب بعض مقاييس النزعة المركزية والتشتت (التوصيف الكمي).
- 2 - لوصف المفردة، مثل وصف مكانة المفردة النسبية (مثل ترتيبها بالنسبة للمجتمع وتعيين موقعها النسبي بالمقارنة مع جميع المفردات في مجال البحث).

2-2- مصادر البيانات الإحصائية :

يمكن أن نميز بين نوعين أساسيين من مصادر البيانات الإحصائية:

I - المصادر الأصلية أو المباشرة :

يتم جمع البيانات بهذا الأسلوب من الوحدات المدروسة مباشرة، مثال ذلك جمع البيانات الخاصة بالحالة التعليمية من الناجحين عن طريق الاستبيان. ويمتاز هذا الأسلوب بإمكانية المراجعة والتحقق من الدقة مع مصدر البيان نفسه. ولكن يعاب عليه كثرة الجهد والتكلفة.

II - المصادر التاريخية أو غير المباشرة (الوثائقية):

يتم الحصول على البيانات بهذا الأسلوب من السجلات التي تقوم هيئات رسمية أو شبه رسمية بإصدارها بصفة دورية، حيث يمكن الرجوع إليها للحصول على تلك البيانات أو المعلومات الإحصائية دون الرجوع إلى الوحدات الأصلية ومن أمثلتها: (سجلات السكان، سجلات المواليد والوفيات، السجلات المالية، سجلات

الطلبة....) ويمتاز هذا الأسلوب بالسهولة والتوفير في الجهد والتكاليف والوقت، بينما يعاب عليه عدم إمكانية التحقق من صحة البيانات .

2-3- أساليب جمع البيانات الإحصائية :

إذا ما تقرر جمع البيانات الإحصائية من مصادرها الأصلية أو المباشرة، فإن عملية جمع البيانات يمكن أن تتم باتباع أحد أسلوبين أساسيين، هما :

I – أسلوب الحصر الشامل : Complete Census

يتم الحصول بهذا الأسلوب على البيانات المطلوبة من جميع أفراد المجتمع موضع البحث دون استثناء، مع مراعاة توفر جميع الإمكانيات اللازمة لجمع البيانات ويمتاز هذا الأسلوب بتوفير المعلومات عن جميع وحدات المجتمع، ويستخدم هذا الأسلوب في الحالات الآتية:

أ – إذا كان المجتمع قيد الدراسة صغيراً.

ب – إذا كان المطلوب الحصول على بيانات على مستوى عالي من الدقة، كما هو الحال في التعدادات العامة للسكان أو الزراعة أو الصناعة أو غيرها.

ج – إذا تعذر الحصول على إطار لمفردات المجتمع، فالإطار هو عبارة عن قوائم أو خرائط دالة لعناصر المجتمع قيد الدراسة.

II – أسلوب العينات : Sampling Method

لقد دلت التجارب أن معظم الأبحاث والدراسات تجري بأسلوب العينات لاعتبارات مادية وفنية، حيث يقتصر هذا الأسلوب على بحث عدد محدود من وحدات المجتمع والخروج بنتائج يمكن تعميمها على المجتمع الذي سحبت منه وفق قواعد علمية معينة، شرط أن تمثل العينة المجتمع الأصلي الذي سحبت منه

أصدق تمثيل. ويتميز أسلوب العينة بتوفير الجهد والوقت والتكاليف، والحد من الخطأ الناتج عن عدم الدقة في القياس وذلك لمحدودية مفردات المجتمع المختارة، كما تعطي معلومات أكثر تفصيلاً وأكثر دقة.

أنواع العينات :

أولاً : العينات الاحتمالية تعتمد على نظرية الاحتمالات في أسلوب السحب وتعميم النتائج وتقدير الأخطاء (في العينات عشوائية، يتم اختيار مفرداتها وفق أسس نظرية الاحتمالات) وتشمل:

I - العينة العشوائية البسيطة Simple Random Sample :

من أبسط أنواع العينات، ويتم اختيار المفردات بشكل عشوائي من المجتمع المدروس ويمكن الحصول عليها بإحدى الطرق الآتية:

الطريقة الأولى : طريقة القرعة، وتتلخص في عمل بطاقة لكل وحدة من وحدات المجتمع ثم نختار بطريقة عشوائية عدداً من تلك الوحدات .

الطريقة الثانية : طريقة جداول الأرقام العشوائية، حيث نعطي مفردات المجتمع المدونة في الإطار أرقاماً متسلسلة ثم نختار أحد الأعمدة أو الصفوف بالجدول العشوائي لاختيار وحدات العينة (انظر جداول الأرقام العشوائية في نهاية الكتاب)، بشرط وجود الرقم المختار في الإطار للمجتمع.

الطريقة الثالثة: دواليب الحظ

الطريقة الرابعة: بواسطة الحاسوب حيث يتم إعطائه حجم المجتمع وحجم العينة فهو يقوم بتحديد المفردات التي تنتمي إلى العينة.

وتتلخص عيوب العينة العشوائية البسيطة بعدم إمكانية التمثيل لكل مجموعات المجتمع خاصة عندما يكون غير متجانس، لكن تعتبر مقياساً للمقارنة مع العينات الأخرى.

II - العينة الطبقية Staratified sample :

تستخدم في حالة المجتمعات غير المتجانسة حيث يمكن تمييز وحدات المجتمع على شكل مجموعات متجانسة فيما بينهما وتسمى كل مجموعة حينئذٍ طبقة، ويتم اختيار وحدات كل طبقة بطريقة عشوائية وتكون العينة عبارة عن إجمالي الوحدات التي تم اختيارها من كل طبقة على حدة، وغالباً ما يتناسب عدد الوحدات التي يتم اختيارها من كل طبقة مع حجم الطبقة، ويفضل تقسيم المجتمع إلى طبقات على أساس عوامل متصلة بموضوع الدراسة مثال ذلك: تقسيم مجتمع العمال حسب الدخل إلى (مرتفع، متوسط، منخفض)، وتقسيم المخازن عند دراسة المبيعات (صغيرة، متوسطة، كبيرة)

عادة يتم تقسيم المجتمع إلى طبقات متجانسة وفق إحدى الطرائق الآتية:

1 - طريقة التخصيص المتساوي: توزع مفردات العينة بالتساوي على طبقات المجتمع.

2 - طريقة التخصيص النسبي (المتناسب مع الحجم): يؤخذ من كل طبقة عدد من المفردات يتناسب وحجم الطبقة إلى حجم المجتمع، لتكون في مجموعها حجم العينة.

مثال: أردنا أن نختار عينة طبقية حجمها $n=100$ من مجتمع $N=1000$ وينقسمون إلى ذكور (400) وإناث (600) فإننا نختار.

$$n_1 = \frac{N_1}{N} * n = \frac{400}{1000} * 100 = 40 \text{ ذكور}$$

$$n_2 = \frac{N_2}{N} * n = \frac{600}{1000} * 100 = 60 \text{ إناث}$$

$$n = 40 + 60 = 100$$

مثال: كان من المرغوب تشكيل عينة من (100) ناخب من (1000) ناخب وكنا نعلم أن المستوى التعليمي لهم على النحو التالي: ابتدائي (50%) إعدادي

(30%)، ثانوي (15%)، جامعيون (5%) في هذه الحالة تشكل العينة الطبقة
بنسبة تساوي نسبتها في مجتمع الدراسة أي (50) ابتدائي، (30) إعدادي، (15)
ثانوي (5) جامعي.

ويتصف هذا النوع من العينات بمايلي :

- أ - ارتفاع مستوى تمثيل العينة للمجتمع المدروس.
- ب - يمكن الحصول على نتائج جيدة الدقة من عينة حجمها صغير نسبياً.
- ج - أكثر كفاءة من غيرها في الأحوال التي تكون فيها مجموعات معينة من المجتمع ذات مواصفات خاصة هي السائدة.
- د - تتطلب معرفة جيدة بالمجتمع المدروس لتحديد الطبقات بشكل مناسب.
- هـ - قد تتصف بشيء من التعقيد عند التعامل مع عدد طبقات كبيراً.

III - العينة المتعددة المراحل Multistage Sample :

تستخدم في حالة المجتمعات المنتشرة على مساحات كبيرة، وفيها يقسم المجتمع الأصلي إلى عدد الوحدات الابتدائية، حيث نختار عدداً منها بطريقة عشوائية كمرحلة أولى، وهكذا تتعاقب مراحل المعاينة ويستخدم ذلك النوع من العينات لغرض توفير في تكاليف الانتقال الباحثين خاصة في حالة المجتمعات الكبيرة، مثال عند دراسة المحاصيل الزراعية في سورية، فيمكن أن نقسم القطر إلى مناطق زراعية متعددة كل منها متجانس فيما بينها، ثم نختار عدداً من هذه المناطق المتعددة كمرحلة أولى، وبعدها نختار عدد من المزارع بطريقة العينة العشوائية من كل منطقة اخترناها - كمرحلة ثانية - ثم يمكن اختيار بعض المساحات المزروعة بطريقة العينة العشوائية كمرحلة ثالثة.

V - العينة المنتظمة Systematic Sample :

تعتمد على قائمة تتضمن مفردات المجتمع وعلى تحديد فترة السحب
وهي أسهل في تطبيقها واستخدامها من العينة العشوائية البسيطة، بالرغم
$$K = \frac{N}{n}$$

من أنها تعطي نتائج مشابهة لها من حيث درجة تمثيل المجتمع المدروس وإمكانية تعميم نتائجها بثقة، مثال: اختيار عدد من رواد السينما أو المسرح بطريقة العينة المنتظمة لمعرفة دور السينما والمسرح في تنمية ثقافة المجتمع.

فإذا كان لدينا مجتمع مكون من 90 وحدة، ومطلوب اختيار عينة من 10 وحدات، فإن فترة السحب $= 9 = 90/10$ ، وقد تم الاختيار لرقم عشوائي بين 9,1 وكان الرقم 5، فعندها تكون وحدات العينة هي الوحدات ذات الرقم المتسلسلة بالإطار 5، 14، 23، 32، 41، 50، 59، 68، 77، 86. وتمتاز طريقة العينة المنتظمة بالسهولة والبساطة في إجرائها، ولكن يعاب عليها عدم صلاحيتها، إذا ما وجدت علاقة دورية في ترتيب الوحدات الأولية المكونة للمجتمع مع طول فترة السحب.

IV - العينة العنقودية: تقوم على تقسيم المجتمع إلى مجموعات جزئية والاختيار يكون على شكل عناقيد وليس مفردة مثال: في إحدى المدن (10,000) مسكن، وطلب منا اختيار عينة من (500) مسكن ففي هذه الحالة يتم تقسيم المدينة إلى (1000) منطقة وبكل منطقة (10) مساكن ومن ثم نختار عشوائياً (50) منطقة من المناطق الـ (1000) وتجمع المعلومات اللازمة عن كل مسكن من المساكن بالمناطق المختارة.

ثانياً - العينات غير الاحتمالية تعتمد على الاختيار غير العشوائي للمفردات وتشمل:

I - عينة الحصص حيث تقسم العينة المطلوبة إلى حصص تشمل كل حصة مجموعة من الوحدات يتم جمع البيانات منها على أن تترك عملية اختيار تلك الوحدات للباحث نفسه واعتماد نسبة معينة من المفردات في كل طبقة وتستخدم بشكل واسع في أبحاث التسويق واستطلاعات الرأي العام.

مثال: إذا حددنا العينة بـ 400 طالب من كلية الاقتصاد بجامعة دمشق موزعين

على التخصصات الأربعة اقتصاد وإدارة ومحاسبة ومصارف، وفرضنا أن نسبة الطلبة من كل تخصص هي : 3:5:1:1 على التوالي، سيكون لدينا 120 طالباً من تخصص اقتصاد 200 من تخصص إدارة، 40 من تخصص المحاسبة، 40 من تخصص المصارف.

ونترك عملية اختيار هؤلاء الطلبة للقائمين بعملية جمع البيانات أنفسهم وبالطبع ستتم عملية المقابلة وجمع البيانات مع الطلبة الذين يمكن مقابلتهم أو من السهل جمع المعلومات منهم. وتمتاز هذه العينة بالبساطة وقلة التكاليف والسرعة في الحصول على النتائج. مثال: إذا أراد باحث أن يدرس تاريخ بلد ما فإنه يختار عينة لجمع المعلومات من الأشخاص الذين عايشوا فترات زمنية مختلفة من تاريخ ذلك البلد أو أساتذة جامعات أو سياسيين.

II – العينة العمدية (المنتقاة) : في هذا النوع من العينات يحدد الباحث هو الذي يحدد المفردات التي ستدخل في العينة وتتوقف جودة النتائج على حكمة الباحث ومهارته في الاختيار.

– العينة الكبيرة الحجم : تعتمد على اختيار عدد كبير من المفردات بشكل غير عشوائي مثال: استطلاعات الرأي العام حول الانتخابات الرئاسية.

III – العينة الملائمة (Convenience Sampling) : يتم اختيار وحدات المعاينة على أساس مدى مناسبتها لأهداف الدراسة ورؤية الباحث المتمثلة في قدرة الوحدات المختارة على الإفادة بمعلومات صادقة ومفيدة للدراسة بشكل مباشر، كأن يقوم الباحث باختيار عينة من مرتادي أسواق تجارية معينة لدراسة رضا المرتادين على الخدمات المقدمة من المحلات التجارية في ذلك السوق معاينة الجزء السهل (Readily Accessible sampling) : يتم اختيار وحدات المعاينة على أساس سهولة الوصول لهم.

وهناك العديد من أساليب المعاينة غير الاحتمالية المختلفة والتي تختلف مسمياتها باختلاف الأسلوب الذي يتم به اختيار وحداتها كالعينة المساحية، والمزدوجة،

والمكررة والتطوعية، والصدقية. ففي هذه الأساليب يمكن للباحث التوصل إلى استنتاجات معينة، ولكن يصعب عليه تعميمها وغالباً ما يتم تطبيقها في المراحل الاستكشافية أو الاستطلاعية، وفيها أيضاً لا يمكن حساب خطأ المعاينة ولا يمكن حساب دقتها ولا يعتمد عليها عادة في تقديرات أخرى مهمة.

كما تم تطوير أساليب معاينة مختلفة لدراسة الخواص النادرة في مجتمع الدراسة بهدف إيجاد التقديرات الإحصائية كتقدير نسبة غير الملتزمين بالدوام الرسمي لدى الدوائر الحكومية كالمعاينة ذات المرحلتين والمعاينة المتعددة الإطار والمعاينة المكررة وحديثاً المعاينة الشبكية.

كما تم تطوير بعض أساليب المعاينة لتقدير بعض خصائص المجتمع كالمعاينة المضاعفة التي يتم فيها اختيار عينة أولية من أجل تحديد بعض المعلومات الإضافية فقط، ومن ثم يتم اختيار عينة جزئية من الأولية ويتم ملاحظة المتغير المراد قياسه في الدراسة. والمعاينة بالقبض وإعادة القبض (capture-recapture sampling): يتم تقدير العدد الكلي للوحدات في المجتمع، بحيث يتم سحب عينة أولية من المجتمع ووضع علامة على الوحدات المختارة ومن ثم يتم اختيار عينة أخرى من المجتمع مستقلة عن العينة الأولية وتحديد عدد الوحدات المألوفة. كما تم طرح طرق عديدة لإجراء التقديرات الإحصائية باستخدام مفهوم المعاينة المكيفة، (Adaptive samling)، بالإضافة إلى بعض طرق المعاينة الإحصائية المشتقة من مفهوم المعاينة المكيفة كالمعاينة العنقودية المكيفة والمعاينة العنقودية المكيفة الطبقية.

• أخطاء العينة:

يتعرض الباحث عند القيام بدراسة ما واختيار أحد أنواع العينات الاحتمالية للحصول على بيانات يمكن تعميم نتائجها إلى نوعين من خطأ المعاينة وهي:

- 1 - الأخطاء الشخصية (التحيز) : وتنتج عن التقصير في خطوات البحث بالعينة، وأسبابها عدم مراعاة الأساليب العلمية في سحب العينات، وتعالج باتباع الأساليب العلمية في السحب.

2 - الأخطاء الاحتمالية (الخط والصدف): تصاحب عملية الاختيار ونتيجة عن دراسة جزء من المجتمع، وتعالج بأخذ عينة أكبر.

• حجم العينة: هناك اعتبارات أو عوامل عديدة تتحكم في اختيار حجم العينة، أهمها:

أ - التجانس: كلما ازداد تجانس عناصر المجتمع وقلت الفروقات بين عناصره، كلما أمكن تصغير حجم العينة.

ب - إجراءات تحديد واختيار العينة: حيث تؤثر نوعية العينة المختارة وطريقة اختيارها على حجم العينة.

ج - الوقت والموارد المادية والبشرية المتاحة للدراسة: كلما ازداد حجم العينة المختارة بشكل صحيح، كلما ازدادت دقة وموثوقية تعميم نتائجها على المجتمع المأخوذة منه.

وأخيراً فإن فرصة اختيار نوع العينة وطريقة سحبها غالباً ما تتحدد بمدى الحاجة إلى تقسيم المجتمع إلى مجموعات أو طبقات ومدى توافر أطر سحب العينات (سوف نتناول لاحقاً بشكل مفصل أكثر تحديد حجم العينة حسابياً).

• أنواع البحوث الإحصائية :

- البحوث الوصفية: وهي التي تجمع المعلومات عن ظاهرة معينة لا لخدمة هدف بذاته محدد سلفاً، وإنما بقصد توفير البيانات التي من الممكن أن تخدم أغراضاً متعددة لباحثين فيما بعد، مثل تعدادات السكان والزراعة والصناعة والتجارة والصحة ... إلخ.

- البحوث الإحصائية التحليلية : وهي التي تجمع فيها المعلومات التي تخدم هدف معين أو تساعد في تفسير مشكلة معينة لاحفلها الباحث أو لاختبار صحة فرض معين.

- البحوث الإحصائية التجريبية : ويستخدم هذا النوع من البحوث في ميسادين مختلفة كالطب والزراعة وغيرها.

4-2- طرائق جمع البيانات الإحصائية :

يمكن جمع البيانات الإحصائية بإحدى الطريقتين الرئيسيتين :

I — الطريقة المباشرة :

ويتم عن طريق الاتصال المباشر بمصدر البيان أو وحدة البحث وذلك بإحدى الطرائق (الاتصال الشخصي، التلّфон، الفاكس، الانترنت) وتمتاز هذه الطريقة بالدقة، وعيوبها تتلخص في كثرة الكلفة والجهد والوقت. وقد يتم الاتصال المباشر بالهاتف أو بالانترنت كما تشمل أيضاً أسلوب الملاحظة وتسجيل المعلومات

II — الطريقة غير المباشرة :

حيث يتم جمع البيانات أو المعلومات من المصدر عن طريق الاستبيان، أو من المصادر التاريخية، ويعاب على هذه الطريقة انخفاض نسبة الردود وعدم إمكانية مراجعتها مع مصدر البيان للتحقق من دقة البيانات، ولنجاح العملية يتم اختيار الكفاءة والخبرة في جامع البيانات.

• الاستمارة الإحصائية (الاستبيان): عبارة عن كشف يتضمن الأسئلة المطروحة من قبل الباحث بهدف جمع البيانات، وهي الوسيلة التي تربط طالب البيانات بمصادر البيانات.

• أنواع الاستمارة الإحصائية (الاستبيان):

- مغلق: هناك خيارات للإجابة (نعم، لا)، (موافق، غير موافق، محايد).
- مفتوح: يعطي الحرية للإجابة على الأسئلة بدلاً من الخيارات وحصر الإجابة.

— مختلط: يتضمن المغلق والمفتوح معاً.

• مكونات الاستمارة الإحصائية (الاستبيان) :

تتكون الاستمارة من أربعة أجزاء رئيسية وهي:

الجزء الأول : ويتضمن صدر الاستمارة وفيه يذكر اسم الجهة المشرفة على البحث كما يتضمن ملخصاً للأهداف المرجوة من البحث.

الجزء الثاني : ويشمل البيانات المميزة لوحدة البحث (الاسم، الجنس، العمر، مكان الإقامة)

الجزء الثالث : ويتضمن مجموعة الأسئلة والاستفسارات المطلوب الإجابة عليها.

الجزء الرابع : ويتعلق بملاحظات جامع البيان إن وجدت، كما قد يتضمن بعض التعليمات والإشارات الخاصة بكيفية استيفاء الاستمارة.

• مزايا الاستبيان:

— كلفة أقل

— إعطاء حرية للفرد في الإجابة عن الأسئلة

— إمكانية إرساله لجميع فئات المجتمع

• سلبيات الاستبيان:

— انخفاض نسبة الردود.

— عدم إمكانية إرساله إلى الأميين.

— عدم ملاحظات ردود انفعالات الأفراد.

• أهم الشروط الواجب مراعاتها عند وضع أسئلة الاستمارة الإحصائية (الاستبيان) :

1 — يجب أن تكون الأسئلة محددة وواضحة ومتوافقة مع الأهداف من البحث.

2 — يجب أن تكون الأسئلة بعيدة عن التأويل والتفسير وهذا يتطلب وضع التعاريف المناسبة التي تبين المقصود من كل سؤال.

3 — يجب أن تكون الأسئلة حديثة لا قديمة.

- 4 - التسلسل المنطقي للأسئلة بحيث تكون مفتاحاً للأسئلة التي تليها ما أمكن.
- 5 - يستحسن تجنب الأسئلة التي تثير حرج المستجوب.
- 6 - يفضل بالنسبة للأسئلة التي تستهدف قياس ظواهر غير رقمية أن لا يترك السؤال مفتوحاً بل وضع احتمالات للإجابة.

• **التدريب واختبار الاستمارة الإحصائية (الاستبيان):** يجب إجراء تدريب للباحثين الذين سيقومون بجمع البيانات، ويشمل هذا التدريب شرحاً وافياً لأهداف البحث ونوع المجتمع الذي سيشمله وشرحاً للتعريف المستخدمة في الأسئلة الواردة في الاستمارة وطريقة الاتصال بالأشخاص المبحوثين وكيفية توجيه الأسئلة إليهم لضمان الحصول على بيانات دقيقة، كما يفضل إجراء بحث تجريبي على عينة من المفردات لتدريب الباحثين على جمع البيانات واختبار الاستمارة الإحصائية والتعرف على ما قد تكون فيها من ثغرات أو مواطن قصور وتعديل الاستمارة على ضوء هذه الخبرة الميدانية (سيتم شرحه بشكل مفصل لاحقاً) أخيراً يجب على الباحث أن يكون متحلياً ببعض الصفات ليستطيع أن يتصرف بحكمة ومن هذه الصفات:

- أن يكسب ثقة من يقابله.
- أن يخلق جواً من الود يشجع على الكلام.
- أن لا يغير موضوع الحديث وأن يتجنب دور المعلم.
- أن يطرح الأسئلة بسهولة.
- أن يتحقق من صحة الأجوبة.
- أن يتصف ببشاشة الوجه وسعة الصدر والصبر وأن يراعي الذوق والأدب في الحديث.
- **التوعية :** تسهياً لمهمة الباحث وتنويراً للرأي العام وكسباً لثقته يتعين على

الجهة القائمة بالبحث أن تقوم بحملة توعية القصد منها إطلاع الرأي العام على الهدف من البحث وقوائمه وشرح طريقة القيام بالعملية والتأكيد على سرية البيانات.

• اختبار " استمارة الاستبيان " :

بعد الانتهاء من إعداد وتصميم أداة الدراسة المتمثلة في استمارة الاستبيان في صورتها المبدئية، وقبل الشروع في توزيعها على مفردات عينة الدراسة، لابد من اختبار صدقها وثباتها أي: مدى صلاحيتها في قياس ما صممت من أجله، ويتم اختبار أداة الدراسة على النحو التالي :

أ - صدق أداة الدراسة " استمارة الاستبيان " :

لغرض التحقق من صدق أداة الدراسة من حيث صحة محتوياتها بالنسبة للأسئلة والعبارات التي جرى تصميمها أو تطويرها، ومدى تمثيلها لمتغيرات الدراسة، يتم تحكيم أداة الدراسة (استمارة الاستبيان) من خلال عرضها على أساتذة مختصين ومحكمين من الأساتذة المتخصصين في نوع الدراسة، بالإضافة إلى عرضها على عدد أساتذة متخصصين بالإحصاء للاطلاع على محتويات أداة الدراسة من الأسئلة والعبارات الواردة بها، للتحقق من مدى قدرتها على قياس متغيرات الدراسة وفرضياتها، لإبداء الرأي فيما يتعلق بالفقرات ومدى موضوعيتها وإجراء التصحيحات اللغوية إن وجدت، وكذلك إعادة صياغة وبناء لبعض الفقرات من الأسئلة والعبارات إذا دعت الحاجة لذلك، وفي بعض الأحيان إضافة أو استبعاد بعضها الآخر والتي تؤكد عدم أهميتها في عملية التحليل والقياس وكذلك إعادة توزيع محتويات أداة الدراسة بالشكل الذي يخدم أغراضها التي وضعت من أجلها، وفضلاً عن ذلك يتم في أغلب الأحيان عرض أداة الدراسة على عينة اختبارية لحجم محدد من المفردات من مجتمع الدراسة الأصلي، بغرض التعرف على مدى وضوح محتوياتها من

الأسئلة والعبارات وفهمهم لها وكذلك تحديد الوقت اللازم للإجابة عليها.
إن معظم هذه الإجراءات التي يتوجب على الباحث القيام بها هي التي تعطي
مؤشراً كافياً على صدق أداة الدراسة.

ب - ثبات أداة الدراسة " استمارة الاستبيان ":

من الضروري التحقق من مدى ثبات أداة الاستبيان (المأمونية) الذي
يتمتع بها، إذ أن صفة الثبات شرط ضروري لا بد من توفره للحكم على
صلاحية أداة الدراسة أو القياس، ويقصد بثبات الأداة الاتساق الداخلي بين
فقراتها من الأسئلة والعبارات، بحيث تعطي نفس النتائج أو على الأقل نتائج
متقاربة، إذا ما استخدمت لقياس الظاهرة نفسها مرات متتالية بالظروف نفسها،
وهذا ما يعرف بالاستقرار الذي ينبغي أن يتوفر فسي أداة الدراسة أو القياس
ولغرض التأكد من ثبات أداة الدراسة يتم استخدام ما يعرف بطريقة الاختبار
 وإعادة الاختبار من خلال تقديمها لعينة من مجتمع الدراسة الأصلي قوامها عدد
معين أو محدد من المفردات مرتين متتاليتين تفصل بينهما فترة زمنية تقدر
بأسبوعين من خلال اختبار معامل ارتباط سبيرمان لإجابات العينة في المرتين
على مجموعة من الأسئلة الواردة باستمارة الاستبيان، وإيجاد قيمة معامل
الارتباط والمقارنة مع مستوى المعنوية المشاهد المقابلة لقيمة معامل ارتباط
سبيرمان (P-value) لتحديد وجود علاقة ارتباط أم لا بين استجابات مفردات
العينة في المرتين لمعرفة ثبات أداة الدراسة واستقرارها.

كما يعزز النتيجة أو الاختيار السابق حساب معامل الثبات باستخدام أسلوب ألفا
كرونباخ (cronbach Alpha) للأسئلة التي تم اختيارها في السابق ويفضل أن
تتجاوز قيمة ألفا كرونباخ (0.60) لكي يعد دليلاً كافياً على وجود اتساق داخلي
بين فقرات تلك الأسئلة، وبالتالي الحكم من خلالها على ثبات أداة الدراسة إلى
الحد الذي يجعل من الممكن الاعتماد عليها.

ويعطى معامل ألفا كرونيباخ بالعلاقة التالية :

$$\alpha = \frac{K\bar{r}}{1 + (K-1)\bar{r}}$$

حيث أن :

K : عدد بنود الاستبيان

$\bar{r}$: يساوي متوسط الارتباط بين هذه البنود وذلك في حال كانت هذه البنود متساوية التباين.

• تحديد مقاييس متغيرات الدراسة:

بعد تحديد المتغيرات التي ستدرس لابد من تحديد المقاييس التي قد تستخدم لقياس هذه المتغيرات، فالمقياس هو عبارة عن أداة أو وسيلة لتصنيف الأشخاص أو الأشياء بناءً على متغير أو متغيرات، بمعنى آخر غالباً ما يقاس مدى الاقتراب من تحقيق أهداف الدراسة من خلال ما يتم التوصل إليه من نتائج علمية، تأتي عادة كنتائج لعملية تحليل البيانات التي تم الحصول عليها أو تجميعها من مفردات عينة الدراسة، التي اختيرت من مجتمع الدراسة الأصلي (المجتمع الإحصائي) وذلك بالاعتماد على استخدام أساليب التحليل الإحصائي الممكنة والملائمة لطبيعة تلك البيانات.

ومن المهم الإشارة إلى أن البيانات التي تعتمد عليها الدراسات تتدرج في معظمها ضمن التصنيفات الرئيسية التالية :

1 - المقاييس الأسمية: وهي تصنيف الأفراد إلى مجموعات ويكون ذلك تبعاً لبعض الخصائص النوعية للجنس والمناطق السكنية والألوان وغيرها من الخصائص وتعطى كل خاصية رقماً خاصاً بها يدل عليها وتفقد هذه الأرقام خصائصها الرياضية المعروفة من عمليات جمع وطرح وضرب وقسمة، وهذا النوع من المقاييس لا يقوم بأكثر من التصنيف إلى مجموعات من أجل التمييز بينها.

2 - المقاييس الرتيبة : وهو يصنف الأفراد أو الأشياء إلى مجموعات مع توضيح درجة الأفضلية حيث يرتبها تصاعدياً أو تنازلياً تبعاً لصفة أو خاصية معينة أي يتم ترتيب المتغيرات من الأفضل إلى الأسوأ أو من الأول إلى الأخير حيث تعطى أرقاماً مثل (1) للأفضل ورقم (2) للأقل تفضيلاً وهكذا حتى يمثل الرقم الأخير ما هو غير مفضل، وهذه الأرقام لا تمثل كميات معينة ومحددة، كما أن المسافات الفاصلة بينها لا يشترط أن تكون متساوية فعند ترتيب خمسة موظفين حسب درجة تعاونهم مع مدرائهم وإعطائهم الرقم (1) لأكثرهم تعاوناً والرقم (5) لأقلهم تعاوناً فهذا لا يعني أن الموظف الذي ترتيبه (2) تعاونه ضعف الموظف الذي ترتيبه (4) أو الفرق في درجة التعاون بين الأول والثاني تساوي الفرق بين درجة تعاون الرابع والخامس فالمقياس الرتبي لا يعطي صورة واضحة عن حجم الفروق الموجودة بين الأفراد المتجاورين فهي أية مجموعة.

3 - المقاييس الفاصلة أو مقاييس المدى: يعتبر هذا النوع من المقاييس أعلى مستوى من المقاييس السابقين كونه يمثل خاصية الفواصل أو المسافات المتساوية التي تفصل بين درجة وأخرى مجاورة لها ووجود هذه المسافة المتساوية بين كل درجتين متجاورتين يعني إمكانية إجراء بعض التعديلات الحسابية كالجمع والطرح ويعتبر الصفر في هذا المقياس نسبياً وليس مطلقاً.

4 - المقاييس النسبية : وهي التي تتميز بخصائص جميع المقاييس السابقة إضافة لوجود الصفر المطلق، وبالتالي يمكن مقارنة الصفات أو الخصائص حسب هذه المقاييس كمياً نظراً لوجود أرقام فعلية لها، ومثال ذلك مقارنة الأفراد فيما بينهم من حيث الدخل الشهري والطول وغيرها.

أخيراً : يمكن القول أن المقياس الأسمي يوضح الفرق فقط بين المجموعات بعد أن تصنف الأشياء أو الأفراد إلى هذه المجموعات وتجري عليها دراسات

إحصائية باستخدام المنوال واختبار كاي مربع. (سنتناول جميع هذه المقاييس والاختبارات لاحقاً).

أما في المقياس الرتبي يوجد توضيح لترتيب المفردات تبعاً لدرجة أفضلية معينة وتجري الدراسات الإحصائية عليه باستخدام معامل الارتباط للرتب والوسيط والانحراف الربيعي.

أما في المقاييس الفاصلة يوجد تركيز على المسافة الفاصلة بين المفردات المرتبة ويمكن أن تستخدم في الدراسات الإحصائية المتعلقة بهذه المقاييس الوسط الحسابي والانحراف المعياري ومعامل الاختلاف والاختبارات الإحصائية المعتمدة على توزيعي فيشر (F) وستودنت (t).

أما في المقاييس النسبية فهي تحقق كافة خصائص المقاييس السابقة كافة ونستطيع استخدامها في الدراسات الإحصائية المتقدمة المستندة على المقاييس والاختبارات الكمية.

ثانياً : تصنيف آخر للمقاييس :

1- مقاييس المعدل: وهي تقيس معدل الميل بالنسبة للأشخاص أو الأشياء أو الأحداث وأهمها:

— المقياس الثنائي : وفيها تكون الإجابة حكماً بنعم أو لا، مثال ذلك:

هل أنت راضٍ عن دراستك في الجامعة؟

— مقياس ليكرت (Likert): وهو من المقاييس الفاصلة وفيه تستخدم العبارات الآتية:

— موافق، محايد، غير موافق (ليكرت الثلاثي).

— غير موافق بشدة، غير موافق، محايد، موافق، موافق بشدة (ليكرت الخماسي).

ومن الممكن أن يكون مقياساً ذو (5) درجات أو ذو ثلاث أو أربع أو ست أو سبع أو ثمان درجات، مثال ذلك:

ما رأيك بتسليم المرأة للمناصب السياسية؟!

2 - المقياس المعدل المحدد الصفات: ويحدد له أرقاماً لعدد من الصفات مرتبة

تصاعدياً أو تنازلياً ويمكن أن يكون ذو أربع أو خمس درجات، مثال ذلك:

بناءً على آخر نتائج امتحان لمجموعة من الطلاب في أحد المقررات يمكن

طرح السؤال التالي: كم كنت راضياً عن نتائجك في الامتحان:

1 - راضياً كلياً 2 - راضياً جداً 3 - راضياً 4 - راضياً إلى حد ما 5

- لست راضياً أبداً .

مثال آخر:

ستحصل على مركز جديد في عملك خلال الأشهر الثلاثة القادمة:

1 - بالتأكيد 2 - احتمال قوي 3 - احتمال قائم 4 - غير وارد 5 -

بالتأكيد لا.

3 - مقاييس الرتبة: وهي مقاييس لبيان درجة التفصيل ومن أهمها مقياس

الاختبار الإجمالي، وهو إعطاء مجموعة من المتغيرات المحددة ويطلب ممن

يجيب على السؤال أن يرتبهم تبعاً للأفضلية تنازلياً مثال ذلك:

رتب الكتب الآتية حسب أهميتها بالنسبة لك مبتدئاً بالرقم (1) لأكثرها تفضيلاً

ومنتهياً بالرقم (6) لأقلها تفضيلاً:

كتب دينية - كتب سياسية - كتب تاريخية - كتب اقتصادية - كتب فكرية -

كتب معلوماتية.

من المهم الإشارة إلى أنه عادة يتم استخدام الأساليب الإحصائية المناسبة دون

غيرها في التحليل وبما يتلائم مع طبيعة البيانات، وغالباً ما يتم إدخال هذه

البيانات وباستخدام برامج الحاسوب المعدة لذلك، إضافة إلى إمكانية القياس

بإجراء عمليات التحليل الإحصائي اللازمة على بيانات الدراسة المعنية المتحصل عليها باستخدام برمجيات الحاسوب الخاصة بالتحليل الإحصائي مثل استخدام الحزمة البرمجية الإحصائية للعلوم الاجتماعية (SPSS) حيث يعد هذا البرنامج من أكثر البرامج الإحصائية شيوعاً واستخداماً في جميع المجالات التربوية والاجتماعية والاقتصادية والطبيعية والهندسية، لإجراء التحليلات الإحصائية والرسوم البيانية اللازمة باستخدام الحاسب الآلي.

خطوات إعداد وتجهيز بيانات الدراسة باستخدام البرنامج (SPSS):

أ - حصر استمارات الاستبيان المتحصل عليها من مفردات عينة الدراسة وتحديد العدد الكلي والنهائي لها والتي ستخضع لعملية التحليل.

ب - تجهيز البيانات وتجهيزها لعملية التحليل، وذلك من خلال إجراء المراجعة والتدقيق اللازمين لمحتويات استمارات الاستبيان المتحصل عليها من مفردات عينة الدراسة، وتوضيح بعض فقراتها وإزالة أي لبس أو غموض فيها بما يسهل إجراء العمليات اللاحقة التي تستلزمها عملية التحليل باستخدام البرنامج الإحصائي (SPSS).

ج - ترميز محتويات استمارات الاستبيان - من الأسئلة والعبارات الواردة فيها - المتحصل عليها من مفردات عينة الدراسة؛ وذلك من خلال إعداد نماذج ترميز لمحتويات استمارة استبيان واحدة، ثم سحبه على جميع استمارات الاستبيان التي خضعت لعملية التحليل، ثم إعطاء حروف وأرقام للمتغيرات الواردة بها، والتي تحل محتوياتها من الأسئلة والعبارات المختلفة، حتى تسهل عملية تفرغها ومن ثم تحليلها باستخدام البرنامج الإحصائي (SPSS).

د - القيام بعملية التفرغ لمحتويات استمارة الاستبيان - التي تم ترميزها وصار من الممكن التعاطي معها كأرقام وليس كصفات - في صفحة

البيانات، أو ما يطلق عليها نافذة محرر البيانات (Data view) بالبرنامج الإحصائي (SPSS) بعد إجراء التعريفات اللازمة لكافة المتغيرات المتضمنة في استمارة الاستبيان من الأسئلة والعبارات؛ ووفق القواعد والتعليمات المعدة لذلك بالبرنامج (SPSS)، من خلال صفحة المتغيرات أو ما يطلق عليها نافذة محرر المتغيرات (Variables view)

5-2-تفريغ وتبويب البيانات الإحصائية :

بعد عملية جمع البيانات سواء من مصادرها الأصلية أو التاريخية، يكون من الصعب فهم مدلولها واتجاهها، لذلك يقوم الباحث بتبويب تلك البيانات وعرضها في جداول إحصائية بحيث يمكن باستعراضها تكوين فكرة عامة عن طبيعة الصفة أو الصفات محل الدراسة وهذه الجداول تعتبر أساساً للتحليل الإحصائي .

• تبويب البيانات :

التبويب هو جمع مفردات البيانات في مجموعات متجانسة، بحيث يعرض أكثر ما يمكن من البيانات بأقل ما يمكن من الاختصار، ونمّيز بين نوعين من التبويب:

1- التبويب اليدوي :

يعتمد على جدول التفريغ الذي يتكون من ثلاثة أعمدة يخصص الأول منها لبيان القيم، والعمود الثاني لوضع الإشارات نتيجة فرز الاستمارات واحدة تلو الأخرى، ويخصص العمود الثالث لوضع عدد تلك الإشارات مقابل الإجابة المعنية، وهذا ما يطلق عليه اسم التكرار ويرمز له بالحرف (f) .

II - التبريد الآلي :

يعتمد على الحاسب الآلي كما أسلفنا سابقاً. وفيه تترجم البيانات المدونة في الاستمارات الإحصائية إلى رموز معينة مقسمة إلى عدة أعمدة يحمل كل عمود منها رقماً مسلسلًا ويقسم كل عمود إلى سطور.

مثال : أخذت عينة عشوائية من (50) مشترك من المشتركين في إحدى الصحف بإحدى المدن بهدف دراسة الحالة التعليمية لهؤلاء المشتركين فكانت النتائج التالية: يقرأ ويكتب، إحصائي، ثانوي ... وهكذا لبقية المشتركين.

المطلوب: تفرغ هذه البيانات في جدول تفرغ.

جدول رقم (1) التوزيع التكراري للمشاركين في إحدى الصحف حسب الحالة التعليمية

التكرار	الإشارات	الحالة التعليمية
10		يقرأ ويكتب
7		إحصائي
3		إحصائي
15		ثانوي
15		جامعي

المصدر: فرضي

العمود الأول والثالث يكونان معاً "جدول التوزيع التكراري" للمتغير وهو يعطينا عدد المرات التي ظهرت فيها كل من قيم المتغير في العينة .

وعندما تكون قيم المتغير كثيرة (كمية أو وصفية) أو يكون المتغير كمياً متصلاً فمن الصعب أن نخصص لكل قيمة من قيم المتغير خانة خاصة بها، ولذلك تجمع القيم المتقاربة في عدد مناسب من الفئات، ومع ذلك ليس هناك قاعدة ثابتة للحصول على العدد الملائم من الفئات ولو أن بعض الإحصائيين ينصحون

باستخدام قاعدة سترجس التالية:

$$(K) = 1 + 3.322 \log (n)$$

حيث n = عدد المفردات (حجم العينة)

ويفضل أن يكون عدد الفئات يتراوح بين (5 و 12) فئة طبقاً لحجم العينة مع مراعاة حدود الفئات من خلال تحديد المدى الذي يعرف بأنه: عبارة عن الفرق بين أكبر قيمة وأصغر قيمة في مجموعة المفردات، ويتم حساب طول الفئة بالعلاقة التالية:

$$(C) \text{ طول الفئة} = \frac{(R) \text{ المدى}}{(K) \text{ عدد الفئات}} \quad (\text{يتم تقريب الناتج إلى أقرب رقم صحيح})$$

تمرين: أخذت عينة عشوائية مكونة من (30) منشأة في إحدى الدول فكان رأسمال كل منشأة من هذه المنشآت على النحو التالي (بالمليون):

21	10	18	17	14
15	12	21	25	22
23	19	20	16	26
23	18	14	19	16
19	19	29	18	20
24	27	22	20	24

والمطلوب: إنشاء جدول تكراري لهذه البيانات في فئات متساوية إذا علمت

أن : $(K=5)$

$$\text{الحل: } R = 29 - 10 = 19$$

$$K=5$$

$$C = \frac{R}{K} = \frac{19}{5} = 3.90 \approx 4$$

ويمكن أيضاً تحديد عدد الفئات المطلوبة لتكوين الجدول، باستخدام علاقة رياضية أخرى تدعى علاقة يول (Yule) على النحو التالي:

$$K = 2.5 \sqrt[4]{n}$$

وبالتالي يمكن استخدام أي من الصيغتين كمؤشر في تحديد العدد المناسب للفئات.

وبالتالي سيكون لدينا الجدول التكراري:

الفئات	التكرارات	وسط الفئات (x)
10-14	2	12
14-18	6	16
18-22	12	20
22-26	7	24
26-30	3	28

من خلال الجدول نجد أن طول الفئة عبارة عن المدى العددي الواقع بين حدي الفئة الأدنى والأعلى، وبالتالي طول الفئة الأولى هو (4) ونقرأ (10 وأقل من 14) وهكذا لبقية الفئات، وتعتبر هذه الصورة العامة للفئات سواء أكان المتغير متصل أم منفصلاً، أما وسط الفئة فهو عبارة عن (الحد الأدنى + الحد الأعلى) مقسوماً على (2). مع العلم أن هناك عدة طرائق في تحديد فئات الجدول التكراري تتفق جميعها في الأسس، ولكنها تختلف في طريقة العرض، ومن أمثلة ذلك:

أ — (10-20) تحتوي على كل البيانات التي أكبر أو تساوي (10) إلى أقل من (20).

ب — (10-20) تحتوي على كل البيانات الأكبر من (10) إلى (20) (بما فيها 20)

ج — التعامل مع الحدود الحقيقية للفترات فمثلاً الفئة (10-20) حدها الأول

الظاهري يساوي (10) وحدها الأعلى الظاهري يساوي (20) بينما الحد الأدنى الحقيقي والأعلى الحقيقي هو :

(9.5-20.5) أي :

الحد الأدنى الحقيقي = الحد الأدنى الظاهري - 0.5 وحدة قياس.

الحد الأعلى الحقيقي = الحد الأعلى الحقيقي + 0.5 وحدة قياس

6-2- عرض البيانات الإحصائية :

تعتبر عملية عرض البيانات الإحصائية امتداداً لعملية التفريغ والتبويب بحيث يصعب الفصل بينهما وكلتا العمليتين تهدفان إلى ترتيب البيانات المجموعة عن الظاهرة محل الدراسة وتنسيقها بطريقة تساعد على فهم مدلولها والاستفادة منها تمهيداً لعرضها وإبراز أهميتها، ويتم إما جدولياً أو بيانياً.

• العرض الجدولي للبيانات الإحصائية :

العرض الجدولي هو: "عبارة عن عرض البيانات على شكل جدول"، والجدول الإحصائي: "عبارة عن ترتيب البيانات العددية في صورة صفوف وأعمدة بحيث يمكن قراءة الجدول بالاتجاهين"، والهدف منه الاختصار وإبراز أهمية البيانات، وتسمى العملية التي يتم فيها تجميع البيانات في مجموعات مميزة ومتجانسة عملية التصنيف، ويوجه عام تصنف البيانات الإحصائية طبقاً لإحدى القواعد التالية :

I - تصنيف جغرافي :

في هذا النوع من التصنيف تُجمع الوحدات التي تشترك في صفة مكانية واحدة في مجموعة مستقلة مثال ذلك: التوزيع النسبي للصادرات والمستوردات السورية حسب الكتل الدولية لعام 2004.

التوزيع النسبي للمصادر والمستوردات السورية

جدول رقم (4)

المستوردات	المصادر	الكتل الدولية
16.4	29.9	البلدان العربية
15.2	53.2	بلدان الاتحاد الأوروبي
17.1	2.3	بلدان أوربية أخرى
9.6	3.5	البلدان الأمريكية
26.5	10.4	بلدان أسيوية مختلفة
15.2	0.7	بلدان أخرى

المصدر: م.م للإحصاء 2004

II - تصنيف تاريخي أو زمني:

حيث تُجمع الوحدات المرتبطة بزمان معين في مجموعة واحدة توضع البيانات حسب زمن حدوثها، وتسمى بيانات ذلك النوع من الجداول باسم السلسلة الزمنية، كما هو موضح في الجدول رقم (5) الآتي:

نسبة موازنة التعليم العالي من الموازنة العامة للدولة من عام (2001-2005)

جدول رقم (5)

السنة	النسبة
2001	2.73
2002	3.16
2003	3.70
2004	3.76
2005	3.46

المصدر: المجموعة الإحصائية السورية ، 2006

III - تصنيف كمي :

ويتم ذلك من خلال تجميع الوحدات المشتركة في درجة الصفة المعينة والتي تأخذ شكلاً رقمياً محدداً في مجموعة واحدة، مثال ذلك : عدد أعضاء الهيئة التدريسية، وإجمالي عدد العاملين وإجمالي عدد الطلاب في جامعة تشرين في عام 2006

جدول رقم (6)

عدد أعضاء هيئة التدريس	عدد العاملين	عدد الطلاب	عدد الطلاب لكل عضو هيئة تدريس	عدد الطلاب لكل عامل
1161	1428	52869	45.54	37.02

المصدر: جامعة تشرين 2006

أنواع الجداول الإحصائية :

يمكن التمييز بين نوعين للجداول الإحصائية:

1 - الجدول البسيط :

عبارة عن الجدول الذي يشمل بياناً واحداً فقط، فهو يحتوي على عمودين أحدهما للبيان موضوع الدراسة يقابله عمود وآخر يظهر التكرارات المقابلة، مثال ذلك: جدول نسبة الأنفاق على وسائل الإعلان الأساسية في سورية خلال (2000-2005).

جدول رقم (7): النسبة المئوية للاتفاق على وسائل الإعلان

وسيلة الإعلان	الإعلان الطرقي	الإعلان المطبوع	التلفاز	الإذاعة	المجموع
النسبة المئوية للاتفاق %	50	8	41	1	100

المصدر: المجموعة الإحصائية السورية م.م. ح 2005

II - الجدول المزدوج :

عبارة عن الجدول الذي صنف في الوحدات حسب ظاهرتين، مثال ذلك: الجدول التكراري التالي ويتضمن توزيع (100) امرأة حسب الحالة التعليمية والحالة الزوجية في إحدى دوائر الأحوال المدنية بإحدى المدن.

جدول رقم (8): توزيع النساء حسب الحالة التعليمية والزواجية

الحالة التعليمية / الحالة الزوجية	أمية	قرأ وكتب	ابتدائية	إعدادية	ثانوية	جامعية	دراسات عليا	المجموع
عازية	4	3	3	2	5	6	2	25
متزوجة	11	7	10	7	4	3	1	43
مطلقة	3	0	2	3	8	2	1	19
أرملة	2	3	2	1	3	2	0	13
المجموع	20	13	17	13	20	13	4	100

المصدر: فرضي

• الشروط الواجب مراعاتها عند إنشاء الجداول الإحصائية :

- 1 - كتابة عنوان الجدول بشكل واضح يبين محتواه.
- 2 - كتابة المصدر الذي أخذت منه البيانات (وزارة التعليم العالي، جمع شخصي، مصدر فرضي، 2000).
- 3 - تسجيل الملاحظات الخاصة في أسفل الجدول مع الإشارة إليها بعلامات خاصة.
- 4 - يجب أن تذكر وحدات القياس ورقم الجدول .

• العرض البياني للبيانات الإحصائية:

نحن نعلم أن الهدف الأساسي لعلم الإحصاء هو جمع وعرض البيانات المجموعة عن ظاهرة معينة بأسلوب يسهل معه إدراك اتجاه الظاهرة، ولذلك إلى جانب العرض الجدولي للبيانات الإحصائية يسعى الباحث غالباً إلى عرضها

بيانياً على شكل مصورات ورسومات بيانية وأشكال هندسية لإبراز خصائص ومميزات البيانات الأقل وضوحاً، وذلك كالتالي:

II - الأشكال المعبرة :

عبارة عن استخدام رسومات أو صور أو أشكال تعبيرية تمثل القسيم المختلفة للظواهر المطلوب عرضها، وهذا الأسلوب يقتصر استخدامه على القيم الإجمالية للظاهرة أو الظواهر محل الدراسة ولغرض المقارنات فقط ، مثال ذلك : طلاب و طالبات إحدى الجامعات .

II - الرسوم الدائرية:

تقوم على افتراض أن القيمة الكلية للظاهرة المدروسة تمثل إجمالي مساحة الدائرة ثم نقسم الدائرة إلى قطاعات جزئية بحيث تناسب مساحات تلك القطاعات مع قيم المجموعات الجزئية للظاهرة وتستخدم عادة هذه الطريقة في إظهار نسب التوزيع الهيكلي للظاهرة.

مثال :

يوضح الجدول التالي عدد طلاب الدراسات العليا في كلية الآداب بإحدى الجامعات العربية خلال إحدى السنوات حسب التخصص والمطلوب عرض تلك البيانات باستخدام أسلوب الدائرة .

جدول رقم (9): يبين عدد الطلبة في الدراسات العليا

عدد الطلبة	التخصص
14	إعلام
12	مكتبات
24	لغات
6	جغرافية
4	علم اجتماع
60	المجموع

المصدر: فرضي

الحل :

- 1 - نرسم دائرة باستخدام مقياس رسم مناسب.
- 2 - نحسب مقدار زوايا القطاعات المختلفة حسب ما هو وارد بالجدول التالي على النحو التالي:

$$\text{زاوية القطاع} = \frac{\text{تكرار الظاهرة}}{360^\circ} \times \text{مجموع التكرارات}$$

جدول رقم (10): كيفية ضرب قيم زاوية القطاعات

التخصص	عدد الطلبة	النسبة المئوية %	قيمة زاوية القطاع
اعلام	14	0.23	84
مكتبات	12	0.20	72
لغات	24	0.40	144
جغرافية	6	0.10	36
علم اجتماع	4	0.06	24

و بالتالي نحصل على دائرة مقسمة إلى قطاعات كما هو موضح بالشكل :



علم اجتماع □ جغرافية ■ لغات □ مكتبات ■ اعلام ■

III - الأعمدة أو المستطيلات :

تستخدم الأعمدة أو المستطيلات ذات القواعد المتساوية لعرض البيانات الإحصائية، والفكرة الأساسية تقوم على رسم أعمدة أو مستطيلات متساوية القاعدة يتناسب ارتفاعها مع قيم الظاهرة المعروضة مع العلم عند استخدام

طريقة المستطيلات أو الأعمدة في عرض البيانات الإحصائية يجب التمييز بين الأنواع التالية :

الأعمدة البسيطة :

تستخدم لتمثيل ظاهرة واحدة في عدة فترات زمنية أو أوجه مختلفة لظاهرة واحدة في فترة زمنية واحدة

مثال : يوضح البيان التالي عدد الخريجين (دكتوراه) من الجامعات السورية خلال الفترة (1999 – 2004).

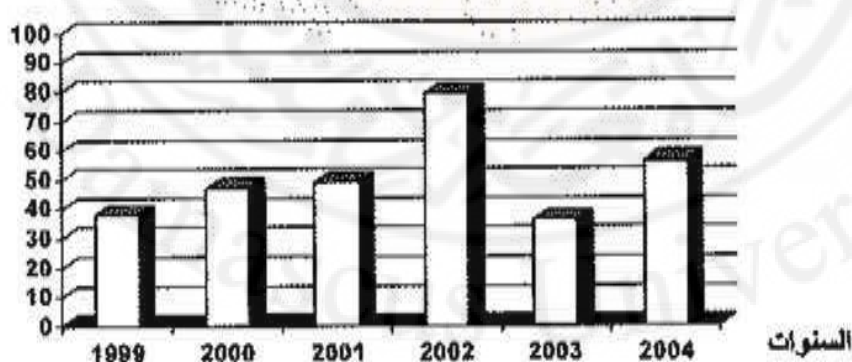
جدول رقم (11): عدد خريجين الدكتوراه

السنوات	عدد الخريجين (دكتوراه)
1999	38
2000	47
2001	49
2002	79
2003	36
2004	56

المصدر: المجموعة الإحصائية السورية لعام 2005

المطلوب: عرض تلك البيانات باستخدام الأعمدة .

الحل : باستخدام ورقة الرسم البياني للسهولة فإننا نقوم برسم المحورين الأفقي والعمودي، ويخصص المحور الأفقي للسنوات والعمودي للقيم. عدد الخريجين



الأعمدة المتلاصقة :

تستخدم عند عرض ظاهرة لعدة فترات زمنية أو عدة أماكن جغرافية، بمعنى آخر عند المقارنة بين ظاهرتين أو أكثر لعدة سنوات، مثال ذلك الصادرات والواردات لنفس الفترة وعدد المواليد وعدد الوفيات لنفس السنة .. الخ .

مثال :

البيان التالي يوضح قيمة الصادرات والواردات بملايين الليرات السورية في الفترة من 1995-1999—

جدول رقم (12)

السنة	الصادرات	الواردات
1995	44562	52856
1996	44887	60385
1997	43953	45211
1998	32443	43725
1999	38880	43010

المصدر: المجموعة الإحصائية السورية (1996-2000)

والمطلوب:

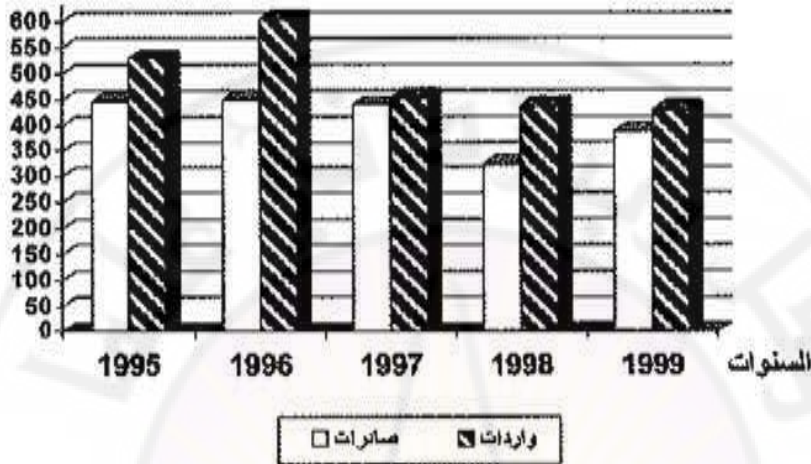
عرض تلك البيانات باستخدام الأعمدة (المستطيلات).

الحل:

يخصص لكل سنة على المحور الأفقي عمودان أحدهما يمثل الصادرات والآخر الملاصق له يمثل الواردات مع ترك مسافات منتظمة بين أعمدة كل سنة.

شكل رقم (14)
تطور قيمة الصادرات والواردات

الصادرات، الواردات



الأعمدة المجزأة :

تستخدم عندما تكون القيمة الإجمالية للظاهرة موزعة على مجموعات فرعية مميزة، أي تجزئة القيم الإجمالية، بحيث يخصص لكل جزء ارتفاع يتناسب مع حجم الظاهرة بالنسبة إلى الحجم الكلي، بمعنى آخر، العمود الذي يمثل إجمالي قيمة الظاهرة يقسم بما يتناسب وقيم المجموعات الجزئية مع التمييز بين أقسام كل عمود بألوان مميزة .

مثال :

يوضح البيان التالي أعداد العاملين في إحدى القنوات الفضائية العربية والوضع الوظيفي خلال الأعوام : 2000، 2001، 2002. والمطلوب : تمثيل تلك البيانات باستخدام أسلوب الأعمدة (المستطيلات) المجزأة.

عدد العاملون حسب الوضع الوظيفي في إحدى القنوات الفضائية العربية
جدول رقم (13)

البيان	السنوات		
	2002	2001	2000
العاملون الدائمون	255	175	120
العاملون بعقود سنوية	340	288	130
العاملون بعقود موسمية	100	90	105
إجمالي	695	553	355

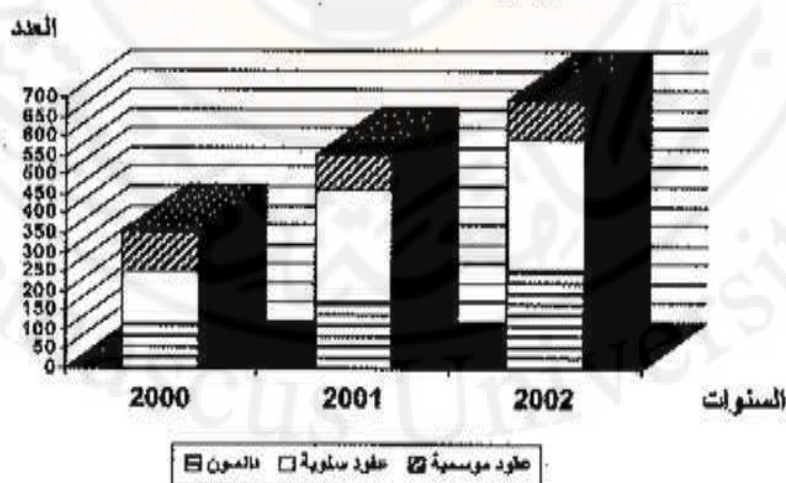
المصدر: فرضي

الحل : لاحظ هنا أن البيان يمثل ظاهرة واحدة وهي عدد العاملين في كل سنة، وهذا العدد (الظاهرة) مجزأة إلى مجموعات جزئية .

1 - نرسم أعمدة تتناسب ارتفاعاتها مع إجمالي قيمة الظاهرة وهي إجمالي عدد العاملين.

2 - نقوم بتجزئة كل عمود بما يتناسب مع عدد العاملين في كل مجموعة جزئية.

والشكل التالي يوضح ذلك:



IV – الخطوط البيانية:

تستخدم في الغالب لمقارنة البيانات المتجانسة عبر الزمن كتمثيل العلاقة بين ظاهرتين أو متغيرين x ، y بخط بياني من خلال رسم النقاط والوصل بينها للحصول على خطوط مستقيمة متصلة.

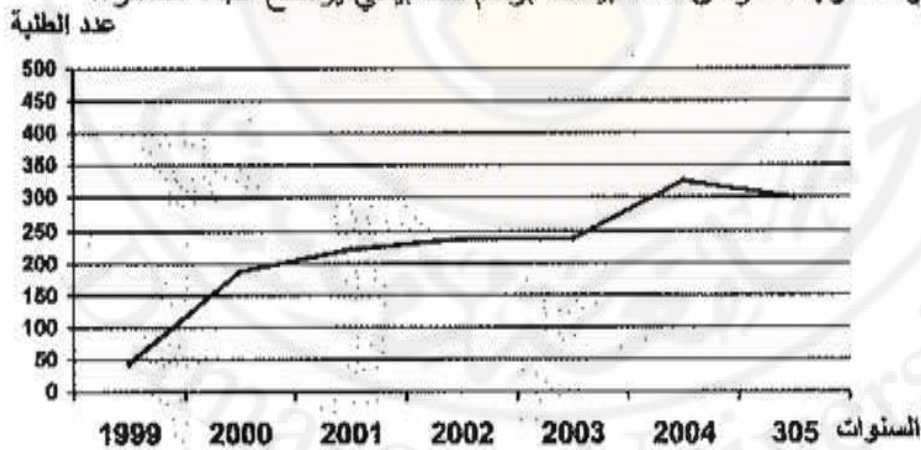
مثال : يوضح الجدول التالي الصادرات السورية : القيمة بمليارات الليرات السورية خلال الفترة من 1999 حتى 2004 :

جدول رقم (14): الصادرات السورية

الصادرات	السنوات
43	1999
187.5	2000
221	2001
236	2002
237	2003
327	2004

المصدر: المجموعة الإحصائية السورية (2006)

والمطلوب : عرض تلك البيانات برسم خط بياني يوضح اتجاه الظاهرة.



وتجدر الإشارة إلى أنه أصبح من السهولة تنفيذ الرسوم البيانية والأعمدة والدائرة وغيرها بواسطة البرامج الحاسوبية المختلفة منها (SPSS, Excel):

* من خلال برنامج (Excel):

— تحديد البيانات المراد تمثيلها ومن ثم النقر على رمز معالج التخطيطات (chart wizard) الموجودة في شريط الأدوات الرئيسي لتظهر لوحة حوار معالج التخطيطات يلي ذلك النقر على نوع التخطيط المطلوب (شريطي، خطي، دائري، مبعثر، مساحي..). ليظهر بأشكاله المتنوعة، ويتم اختيار معالج التخطيطات من شريط الأدوات ليتم اختيار النمط المطلوب (أعمدة ثلاثية الأبعاد مثلاً).

أخيراً يتم كتابة العنوان للشكل وعناوين فرعية للفئات والمشاهدات ثم أمر (انتهاء) فيظهر الرسم المطلوب من الزاوية اليسرى من الشاشة ومن الممكن استخدام الأعمدة الثنائية الأبعاد كما يمكن اختيار أعمدة أفقية أو رأسية وبسفس الأسلوب للأعمدة البيانية المركبة والمجزأة والدائرية والرسوم البيانية.

* من خلال برنامج (SPSS):

— اختر الأمر (Bar) عمود من شريط الأدوات فيظهر على الشاشة (Bar Charts) ثم اختر الشكل البسيط (Simple) أو المركب (Clusterce) أو المجزأ (Stached) وبعدها حدد (Define) البيانات، بحيث ينقل المتغير (العدد) إلى مستطيل تمثيل الأعمدة كما ينقل متغير الظاهرة إلى مستطيل المتغير وبعدها ينقر على موافق (OK) ليظهر الرسم البياني في ملف المخرجات. مثال: لدينا الجدول التالي المتضمن أعداد الطلبة المقبولين في كلية الإعلام في إحدى الجامعات خلال السنوات (2000-2003). اعرض هذه البيانات بطريقة الأعمدة المستطيلات

السنة	الذكور	الإناث	المجموع
2000/2001	300	200	500
2001/2002	200	350	550
2002/2003	450	400	850
2003/2004	350	300	650

هذا المثال باستعمال (SPSS) نتبع الخطوات التالية:

- 1 - لتشغيل SPSS أشر على Start ثم Programs واختبر SPSS for Windows ثم أشر على Type in data ثم انقر OK.
- 2 - عرف المتغيرات وذلك بالنقر على variable view، وأكتب كلمة السنة year تحت name وحدد نوع المتغير بالانتقال على خانة type ثم انقر الزر الرمادي الموجود في تلك الخانة. تأكد أنك اخترت نوع Numeric.
- 3 - عرف المتغيرات الأخرى وهي الإناث females، الذكور males، المجموع total، مع التأكد أن هذه المتغيرات هي من نوع Numeric، وبما أن الأعداد في بيانات هذا المثال لا تحتوي على كسور ضع (0) صفراً في خانة Decimal places ثم انقر Ok.
- 4 - انقر Data view فيظهر لك الشكل الذي فيه السنة في رأس العمود الثاني، الطالبات على رأس العمود الثالث، الطلاب على رأس العمود الرابع والمجموع على رأس العمود الخامس.
- 5 - أدخل بيانات الجدول في هذه الأعمدة.
- 6 - احفظ save أو اطبع print فتظهر لك النتيجة كما في الجدول (حفظت البيانات في الملف).
- 7 - من menu bar اختر (graphs) ثم bar فيظهر لك صندوق bar charts اختر منه clustered أو stached حسب رغبتك فيكون

الأعمدة بجانب بعضها أو فوق بعضها على التوالي، أما في حالة عرض ظاهرة واحدة فاختر simple.

8 - في مربع الخيارات الذي يظهر، انقل year إلى categorg axis وذلك بتظليل year ثم انقر على السهم.

	years	Females	males	total
1	2000	300	200	500
2	2001	200	350	550
3	2002	450	400	850
4	2003	350	300	650

9 - ظلل الظواهر التي تريد عرضها واحدة تلو الأخرى.

10 - انقر OK فيظهر لك عرض البيانات المطلوبة حسب الزمن

* مزايا العرض البياني:

— البساطة في قراءة البيانات وخاصة إذا كان عدد المشاهدات كبيراً.

— سهولة تذكر النتائج.

— إمكانية توضيح أو تأكيد بيان ما، وذلك باستخدام الألوان.

— جذب الانتباه.

* مساوئ العرض البياني:

— التوضيحية في دقة البيانات، باعتبار الأشكال توضح التغيرات العامة ولا تدخل في التفاصيل.

— أحياناً قد تكون معقدة إذا كان هناك بيانات متعددة ومختلفة.

6-2- التوزيعات التكرارية :

تعتبر جداول التوزيع التكراري للبيانات الكمية أساساً لمعظم العمليات الإحصائية الوصفية والتحليلية، وجدول التوزيع التكراري بالتعريف هو عبارة عن عرض البيانات العددية في فئات متسلسلة يظهر أمام كل منها ما يقابلها من تكرارات في بيانها الأصلي.

مثال: أخذت عينة عشوائية من مراكز بيع الصحف والمجلات بإحدى الدول فكان التوزيع التكراري لفئات المبيعات الشهرية لهذه المراكز كالتالي (المبيعات بالآلاف الوحدات النقدية):

جدول رقم (15): الجدول التكراري للمبيعات الشهرية في مراكز بيع الصحف للمجلات (بالآلاف الوحدات المقدمة)

فئات المبيعات	45-35	55-45	65-55	75-65	85-75	95-85	105-95	المجموع
عدد المراكز	3	4	7	14	15	6	1	50

يتضح من الجدول رقم (15) أن عدد المراكز في الفئة الأولى هو (3) والثانية هو (4) وهكذا...

وتجدر الإشارة إلى أنه في حالة الظواهر التي تتغير تغيراً متصلاً كأطوال الأشخاص أو أعمارهم، فإن قيمتها تنتشر في المدى بين القميتين الصغرى والكبرى في المجموعة بالتدرج بدون أن تتجمع في بعض النقاط دون الأخرى لأن المتغير المتصل معناه أن القيم لا تقفز من قيمة لأخرى أكبر منها، وإنما بالتدرج بحيث لا تمر بكل القيم المتوسطة

• أنواع التوزيعات التكرارية: ويمكننا التمييز بين:

— توزيعات تكرارية منتظمة: أي أطوال فئاته متساوية ويمكن ملاحظة تمركز مفرداته.

— توزيعات تكرارية غير منتظمة: أي أطوال فئاته غير متساوية وصعوبة تمركز مفرداته.

• أشكال التوزيعات التكرارية:

— مغلقة: أي محدد البداية والنهاية.

— مفتوحة: أي فئاته الحدية مفتوحة.

• الجداول التكرارية المزدوجة:

يستخدم هذا النوع من الجداول في وصف وتلخيص البيانات المتعلقة بدراسة ظاهرتين معاً وقد يكون الجدول المزدوج كمياً أو نوعياً أو خليطاً ومن أمثلة ذلك الجداول التالية:

أ - جدول تكراري مزدوج كمّي :

المجموع	180-160	160-140	140-120	الطول الوزن
24	6	8	10	40-20
47	10	22	15	60-40
33	4	17	12	80-60
16	2	6	8	100-80
120	22	53	45	المجموع

ب - جدول تكراري مزدوج نوعي :

المجموع	غير مدخن	مدخن	التدخين الإصابة
340	40	300	مصاب
260	250	10	سليم
600	290	310	المجموع

ج - جدول تكراري خليط (كمّي ونوعي):

عدد الأطفال	2-0	4-2	6-4
المستوى التطبيقي للأب	أمي	ابنتاتي وإعداداتي	ثانوي
10	5	6	4
6	3	8	2
4	3	2	3
2	3	3	2

مفهوم التكرار التجميعي الصاعد والهابط :

يستخدم هذا النوع من التكرارات لمعرفة عدد المفردات في التوزيع الأكبر أو الأقل من قيمة معينة هي الحدود الدنيا أو العليا للفئات التي يتكون منها جدول التوزيع التكراري الأصلي.

ولتوضيح ذلك نعود إلى بيانات الجدول رقم (15) والمتعلق بتوزيع المراكز حسب فئات المبيعات فإنه يمكن استنتاج التكرار التجميعي الصاعد والتكرار التجميعي الهابط على النحو التالي :

جدول التكرار المتجمع الصاعد والهابط

جدول رقم (16)

التكرار التجميعي الهابط	الحدود	التكرار التجميعي الصاعد		التكرار (عدد المراكز f	فئات المبيعات بالآلاف الوحدات النقدية
		التكرار المتجمع	الحدود		
50	35 فأكثر	صفر	أقل من 35	3	45-35
47	45 فأكثر	3	أقل من 45	4	55-45
43	55 فأكثر	7	أقل من 55	7	65-55
36	65 فأكثر	14	أقل من 65	14	75-65
22	75 فأكثر	28	أقل من 75	15	85-75
7	85 فأكثر	43	أقل من 85	6	95-85
1	95 فأكثر	49	أقل من 95	1	105-95
صفر	105 فأكثر	50	أقل من 105		
				50	المجموع

وبذلك يمكن استخدام الجدول التكراري التجميعي الصاعد لحساب تقدير لعدد المفردات التي تحمل قيمة الظاهرة عند حدود تقل عن قيمة معينة، فمثلاً نلاحظ أن عدد المراكز التي تقل مبيعاتها عن (45) ألف وحدة نقدية هم ثلاثة مراكز، وبذلك تكون نسبتهم إلى إجمالي المراكز هي $(3/50) \times 100 = 6\%$ من المراكز،

وكذلك فإن عدد المراكز التي مبيعاتها أقل من (65) ألف يساوي (14) ونسبتهم $= 14/50 = 0.28$ أو (28%) من إجمالي المراكز.

— مفهوم التكرار النسبي :

يتم تحديد التكرار النسبي الخاص بكل فئة من خلال قسمة تكرار كل فئة

$$\text{على مجموع التكرارات، } (fc = \frac{f}{\sum f})$$

مثال:

الجدول التكراري المرفق يوضح فئات الودائع وعدد المودعين في إحدى المصارف الخاصة (بالمليون ليرة).

المطلوب: حساب التكرارات النسبية

جدول رقم (17)

التكرار النسبي	المودعين	فئات الودائع
0.03	1	15-10
0.10	3	20-15
0.20	6	25-20
0.46	14	30-25
0.03	3	35-30
0.06	2	40-35
0.10	1	45-40
≈ 1.000	30	المجموع

من خلال الجدول رقم (17) نجد أن نسبة المودعين ما بين 10 إلى 15 مليون تعادل 0.03 من إجمالي المودعين، ونسبة المودعين الذين لدى كل منهم من 15 إلى 20 مليون تبلغ 0.10 من إجمالي المودعين في ذلك المصرف وهكذا.

V - التمثيل الهندسي والبياني للجدول التكرارية :

يمكن تمثيل بيانات الجدول التكراري عن طريق رسم كل من :

- (1) المدرج التكراري : عبارة عن مجموعة من المستطيلات المتلاصقة، قواعدها في الفئات وارتفاعها هي التكرارات.
- (2) المضلع التكراري : عبارة عن الوصل بين نقاط أوساط الفئات للمستطيلات بخطوط مستقيمة.
- (3) المنحني التكراري : عبارة عن التمهيد للمضلع التكراري.

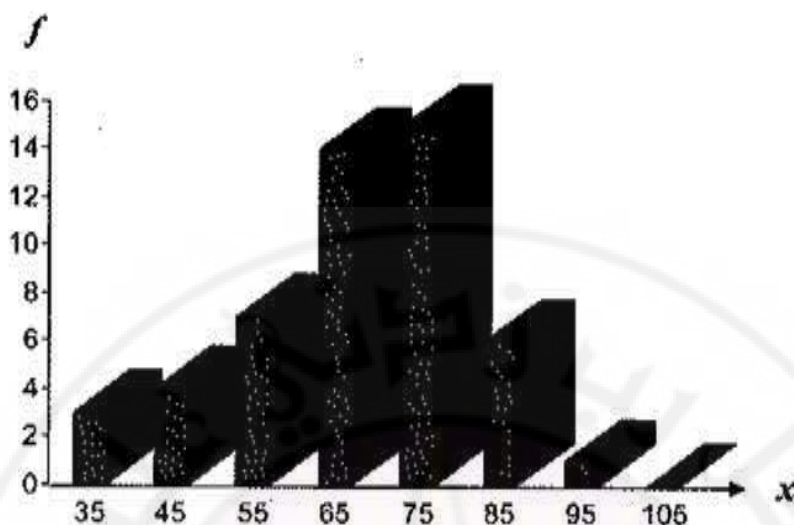
مع العلم عند رسم المدرج التكراري ينبغي مراعاة مايلي:

- 1 - أن تكون الفئات متصلة بمعنى نهاية أي فئة هي بداية الفئة التي تليها وإن لم تكن كذلك يتم التعامل مع الحدود الحقيقية للفئات.
- 2 - أن تكون أطوال الفئات ذات أطوال متساوية.

مع العلم أن المساحة الكلية للبيانات تقع تحت المدرج التكراري وكل مستطيل من مستطيلات المدرج التكراري يمثل جزء من هذه المساحة، وتتم المقارنة بين المستطيلات المختلفة من حيث كمية المعلومات التي يتضمنها من خلال حساب مساحة كل مستطيل على حدة وهذه المقارنة صحيحة إذا كانت الفئات متساوية الأطوال.

ولرسم المدرج التكراري يخصص المحور الأفقي لفئات الظاهرة، ويخصص المحور العمودي للتكرارات المقابلة. ولتوضيح ذلك نعود لبيانات الجدول رقم (15) المتعلقة بمبيعات مراكز بيع الصحف والمجلات.

يوضح الشكل التالي نرسم الشكل التالي ليوضح المدرج التكراري لتوزيع (50) مركز بيع للصحف والمجلات بإحدى الدول



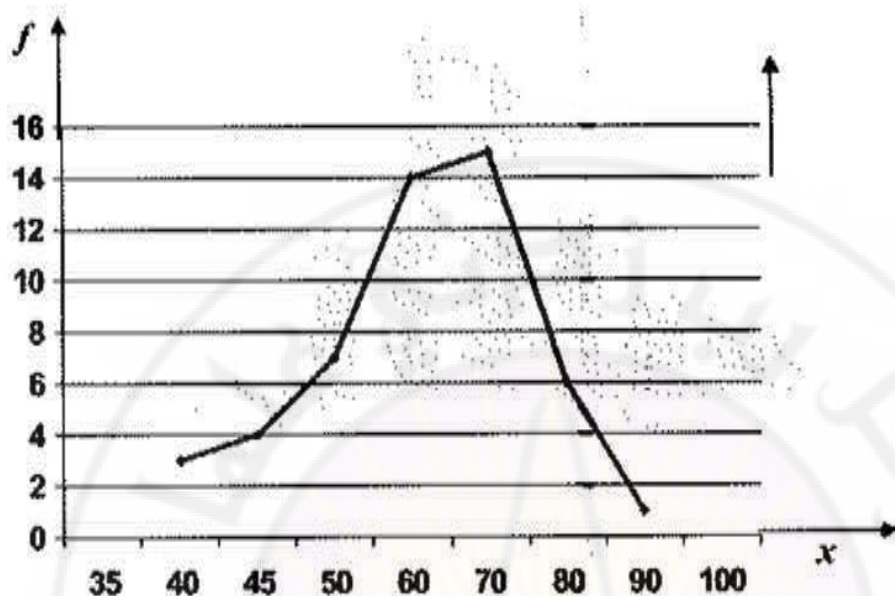
المدرج التكراري لتوزيع مراكز البيع

ولرسم المضلع التكراري نأخذ مراكز الفئات (مركز الفئة هو نصف مجموع حدي الفئة الأدنى والأعلى)، والتكرارات المناظرة لها، ثم نوصل هذه النقاط بخط منكسر فنحصل على المضلع التكراري كما هو موضح في الجدول التالي:

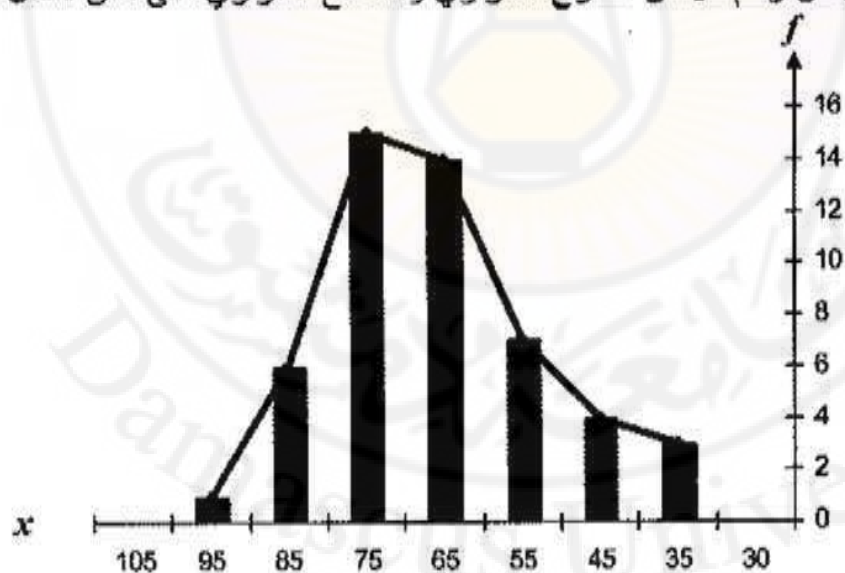
جدول رقم (21)

105-95	95-85	85-75	75-65	65-55	55-45	45-35	فئات
100	90	80	70	60	50	40	مركز الفئة
1	6	15	14	7	4	3	التكرارات

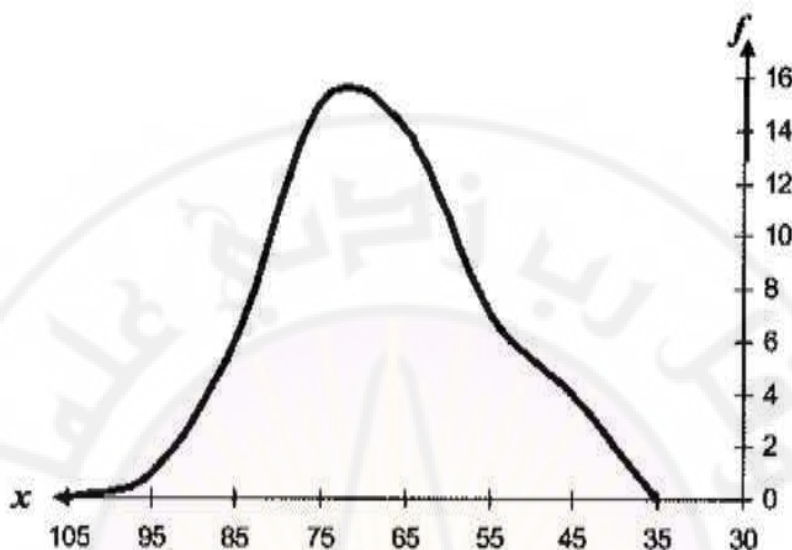
وبذلك نحصل على الشكل التالي للمضلع التكراري:



كما يمكن رسم كل من المدرج التكراري والمضلع التكراري على نفس الشكل :



وبإجراء عملية تهذيب للمضلع التكراري فإنه يمكن الحصول على المنحني التكراري حيث لا يشترط هنا أن يمر المنحني بجميع نقاط المضلع التكراري .



المنحني التكراري لتوزيع المراكز

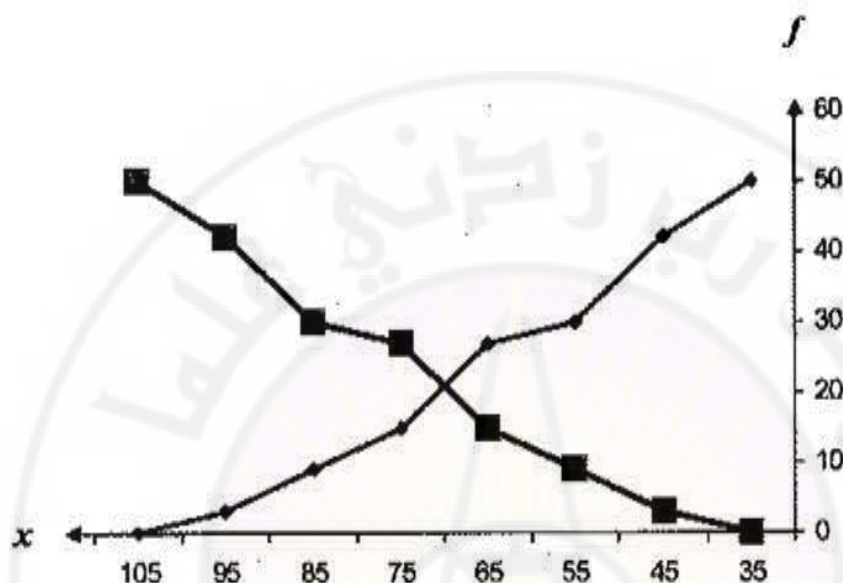
وفي حالة عرض الجدول التكراري في فئات غير منتظمة، فإنه لابد من تعديل التكرارات المقابلة لكل فئة، حيث أنه سبق القول: أن المساحة المخصصة لكل عمود تعبر عن حجم الظاهرة المقابلة لكل فئة من فئات الجدول مع مراعاة تفاوت أطوال القواعد للأعمدة الناتجة.

التمثيل البياني للجدول التكراري التجميعي الصاعد والهابط :

يمكن تمثيل الجدول التكراري التجميعي الصاعد بيانياً بواسطة تحديد نقط المنحني بنفس الأسلوب السابق، حيث يخصص المحور الأفقي لحدود قيم الظاهرة والمحور العمودي للتكرارات المتجمعة، ويمكن تمثيل كل من منحني التكرار التجميعي الصاعد والهابط على نفس الرسم .

مثال :

رسم منحني التكرار التجميعي الصاعد والمنحني التكراري التجميعي الهابط لنفس المثال السابق المتعلق بمراكز بيع الصحف والمجلات.



VI - خصائص التوزيعات التكرارية:

- 1 - تظهر المفردات ميلاً للتركز حول قيمة معينة والتناقص على جانبي هذه القيمة، تدعى هذه القيمة النزعة المركزية وتقاس بالمقاييس التالية: الوسط الحسابي، الوسيط، المنوال، الوسط الهندسي، الوسط للتوافقي.
- 2 - تبتعد القيم عن نقطة التركيز هذا الابتعاد يدعى، التشتت، ويقاس بالمقاييس التالية: المدى، الانحراف الربيعي، الانحراف المتوسط، الانحراف المعياري.
- 3 - تختلف التوزيعات التكرارية عن بعضها البعض من حيث درجة التماثل، أي:

- (a) - منحني متماثل ينطبق على المنحني الطبيعي
- (b) - منحني غير متماثل وبالتالي لا ينطبق على المنحني الطبيعي ويكون ملتوياً نحو اليمين أو نحو اليسار.

وهذا يقودنا إلى التعرف على مفهوم الالتواء:

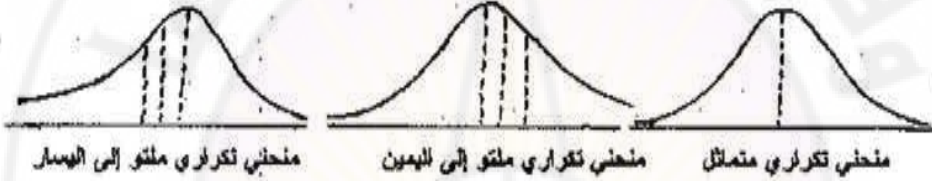
فالالتواء هو تماثل أو درجة التماثل أو البعد عنها، ونحن نميز بين:

(a) المنحني التكراري المتماثل Symmetrical Curve

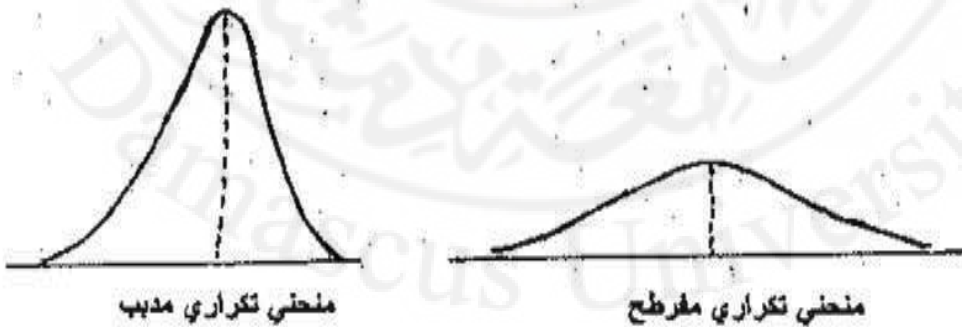
يشبه الجرس المقلوب وله محور تماثل عامودي يمر بنقطة النهاية العظمى له (قمة المنحني) ويقسمه إلى قسمين متساويين ويتحقق عندما ينطبق على منحني للتوزيع.

(b) المنحني التكراري غير المتماثل Skewed Curve :

وهو عبارة عن منحني تكراري لا ينطبق على المنحني الطبيعي فقد يكون ملتوياً نحو اليمين أو ملتوياً نحو اليسار.



4- تختلف المنحنيات التكرارية ذات القمة الواحدة من حيث مدى الاتساع فقد يكون المنحني التكراري مدبباً حيث تضيق قمته بشكل ملحوظ ويسمى منحني مدبب وقد تتسع قمة المنحني فيقال له حينئذ منحني مفرطح وبشكل عام يعرف المفرطح بأنه قياس درجة علو أو انخفاض قمة التوزيع بالنسبة للتوزيع المتماثل.



أسئلة وتمارين غير محلولة

- 1 — ما هي أساليب جمع البيانات الإحصائية، وما هي مزايا وعيوب كل أسلوب؟
- 2 — ما هي المصادر الأساسية للبيانات الإحصائية مع الشرح؟
- 3 — اذكر ثلاثة شروط هامة يجب مراعاتها عند صياغة أسئلة الاستمارة الإحصائية؟
- 4 — ما هي التصنيفات المختلفة للبيانات الإحصائية؟ اذكر مثلاً لكل تصنيف؟
- 5 — ما الفرق بين الإحصاء الوصفي والإحصاء الاستدلالي؟
- 6 — ميز بين المفاهيم الآتية : الجدول الإحصائي، المدرج التكراري، التوزيع التكراري.
- 7 — عرف مع الشرح المصطلحات الإحصائية الآتية :
مجتمع البحث، وحدة البحث، المتغير المتصل والمنقطع، المتغير الكمي، المتغير النوعي، الاستبيان، الرقم الإحصائي، البحث الوصفي، البحث الإحصائي التحليلي
- 8 — اذكر مع الشرح نوعين من العينات العشوائية موضحاً ظروف استخدام كل منهما.
- 9 — اذكر مع الشرح الشروط الهامة التي يجب مراعاتها عند إعداد الجدول الإحصائي.
- 10 — فيما يلي بيان بتطور المعدل اليومي لأقصى استهلاك من الكهرباء بالمليون واط ساعي في إحدى الدول .

السنة	1996	1997	1998	1999	2000
المعدل	4.5	7	18	24	32

والمطلوب عرض تلك البيانات بالأسلوب الذي تراه مناسباً .

11 - يوضح البيان التالي أعداد الأطباء بإحدى المدن حسب التخصص في عامي (2000, 2004) .

التخصص	عام 2000	عام 2004
بشري	208	310
أسنان	24	86
علاج طبيعي	9	23
المجموع	241	419

والمطلوب عرض تلك البيانات بالأسلوب الذي نراه مناسباً.

12 - يوضح البيان التالي أعداد الطلاب والطالبات (بسالالوف) بإحدى الجامعات موزعين حسب النوع خلال أربعة أعوام.

السنة	طلاب	طالبات	إجمالي
1998	18	22	40
1999	24	18	42
2000	28	16	44
2001	32	24	56

والمطلوب عرض تلك البيانات بأسلوبين مختلفين .

13 - توضح البيانات التالية أعداد الخريجين من الجامعات السورية لدرجة الماجستير والدكتوراه خلال الفترة (2002، 2003، 2004).

أعداد الخريجين (الماجستير والدكتوراه) من الجامعات السورية خلال 2002 - 2004

السنة \ خريج	2002	2003	2004
ماجستير	548	494	448
دكتوراه	79	36	56

المصدر: المجموعة الإحصائية السورية لعام 2005

والمطلوب عرض تلك البيانات بأسلوب الأعمدة البيانية .

14 — توضح البيانات التالية معدل الناتج المحلي الإجمالي في سورية بالأسعار الجارية للفترة (2000 حتى 2004).

السنوات	2000	2001	2002	2003	2004
بالأسعار الجارية	10.4	5.3	6.6	4.4	9.4

والمطلوب :

عرض تلك البيانات بأسلوب الخطوط البيانية .

15 ... يوضح الجدول التالي توزيع أطوال عينة عشوائية من طلبة إحدى الكليات في جامعة دمشق.

فئات الأطوال	165-160	170-165	175-170	180-175	185-180	المجموع
عدد الطلبة	10	20	40	25	5	100

والمطلوب :

1 — ارسم المنحنى التكراري الصاعد والهابط لبيانات الجدول السابق في نفس الشكل .

2 — احسب نسبة الطلاب الذين أطوالهم أقل من 170 سم .

3 — احسب نسبة الطلاب الذين أطوالهم أكثر من 180 سم .

4 — ما هو عدد الطلاب الذين أطوالهم ما بين (165 و 170).

13 ... أخذت عينة عشوائية مكونة من (50) منشأة صغيرة الحجم في إحدى

الدول فكان رأس مال هذه المنشآت على النحو التالي (بملايين الوحدات نقدية).

25	49	43	28	30	38	34	48	39	37
38	28	34	38	48	25	39	37	43	38
39	43	38	43	28	43	37	39	48	30
43	30	39	34	37	39	49	34	43	38
43	25	43	37	39	38	48	43	38	34

والمطلوب: إنشاء جدول تكراري بفئات متساوية الأطوال وبفرض أن : $K=5$ (عدد الفئات)

14 — أخذت عينة عشوائية مكونة من (50) شركة من شركات القطاع الخاص في إحدى الدول، وكان التوزيع التكراري لصادفي الأرباح السنوية لهذه الشركات (بملايين الوحدات النقدية) على النحو التالي:

فئات مصادفي الأرباح	0-5	5-10	10-15	15-20	20-25
عدد الشركات	2	8	28	10	2

والمطلوب:

- 1 — ما هو عدد الشركات التي صادفي أرباحها أقل من (20) مليون.
- 2 — ما هو عدد الشركات التي صادفي أرباحها أكثر من (10) مليون وما هي نسبتهم.
- 3 — ما هو نسبة الشركات التي صادفي أرباحها ما بين (10-15) مليون.
- 15 — في الجدول التكراري المرفق، وجد أن هناك ثلاث مفردات مفقودة، وكانت قيمة هذه المفردات هي (40، 17، 10)؟

الفئات	0-10	10-20	20-30	30-40	40-50
للتكرارات	5	7	13	9	6

والمطلوب: بناء الجدول التكراري بعد إضافة المفردات المفقودة إليه.

- 16 — البيانات التالية تمثل الأنفاق السنوي على وسائل الإعلام في إحدى الدول (بالمليون)

وسيلة الإعلام	الجراند	للتلفزيون	الراديو	المجلات	البريد المباشر
الأفانق	38	80	18	13	3

المطلوب: عرض تلك البيانات بأسلوب الدائرة.

17 - تم مقابلة (500) ناخب فور إدلائهم بأصواتهم في إحدى الدوائر الانتخابية بإحدى المدن العربية، فكان الجدول التالي :

نوع المهنة	ربات منازل	موظفون	عمال	مدرسون	آخرون
العدد	130	250	50	20	50

المطلوب : عرض تلك البيانات بأسلوبين مختلفين.

18 - أخذت عينة عشوائية حجمها (100) امرأة لدراسة العلاقة بين المستوى التعليمي والحالة الزوجية في إحدى المدن، فكانت النتائج الآتية:

المجموع	ثانوية وما بعدها	إعدادية	ابتدائية	تقرأ وتكتب	الحالة التعليمية
65	11	18	12	24	العدد
المجموع	أرملة	مطلقة	متزوجة	عازبة	الحالة الزوجية
65	17	12	14	22	العدد

المطلوب: إعداد جدول إحصائي مزدوج .

19 - أخذت عينة عشوائية من (60) جريمة من الجرائم المسجلة في سجلات البحث الجنائي في إحدى المدن، فكان الجدول التالي :

المجموع	جرائم أخرى	الانتحار	التزوير	الرشوة	القتل	السرقه	نوع الجريمة
60	8	4	7	18	8	15	العدد

المطلوب :

- 1 — احسب نسبة القتل والرشوة والسرقة.
- 3 — عرض بيانات الجدول السابق بأسلوب الدائرة.

20 — أخذت عينة عشوائية من نزلاء السجون حجمها (50) سجيناً لدراسة الحالة التعليمية لهم، فكانت النتائج التالية :

(15) أمي، (10) يقرأ ويكتب، (8) ابتدائي، (12) إعدادي (3) ثانوي، (2) جامعي.

المطلوب:

- 1 — إعداد جدول تفرغ للبيانات السابقة.
- 2 — إعداد جدول تكراري يبين الحالة التعليمية لهؤلاء السجناء .

21 — لدينا القياسات التالية:

30	11	42	8	30	18	25	25	17	30
29	21	23	25	15	35	26	13	21	30

المطلوب: إعداد جدول تكراري في فئات متساوية.



الفصل الثالث

مقاييس النزعة المركزية

— مقاييس النزعة المركزية

- المنوال
- الوسيط
- الوسط الحسابي
- الوسط الهندسي
- الوسط التوافقي

— العلاقة بين المنوال والوسيط والوسط الحسابي



مقاييس النزعة المركزية

3-1 مقدمة :

استعرضنا في الفصل السابق طرائق جمع البيانات عن المجتمع موضع الدراسة طرائق عرضها هذه البيانات جدولياً وبيانياً، وبالرغم من القيمة الكبيرة لطرائق العرض هذه، فإنها عموماً لا تساعدنا إلا في الحصول على انطباع عن الظاهرة المدروسة. ولكن إذا احتاج الأمر إلى صورة أكثر وضوحاً للظاهرة موضع الدراسة نلجأ إلى التوصيف الكمي للبيانات ، حيث نستخدم مقاييس أو مؤشرات كمية محسوبة من البيانات المتاحة بحيث تظهر أهم الخصائص التي نريد معرفتها عن ظاهرة ما في المجتمع الذي ندرسه. مثل النصب المئوية أو التكرارات والتي تعتبر ملائمة لوصف بيانات ذات الطبيعة الأسمية في معظمها، وكذلك الترتيبية منها، وتجدر الإشارة هنا، إلى أن جميع مقاييس النزعة المركزية والنشئت لا تصلح إلا للبيانات التي تصف متغيرات يمكن قياسها بمقاييس من نوع ذي مسافات متساوية. أما البيانات الترتيبية كمدخل الأسرة الشهري وعدد أفراد الأسرة وغيرها، فإنه بالإمكان استخدام بعض مقاييس النزعة المركزية والنشئت في وصف تلك البيانات.

3-2 مقاييس النزعة المركزية:

من المعروف أن القيم التي تأخذها الظواهر محل الدراسة غالبيتها قريبة من بعضها البعض حيث نجد أن عدداً كبيراً من تلك القيم يميل إلى التجمع حول قيمة معينة تقع في وسط البيانات وتعمل على جذب القيم إليها وكأن هناك نزعة عند البيانات للتجمع حول تلك القيمة، ويقل عدد البيانات تدريجياً كلما ابتعدت البيانات عن تلك القيمة المتوسطة، لذلك سميت بالنزعة المركزية، كونها تتوسط تجمع البيانات ومن خصائص المتوسط الجيد:

أ — أن يكون معرفاً بشكل دقيق وقيمه تتوقف على الأعداد المستخرج منها.

ب - أن يأخذ بالحسبان جميع القيم بالمجموعة.

ج - أن يكون سهلاً وسريعاً في الحساب.

د - أن لا يتأثر كثيراً بالمتقلبات في قيم مفردات العينة.

هـ - أن يخضع للعمليات الجبرية.

وهناك عدة أنواع من المتوسطات ونحن سنكتفي هنا بثلاثة مقاييس للنزعة المركزية تجمع بين الشبوع وبساطة الحساب ووضوح المعنى في كل منها .

• المنوال Mode :

يعرف المنوال: "بأنه القيمة الأكثر شيوعاً (أو الأكثر تكراراً) في البيانات، موضوع الدراسة".

حساب المنوال لبيانات غير مبوبة:

لدينا مجموعة البيانات التالية :

119 118 115 119 116 123 115 114

نجد أن قيمة المنوال هي 115 إذ (ظهرت مرتين).

حساب المنوال لبيانات مبوبة:

إذا كانت البيانات مبوبة، فإنه لا يمكن إيجاد قيمة المنوال بمجرد النظر، وإنما يمكن افتراض أن قيمة المنوال هي إحدى قيم المتغير داخل الفئة ذات أكبر تكرار والتي تسمى بالفئة المنوالية ولتوضيح ذلك لدينا التمرين التالي:

مثال: : يوضح الجدول التكراري التالي توزيع عينة عشوائية من 200 مرشح لمجلس الإدارة المحلية تبعاً لفئات الأعمار في إحدى الدول.

فئات الأعمار	40-35	45-40	50-45	55-50	60-55	65-60	المجموع
المرشحين	20	35	80	43	16	6	200

نلاحظ أن أكبر تكرار في هذا الجدول هو 80 وهو المقابل للفئة 45 أقل من 50 وتسمى هذه الفئة بالفئة المنوالية، ومن ثم يمكن القول: أن المنوال هو أحد قيم

هذه الفئة، ولأنه من الصعب الوصول إلى القيمة الحقيقية للمنوال، فإننا نلجأ إلى تقديرها بإحدى الطرق التالية:

I - مركز الفئة المنوالية :

أسهل تقريب لقيمة المنوال هو مركز الفئة المنوالية، وفي مثالنا هذا، يكون المنوال طبقاً لهذه الطريقة هو 47.5 ولكن هذه الطريقة لا تراعي شكل التوزيع وتفضل عليها طريقة الفروق وطريقة العزوم، وسوف نتناول طريقة الفروق باعتبارها الأكثر شيوعاً واستخداماً.

II - طريقة الفروق :

تعتمد طريقة الفروق على تكرار الفئة المنوالية وتكرار ما يسبقها وما يليها، كأساس لحساب المنوال، ويتم حساب المنوال بأسلوب الفروق من خلال العلاقة التالية :

$$Mod = L_1 + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2} \right) \cdot C$$

حيث أن :

L_1 : الحد الأدنى للفئة المنوالية.

Δ_1 : الفرق بين تكرار الفئة المنوالية وتكرار الفئة التي تسبقها .

Δ_2 : الفرق بين تكرار الفئة المنوالية وتكرار الفئة اللاحقة.

C : طول الفئة المنوالية .

لو عدنا لبيانات المثال السابق والمتعلق بأعمار المرشحين، فإن منسوال أعمار

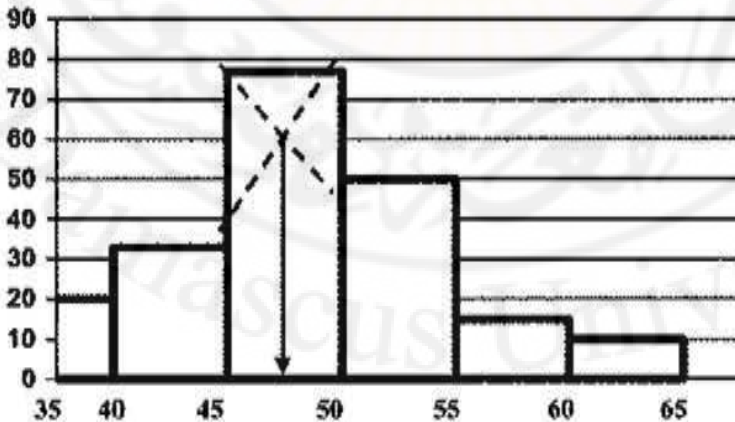
هؤلاء المرشحين هو :

$$\begin{aligned} Mod &= 45 + \left(\frac{80 - 35}{(80 - 35) + (80 - 43)} \right) \cdot 5 \\ &= 45 + \left(\frac{45}{45 + 37} \right) \cdot 5 \\ &= 45 + 2.70 = 47.70 \quad \text{سنة} \end{aligned}$$

التفسير: إن العمر الأكثر شيوعاً أو تكراراً بين هؤلاء المرشحين هو 47.70 وبالتقريب (48) سنة تعطينا هذه الطريقة أحسن تقريب لقيمة المنوال ويمكن الوصول إلى نفس القيمة بيانياً كما يلي :

تحديد المنوال بيانياً:

ويتم من خلال رسم المدرج التكراري وتحديد الفئة المنوالية ومن ثم نقوم بمتوصيل نقطة الحد الأعلى للفئة السابقة بالحد الأعلى للفئة المنوالية والحد الأدنى للفئة المنوالية بالحد الأدنى للفئة اللاحقة لها فإن العمود النازل من نقطة التقاطع يحدد قيمة المنوال، وبالعودة إلى التمرين السابق، نجد من الرسم أن قيمة المنوال هي 48 سنة تقريباً (وبالدقة تساوي 47.70).



ملاحظة :

عند إيجاد قيمة المنوال من جدول تكراري ذي فئات غير متساوية يجب أولاً التخلص من أثر اختلاف أطوال الفئات أي أنه يجب أولاً الحصول على قيمة التكرارات المعدلة ثم استخدامها لإيجاد المنوال بالطريقة السابقة، ويمكن حساب التكرار المعدل بالعلاقة التالية:

$$fc = \frac{f}{C} \quad (\text{التكرار المعدل})$$

— خصائص المنوال :

- 1 — قيمة المنوال على المحور (x) تقابل أعلى نقطة على المنحني التكراري، لذلك يستخدم معيار لتحديد طبيعة التوزيع.
 - 2 — لا يتأثر بالقيم المتطرفة، لأنه يعتمد في حسابه على الفئتين السابقة واللاحقة.
 - 3 — يمكن حسابه في التوزيعات التكرارية المفتوحة .
 - 4 — طرق حسابه تقريبية، يعطي نتائج مختلفة عند حسابه من جدول تكراري واحد، لأن قيمة المنوال تتوقف على كيفية إعداد الجدول التكراري سواء من حيث طول الفئة وحدودها لذلك فهو مقياساً غير مستقراً للترعة المركزية.
 - 5 — قليل الاستخدام بالمقارنة مع الوسط الحسابي في الدراسات الإحصائية.
- ملاحظة: في بعض الأحيان قد يكون لبعض البيانات أكثر من منوال واحد، وهذا دليل على أن هذه البيانات إما أن تكون غير متجانسة أو أنها قد سحبت من مجتمع إحصائي آخر.

• الوسيط Med :

الوسيط هو القيمة التي تقع في منتصف التوزيع للقيم، بحيث نصف عدد القيم أصغر منها والنصف الآخر أكبر منها، أي يكون عدد المفردات الأصغر

منها مساوياً لعدد المفردات الأكبر منها، وذلك عند ترتيب القسم تصاعدياً أو تنازلياً.

• حساب الوسيط لبيانات غير مبوبة:

لإيجاد قيمة الوسيط يجب أولاً ترتيب البيانات ترتيباً تصاعدياً أو تنازلياً، ثم نحدد مكان المفردة الوسيطة ويسمى ترتيب الوسيط. بعدها نوجد قيمة الوسيط.

لإيجاد الوسيط من مجموعة البيانات التالية :

119 118 115 119 115 116 123 115 114

نبدأ بترتيبها تصاعدياً أي :

123 119 118 118 116 115 115 115 114

ويلاحظ هنا أن ترتيب الوسيط هو (5)، حيث أن هناك (4) قيم أصغر منها و (4) قيم أكبر منها وبالتالي فإن قيمة الوسيط هي قيمة المفردة رقم (5) في البيانات المرتبة أي أن: الوسيط = 116.

يلاحظ في هذا المثال أن عدد المفردات هو 9 وأن ترتيب الوسيط هو (5) يمكن الحصول عليه على الصورة من العلاقة $\frac{n+1}{2}$ أي $\frac{1+9}{2}$ وإذا حذفنا القيمة الأخيرة وهي (123) من البيانات أصبح عدد المفردات (8) ونجد أنه في هذه الحالة يقع الوسيط بين قيمة المفردتين رقم 4 و 5، إذ أن ترتيب الوسيط في هذه الحال $\frac{1+8}{2} = (4.5)$ أي بين القيمتين (116, 115) فيكون الوسيط $\frac{116+115}{2} = (115.5)$ ويمكن تلخيص خطوات حساب الوسيط في الخطوات

الثلاث التالية:

1 - ترتيب القيم تصاعدياً أو تنازلياً.

2 - إيجاد ترتيب الوسيط $= \frac{n+1}{2}$ ، حيث أن n عدد المفردات .

3 - إيجاد قيمة الوسيط وهو قيمة المفردة المقابلة لترتيب $= \frac{n+1}{2}$ ، فيما إذا

كان عدد المفردات فردياً، أما إذا كان عددها زوجياً فيؤخذ متوسط المفردتين اللتان يقع بينهما.

— حساب الوسيط لبيانات مبوبة (جدول تكراري)

ويتم على النحو التالي :

- 1 — تكوين الجدول التكراري المتجمع الصاعد (أو الهابط).
- 2 — إيجاد ترتيب الوسيط وهو $\frac{\sum f}{2}$ ثم البحث عن القيمة $\sum f/2$ ما يساويها أو أكبر منها مباشرة لتحديد الفئة الوسيطة.
- 3 — إيجاد قيمة الوسيط، ويتم ذلك من خلال تطبيق العلاقة التالية:

$$Med = L_1 + \left(\frac{\sum f/2 - \sum f_i}{f_{me}} \right) \cdot C$$

حيث أن :

L_1 : الحد الأدنى للفئة الوسيطة .

$\sum f_i$: مجموع التكرارات التي تسبق تكرار الفئة الوسيطة f_{me}

f_{me} : تكرار الفئة الوسيطة

C : طول الفئة الوسيطة .

وبالعودة إلى التمرين السابق والمتعلق بالمرشحين لمجالس الإدارة المحلية نجد أن وسيط أعمار هؤلاء المرشحين هو :

$$\sum f/2 = \frac{200}{2} = 100$$

$$Med = 45 + \left(\frac{100 - 55}{80} \right) \cdot 5$$

$$= 45 + 2.81 = 47.81 \text{ سنة}$$

الفئات	f	ت.ت.ص
40-35	20	20
45-40	35	55
50-45	80	135
55-50	43	178
60-55	16	194
65-60	6	200

التفسير : إن نصف المرشحين كانت أعمارهم أقل من (47.81) سنة وإن نصف المرشحين الآخر كانت أعمارهم أكبر من هذا العمر .

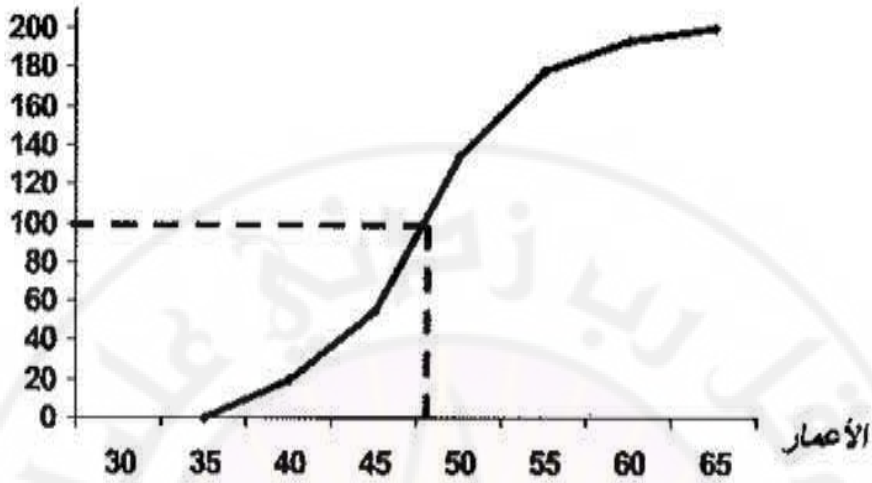
II - تحديد الوسيط بيانياً:

نرسم المنحنى المتجمع الصاعد (أو الهابط) ثم نوجد القيمة $(\sum f/2)$ على المحور العمودي (محور ت.ت.ص أو ت.ت.هـ) ثم نرسم من النقطة $(\sum f/2)$ خطاً أفقياً موازياً لمحور الفئات إلى أن يلاقي المنحنى الصاعد أو الهابط ثم نسقط هذه النقطة على محور الفئات وتكون هي قيمة الوسيط.

مثال : اوجد الوسيط بيانياً لبيانات الجدول السابق والمتعلق بأعمار المرشحين لمجالس الإدارة المحلية.

حدود الفئات	التكرار المتجمع الصاعد
أقل من 35	—
أقل من 40	20
أقل من 45	55
أقل من 50	135
أقل من 55	178
أقل من 60	194
أقل من 65	200

عدد
المرشحين



نوجد ترتيب الوسيط من العلاقة $\frac{\sum f}{2} =$ فيكون ترتيب الوسيط $100 = \frac{200}{2}$ من الرسم نوجد القيمة المقابلة لتكرارات 100 فتكون قيمة الوسيط هي (47.81) سنة.

ملاحظة : يمكن تحديد قيمة الوسيط بيانياً أيضاً من خلال رسم منحنى التكرار التجميعي الهابط أو الالئيين معاً (الصاعد والهابط) نقطة التقاطع وتسقط علسي المحور الأفقي فتكون قيمة الوسيط .

• خصائص الوسيط :

- 1 - يمكن حسابه من الجداول التكرارية المفتوحة .
- 2 - لا يتأثر بالقيم المتطرفة.
- 3 - قيمته غير ثابتة عندما يكون حجم العينة صغيراً.

ملاحظة : يفضل استخدام الوسيط:

- 1 - إذا أردنا إيجاد قيمة ممثلة، أي قيمة نموذجية بدلاً من الاهتمام بالمجموع الكلي.

2 - إذا فقدت بعض القيم وعرف ترتيبها (حيث لا يمكن حساب الوسط الحسابي مباشرة)

ونشير هنا إلى أنه يفضل استخدام الوسيط في البيانات التي نرغب بمعرفة ترتيبها ولا نعرف قيمها، وكذلك في البيانات الناقصة، أي في البيانات التي فقدت أو سقطت قيمها بعض المفردات ولكننا نعرف أين تقع، وأخيراً أن قيمة الوسيط تنجذب باتجاه القيم المتطرفة وتكون أقرب إلى نقطة التركز العظمى بالمقارنة مع الوسط الحسابي، وذلك إذا كان مقدار الالتواء شديداً.

• الوسط الحسابي :

يعرف الوسط الحسابي لمجموعة من القيم بأنه: "عبارة عن حاصل جمع هذه القيم مقسوماً على عددها"، أي القيمة التي لو أعطيت لكل مفردة فسيان مجموعها لا يتغير.

وبصورة عامة إذا كان المتغير x يأخذ القيم $x_1, x_2, x_3, \dots, x_n$ فإن الوسط الحسابي لهذه القيم يرمز له بالرمز $\bar{x}$ ويعبر عنه على أنه :

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$

ويمكن كتابة ذلك في صورة مختصرة :

$$\bar{x} = \frac{\sum x}{n}$$

(حيث n عدد المفردات الظاهرة أو المتغير x)

مثال : أوجد الوسط الحسابي لأجور (5) عمال من عمال موزعي الصحف والمجلات كانت أجورهم الأسبوعية على النحو التالي:

2150 2520 1950 1875 2125 (ليرة سورية)

الأسلوب المباشرة :

ويتم من خلال تطبيق العلاقة التالية : $\bar{x} = \frac{\sum x}{n}$

$$= \frac{1}{5}(2150 + 2520 + \dots) = \frac{10620}{5} = 2124 \text{ ل.س}$$

أي الوسط الحسابي لأجر العامل = 2124 ليرة، بمعنى لو تساوت أجور العمال فيما بينهم لكان أجر العامل وسطياً (2124 ليرة سورية).

ملاحظة :

هناك أسلوب آخر لحساب الوسط الحسابي لبيانات غير مبوبة، ونحن تناولنا الأسلوب المباشر، علماً أن كلا الأسلوبين يعطيان نفس النتائج.

— حساب الوسط الحسابي لبيانات مبوبة (جدول تكراري):

ويتم ذلك على النحو التالي:

— نحسب مراكز الفئات

— نحسب (حاصل جداء مراكز الفئات بالتكرارات المقابلة لها ولكل فئة،

أي $(f \cdot x)$:

$$\bar{X} = \frac{\sum f \cdot x}{\sum f}$$

ولتوضيح ذلك لدينا التمرين التالي:

تمرين : أخذت عينة عشوائية مكونة من (200) عامل من عمال إحدى مؤسسات الإعلان التابعة لوزارة الإعلام في إحدى الدول، فكانت أجور العمال الأسبوعية موزعة في الجدول التكراري التالي (بالدولار)

فئات الأجور	40-20	60-40	80-60	100-80	120-100	140-120	المجموع
عدد العمال	9	22	68	64	27	10	200

الحل :

من تعريف الوسط الحسابي نعلم أن :

مجموع الأجور الشهرية

$$\bar{X} = \frac{\sum fx}{\sum f} = \frac{\text{الوسط الحسابي}}{\text{عدد العمال}}$$

(أجر العامل في الشهر)

حيث أن (x) مركز الفئات و (f) التكرارات.

من الجدول نلاحظ أن هناك (9) عمال كل منهم أجره يتراوح بين (20)، (40)

دولار، أي أن كل منهم أجره في المتوسط (30) وحدة $= \frac{40+20}{2}$ (وهو

مركز الفئة الأولى)، وبالتالي فإن الأجور في هذه الفئة الأولى هي

(30)، (90) = 270 دولار، كذلك نجد أن هناك (22) عاملاً أجرهم يتراوح بين

(40)، (60) دولار أي أن أجر العامل منهم هو (50) في المتوسط

$= \frac{60+40}{2}$ أي مركز (الفئة الثانية) فيكون بذلك مجموع الأجور في الفئة

الثانية هو (50)(22) = 1100.

وبالمثل يمكن إيجاد باقي فئات الجدول، كما الآتي:

فئات الأجور	عدد العمال (f)	مركز الفئات (x)	fx
20 - 40	9	30	270
40 - 60	22	50	1100
60 - 80	68	70	4760
80 - 100	64	90	5760
100 - 120	27	110	2970
120 - 140	10	130	1300
المجموع	200		16160

ويكون الوسط الحسابي $= \frac{16160}{200} = 80.8$ دولار أي أن متوسط الأجور هو (81)

دولار تقريباً.

ملاحظات :

- لإيجاد قيمة الوسط الحسابي في حالة جدول تكراري ذي فئات غير متساوية، فإننا نبتع نفس الطريقة كما في الأمثلة السابقة (الفئات المتساوية) دون أي اختلاف.

- هناك عدة أساليب لحساب الوسط الحسابي لبيانات مبوبة، ونحن نتاولنا الأسلوب المباشر، باعتبار جميع الأساليب تعطي نفس النتائج.

- في حال كان لدينا جدول تكراري يتضمن توزيع عينتين معاً أو توزيع تكراري وعينة أخرى معلوم منها الوسط الحسابي، ورغبنا في حساب متوسط العينتين معاً، فلنح في هذه الحالة نميز بين حالتين :

أ - إذا كانت $(n_1 = n_2)$ فإن الوسط الحسابي للعينتين معاً يكون على النحو التالي:

$$\bar{X} = \frac{\bar{X}_1 + \bar{X}_2}{2}$$

ب - إذا كانت $(n_1 \neq n_2)$ فإن الوسط الحسابي للعينتين معاً يكون على النحو التالي:

$$\bar{X} = \frac{\bar{X}_1 \sum f_1 + \bar{X}_2 \sum f_2}{\sum f_1 + \sum f_2}$$

مثال : لدينا البيانات الآتية:

$$n_1 = 40$$

$$\bar{x}_1 = 37.5$$

$$n_2 = 60$$

$$\bar{x}_2 = 35.5$$

المطلوب : حساب الوسط الحسابي للعينتين معاً

$$\bar{X} = \frac{37.5 + 35.5}{2} = 36.5$$

مثال : لدينا البيانات الآتية :

$$n_1 = 40 \quad \bar{x}_1 = 37.5$$

$$n_2 = 60 \quad \bar{x}_2 = 30.5$$

المطلوب: حساب الوسط الحسابي للعينتين معاً

$$\bar{X} = \frac{(37.5)(40) + (30.5)(60)}{40 + 60} = 33.3$$

... الوسط الحسابي المرجح :

في بعض الأحيان يكون من المهم ترجيح بعض القيم $x_1, x_2, \dots, x_n$ بأوزان $w_1, w_2, \dots, w_n$ وهذه تعتمد على الأهمية المرتبطة بهذه القيم فسي الدراسة ويعرف الوسط الحسابي هنا بالوسط الحسابي المرجح ويحسب بالعلاقة التالية:

$$\bar{x} = \frac{x_1 w_1 + x_2 w_2 + \dots + x_n w_n}{w_1 + w_2 + \dots + w_n} = \frac{\sum x w}{\sum w}$$

مثال :

إذا أعطى الامتحان النهائي في مقرر الإحصاء وزناً يساوي ثلاثة أمثال الامتحانات الشهرية، وإذا حصل طالب في الامتحان النهائي على 85 درجة وفي الامتحانات الشهرية على 70 و 90 درجة، فإن متوسط درجاته في الإحصاء يكون :

$$\bar{x} = \frac{(70)(1) + (90)(1) + (85)(3)}{1 + 1 + 3} = 83 \text{ درجة}$$

مثال : إذا كان لدينا ثلاث عينات : $n_1=15$ $n_2=20$ $n_3=25$

$$\bar{x}_1 = 45 \quad \bar{x}_2 = 75 \quad \bar{x}_3 = 60 \text{ أوساطها}$$

الحسابية.

وقد دمجت هذه العينات الثلاث معاً، والمطلوب: أوجد الوسط الحسابي لهذه المجموعات بعد الدمج .

$$\bar{x} = \frac{(\bar{x}_1)(n_1) + (\bar{x}_2)(n_2) + (\bar{x}_3)(n_3)}{n_1 + n_2 + n_3}$$

$$\bar{x} = \frac{(15)(45) + (20)(75) + (25)(60)}{60} = \frac{3675}{60} = 61.25$$

• خصائص الوسط الحسابي :

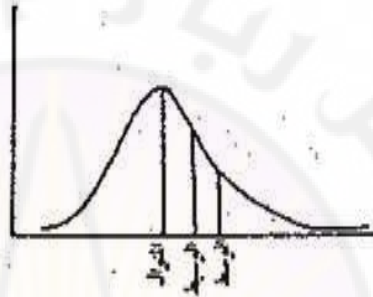
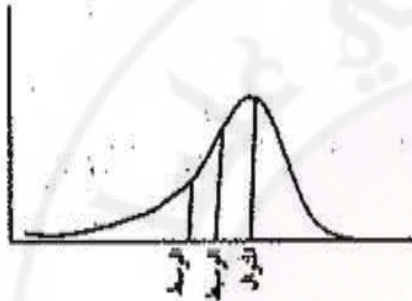
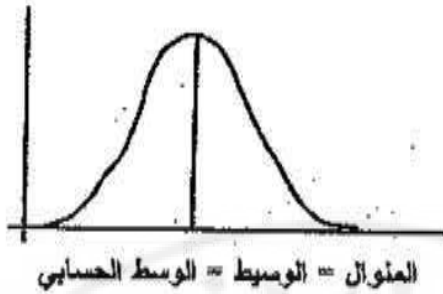
- 1 — يعتبر الوسط الحسابي من أكثر المقاييس أمانة ومصداقية وتمثيلاً لمفهوم النزعة المركزية، لأنه يأخذ بالحسبان جميع القيم عند حسابه.
- 2 — يتأثر بالقيم المتطرفة على جانبي التوزيع.
- 3 — صعوبة حسابه من جدول تكراري مفتوح .
- 4 — معيار للمقارنة.
- 5 — أقل تذبذباً من بقية المتوسطات في تقدير متوسط المجتمع.
- 6 — يفضل استخدامه في الحالات التي يكون فيها التوزيع متماثلاً أو قريباً من التماثل، وإذا كان الاهتمام منصباً على القيمة العددية لجميع البيانات.
- 7 — الوسط الحسابي مقياس قيم.

3-4 : العلاقة بين المنوال والوسيط والوسط الحسابي :

يلاحظ أنه في التوزيعات التكرارية المتماثلة (الشكل المرفق) تتطابق قيم كل من المنوال والوسيط والوسط الحسابي، ومن الناحية التطبيقية يندر وجود التوزيعات المتماثلة، فقد نحصل على توزيعات ملتوية نحو اليمين أو نحو اليسار، وفي التوزيعات التكرارية وحيدة المنوال المتماثلة والقريبة من التماثل نتحقق العلاقة التجريبية التالية:

$$\bar{x} - Mod = 3(\bar{x} - Med)$$

أما في التوزيعات التكرارية الملتوية التواء شديداً فلا تتحقق هذه العلاقة. الأشكال المرفقة توضح الموضع النسبي للوسط والوسيط والمنوال للمنحنيات التكرارية الملتوية إلى اليمين والملتوية إلى اليسار على الترتيب.



تستخدم هذه العلاقة لإيجاد قيمة أي مقياس مجهول، إذا كان معلوماً لدينا قيمة مقياسين، لاسيما الوسط الحسابي عندما تكون نهايتي أو إحدى نهايتي التوزيع مفتوحة.

• الوسط الهندسي :

يعرف الوسط الهندسي لمجموعة من القيم $x_1, x_2, \dots, x_n$ على أنه الجذر النوني لحاصل ضرب هذه القيم بعضها ببعض أي:

$$\text{الوسط الهندسي} = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdot \dots \cdot x_n}$$

أغلب استخداماته في حساب معدلات النمو ومعدلات الفائدة وحساب الأرقام القياسية .

• خصائص الوسط الهندسي:

- 1- يتمتع بخواص رياضية
- 2- يعطي قيمة أكثر تمثيلاً من الوسط الحسابي عند التعامل مع توزيعات شديدة الالتواء نحو اليمين.

• الوسط التوافقي :

الوسط التوافقي لمجموعة من القيم هو مقلوب الوسط الحسابي لمقلوبات هذه القيم، أي أن:

$$\text{الوسط التوافقي} = \frac{n}{\sum \frac{1}{x}}$$

أغلب استخداماته في مجال إنتاجية العمل والأسعار والعلاقة الثلاثية (السرعة، الزمن، المسافة) ويتميز: بأخذ جميع القيم بالحسبان عند حسابه.
عيوبه:

— لا يمكن حسابه من جدول تكراري مفتوح.



الفصل الرابع

مقاييس التشتت

— مقدمة

— مقاييس التشتت (المدى المطلق، نصف المدى الربيعي، الانحراف

المتوسط، الانحراف المعياري)

— قياس الالتواء

— قياس التفرطح

— مفهوم الدرجة المعيارية

— أهمية التوزيع الطبيعي



مقاييس التشتت

1-4- مقدمة:

لقد استعرضنا كيفية تحليل البيانات الإحصائية حسابياً عن طريق المتوسطات لكن هذه المتوسطات (مقاييس الموضع) لا تعطي وصفاً كاملاً للبيانات، إذ أنها تبين القيمة التي تتركز حولها بقية قيم العينة فقط ولا تعطي أي إيضاح عن مدى تباعد أو تقارب باقي القيم من هذه القيمة، أي أنها لا توضح مدى التجانس في البيانات عن قيمة النزعة المركزية، أو مدى تشتتها فمثلاً إذا كان لدينا المجموعتان التاليتان من القيم:

14، 0، 7

7، 6، 8

فإن المتوسط لكل منهما يساوي (7)، فإذا اكتفينا بهذا المقياس فإننا نقرر أن المجموعتين متشابهتان، لكن في الحقيقة أن قيم المجموعة الأولى أكثر تباعداً من قيم المجموعة الثانية.

لذا فإننا نحتاج إلى مقياس آخر يوضح مدى تقارب أو تباعد البيانات وهذه المقاييس تسمى بمقاييس التشتت ومنها: المدى المطلق، نصف المدى الربيعي — الانحراف المتوسط — الانحراف المعياري.

• المدى المطلق (R):

هو الفرق بين أكبر وأصغر قيمة في مجموعة البيانات، ويعطى بالعلاقة التالية:

$$R = X_{max} - X_{min}$$

ومن عيوب هذا المقياس أنه يعتمد على قيمتين فقط من البيانات ويتجاهل باقي القيم، وأغلب استخدامات هذا المقياس في وصف الأحوال الجوية وقياس تشتت الأسعار اليومية للسندات والأسهم والأوراق المالية ومراقبة جودة الإنتاج .

خصائص المدى :

- مقياس مضلل في حالة وجود قيم شاذة.
- يتأثر بالقيم الحدية فقط.
- لا يمكن حسابه من جدول تكراري مفتوح
- يستخدم في رسم الخرائط الإحصائية لمراقبة جودة الإنتاج.
- المدى النسبي ($R\%$):
- مقياس من مقاييس التشتت النسبية يعطى بالعلاقة التالية:

$$R\% = \frac{R}{X} * 100$$

• نصف المدى الربيعي (Q.D) :

يعرف على أنه: "نصف الفرق بين أكبر وأصغر قيمة من القيم المتبقية بعد استبعاد الربع الأدنى والربع الأعلى من البيانات (الفرق بين الربيع الثالث والربيع الأول مقسوماً على (2))."

حساب نصف المدى الربيعي (Q.D) لبيانات غير مبوبة:
للحصول على نصف المدى الربيعي نتبع الخطوات التالية :
(1) ترتيب البيانات ترتيباً تصاعدياً (أو تنازلياً).

(2) تحديد ترتيب (موضع) القيمة التي تقع في آخر الربع الأول : $K_1 = \frac{n+1}{4}$

(3) إيجاد قيمة (Q_1).

(4) تحديد ترتيب القيمة التي تقع في آخر الربع الثالث : $K_3 = \frac{3(n+1)}{4}$

(5) إيجاد قيمة Q_3

(6) $Q.D = \frac{Q_3 - Q_1}{2}$ نصف المدى الربيعي

حساب المدى الربيعي لبيانات مبوبة :

إن حساب الربيعيات لبيانات مبوبة مشابهة تماماً لطريقة حساب الوسيط، ويتم ذلك على النحو التالي :

... إنشاء عمود التكرار التجميعي المساعد.

— تحديد ترتيب الربع الأول من خلال $(\sum f/4)$ لتحديد فئة الربع الأول

— نبحث عن قيمة $(\sum f/4)$ في عمود (ت.ت.ص) ما يساويها أو أكبر منها مباشرة فتكون الفئة المقابلة "فئة الربع الأول".

— نطبق القانون:

$$Q_1 = L_1 + \left(\frac{\sum f/4 - \sum f_1}{f_{Q_1}} \right) \cdot C$$

حيث أن :

L_1 : تشير إلى الحد الأدنى لفئة الربع الأول .

$\sum f/4$: ربع مجموعة التكرارات (ترتيب الربع الأول) .

$\sum f_1$: مجموع التكرارات السابقة لتكرار فئة الربع الأول .

f_{Q_1} : تكرار فئة الربع الأول .

C : تكرار فئة الربع الأول.

— لتحديد فئة الربع الثالث نتبع نفس الخطوات السابقة مع ملاحظة أن ترتيب

Q_3 يصبح $(3\sum f/4)$ ونطبق القانون التالي:

$$Q_3 = L_3 + \left(\frac{3\sum f/4 - \sum f_1}{f_{Q_3}} \right) \cdot C$$

حيث أن :

L_3 : نشير إلى الحد الأدنى لفئة الربع الثالث .

$3\sum f/4$: ترتيب الربع الثالث .

$\sum f$: مجموع التكرارات السابقة لتكرار فئة الربع الثالث.

h : تكرار فئة الربع الثالث .

وبالتطبيق على المثال السابق المتعلق بأعمار المرشحين للمجالس المحلية (بحث الوسيط) نجد أن:

$$\sum f/4 = \frac{200}{4} = 50$$

$$3\sum f/4 = \frac{3(200)}{4} = 150$$

الفئات	f	ت.ت.ص
40-35	20	20
45-40	30	55
50-45	80	135
55-50	43	178
60-55	16	194
65-60	6	200

ل الربع الأول :

$$Q_1 = 40 + \left(\frac{50 - 20}{30} \right) \cdot 5$$

$$40 + 5 = 45 \text{ سنة}$$

التفسير : أن 25% من المرشحين أعمارهم كانت أقل (45) سنة وأن 75% من هؤلاء المرشحين كانت أعمارهم أكبر من (45) سنة.

الربيع الثالث :

$$Q_3 = 50 + \left(\frac{150 - 135}{43} \right) \cdot 5$$

$$= 50 + 1.70 = 51.70 \text{ سنة}$$

التفسير : أن 75% من المرشحين أعمارهم كانت أقل من (51.70) سنة وأن 25% من هؤلاء المرشحين كانت أعمارهم أكبر من (51.70) سنة .
نصف المدى الربيعي :

$$Q.D = \frac{51.7 - 45}{2} = 3.35$$

— خصائص الانحراف الربيعي (نصف المدى الربيعي) :

- 1 — يمكن حسابه من التوزيعات التكرارية المفتوحة
- 2 — مقياس النزعة المركزية المقابل له هو الوسيط .
- 3 — هو عبارة عن فترة تحتوي على (50%) من البيانات.

— الانحراف الربيعي النسبي ($Q.D\%$) :

يعتبر من مقاييس التشتت النسبية له أهمية في التحليل الإحصائي

لمقارنة تشتت توزيعين مختلفين، ويعطى بالعلاقة التالية :

$$Q.D\% = \frac{Q.D}{Med} \cdot 100$$

أو

$$Q.D\% = \frac{Q_3 - Q_1}{2Med} \cdot 100$$

وبالتطبيق على التمرين السابق والمتعلق بأعمار عينة عشوائية من المرشحين نجد أن :

$$Med = 47.81, Q_1 = 45, Q_3 = 51.70, Q.D = 3.35$$

$$Q.D\% = \frac{3.35}{47.81} \cdot 100 = 7\%$$

وبفرض أخذت عينة عشوائية من حجم مماثل من المرشحين في دولة مجاورة وكان الوسيط للأعمار (40) سنة وانحراف ربعي مقداره (4)، سنة والمطلوب :

أيهما يعرض تشتتاً أكبر أعمار المرشحين في هذه الدولة أم أعمار المرشحين في الدولة المجاورة؟ برر إجابتك إحصائياً. $Q.D\% = \frac{4}{40} \cdot 100 = 10\%$
(الدولة المجاورة)

من خلال المقارنة نجد أن $Q.D\% (10\%) > Q.D\% (7\%)$ وبالتالي يمكن القول: أن أعمار المرشحين في الدولة المجاورة تعرض تشتتاً أكبر من أعمار المرشحين لهذه الدولة .

• الانحراف المتوسط أو متوسط الانحرافات (M.D) :

يعرف الانحراف المتوسط لمجموعة القيم $x_1, x_2, \dots, x_n$ كالتالي:

$$M.D = \frac{\sum |xi - \bar{x}|}{n}$$

أي أن متوسط القيم المطلقة لانحرافات القيم عن وسطها الحسابي (القيمة المطلقة لرقم هي الرقم بإشارة موجبة ويعبر عنها بخطتين حول الرقم فتكون القيمة المطلقة للرقم)

مثال : أوجد الانحراف المتوسط لمجموعة القيم :

2 3 6 8 11

الحل :

$$x = \frac{2+3+6+8+11}{5} = 6 \text{ الوسط الحسابي}$$

$$\text{الانحراف المتوسط} = \frac{1}{5} [|2-6| + |3-6| + \dots] = \frac{4+3+0+2+5}{5} = 2.8$$

إذا كانت القيم $x_1, x_2, \dots, x_n$ تحدث بتكرارات $f_1, f_2, \dots, f_n$ فالانحراف المتوسط يمكن كتابته على صورة

$$M.D = \frac{1}{n} \sum f |x_i - \bar{x}|$$

ملاحظة: كلما كانت قيمة الانحراف المتوسط صغيرة، دل ذلك على اقتراب المفردات من الوسط الحسابي وبالعكس.

ملاحظة: عندما يكون التوزيع بعيداً عن التماثل، يفضل حساب الانحرافات عن الوسيط، لأن مجموع الانحرافات المطلقة عن الوسيط دوماً أقل من قيمة معينة.

مزاياه:

— يتأثر بالقيم الشاذة.

— مقياس تشتت مرضي بالمقارنة مع المدى والانحراف الربيعي

— سهل الحساب

عيوبه:

— عدم اتصافه بالمفهوم الجبري

— لا يمكن حسابه من جدول تكراري مفتوح

— قليل الاستخدام في الاستدلال الإحصائي لعدم قابليته للعمليات الجبرية.

• الانحراف المعياري (S):

هو الجذر التربيعي لحاصل قسمة مربعات انحرافات القيم عن وسطها

الحسابي على عدد القيم ومربعه هو عبارة عن التباين.

يعتبر الوسط الحسابي والانحراف المعياري من أكثر المقاييس الإحصائية أهمية

في التحليل الإحصائي فالوسط الحسابي يحدد لنا مركز التوزيع الطبيعي

والانحراف المعياري يحدد لنا كيفية الانتشار على جانبي الوسط الحسابي.

• حساب الانحراف المعياري لبيانات غير مبوبة :

— الأسلوب المباشر :

$$S = \sqrt{\frac{\sum (xi - \bar{x})^2}{n}}$$

$$\bar{x} = \frac{\sum xi}{n} \text{ : حيث أن}$$

ويمكن تبسيط العلاقة السابقة فنحصل على الصيغة التالية :

$$S = \sqrt{\frac{\sum xi^2}{n} - \left(\frac{\sum xi}{n}\right)^2}$$

— حساب الانحراف المعياري لبيانات مبوبة (جدول تكراري):

1 — الأسلوب المباشر :

ويتم من خلال العلاقة التالية :

$$S = \sqrt{\frac{\sum fi(xi - \bar{x})^2}{\sum f}}$$

2 — أسلوب استخدام مراكز الفئات :

$$S = \sqrt{\frac{\sum fixi^2}{\sum fi} - \left(\frac{\sum fixi}{\sum fi}\right)^2}$$

علماً أن جميع هذه الأساليب لحساب الانحراف المعياري تعطي نفس النتائج وبالتطبيق على المثال في بحث الوسط الحسابي والمتعلق بأجور العمال الشهرية في إحدى مؤسسات الإعلان نجد:

فئات أجور العمال الشهرية	عدد العمال	مراكز الفئات (x)	fx	fx ²
40-20	9	30	270	8100
60-40	22	50	1100	55000
80-60	68	70	4760	333200
100-80	64	90	5760	518400
120-100	27	110	2970	326700
140-120	10	130	1300	169000
المجموع	200		16160	1410400

$$S = \sqrt{\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2}$$

$$S = \sqrt{\frac{1410400}{200} - (80.8)^2} = 22.87$$

• خصائص الانحراف المعياري :

- 1 — ليس له مدلول إلا عند مقارنته بانحرافات معيارية لمجموعات أخرى من القيم.
- 2 — لا يمكن حسابه من جداول تكرارية مفتوحة .
- 3 — يتأثر بضرب وقسمة المفردات ولا يتأثر بالجمع والطرح.
- 4 — مقياس النزعة المركزية المقابل له هو الوسط الحسابي.
- 5 — أخطاء المعاينة له أصغر من أي مقياس آخر للنشئت أي القيمة المقدرة من العينة تتجه للانحراف عن القيمة الحقيقية في المجتمع بأصغر نسبة.

• معامل الاختلاف (الانحراف المعياري النسبي) (C.D%)

نحن نعلم أن مقياس التشتت يستخدم للمقارنة بين مدى تقارب البيانات من مقاييس النزعة لها، أما إذا تساوت مقاييس النزعة للمجموعات المختلفة، فإنه يمكننا مقارنة درجات التجانس في هذه المجموعات عن طريق مقاييس التشتت، أما إذا اختلف قيمة مقياس النزعة لا نستطيع استخدام مقاييس التشتت مباشرة، نلجأ إلى مقياس التشتت النسبي، هذا المقياس يسمى (معامل الاختلاف) ويعطى بالعلاقة التالية:

$$C.V\% = \frac{S}{\bar{x}} \times 100$$

يستخدم هذا المقياس بشكل واسع في التحليل الإحصائي لمقارنة تشتت توزيعين مختلفين ولتوضيح ذلك لدينا التمرين التالي:

تمرين (1):

لدراسة الجودة لسلعتين كان الانحراف المعياري للسلعة (A) هو (10) ومتوسط (50) والسلعة (B) كان المتوسط (30) بانحراف معياري (6) فهل نستطيع الحكم على أن السلعة (A) أجود من السلعة (B) أم لا؟ ولماذا؟

الحل: من الطبيعي أننا لا نستطيع الحكم مباشرة بذلك نظراً لاختلاف المتوسط إذ لا نستطيع المقارنة مباشرة بين الرقمين (10)، (6)، لذلك نحسب معامل الاختلاف لكل منهما فيكون:

$$\text{معامل اختلاف السلعة (A)} = 100 \times \frac{10}{50} = 20\%$$

$$\text{معامل اختلاف السلعة (B)} = 100 \times \frac{6}{30} = 20\%$$

وعلى ذلك يكون التجانس في السلعتين من نفس الدرجة، وعليه نستنتج أن السلعة (A) أفضل من السلعة (B) نظراً لارتفاع المتوسط في السلعة (A) عنها في السلعة (B).

تمرين (2):

إذا علمت أن متوسط درجة الحرارة انحرافات المعيارية في شهر أيار من السنة الماضية في مدينتي دمشق وحلب، كالآتي:

$$\text{دمشق } \bar{X}_1 = 24 \quad S_1 = 3$$

$$\text{حلب } \bar{X}_2 = 34 \quad S_2 = 4$$

والمطلوب: أي المدينتين أقل تشتتاً (أكبر تجانساً) من حيث درجة الحرارة
الحل:

$$CV_1 = \frac{S_1}{\bar{X}_1} * 100$$

$$= \frac{3}{24} * 100 = 12.5\%$$

$$CV_2 = \frac{S_2}{\bar{X}_2} * 100$$

$$= \frac{4}{34} * 100 = 11.77\%$$

وبالتالي فإن درجات الحرارة في مدينة حلب أكثر تجانساً (أقل اختلافاً) من درجات الحرارة في مدينة دمشق.

4-2- قياس الالتواء :

يعرف الالتواء على أنه درجة التماثل أو البعد عن التماثل للتوزيع، ويقاس بالعلاقة التالية:

$$SK = \frac{\bar{x} - Mod}{S}$$

وفي حالة استخدام الوسيط تصبح العلاقة:

$$SK = \frac{3(\bar{x} - Med)}{S}$$

وإذا كان معامل الالتواء مساوياً صفرأ كان التوزيع متماثلاً بينما إذا كان :
 $SK > 0$ هذا يعني أن التوزيع نحو اليمين وبالعكس إذا كان : $SK < 0$ فإن
 التوزيع ملتو نحو اليسار.

مثال :

أوجد معامل بيرسون للالتواء باستخدام المنوال ومرة أخرى باستخدام الوسيط
 لأجور مجموعة من العمال إذا علمت أن متوسط الأجر (97.76)، والوسيط =
 (79.06)، والانحراف المعياري = (15.6)، والمنوال = (77.5)

الحل :

$$SK = \frac{(\bar{x} - Med)}{S} = \frac{97.76 - 77.5}{15.6} = 0.14$$

$$0.14 = \frac{2.26}{15.6} , SK = \frac{3(\bar{x} - Med)}{S}$$

$$\frac{(79.06 - 97.76)3}{15.6} = \text{المعامل الثاني للالتواء}$$

$$0.13 = \frac{2.1}{15.6} = \frac{(0.7)3}{15.6}$$

ونذكر هنا أن قيمة معامل الالتواء قد تختلف بحسب الأسلوب المستخدم لحسابه
 وعلى البيانات المتوفرة.

— معامل الالتواء الربيعي :

هناك مقاييس أخرى للالتواء منها معامل الالتواء الربيعي ويعرف كالآتي:

$$\frac{Q_3 + Q_1}{Q_3 - Q_1} \text{ أو } \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{Q_3 - Q_1} = \text{معامل الالتواء الربيعي}$$

أوجد معامل الالتواء الربيعي إذا علمت أن $Q_3=52$ ، $Q_2=48$ ، $Q_1=44$

الحل :

$$0 = \frac{(52-48) - (48-44)}{(52-44)} = \text{معامل الالتواء الربيعي}$$

أي أن التوزيع متمثل في هذه الحالة.

3-4- قياس التفرطح :

يعرف التفرطح: "بأنه مقياس لقياس درجة علو أو انخفاض قمة التوزيع

$$K = \frac{m^4}{S^4} \text{ بالنسبة للتوزيع الطبيعي ويعطى بالعلاقة الآتية :}$$

حيث أن : m^4 : العزم الرابع حول الوسط الحسابي .

S^4 : مربع التباين .

وإذا كان ($k=3$) كان التوزيع متمثلاً، بينما إذا كان ($k<3$) كان التوزيع منبسجاً أما إذا كان ($k>3$) فإن التوزيع مدبب.

4-4- الدرجة المعيارية :

تعرف الدرجة المعيارية: "بأنها النسبة الناتجة عن قسمة انحراف القيمة

الخام عن المتوسط الحسابي للتوزيع التكراري على الانحراف المعياري، وتعطى بالعلاقة التالية:

$$Z = \frac{|xi - \bar{x}|}{s}$$

ويمكن القول: أن الدرجة المعيارية عبارة عن وحدة لقياس الانحراف

المعياري عن الوسط الحسابي، وتستخدم الدرجات المعيارية المقدرة بوحدات

الانحرافات المعيارية لمقارنة قيمتين مأخوذتين من مجموعتين إحصائيتين

مختلفتين ولا يجوز مقارنة القيمتين على أساس القيم الخام .

تمرين : كانت درجتا طالبين في أحد المقررات في مجموعتين مختلفتين على النحو التالي:

المقاييس	الطالب الأول (A)	الطالب الثاني (B)
درجة الطالب	86	64
متوسط درجات الطلاب	75	58
الانحراف المعياري	10	4

والمطلوب : أيهما كان تحصيله أفضل بالنسبة لمستوى الطلاب ؟

$$Z = \frac{|x_i - \bar{x}|}{s}$$

$$Z_{(A)} = \frac{|x_i - \bar{x}|}{s} = \frac{|86 - 75|}{10} = 1.1$$

$$Z_{(B)} = \frac{|x - \bar{x}|}{s} = \frac{|64 - 58|}{4} = 1.5$$

$$Z_{(B)} > Z_{(A)}$$

وبالتالي فإن تحصيل الطالب (B) أفضل من تحصيل الطالب (A) على الرغم أنه قبل التحويل إلى درجات معيارية كانت درجات الطالب (A) أفضل من درجات الطالب (B)

تمرين :

رغبت إحدى الوزارات أن توفد شخص للخارج لإتباع دورة تدريبية، فتقدم (6) أشخاص من حملة الشهادة الجامعية للاستفادة من هذه الدورة، أجرت الوزارة ثلاثة اختبارات لكل من المتقدمين في مجال تخصصه وفي اللغة الإنكليزية وفي الحاسوب، فكانت النتائج الآتية:

المتقدمين	المعلومات ضمن التخصص	اللغة E	الحاسوب
A	10	9	5
B	9	9	6
C	5	2	0
D	5	9	10
E	4	0	6
F	3	1	9

والمطلوب:

- 1 - حول الدرجات إلى درجات معيارية في كل من الاختبارات الثلاثة.
- 2 - من هو الفائز بهذه الدورة، علماً أن صاحب أكبر مجموع بالدرجات المعيارية هو الفائز.

(الحل يترك للطالب)

تمرين:

تقدم طالب من طلاب السنة الأولى في قسم الإعلام إلى ثلاثة امتحانات، فكانت النتائج على النحو التالي:

المادة	الدرجة	$\bar{X}$	S
A	72	70.8	0.6
B	61	51.0	6.0
C	50	30.0	12.0

والمطلوب: في مادة أي امتحان كانت نتيجة الطالب أفضل بالمقارنة.

(الحل يترك للطالب)

4-5- أهمية التوزيع الطبيعي The normal Distribution

بعد التوزيع الطبيعي من أهم التوزيعات الاحتمالية في الإحصاء، ويرجع ذلك للأسباب الآتية: أولاً: تتوزع كثيراً من الظواهر التي تظهر في التجارب العملية توزيعاً طبيعياً.

ثانياً: في بعض الأحيان قد لا يكون المتغير العشوائي موزعاً توزيعاً طبيعياً ولكن يمكن أن يحول إلى متغير عشوائي يتوزع توزيعاً طبيعياً، وذلك تحت شروط معينة.

ثالثاً: قد توجد توزيعات معقدة وبالتالي يمكن أن يستخدم التوزيع الطبيعي تقريباً لها.

وأخيراً إن كافة التوزيعات الاحتمالية منفصلة كانت أم متصلة يتقارب توزيعها من التوزيع الطبيعي وذلك من خلال نظرية النهاية المركزية Central Limit Theorem.

يدخل التوزيع الطبيعي في كافة المجالات الصناعية والزراعية والاقتصادية وغيرها من المجالات الأخرى، باعتبار جميع المقاييس الإحصائية المحسوبة من مجتمع إحصائي له شكل توزيع طبيعي تقبل التفسير والتحليل، وإذا كان X متغيراً عشوائياً يتوزع وفق التوزيع الطبيعي دالة كثافة احتماله له تأخذ الصيغة الآتية:

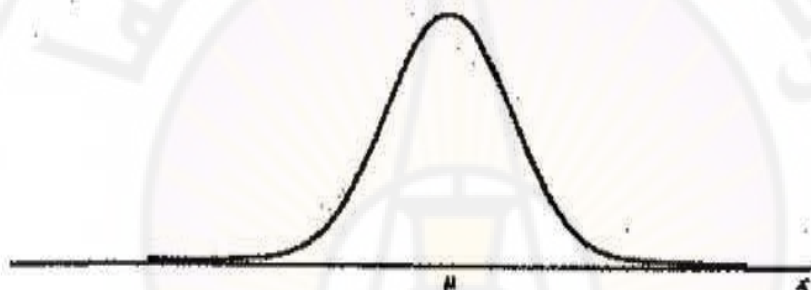
$$f_x(X) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, -\infty < X < \infty$$

حيث $(-\infty < \mu < \infty)$ و $(\sigma^2 > 0)$ يمثلان معلمتي التوزيع وهما المتوسط والتباين على التوالي (كما سنرى فيما بعد).

ويرمز للتوزيع الطبيعي عادة بالرمز $N(\mu, \sigma^2)$ ، ونقول بأن $X \sim N(\mu, \sigma^2)$.

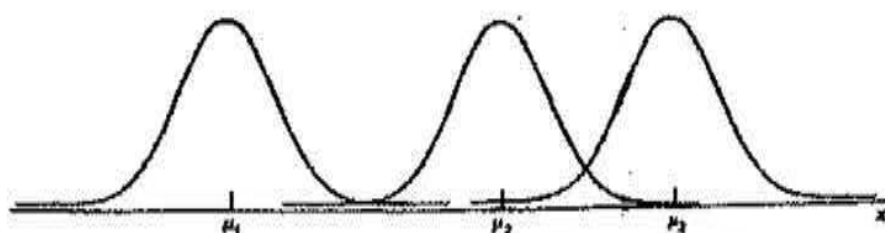
يخضع للتوزيع الطبيعي، وبالتمعن في دالة الكثافة الاحتمالية وتمثيلها البياني يتضح أن:

- 1- منحنى التوزيع الطبيعي ناقوسي الشكل ومتماثل وتقع قمته فوق المتوسط ويمتد طرفاه إلى ما لا نهاية من الجانبين دون أن يلامسا المحور الأفقي. وبالتحديد إن دالة الكثافة متماثلة حول $X=\mu$ أي $f(\mu+x)=f(\mu-x)$ وهذا يعني أن المتوسط يساوي الوسيط ويساوي المنوال وأن نقطتي الانقلاب في منحنى الدالة هما $x_1=\mu-\sigma$ و $x_2=\mu+\sigma$ وأنهما يقعان على بعد ثابت على يمين ويسار المنوال، كما في شكل (1).

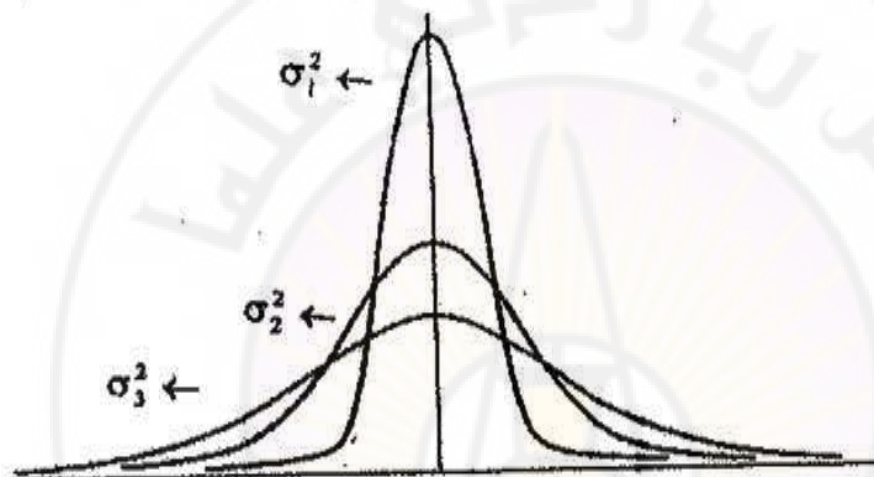


شكل (1) : دالة كثافة الاحتمال للتوزيع الطبيعي

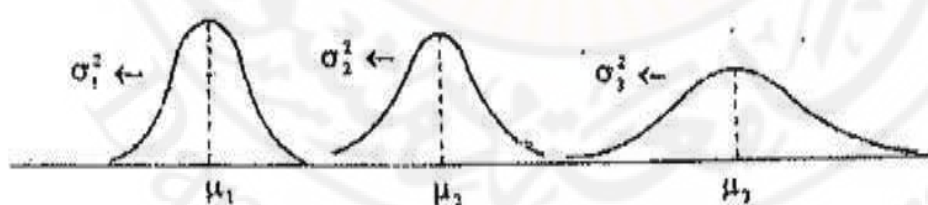
- 2- يتحدد التوزيع بمعلومية μ و σ^2 ، ويختلف التوزيع إذا اختلفت μ أو σ^2 أو كلاهما. حيث أن قيمة μ تحدد موقع التوزيع الطبيعي على الخط الحقيقي، فكلما زادت هذه القيمة تغير موقع المنحنى في الاتجاه الأيمن والعكس صحيح. بينما σ^2 تبين مقدار تشتت وتفرطح منحنى الدالة، فكلما كانت σ^2 صغيرة كلما كان المنحنى مدبباً، وكلما كانت كبيرة كلما كان المنحنى متفرطح، بمعنى آخر كلما قل التباين ارتفعت قمة المنحنى وزاد تقارب الطرفين والعكس صحيح، وذلك كما هو موضح في الأشكال (2) و (3) و (4).



شكل (2) : ثلاث توزيعات طبيعية متساوية التباينات ومختلفة المتوسطات
(دالة كثافة الاحتمال لثلاث توزيعات طبيعية عندما $\mu_1 < \mu_2 < \mu_3$ بينما $\sigma_1^2 = \sigma_2^2 = \sigma_3^2$)



شكل (3) : ثلاث توزيعات طبيعية متساوية المتوسطات ومختلفة التباين
(دالة كثافة الاحتمال لثلاث توزيعات طبيعية عندما $\mu_1 = \mu_2 = \mu_3 = \mu$ و $\sigma_1^2 \neq \sigma_2^2 \neq \sigma_3^2$)

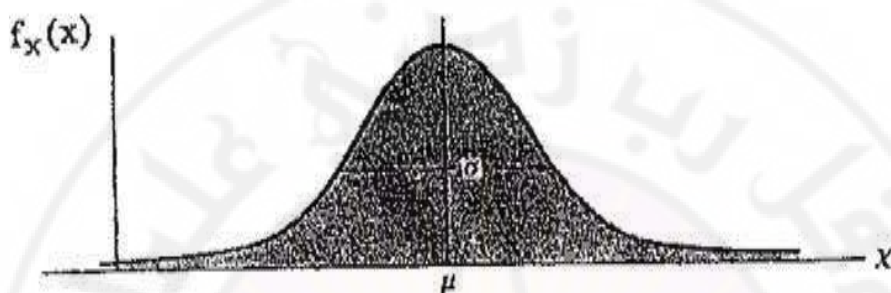


شكل (4) : ثلاث توزيعات طبيعية مختلفة المتوسطات والتباينات
(دالة كثافة الاحتمال لثلاث توزيعات طبيعية عندما $\mu_1 < \mu_2 < \mu_3$ و $\sigma_1^2 \neq \sigma_2^2 \neq \sigma_3^2$)

3 — المساحة تحت منحنى التوزيع الطبيعي تساوي واحد صحيح، أي أن :

$$\int_{-\infty}^{\infty} f_X(x) dx = 1$$

وذلك كما هو موضح في شكل (5) أدناه:



شكل (5) : المساحة المظللة = 1

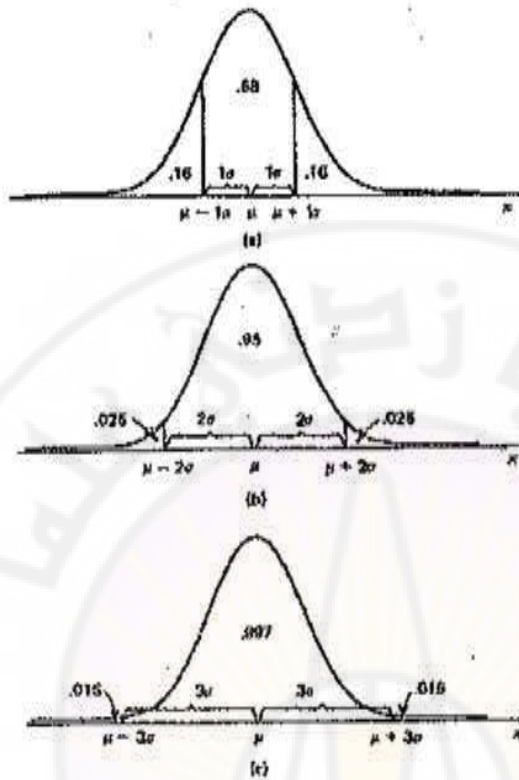
4 — توجد نسب معينة من المساحة الواقعة ضمن أي عدد من الانحراف المعياري فمثلاً:

(a) المساحة الواقعة ضمن انحراف معياري واحد عن المتوسط تساوي 68.27% من المساحة الكلية.

(b) المساحة الواقعة ضمن انحرافين معياريين عن المتوسط تساوي 95.45% من المساحة الكلية.

(c) المساحة الواقعة ضمن ثلاثة انحرافات معيارية عن المتوسط تساوي 99.73% من المساحة الكلية ويمكن توضيح الحالات الثلاثة السابقة

ببياناً كما في شكل (6) التالي:



- شكل (6) النسب المتوقعة من المساحة الواقعة ضمن عدد معين من الانحراف المعياري
- 5 — المعامل العزمي لثلاثاء يساوي صفر لجميع المنحنيات الطبيعية، وذلك لكونها متماثلة ويقع محور تماثلها تحت قمة المنحنى.
- 6 — المعامل العزمي للتفرطح يساوي 3 لجميع المنحنيات الطبيعية متوسطة التفرطح.

• التوزيع الطبيعي المعياري The Standard Normal Distribution

عندما يكون متوسط التوزيع الطبيعي يساوي صفراً والتباين يساوي واحداً، فإنه يسمى التوزيع الطبيعي المعياري، وعادة يرمز لدالة كثافة احتمال

التوزيع الطبيعي المعياري بالرمز ϕ ولدالة التوزيع التراكمي بالرمز Φ . وعليه إذا كان X متغيراً عشوائياً طبيعياً معيارياً، فإن:

$$\phi(x) = f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad -\infty < X < \infty$$

$$\Phi(x) = F_X(X) = \int_{-\infty}^x \phi(x) dx, \quad -\infty < X < \infty$$

إن دالة التوزيع التراكمي $\Phi(x)$ لا يمكن وضعها في صيغة سهلة، وبالتالي فإن احتمالات التوزيع الطبيعي المعياري أو أي توزيع طبيعي آخر يتم حسابها باستخدام القيم الجدولية لدالة $\Phi(x)$ كذلك الموجودة في آخر هذا الكتاب مع ملاحظة أنه لا يمكن استخدام هذه الجداول مباشرة إلا بعد تحويل قيمة التوزيع المراد استخدام الجداول له إلى قيم معيارية باستخدام المعادلة التالية:

$$Z = \frac{|x - \bar{x}|}{S}$$

ولتقدير قيمة أي متغير (x) طبقاً لهذا الجدول نقرأ (Z) المطلوبة من الجدول مباشرة وتكون هي قيمة التكرار النسبي أو احتمال أن تقع القيمة بين المتوسط الحسابي وقيمة (Z) .

إن الهدف من الاستعراض السابق هو معرفة نسبة أو احتمال المفردات الواقعة في مجال محدد أو خارجه تحت المنحنى الطبيعي من خلال التوزيع الطبيعي المعياري بطريقة رياضية تعتمد على تحويل الانحرافات المعيارية إلى درجات معيارية (Z) ونحن عادة نكتفي بالقيم الشهيرة.

Z	1	1.96	2	3
المساحة المقابلة لنصف المنحني الموجب	0.34135	0.47500	0.47725	0.49865

تمرين : قامت إدارة إحدى الكليات في الجامعة بقياس أطوال الطلاب المقبولين في السنة الأولى وعددهم (2500) طالب، فوجد أن الوسط الحسابي ($\mu=170$) سم والانحراف المعياري ($\sigma=8$)، ويفرض أن أطوال الطلاب يتبع التوزيع الطبيعي. المطلوب:

- 1 - ما هو عدد الطلاب الذين يتراوح طولهم بين (162، 178) سم.
- 2 - ما هو عدد الطلاب الذين يتراوح طولهم بين (146، 154) سم.
- 3 - ما هو عدد الطلاب الذين طولهم أكبر من (186) سم.

الحل:

إن المساحة تحت المنحني الطبيعي تمثل عدد الطلاب جميعاً، والوسط الحسابي يقسم هذا المنحني إلى قسمين متساويين (50%) أقل من (170) سم و (50%) أكبر من (170) سم والانحراف المعياري (8) سم.

- 1 - عدد الطلاب الذين يتراوح طولهم بين (162 و 178) سم أي

$$P(162 \leq X \leq 178)$$

$$Z_{(1)} = \left| \frac{162 - 170}{8} \right| = | -1 |$$

$$Z_{(2)} = \left| \frac{178 - 170}{8} \right| = | 1 |$$

من جدول (Z) للمساحات للقيم الشهيرة نجد أن (1) تقابل مساحة (0.34135)

— الاحتمال المطلوب :

$$0.34135 + 0.34135 = 0.6827$$

— النسبة المئوية :

$$0.6827 * 100 = 68.27\%$$

— عدد الطلاب :

$$0.6827 * 2500 = 1706.75 = 1707 \text{ طالباً}$$

2 — عدد الطلاب الذين يتراوح طولهم مسابين (146 و 154) سم أي :

$$P(146 \leq X \leq 154)$$

$$Z_{(1)} = \left| \frac{146 - 170}{8} \right| = |-3|$$

$$Z_{(2)} = \left| \frac{154 - 170}{8} \right| = |-2|$$

من جداول (Z) للمساحات للقيم الشهيرة نجد أن (3) تقابل (0.49865) و (2) تقابل (0.47725) .

— الاحتمال المطلوب: $Z = 0.49865 - 0.47725 = 0.0214$

— النسبة المئوية : $0.0214 * 100 = 2.14\%$

— عدد الطلاب : طالب 54 = $0.0214 * 2500 = 53.5$

3 — عدد الطلاب الذين أطولهم أكبر من (186) سم أي $P(X \geq 186)$

$$Z_{(1)} = \left| \frac{186 - 170}{8} \right| = 2$$

من جداول (Z) للمساحات للقيم الشهيرة نجد أن (2) تقابل (0.47725) ونحن نعلم أن مساحة نصف المنحني تعادل (0.5) لذلك نجد أن :

— الاحتمال المطلوب:

$$0.5 - 0.47725 = 0.0227$$

— النسبة المئوية :

$$0.0227 * 100 = 2.27\%$$

— عدد الطلاب:

$$0.0227 * 2500 = 56.75 = 57 \text{ طالب}$$

تمارين غير محلولة

- 1 - أخذت عينة عشوائية مكونة من (200) مودع من المودعين لدى أحد المصارف الخاصة في مدينة دمشق فكانت الودائع موزعة في الجدول التكراري المرفق (الودائع بالملايين) بتاريخ 2004/12/31 :

فئات الودائع	10-0	20-10	40-20	50-40	80-50	المجموع
عدد المودعين	7	42	75	60	16	200

المطلوب :

- 1 - حساب الوسط الحسابي والانحراف المعياري ومعامل الاختلاف لودائع هؤلاء المودعين.
- 2 - حساب المنوال للودائع بطريقة الفروق ثم استخدم النتائج في حساب معامل الالتواء للودائع .

- 2 - أخذت عينة عشوائية مكونة من (200) ناخب في إحدى الدوائر الانتخابية في مدينة دمشق، فكانت فئات الأعمار موزعة وفق الجدول التكراري المرفق :

الأعمار بالسن	25-15	35-25	45-35	55-45	65-55	المجموع
عدد الناخبين	12	14	100	48	26	200

والمطلوب :

- 1 - حساب منوال العمر.
- 2 - باستخدام المنوال احسب معامل الالتواء للجدول التكراري أعلاه.
- 3 - فيما يلي توزيع تكراري لعينة عشوائية مكونة من (100) عامل موزعين حسب فئات الأجور بالدولار .

فئات الأجور	S و أقل من 10	20-10	30-20	40-30	50-40	60-50	60 فأكثر	المجموع
عدد العمال	3	5	10	25	30	20	7	100

والمطلوب:

- 1 - احسب مقياسين من مقاييس النزعة المركزية لبيانات الجدول السابق.
- 2 - إذا افترضنا أن الفئة الأخيرة كانت 60 وأقل من 70 فأوجد الانحراف المعياري للتوزيع وكذلك معامل الاختلاف.
- 3 - حدد طبيعة التوزيع لبيانات الجدول بعد إضافة الفئة الأخيرة.
- 4 - أخذت عينة عشوائية مكونة من (50) عضو هيئة تدريسية في إحدى الجامعات بإحدى الدول فكانت فئات أعمارهم موزعة في الجدول التكراري التالي:

فئات الأعمار	25-35	35-40	40-50	50-55	55-65
عدد الأعضاء	5	10	25	8	2

والمطلوب:

- 1 - احسب متوسط ومنوال أعمار هؤلاء الأعضاء.
- 2 - احسب عدد الأعضاء الذين أعمارهم أقل من (50) سنة.
- 3 - احسب عدد الأعضاء الذين أعمارهم تزيد عن (40) سنة.
- 4 - نسبة الأعضاء الذين أعمارهم ما بين (55 و 65).
- 5 - أخذت عينة عشوائية من (30) مدخن من الطلاب المدخنين في إحدى الكليات، فكانت البيانات موزعة في الجدول التكراري التالي:

الفئات	0-8	8-10	10-12	12-16	16-20
عدد المدخنين	4	7	10	6	3

المطلوب:

- 1 - احسب الوسط الحسابي لبيانات الجدول السابق.
- 2 - احسب معامل الاختلاف
- 3 - ما هو أفضل مقياس للنزعة المركزية لبيانات الجدول السابق، ولماذا؟

6 - أخذت عينة عشوائية مكونة من (50) شخص مقبوض عليهم بتهمة التزوير (من واقع سجلات البحث الجنائي) في إحدى المدن، فكانت فئات الأعمار موزعة في الجدول التكراري التالي:

فئات الأعمار	أقل من 25-	25-35	35-40	40-50	50-60	60-65
عدد الأشخاص	2	8	28	7	4	1

المطلوب:

- 1 - ما هو أفضل مقياس للتشتت لبيانات الجدول السابق، وما هو مقياس النزعة المركزية المقابل له.
- 2 - احسب الانحراف الربيعي النسبي لبيانات الجدول السابق بإجراء الحسابات اللازمة.
- 3 - حدد طبيعة التوزيع لبيانات الجدول السابق.

7 - أخذت عينة عشوائية مكونة من (35) أسرة بإحدى المدن، فكانت بيانات الأنفاق الشهري، كما هو موضح بالجدول التكراري (بألف الليرات).

فئات الأنفاق	4-6	6-8	8-10	10-12	12-14	14-16	16-18
عدد الأسر	2	3	8	10	6	4	2

المطلوب:

- 1 - احسب الأنفاق الأكثر شيوعاً.
- 2 - حدد طبيعة للتوزيع لبيانات الجدول السابق بإجراء كافة الحسابات اللازمة.
- 3 - حدد المنوال بيانياً.
- 8 - بفرض أن معامل الذكاء (العلاقة بين السن الفكري والسن الحقيقي) للمسجلين في إحدى الكليات تخضع للتوزيع الطبيعي قدره (0.9) وانحراف معياري قدره (0.4). إذا كان عدد الطلاب في هذه الكلية (1000) طالباً. المطلوب.
- 9 - احسب عدد الطلاب الذين لهم معدل الذكاء أقل من (1.3) اختيرت عينة عشوائية من (100) مقال فكان (0.10) من المقالات المكتوبة في الصحف اليومية بها أخطاء لغوية بانحراف معياري قدره (0.02) وأن الأخطاء تخضع للتوزيع الطبيعي المطلوب:
- 1 - احسب عدد المقالات المكتوبة في الصحف اليومية التي بها أخطاء أقل من (12%) .
- 2 - احسب عدد المقالات التي بها أخطاء أكثر من (16%)
- 3 - احسب عدد المقالات التي بها أخطاء ما بين (12% و 16%).
- 10 - إذا علمت أن فترة صلاحية دواء معين تخضع للتوزيع الطبيعي بمتوسط قدره (4) سنوات وانحراف معياري (0.7) وأن مدة صلاحية هذا الدواء تنتهي في (3.3) سنة. المطلوب: ما هو احتمال أن يظل هذا الدواء صالحاً.
- 11 - بفرض أن لنوع معين من المصابيح الكهربائية عمراً يتبع التوزيع الطبيعي بمتوسط (750) ساعة وانحراف معياري قدره (60) ساعة، مع العلم أن مدة تشغيل هذه المصابيح (870) ساعة. المطلوب: ما احتمال

احتراق مصباح من هذه المصابيح بين (690، 870) ساعة من تشغيلها.
12 — إذا علمت أن الإنفاق الشهري لأفراد مجتمع مكون من (25000) فرد
يتبع التوزيع الطبيعي بمتوسط قدره (17000) ليرة وانحراف معياري
(8000) ليرة المطلوب:
— ما هو عدد الأفراد الذي يتراوح إنفاقهم ما بين (16200) و
(17800) ليرة.

الفصل الخامس

الارتباط والاتحدار الخطي البسيط

- مقدمة
- العلاقة الارتباطية
- خطوات دراسة العلاقة الارتباطية (رسم شكل الانتشار، تحديد المتغيرات)
- الارتباط الخطي البسيط
- معامل بيرسون للارتباط
- معامل ارتباط سبيرمان
- اختبار معنوية معامل الارتباط
- معامل التحديد
- الاتحدار الخطي البسيط
- معامل الاقتران
- معامل التوافق
- نماذج غير محلولة

١٣٢٠

١٣٢٠



الارتباط والانحدار الخطي البسيط

5-1- مقدمة :

من المعروف أن كل ظاهرة من الظواهر، لا تنشأ من تلقاء نفسها بدون الغير، وإنما تقع ضمن نطاق تأثيرات الظواهر الأخرى، وهذا يعني أن هناك علاقة بين الظواهر.

يطلق على العلاقة التي تربط المتغير العشوائي (y) بالمتغير العشوائي (x) اسم الانحدار، والعلاقة الرياضية بينهما اسم معادلة الانحدار، كما يطلق على وصف طبيعة تلك العلاقة بين المتغيرين المذكورين اسم تحليل الانحدار، بينما يطلق على قوة أو ضعف العلاقة بينهما اسم تحليل الارتباط، لذلك عند دراسة البيانات يجب الإجابة على السؤالين التاليين:

- أ - هل هناك علاقة بين المتغيرين؟
 - ب - إذا كان هناك علاقة بين المتغيرين فكيف نعبر عنها بمعادلة؟
- إن الهدف في العديد من الأبحاث الإحصائية هو إمكانية إيجاد علاقة ما بين متغيرين أو أكثر فمثلاً دراسة :

— العلاقة ما بين دخل الأسرة وإنفاقها على الغذاء.

— العلاقة بين كمية المبيعات في سلعة ما وسعرها.

— العلاقة بين دخل الأسرة وإنفاقها على الثقافة.

— العلاقة بين تكاليف الدعاية وحجم المبيعات لسلعة ما.

— العلاقة بين الوزن والطول.

— العلاقة بين ضغط الدم والعمر.

— العلاقة بين معدل الجريمة ومعدل البطالة.

2-5- العلاقة الارتباطية :

إن دراسة العلاقة الكائنة بين المتغيرين أو أكثر هدفها معرفة فيما إذا كان هناك ارتباط (علاقة) بين المتغيرين وقوة هذه العلاقة، وعند دراسة العلاقة الارتباطية يجب توافر الشروط التالية:

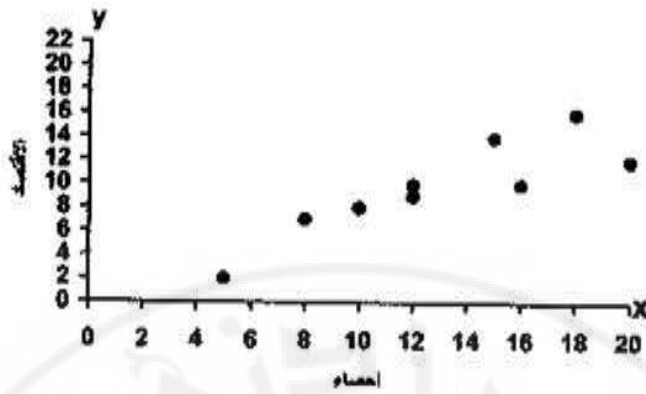
- 1 - أن تكون العلاقة بين المتغيرين منطقية.
- 2 - أن تكون أحدهما مسببة والآخرى ناتجة.
- 3 - أن تكونا قابلتين للقياس بوحدة قياس معينة.
- 4 - أن تكون للقياسات متقابلة من حيث الزمان أو المكان.

3-5- خطوات دراسة العلاقة الارتباطية :

1 - شكل الانتشار Scatter Diagram :

ونحصل عليه من خلال التمثيل البياني لقيم (x) و (y) كما هو موضح في المثال التالي الذي يمثل درجات عشرة طلاب في مادتي الإحصاء والحاسوب.

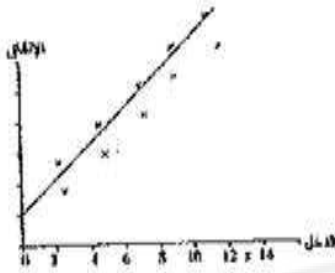
درجات الإحصاء	درجة الحاسوب
20	12
18	16
16	10
15	14
14	12
12	10
12	9
10	8
8	7
5	2



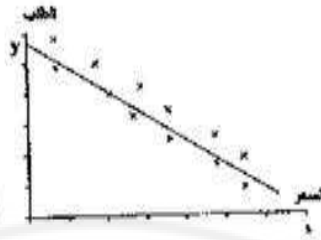
بالنظر إلى الشكل المرفق نلاحظ أن النقاط تنتشر تقريباً حول خط مستقيم مما يوحي بوجود ارتباط بين المتغيرين فالقيم الكبيرة في كلا المتغيرين تتواجد سوياً وكذلك القيم الصغيرة أي أن هناك علاقة طردية بين المتغيرين. وبشكل عام لو رسمنا المشاهدات بهذه الطريقة فإننا نحصل على مجموعة من النقاط تعطينا انطباعاً بوجود علاقة أو عدم وجودها والآن لننظر إلى الأشكال التالية وننظر ماذا يمكن استنتاجه منها.

واضح من الشكلين في (A) و (B) أنهما يمثلان علاقة ارتباطية والارتباط بينهما، في الشكل (A) يمثل علاقة طردية و (B) يمثل علاقة عكسية .

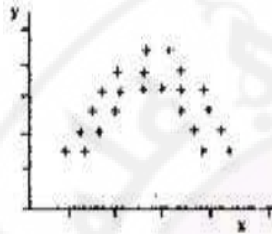
أما في الشكل (C)، فإن النقاط مبعثرة بطريقة لا توحي بوجود أي علاقة بين المتغيرين فالقيم (x) تصاحبها قيم صغيرة وقيم كبيرة للمتغير (y) وكذلك القيم الكبيرة لنفس المتغير (x) تصاحبها أيضاً قيم صغيرة وقيم كبيرة للمتغير (y) .. ومن ثم فإنه من الصعب استنتاج وجود علاقة بين المتغيرين .



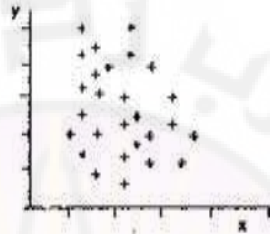
الشكل (A)



الشكل (B)



الشكل (D)



الشكل (C)

أما الشكل (D) فهو يمثل علاقة غير خطية من نوع خاص لن نهتم بها في دراستنا الحالية وإنما سيكون اهتمامنا بالعلاقات الخطية فقط. مما سبق يمكن القول: إن شكل الانتشار مهمته تحليل سلوك المتغيرين والهدف منه تحديد نوع وطبيعة العلاقة بين المتغيرين.

2 - تحديد طبيعة المتغيرات : نحن نعلم أن المتغير هو عبارة عن القيم للملاحظات التي تخص الظاهرة المدروسة خلال فترة معينة (كمية أو نوعية)، ويمكن تقسيم المتغير إلى نوعين من المتغيرات هي:

- المتغير المستقل (x) : هو المتغير الذي تكون تغيراته غير مرتبطة بتغيرات المتغير الآخر.

- المتغير التابع (y) : هو المتغير الذي تغيراته مرتبطة بتغيرات المتغير المستقل، ويتم تحديد المتغير المستقل من خلال: "طريقة الأسبقية الزمنية" : ونقوم على فكرة المتغير الأسبق حدوث زمنياً هو المتغير المستقل مثل: الدخل والإنفاق على الغذاء، فالدخل هنا هو الأسبق زمنياً

لذلك هو المتغير المستقل والإنفاق هو المتغير التابع، ومع ذلك تبقى طريقة الأسبقية الزمنية هي الطريقة الأكثر استخداماً وشيوعاً.

4-5- الارتباط الخطي البسيط Linear Correlation :

معامل الارتباط (r) :

إذا كانت هناك علاقة إحصائية بين متغيرين فإننا نسمي هذا ارتباطاً، والارتباط قد يكون طردياً أو عكسياً والأهم من ذلك هو "درجة" أو قوة هذا الارتباط والتي تتراوح ما بين (+1) و (-1) أي $r \in [-1, +1]$ وكلما اقتربت (r) من الواحد الصحيح، دل ذلك على قوة العلاقة بين المتغيرين وبالعكس ومثال ذلك العلاقة بين الدخل والإنفاق على الثقافة وغيرها من العلاقات .

4-5- معامل بيرسون للارتباط Pearson's Correlation Coefficient :

يعرف معامل بيرسون للارتباط بالمعادلة التالية :

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{n \cdot s_x \cdot s_y} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}$$

ولتوضيح كيفية حساب معامل الارتباط، نعود للتمرين السابق والمتعلق بدرجات الطلاب في مقرري الإحصاء والحاسوب. المطلوب: احسب معامل الارتباط.

الحل:

بعد حساب المتوسطات، $\bar{y} = 10$ ، $\bar{x} = 13$ على الترتيب نقوم بإنشاء الجدول المساعد التالي لحساب الانحرافات:

$(y - \bar{y})^2$	$(x - \bar{x})(y - \bar{y})$	$(x - \bar{x})^2$	$y - \bar{y}$	$x - \bar{x}$
4	14	49	2	7
36	30	25	6	5
0	0	9	0	3
16	8	4	4	2
4	2	1	2	+1
0	0	1	0	-1
1	1	1	-1	-1
4	6	9	-2	-3
9	15	25	-3	-5
64	64	64	-8	-8
138	140	188	0	0

$$r = \frac{140}{\sqrt{(188)(138)}} = 0.87$$

وبالتالي يمكن القول: إن ثمة علاقة قوية وطرادية بين المتغيرين لأن الإشارة موجبة.

ويمكن أن نستخدم صيغة أبسط في حساب هذا المعامل وهي:

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} = 0.87$$

$$r = \frac{\sum xy - n\bar{x}\bar{y}}{\sqrt{[\sum x^2 - n(\bar{x})^2][\sum y^2 - n(\bar{y})^2]}} = 0.87$$

ملاحظة: إن معامل الارتباط لا يتأثر بإضافة (أو طرح) أو القسمة على مقدار ثابت على جميع قيم أي من المتغيرين

5-5- معامل ارتباط سبيرمان (ارتباط الرتب) Rank Correlation (r_s) :

يستخدم هذا المعامل لقياس المتغيرات النوعية (الوصفية) على شكل رتب، ويعطى هذا المعامل بالعلاقة التالية:

$$r = 1 - \frac{6\sum d^2}{n(n^2 - 1)}$$

مثال (1) :

أخذت عينة عشوائية مكونة من (5) طلاب (خريجين ومستجدين) من إحدى الكليات لمعرفة كيفية ترتيب أولياتهم في كيفية قضائهم لأوقات فراغهم، فكان الجدول الآتي:

الأنشطة	تلفزيون	أفلام	كتب	مجلات	انترنت
رتب الخريجين	1	2	3	4	5
رتب المستجدين	2	3	1	4	5

والمطلوب :

قياس شدة العلاقة بين رأيي الطلاب حول الأنشطة وذلك لمعرفة تقارب وجهات نظر الطلاب حول قضاء وقت فراغهم أو (قياس تقارب وجهات النظر بين الطلاب)

	الخريجون	المستجدون		
الأنشطة	رتب (x)	رتب (y)	d	d ²
تلفزيون	1	2	-1	1
أفلام	2	3	-1	1
كتب	3	1	2	4
مجلات	4	4	0	0
انترنت	5	5	0	0

$$r_s = 1 - \frac{6\sum d^2}{n(n^2 - 1)}$$

$$= 1 - \frac{6(6)}{5(25 - 1)} = 0.70$$

والعلاقة جيدة، ونرى أن المتخرجين والمستجدين اتفقوا في الرأي على شئئين فقط من حيث الأفضلية في قضاء وقت الفراغ (المجلات والانترنت).

مثال (2): في أحد المنافسات تم استعراض (5) لوحات فنية لتقييمها من قبل لجنتي تحكيم، فإذا كانت الرتب الممنوحة للوحات الخمس على النحو التالي:

اللوحات	A	B	C	D	E
اللجنة A	1	2	3	4	5
اللجنة B	3	2	1	5	4

والمطلوب: أحسب معامل الارتباط سبيرمان.
الحل:

اللوحات	رتب اللجنة (A)	رتب اللجنة (B)	d	d ²
A	1	3	-2	4
B	2	2	0	0
C	3	1	-2	4
D	4	5	-1	1
E	5	4	+1	1

$$r_s = 1 - \frac{6(10)}{5(25 - 1)} = 0.5$$

وبلاحظ أن هناك ارتباطاً إيجابياً في التقييم بين اللجنتين.

5-6- اختبار معنوية معامل الارتباط :

لإجراء هذا الاختبار نعتبر المجتمع الذي أخذت منه العينة (n)، ومعامل الارتباط (R) هو تقدير غير متحيز لمعامل الارتباط (r) وأن المجتمع الأصلي يخضع للتوزيع الطبيعي، ويمكن تحديد فيما إذا كان معامل الارتباط (r) لعينة حجمها (n) دالاً على وجود علاقة خطية بين المتغيرين، ويمكن إجراء اختبار معنوية معامل الارتباط بالاعتماد على توزيع (Z) الطبيعي وحساب الخطأ المعياري لمعامل الارتباط (S_r) وذلك على النحو التالي :

$$S_r = \sqrt{\frac{1-r^2}{n-2}}$$

حيث أن : (r^2) معامل التحديد (مربع معامل الارتباط).
ونميز حالتين :

I — $r \geq 3S_r$ وبالتالي العينة المسحوبة عشوائية وتمثل المجتمع الإحصائي المحسوبة منه ولمعامل الارتباط أهمية إحصائية.

II — $r < 3S_r$ العينة المسحوبة غير عشوائية ولا تمثل المجتمع الإحصائي المحسوبة منه وليس لمعامل الارتباط أهمية إحصائية.

مثال: نعود إلى التمرين السابق والمتعلق بدرجات الإحصاء والحاسوب للطلاب والمطلوب: بيان الأهمية الإحصائية لمعامل الارتباط باستخدام التوزيع الطبيعي.

الحل :

$$S_r = \sqrt{\frac{1-(0.87)^2}{10-2}} = 0.17$$

$$0.87 > 3 (0.17)$$

بالتالي العينة المسحوبة عشوائية وتمثل المجتمع الإحصائي المسحوبة منه ولمعامل الارتباط أهمية إحصائية.

5-7- معامل التحديد (r^2) :

هو عبارة عن مربع معامل الارتباط وتتراوح قيمته أو درجته بين الصفر والواحد الصحيح أي : $r^2 \in [0,1]$ وهو يقيس التحسن الناتج عن استخدام خط الانحدار بدلاً من استخدام الوسط الحسابي كأساس لتقدير قيم المتغير التابع أو يقيس النسبة المئوية من تباين المتغير التابع التي يفسرها المتغير المستقل أو النسبة المئوية من التباين في المتغير التابع الذي تفسره معادلة خط الانحدار والباقي لا تستطيع معادلة خط الانحدار تفسيره وبالعودة إلى التمرين السابق المتعلق بدرجات الطلاب في مقرري الإحصاء والحاسوب نجد أن : $r^2 = (0.87)^2 = 0.75$ ويساوي (0.75) وهذا يعني أن (75%) من تباين المتغير التابع استطاعت معادلة خط الانحدار تفسيره والباقي (25%) لم تستطع تفسيره.

5-8- الانحدار الخطي البسيط Regression line :

عند دراسة الارتباط كان الاهتمام بالعلاقة الكائنة بين (x) و (y) وبدرجة قوة هذه العلاقة، أما الانحدار يهدف إلى تمثيل هذه العلاقة بمعادلة رياضية بحيث يمكن استخدامها للتنبؤ بقيمة أحد المتغيرين إذا علمنا قيمة المتغير الآخر. مثال: التنبؤ بإنفاق أسرة ما إذا علم دخلها والتنبؤ بكمية المبيعات من سلعة ما إذا علم مصاريف الدعاية عن تلك السلعة.

— معادلة خط الانحدار : هي المعادلة الرياضية التي تصف العلاقة بين متغيرين موضوع الدراسة، والغاية منها تقدير القيم التي يمكن أن يأخذها المتغير التابع بدلالة المتغير المستقل، وتعطى هذه المعادلة بالعلاقة التالية :

$$\hat{y} = a + bx + e$$

حيث أن :

$\hat{Y}$: القيمة المقدرة للمتغير التابع (القيمة الوسطية).

a : ثابت خط الانحدار

b : معامل الانحدار (ميل خط الانحدار)

e : الخطأ العشوائي

— ويتم إيجاد هذه المعادلة بطريقتين :

1) طريقة الرسم الحر (طريقة الرسم البياني):

تعتمد هذه الطريقة على رسم شكل الانتشار للظاهرة المدروسة باعتبار المحور (x) يمثل المتغير المستقل والمحور الأفقي (y) المتغير التابع ويتم رسم خط الانحدار بحيث يمر بين منتصف نقاط الانتشار ويتم تحديد معادلة خط الانحدار على النحو التالي :

$$\hat{y} = a + bx$$

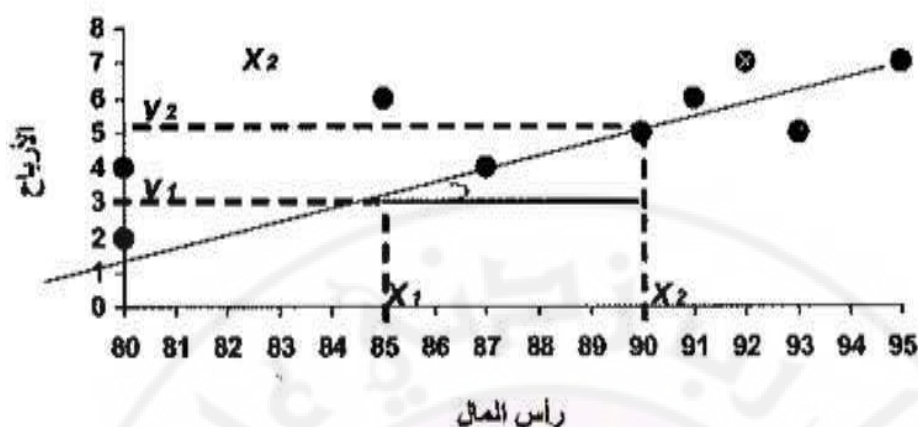
a : عبارة عن نقطة تقاطع خط الانحدار مع المحور (y)

b : عبارة عن ميل خط الانحدار (ظل الزاوية) مع المحور (x)

ويتم حساب ظل الزاوية من خلال أخذ نقطتين على خط الانحدار وإجراء عملية إسقاط منهما على المحور (x) والمحور (y)، كما هو موضح من خلال التمرين التالي:

أخذت عينة عشوائية مكونة من (10) شركات للدعاية والإعلان لدراسة العلاقة بين رأس المال والأرباح، فحصلنا على النتائج الآتية (بالمليون).

رقم الشركات	1	2	3	4	5	6	7	8	9	10
رأس المال	90	80	80	85	87	92	90	95	93	85
الأرباح	5	2	4	6	4	7	6	7	5	5



$a = +1$ (نقطة التقاطع مع المحور y)

$$b = \frac{y_2 - y_1}{x_2 - x_1} = \frac{5.5 - 3}{90 - 85} = 0.5$$

(أو) (ظل الزاوية = المقابل المجاور)

وبالتالي معادلة خط الانحدار :

$$\hat{y} = 1 + 0.5x$$

يعاب على هذه الطريقة، بأنها تقريبية وتختلف من شخص لآخر بسبب اختلاف أسلوب الرسم، إلا أنها تعطي فكرة عن معادلة خط الانحدار .

(2) - الطريقة الرياضية : وتعتبر أدق من الطريقة السابقة، باعتبارها تعتمد

الأسلوب الرياضي وإيجاد ثوابتها المعادلة نعتمد أحد الأسلوبين :

(a) أسلوب المربعات الصغرى : وتتلخص هذه الطريقة بإيجاد (a) و (b) التي

تجعل مجموع مربعات الانحرافات أصغر ما يمكن، أي إيجاد (a) و (b) التي

تجعل المقدار $\sum (y - \hat{y})^2 \rightarrow \min$ أي أصغر ما يمكن، وبالتالي :

$$(y - a - bx)^2 = 0$$

وبالاشتقاق الجزئي الأول بالنسبة لكل من (a) و (b)

وبعدم هذه المشتقات، سنحصل على جملة المعادلتين التاليتين :

$$(1) \sum y = na + b \sum x$$

$$(2) \sum xy = a \sum x + b \sum x^2$$

وبحل جملة المعادلتين، نحصل على قيم الثوابت (a)، (b) وبالتالي على معادلة خط الانحدار.

(b) - الأسلوب المباشر: ويقوم هذا الأسلوب على حل المعادلتين السابقتين لنحصل على الثوابت:

$$b = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2} = \frac{\sum xy - n \bar{x} \bar{y}}{\sum x^2 - n \bar{x}^2}$$

$$b = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2}, \quad a = \bar{y} - b \bar{x}$$

ومع ذلك نجد أن الأسلوب المباشر هو الأسلوب الأكثر استخداماً في التطبيقات.
تمرين : أخذت عينة عشوائية مكونة من (7) أسر في إحدى المدن، لدراسة العلاقة بين الدخل والإنفاق الشهري لهذه الأسر على الغذاء، فكانت البيانات التالية (بآلات الوحدات النقدية).

الدخل (x)	38	32	42	48	40	44	50
الإنفاق (y)	24	21	27	30	27	33	36

والمطلوب : احسب معادلة خط الانحدار.

y^2	x^2	xy	y	x
576	1444	912	24	38
441	1024	672	21	32
729	1764	1134	27	42
500	2304	1440	30	48
729	1600	1080	27	40
1089	1936	1452	33	44
1296	2500	1800	36	50
5760	12572	8490	198	294

الحل: بطريقة المربعات الصغرى، واعتماد طريقة الحذف بالتعويض لحل المعادلتين فنجد حل المعادلات وبالتعويض بالمعادلتين نجد أن:

$$(1) \quad 198 = 7a + b294$$

$$(2) \quad 8490 = 294a + b12572$$

وبضرب المعادلة (1) في (42) نحصل على :

$$8316 = 294a + 12348b \quad (3)$$

$$8490 = 294a + 12572b \quad (2)$$

وبطرح المعادلة (1) من (3) نحصل

$$\begin{array}{r} -174 = 0 \\ -224b \end{array}$$

وبالتالي :

$$b = \frac{-174}{-224} = 0.78$$

نعوض في المعادلة (1) نجد أن :

$$198 = 7a + 0.78(294)$$

$$a = -4.47$$

وبذلك تكون المعادلة المطلوبة هي : $\hat{y} = -4.47 + 0.78x$

أي أنه كلما زاد الدخل بمقدار وحدة واحدة (ألف وحدة نقدية) زاد حجم الإنفاق بمقدار (0.78) أي (780) وحدة نقدية.

ويمكن استخدام هذه المعادلة لتقدير حجم الإنفاق عند أي قيمة من الدخل، فمثلاً إذا كان هذا الدخل هو (45) مثلاً فإن :

$$\hat{y} = -4.47 + 0.78(45) = 30.60$$

أي أن حجم الإنفاق المتوقعة هو : (30600) وحدة نقدية.

• رسم خط الانحدار: لتحديد أي خط مستقيم يكفي تحديد نقطتين وبالعودة

للتمرين السابق والمطلوب:

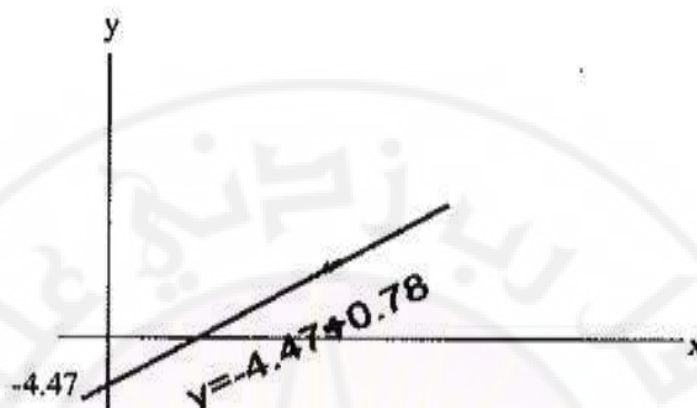
أرسم خط الانحدار بيانياً.

$$x = 0 \Rightarrow y = -4.47$$

$$y = 0 \Rightarrow -4.47 + 0.78x = 0$$

$$x = 5.25$$

الحل :



• تفسير الثوابت : إن (b) تمثل معدل التغير في y المناظر لتغير في (x) مقدار وحدة واحدة. هذا يعني عندما يزداد الإنفاق بمقدار (0.77) يزداد الدخل وحدة واحدة (ألف وحدة).

إن (a) تمثل مقدار الانفاق عندما يكون الدخل معدوماً. ويمكن حساب الثوابت بالطريقة المباشرة وفق السابقة:

$$b = \frac{\sum xy - n\bar{x}\bar{y}}{\sum x^2 - n(\bar{x})^2}$$

$$a = \bar{y} - b\bar{x}$$

وبالتطبيق ينتج لدينا:

$$b = \frac{8490 - 7(42)(28.29)}{12572 - 7(42)^2} = 0.78$$

$$a = 28.28 - 0.78(42) = -4.47$$

وبالتالي ستكون نفس المعادلة السابقة.

• العلاقة بين ميل خط الانحدار ومعامل الارتباط :

توجد علاقة وثيقة بين ميل خط الانحدار (b) ومعامل الارتباط (r) والتي يمكن الحصول عليها بالصيغة التالية :

$$r = b \frac{S_x}{S_y}$$

• خط انحدار x على y :

يمكن التعبير عن ذلك بالمعادلة التالية :

$$\hat{x} = a' + b'y + e$$

وبحل جملة المعادلتين نحصل على الثوابت:

$$(1) \sum x = na' + b'\sum y$$

$$(2) \sum xy = a'\sum y + b'\sum y^2$$

ويسمى الخط الناتج بخط انحدار x على y والميل b' يرتبط بمعامل الارتباط بالعلاقة:

$$b' = r \frac{S_y}{S_x}$$

مثال :

تجدر الإشارة هنا إلى أنه يمكن الحصول على تقديرات (b') ، (a') بالأسلوب المباشر:

$$b' = \frac{n\sum xy - \sum x \sum y}{n\sum y^2 - (\sum y)^2}$$

$$a' = \bar{x} - b'\bar{y}$$

ومن خلال الاستعراض السابق نلاحظ أن :

I — كلا خطي الانحدار يمران بالنقطة ($\bar{x}$) ، ($\bar{y}$) .

$$r = \sqrt{b.b'} \text{ — II}$$

ونلاحظ أن : (b , b') يكون دائماً موجباً، لكن نحن نعلم أن إشارة (r) من إشارة

(b) و (b').

وبالتطبيق على المثال السابق سنجد أن:

$$r = \sqrt{b.b'} = 0.92$$

ويمكن التأكد من ذلك من خلال الصيغة المباشرة لحساب معامل الارتباط:

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} = \frac{7(8490) - (582)(12)}{\sqrt{(1568)(1110)}} = 0.92$$

ونلاحظ أن الإشارة موجبة لأن كلا من (b) و (b') موجب أما لو كان كل من (b) و (b') سالباً فإنه يجب اختيار الإشارة السالبة.

ويمكننا هنا الإشارة إلى إمكانية استخدام معادلة خط الانحدار (x/y) في التنبؤ.

5-9- الخطأ المعياري للتقدير :

يمكن استخدام معادلة خط الانحدار في التنبؤ لبيانات داخل حدود العينة وبيانات خارج حدود العينة مع افتراض ثبات العلاقة بين المتغيرين (أي ستبقى العلاقة خطية بين المتغيرين) وعكس ذلك لا يمكن إجراء هذا التقدير، ولمعرفة مدى الثقة بالتقدير يجب معرفة الخطأ المعياري للتقدير والذي يعرف بأنه: "عبارة عن الجذر التربيعي لمتوسط مجموع مربع الحرافات القيم المقدرة عن القيم الفعلية للمتغير التابع، وهو يشابه الانحراف المعياري لتوزيع تكراري الذي يقيس مدى التشتت حول الوسط الحسابي، بينما الخطأ المعياري يقيس مدى التشتت حول خط الانحدار"، ويبين هذا المقياس مقدار انحراف القيم الفعلية عن القيم المقدرة للمتغير التابع بواسطة معادلة خط الانحدار، وإذا كانت قيمته مساوية للصفر، هذا يعني أنه ليس هناك تشتتاً أو انتشاراً وأن جميع النقاط تقع

على خط الانحدار ، وبالتالي يكون الارتباط بين المتغيرين تاماً ، ولتوضيح ذلك ،
نعود للتمرين السابق والمتعلق بالدخل والإنفاق على الغذاء .
والمطلوب : احسب الخطأ المعياري للتقدير وبين ماذا تقيس هذه القيمة ؟

رقم الأمرة	x	y	$\hat{y} = -4.47 + 0.78x$	$(y - \hat{y})^2$	$(\hat{y} - \bar{y})^2$
1	38	24	25.21	1.464	9.42
2	32	21	20.59	0.1681	59.13
3	42	24	28.29	1.6641	0.001
4	48	30	32.91	8.4681	21.43
5	40	27	26.75	0.0625	2.34
6	44	33	29.83	10.0489	2.4
7	50	36	34.45	2.4025	38.06
	294	198	203.03	24.278	132.781

إن معادلة خط الانحدار للتمرين أعلاه هي :

$$\hat{y} = -4.47 + 0.78x$$

حصلنا في الجدول أعلاه، على القيم ($\hat{y}$) من خلال معادلة خط الانحدار ، علماً :
أن مجموع مربعات الانحرافات عن المتوسط هي (132.78) ومجموع مربع
انحراف القيم الفعلية عن القيم المقدرة يعطى بالعلاقة :

$$S(\hat{y}) = \sqrt{\frac{\sum (y - \hat{y})^2}{n}}$$

بالتطبيق على بيانات التمرين :

$$\sqrt{\frac{24.2783}{7}} = 1.862$$

التفسير : هذا المقياس لا يقبل التفسير لكن يمكن القول أنه كلما صغرت هذه
القيمة كانت الثقة أكثر بمعادلة خط الانحدار .

ونشير هنا، إلى أنه يمكن حساب الخطأ المعياري للتقدير بالعلاقات التالية :

$$S_{\hat{y}} = \sqrt{\frac{\sum y^2 - (a\sum y + b\sum xy)}{n}}$$

$$S_{\hat{y}} = sy\sqrt{1-r^2} \quad (\text{خط انحدار } y/x)$$

$$S_{\hat{y}} = sx\sqrt{1-r^2} \quad (\text{خط انحدار } x/y)$$

وبتطبيق العلاقة الأولى نجد أن :

$$\sum xy = 8490, \sum y^2 = 5760$$

$$S_{\hat{y}} = \sqrt{\frac{5760 - (4.47(198) + 0.78(8490))}{7}} = 1.862$$

10-5- التباين المفسر والتباين غير المفسر :

نحن نعلم أن: التباين الكلي = التباين المفسر + التباين غير المفسر.

ويعرف التباين المفسر r^2 بأنه: "عبارة عن جزء من التباين الكلي وهو التباين الذي يمكن تفسيره بمعادلة خط الانحدار، وكلما زادت قيمته، دل ذلك على إمكانية تحديد قيم (y) بدلالة (x) بدقة أكثر".

أما التباين غير المفسر فهو أيضاً: جزء من التباين الكلي (مربع الخطأ المعياري للتقدير) وهو التباين الذي لا يمكن تفسيره بمعادلة خط الانحدار، وكلما صغرت قيمته زادت الثقة أكثر بمعادلة خط الانحدار.

ويمكن التعبير عما سبق، على النحو التالي :

$$SST = SSR + SSE$$

حيث إن :

SST : مجموع مربعات انحرافات قيم (y) الفعلية عن وسطها الحسابي ($\bar{y}$)

ويعطى بالعلاقة التالية: $\sum (y - \bar{y})^2$

SSR : مجموع مربعات انحرافات قيم ($\hat{y}$) المقدرة عن الوسط الحسابي ($\bar{y}$)

ويعطى بالعلاقة التالية: $\sum (\hat{y} - \bar{y})^2$

SSE : مجموع مربعات انحرافات قيم (y) الفعلية عن القيم المقدرة ($\hat{y}$)
 ويعطى بالعلاقة التالية: $\Sigma(y - \hat{y})^2$

$$\frac{\Sigma(y - \bar{y})^2}{n} = \frac{\Sigma(\hat{y} - \bar{y})^2}{n} + \frac{\Sigma(y - \hat{y})^2}{n}$$

$$\frac{\Sigma y^2}{n} - \left(\frac{\Sigma y}{n} \right)^2 = \frac{b(\Sigma xy - \bar{x}\Sigma y)}{n} + \frac{\Sigma y^2 - (a\Sigma y + b\Sigma xy)}{n}$$

وبالتطبيق على المثال السابق، وباستخدام صيغ التباين المفسر وغير المفسر
 وبدلالة ثوابت معادلة خط الانحدار، نجد أن :

$$SSR = \frac{0.78(8490 - 42(198))}{7} = 18.969$$

$$SSE = (1.862)^2 = 3.467$$

$$SST = \frac{5760}{7} - \left(\frac{198}{7} \right)^2 = 22.436$$

$$SST = 18.96 + 3.46 = 22.43$$

• العلاقة بين التباينات (المفسر وغير المفسر) ومعامل التحديد (r^2):
 يمكن إيجاد معامل التحديد (r^2) من خلال هذه العلاقة :

$$r^2 = \frac{s^2 \hat{y}}{s^2 y} = \frac{\text{التباين المفسر}}{\text{التباين الكلي}}$$

$$r^2 = 1 - \frac{s^2 y \hat{y}}{s^2 y}$$

وبالتطبيق على التمرين السابق نجد أن :

$$r^2 = \frac{18.469}{22.436} = 0.84 \quad (\text{بالاعتماد على الصيغ المباشرة}) \quad \text{أو} \quad r^2 = 1 - \frac{3.467}{22.436} = 0.84$$

أو

$$r^2 = (0.92)^2 = 0.84 \quad (\text{بالاعتماد على الصيغ التي تتضمن ثوابت المعادلة})$$

• ملاحظات :

(1) — كلما اقتربت القيم الفعلية (y) من خط الانحدار، صغر الخطأ المعياري للتقدير.

(2) — عندما يمر خط الانحدار بمركز الإحداثيات، هذا يعني أن جميع نقاط الانتشار وقعت على هذا الخط وبالتالي الارتباط تام بين المتغيرين.

(3) — عندما يكون معامل الانحدار ($b=0$) فإن خط الانحدار ينطبق على خط ($\bar{y}$) وبالتالي الارتباط معدوم بين المتغيرين وأصبح ($s^2 \hat{y} = 0$) الثباين المفسر.

(4) — عندما تكون ($y = \hat{y}$)، فإن جميع نقاط الانتشار تقع على خط الانحدار، أي أن : ($s^2 \hat{y} = 0$) والارتباط تام بين المتغيرين.

(5) — إذا كان : ($s^2 \hat{y} = s^2 y$) لا يوجد علاقة ارتباطية بين (y, x).

(6) — إذا كان : ($s^2 \hat{y} = 0$) العلاقة تامة بين (y, x) أي ($r=1$).

5-11- معامل الاقتران ومعامل التوافق :

— معامل الاقتران بأنه : هو عبارة عن مقياس يقيس شدة العلاقة الارتباطية بين متغيرين وكل منهما ينقسم إلى صفتين، كما في الجداول (2×2)، ويعطى بالعلاقة التالية :

$$r_c = \frac{AD - BC}{AD + BC}$$

كلما اقتربت قيمة (r_c) من الواحد كانت العلاقة بين المتغيرين قوية جداً، وإذا

كانت ($r_s > 0.5$) هذا يعني أن هناك ترابطاً (اقتتراناً) بين المتغيرين (الظاهريين)، وبالعكس إذا كانت ($r_s < 0.5$) العلاقة ضعيفة بين المتغيرين وتأخذ عادة (r_s) بالقيمة المطلقة.

تمرين (1) : يبين الجدول التالي البيانات المستخدمة في دراسة العلاقة بين رضا أو عدم رضا مجموعة من المتعاملين مع أحد المصارف وكونهم موظفين في القطاع العام أو في القطاع الخاص.

المجموع	في القطاع الخاص	في القطاع العام	موقع الوظيفة درجة الرضا
85	35	50	راض
85	45	40	غير راض

والمطلوب: قياس مقدار الترابط بين رضا أو عدم رضا المتعامل مع هذا المصرف كونه في القطاع العام أو القطاع الخاص.

$$r_s = \frac{(50)(45) - (35)(40)}{(50)(45) + (35)(40)} = 0.23$$

وعليه فإن العلاقة ضعيفة بين المتغيرين عدم وجود اقتتران بين موقع الوظيفة، (قطاع عام، قطاع خاص) ودرجة الرضا.

— معامل التوافق :

يستخدم هذا المعامل عند دراسة العلاقة الارتباطية بين الظواهر التي تحمل أكثر من صفتين، لذا يُعد هذا المعامل أعم من معامل الاقتتران ويتم حسابه من خلال العلاقة التالية :

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

$$G = \sum \left(\frac{f_{ij}^2}{f_{i.} f_{.j}} \right)$$

إذا كانت ($r_s > 0.5$) كان هناك علاقة ارتباطية (توافق)، وكلما اقتربت من الواحد دل ذلك على متانة العلاقة الارتباطية (التوافق).

تمرين (2) : أعيد تصنيف المتعاملين مع المصرف في التمرين السابق حسب مستوياتهم التعليمي، فكانت النتائج على النحو التالي :

المستوى التعليمي / درجة الرضا	ثانوية وإعدادية	جامعة أو معهد	فوق جامعة	Σ
راض	20	25	50	95
غير راض	30	25	20	75
Σ	50	50	70	170

والمطلوب: هل من علاقة بين المستوى التعليمي ودرجة الرضا "معامل التوافق" للبيانات السابقة :

$$G = \Sigma \left(\frac{f_{ij}^2}{f_i f_j} \right)$$

$$= 0.08 + 0.24 + 0.13 + 0.167 + 0.376 + 0.08 = 1.073$$

$$rA = \sqrt{\frac{1.073 - 1}{1.073}} = 0.26$$

وبالتالي العلاقة ضعيفة بين المتغيرين (عدم وجود توافق بين المستوى التعليمي ودرجة الرضا).

تمرين: في الجدول التالي (188) إعلامي من الإعلاميين العاملين في وزارة الإعلام في إحدى الدول وموزعين حسب المشاركة في الدورات التدريبية ومستوى الأداء

المشاركة / الأداء	شارك	لم يشارك
تحسن	90	14
كما هو	22	62

والمطلوب : قياس شدة العلاقة بين المشاركة والأداء (يترك الحل للطالب).
تمرين : في دراسة طبية للقاح ضد شلل الأطفال في مجتمع ما، حصلنا على النتائج الآتية:

التحصين \ الإصابة	حُصن ضد الشلل	لم يُحصن ضد الشلل
أصيب بالشلل	39	65
لم يصب	159	26

والمطلوب: حساب معامل الارتباط (الاقتران) (يترك الحل للطالب).
تمرين : لدراسة تأثير الحالة التعليمية للأُم على مستوى الابن في المدرسة الابتدائية، تم جمع البيانات الآتية :

الحالة التعليمية للأُم \ مستوى الابن بالمدرسة	أمية وابتدائية	إعدادية وثانوية	فوق الثانوية
متفوق	5	15	30
وسط	5	40	15
ضعيف	70	15	5

والمطلوب: قياس شدة العلاقة بين المستوى التعليمي للأُم ومستوى الابن في المدرسة الابتدائية (يترك الحل للطالب).

تمارين غير محلولة

1 — كانت حركة القروض والودائع (بملايين الوحدات النقدية) في مجموعة من البنوك التجارية كما يلي :

110	77	84	89	75	القروض (x)
85	90	116	112	92	الودائع (y)

1 — أوجد معادلتين خطيتين الانحدار (y/x) ، (x/y) .

2 — أوجد معامل الارتباط بين المتغيرين (y, x)

2 — في دراسة للعلاقة بين المتغيرين y, x تبين أن :

S	$\bar{x}$	المقياس
7	100	المتغير (x)
5	60	المتغير (y)

وبفرض أن معامل الارتباط بين المتغيرين $= 0.7$
والمطلوب :

1 — أوجد معادلة خط انحدار y/x

2 — قدر قيمة (y)، إذا علمت أن $x=20$

3 — حكمان في مسابقة طلب منهما تقييم 10 متسابقين قدموا الدرجات الآتية :

8	7	6	5	6	7	8	9	8	10	الحكم الأول
4	5	8	9	10	8	6	7	7	5	الحكم الثاني

والمطلوب :

إيجاد معامل ارتباط الرتب (سبيرمان) بين تقديرات الحكمين.

4 - البيانات التالية عبارة عن نتائج دراسة حول تسلم المرأة لمناصب سياسية:

الرأي \ الجنس	مؤيد	معارض	محايد
ذكور	40	32	18
إناث	60	40	20

والمطلوب: قياس شدة العلاقة بين المتغيرين وفسر النتيجة.

5 - ظاهرتان x و y توفرت عنهما البيانات التالية :

5	6	8	7	7	6	4	8	5	4	x
5	4	9	7	8	5	4	5	4	3	y

المطلوب:

- 1 - حساب معادلة انحدار y/x
- 2 - التنبؤ بقيمة y عندما تكون $x=12$ ، تحت افتراض استمرار خطية العلاقة؟
- 3 - حساب معامل الارتباط الخطي البسيط بين الظاهرتين.
- 4 - احسب الخطأ المعياري للتقدير
- 6 - في دراسة إحصائية عن (20) شركة تتناول العلاقة بين المبيعات (y)

ومصاريف الإعلان (x)، توفرت لديك البيانات الآتية : (بآلاف الوحدات النقدية)

$$\begin{aligned} \sum x^2 &= 80 & \sum x &= 10 \\ \sum xy &= 60 & \sum y^2 &= 50 & \sum y &= 20 \end{aligned}$$

المطلوب:

- 1 - أوجد معامل الارتباط الخطي (بيرسون) بين قيم (x)، (y)
- 2 - أوجد معادلة خط الانحدار بفرض أن العلاقة خطية.

7 - تم التوصل من خلال دراسة أجريت عن العلاقة بين درجات الإحصاء (x) ودرجات الرياضيات (y) لعينة مكونة من (25) طالب إلى النتائج الآتية :

$$\begin{aligned}\sum x^2 &= 106250 & \sum x &= 1625 \\ \sum xy &= 98300 & & \\ \sum y^2 &= 97750 & \sum y &= 1500\end{aligned}$$

والمطلوب :

- (1) أوجد معادلة خط الانحدار وما الغاية من هذه المعادلة.
- (2) ماهر تقديرك لدرجات امتحان الرياضيات لطالب حصل على (65) درجة في الإحصاء.
- (3) احسب التباين المفتر والتباين غير المفتر.

8 - إذا كانت درجات ثمانية من الطلبة في مادتي الرياضيات والاقتصاد كالآتي:

75	60	75	80	75	65	60	70	الرياضيات
70	75	65	70	60	75	80	50	الاقتصاد

والمطلوب:

- 1- أوجد معامل ارتباط الرتب (سبيرمان).
- 2 - أوجد معادلة خط الانحدار
- 3 - بيان الأهمية الإحصائية لمعامل الارتباط

9 - أخذت عينة عشوائية مكونة من (10) أسر في إحدى المدن، لدراسة العلاقة بين الدخل والإنفاق على الغذاء، فكانت النتائج الآتية (بآلاف الوحدات النقدية):

$$\begin{aligned}\sum x &= 63 & \sum x^2 &= 465 \\ \sum y &= 50 & \sum y^2 &= 294 \\ \sum xy &= 366\end{aligned}$$

والمطلوب:

- 1 - أوجد معادلة خط الانحدار وفسر الثوابت.
- 2 - احسب معامل الارتباط والتحديد وفسرهما.
- 3 - قدر قيمة الإنفاق المتوقع لأسرة دخلها (10) آلاف وحدة نقدية، بفرض ثبات العلاقة الخطية.

10 - أخذت عينة عشوائية مكونة من (10) مخازن لتجارة التجزئة في إحدى المدن لدراسة العلاقة بين رأس المال والأرباح، فكانت البيانات على النحو التالي (بالمليون وحدة نقدية).

رقم المخزن	1	2	3	4	5	6	7	8	9	10
رأس المال	15	18	9	35	30	35	9	25	60	48
الأرباح	1.5	2.5	1	5	3	3.5	1.2	2.5	6	4.5

والمطلوب:

- 1 - أوجد قيمة معامل الانحدار ومعامل التحديد.
- 2 - قدر أرباح مخزن ما إذا كان رأس ماله (55) مليون.
- 3 - ارسم خط الانحدار بيانياً.

4 - أحسب التباين المفسر والتباين غير المفسر .

5 - بيان الأهمية الإحصائية لمعامل الارتباط.

11 - قام أحد الباحثين باختيار عينة عشوائية لدراسة العلاقة بين التدخين والإصابة بالتهاب الرئة، فكانت النتائج الآتية.

التدخين \ الإصابة	مدخن	غير مدخن
مصاب	10	9
غير مصاب	2	10

والمطلوب : قياس شدة العلاقة بين التدخين والإصابة.

12 - قام خبيران بترتيب (6) منتجات من حيث الجودة، فكانت النتائج في الجدول التالي:

رقم المنتج	1	2	3	4	5	6
رأي الخبير A	متوسط الجودة	جيد	كامل الجودة	عديم الجودة	فقر للجودة	فقر للجودة
رأي الخبير B	عديم الجودة	جيد	متوسط الجودة	فقر للجودة	فقر للجودة	كامل الجودة

المطلوب: حساب معامل الارتباط.

13 - لدينا البيانات التالية والمتعلقة بعدد الأطفال والمستوى التعليمي لكل لعدد من الأسر في إحدى المدن:

رقم الأسرة	1	2	3	4	5
المستوى التعليمي	أمية	إعدادية	جامعية	ابتدائية	ثانوية
عدد الأطفال	12	9	3	10	7

والمطلوب: قياس متانة العلاقة بين المستوى التعليمي للأم وعدد الأطفال لهذه الأسر.

14 - أخذت عينة عشوائية من (8) أشخاص من مجتمع كبير من الأشخاص ووجه لكل واحد منهم سؤال عن عمر أبيه وعمر أمه عند ولادته، فكانت النتائج الآتية :

عمر الأم (x)	27	23	18	29	22	19	25	20
عمر الأب (y)	32	24	18	30	20	29	25	23

المطلوب : من خلال هذه البيانات، هل يمكن القول أن هناك علاقة ارتباطية خطية ما بين عمر الأب وعمر الأم؟

الفصل السادس

مبادئ الاحتمالات

— مقدمة

— تعاريف هامة (الاجتماع، التقاطع، المجموعة الخالية، المجموعة الجزئية، العمليات على المجموعات).

— خواص الاحتمالات

— الاحتمال الشرطي

— نظرية بايز

— تمارين غير محلولة



مبادئ الاحتمالات

1-6- مقدمة :

يدخل مفهوم الاحتمال في حياتنا ومعاملاتنا اليومية، كأن نقول، مثلاً: " أنه من المحتمل أن تمطر السماء غداً" أو " أنه من المرجح وصول الطائرات متأخرة هذا المساء"، أو "إن الفرصة جيدة أمام الطالب للنجاح في مادة الإحصاء". إن كل هذه التعبيرات مبنية على مفهوم الاحتمال، أي ترجيح حدوث حدث معين في المستقبل غير مؤكد الوقوع، وبالتالي من خلال الاحتمال نعبر عن مقدار ثقتنا في وقوع هذا الحدث مستقبلاً، فمصطلح "الاحتمال" يعني إذن مقدار ثقتنا في إمكانية حدوث شيء غير مؤكد الوقوع، وسوف نتعرض إلى ثلاث تفسيرات مختلفة لمفهوم الاحتمال .

ـ تفسير الاحتمال بأنه + تكراراً نسبياً:

هذا المفهوم يعني إذا كان عدد الحالات التي يتحقق فيها الحدث (A) هو (m)، وكانت (n) هي العدد الكلي للحالات الممكنة لحدوث الحدث (A)، وبالتالي الاحتمال يعبر عنه بالعلاقة التالية:

$$P(A) = \lim_{n \rightarrow \infty} \frac{m}{n}$$

ويسمى هذا التعريف أحياناً بالتعريف التجريبي لمصطلح "الاحتمال" ، فمثلاً إذا رمينا قطعة نقود (100) مرة وظهرت على الصورة (45) مرة فإن احتمال الحصول على الصورة في هذه الحالة يساوي $\left(\frac{45}{100}\right)$

— التفسير التقليدي للاحتمال:

يبني هذا التفسير على أساس مفهوم النتائج ذات الفرص المتساوية، ويعطي بالعلاقة التالية:

$$p(A) = \frac{\text{عدد الحالات الملائمة}}{\text{عدد الحالات الممكنة}}$$

فمثلاً عدد إلقاء قطعة نقود معدنية متزنة مرة واحدة يكون هناك نتيجتان ممكنتان، إما صورة أو كتابة، وإن مجموع الاحتمالات يجب أن يكون مساوياً الواحد الصحيح.

— التفسير الذاتي للاحتمال:

ينطوي هذا التفسير على أن الشخص الذي يعطي احتمالاً معيناً لنتيجة ما في تجربة معينة، يمثل وجهة نظره من حيث ترجيح حدوث تلك النتيجة، أي أن الحكم مبني على اعتقاد شخصي ومعلومات عن تلك التجربة.

تقوم نظرية الاحتمال عموماً على مفهومين هما:

- التجربة الاحتمالية أو العشوائية (Random Experiment) : وهي كل عملية تؤدي إلى ملاحظة أو قياس ظاهرة ما.
- الحادثة (Event) : هي الناتج من التجربة العشوائية والتي لا يمكن معرفة نتائجها قبل إجرائها.

تعريف :

مجموعة كل النتائج الممكنة لتجربة عشوائية ما تسمى فراغ العينة ويرمز لها بالرمز (S)، والحادث هو مجموعة جزئية من فراغ العينة. ولتوضيح هذا التعريف لدينا المثال التالي :

مثال (1) :

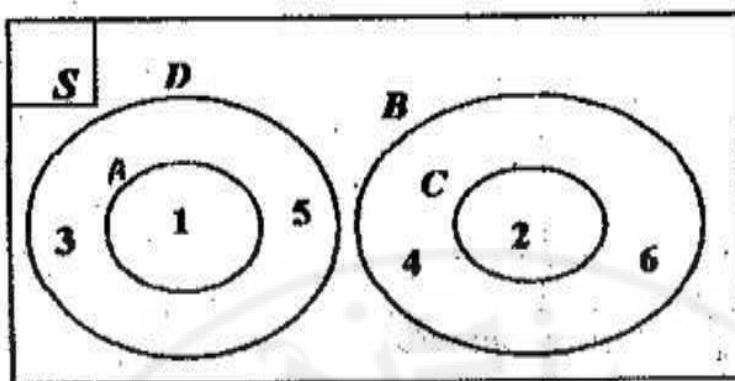
تؤدي تجربة قذف قطعة نقد متوازنة مرة واحدة، إلى إحدى نتيجتين: إما صورة النقد (Head) ويرمز لها عادة بالرمز H أو إلى الكتابة على النقد (Tail) ويرمز لها عادة بالرمز T. عندئذ فإن فراغ العينة لهذه التجربة هو: $S = \{H, T\}$ وبالتالي ستكون نتيجة التجربة إحدى الحالات التالية :

- (a) H هي
- (b) إما H أو T
- (c) كلاهما H, T (وهذا غير محتمل حدوثه) .
- (d) ليست H. أي (ل)

مثال (2) :

نعلم في تجربة قذف حجر النرد، أن هناك 6 نتائج ممكنة، أي أن فراغ العينة لهذه التجربة هو : $S = \{1, 2, 3, 4, 5, 6\}$.
يمكن أن نعرف الحوادث الآتية على فراغ هذه التجربة :
A هي حادثة الحصول على العدد 1 و B هي حادثة الحصول على عدد زوجي و C هي حادثة الحصول على عدد زوجي لا تتجاوز 3 وكذلك D هي حادثة عدد فردي، ويمكن تمثيل تجربة قذف حجر النرد في شكل فن المبين في الشكل المرفق وعلى ذلك يكون لدينا:

- | | | | |
|-----|-------------------|-----|-------------|
| (2) | $B = \{2, 4, 6\}$ | (1) | $A = \{1\}$ |
| (4) | $D = \{1, 3, 5\}$ | (3) | $C = \{2\}$ |



6-2- تعاريف هامة :

— الحادث البسيط : عبارة عن الحادث الذي يحتوي على عنصر واحد من عناصر فضاء العينة.

— الحادث المركب : عبارة عن الحادث الذي يحتوي على عنصرين أو أكثر من عناصر فضاء العينة.

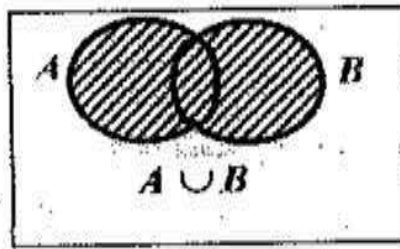
— الحادث الأكيد : عبارة عن الحادث الذي يظهر عند إجراء التجربة (ظهور رقم (5)) عند رمي حجر النرد.

— الحادث المستحيل : عبارة عن الحادث الذي لا يظهر أبداً عند إجراء التجربة ويرمز له (ϕ) (ظهور رقم (7) عند رمي حجر النرد).

— الاجتماع union ورمزه U :

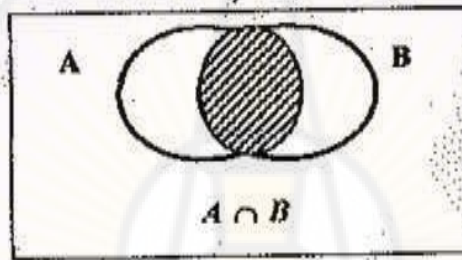
إن اجتماع مجموعتين A و B هو مجموعة العناصر التي تنتمي إلى A أو إلى B أو إليهما معاً. ويرمز لذلك بالرمز $A \cup B$ ، حيث إن $A \cup B$ هو مجموعة العناصر التي تنتمي إلى واحدة منهما على الأقل، ففي مثالنا السابق نجد أن :

$$A \cup B = \{1, 2, 4, 6\} \text{ و } B \cup C = \{2, 4, 6\}$$



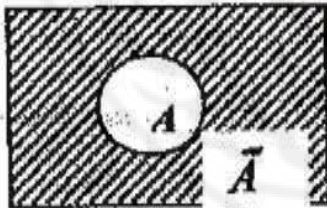
— التقاطع Intersection ورمزہ $\cap$:

إن تقاطع مجموعتين A و B هو مجموعة العناصر التي تنتمي إلى A وإلى B بنفس الوقت ويرمز له بالرمز $A \cap B$ ، ولو عدنا للمثال السابق لوجدنا أن :
 $B \cap C = \{2\}$ و $A \cap B = \phi$ (أحياناً تكتب $A \cap B$ بالصورة AB).



— متممة مجموعة (أو تنمة مجموعة) :

إن متممة مجموعة A ويرمز لها بالرمز A^c (أو $\bar{A}$) .



هي مجموعة العناصر الموجودة في المجموعة الشاملة ولا تنتمي إلى A ، أي أن $\bar{A}$ هي نفي A (أو ليس A).

من الواضح أنه إذا كانت S هي المجموعة الشاملة فإن $A \cup \bar{A} = S$. ففي مثالنا السابق تلاحظ أن $\bar{A} = \{2,3,4,5,6\}$.

— المجموعة الخالية :

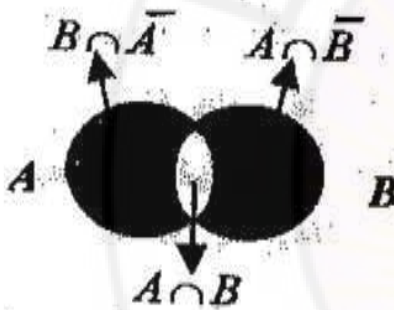
هي تلك المجموعة التي لا تتضمن أي عناصر فيها ويرمز لها بالرمز ϕ أو بالرمز $\{\}$ ونجد في المثال السابق أن $D \cap C = \phi$.

— المجموعات الجزئية :

إذا كان كل عنصر ينتمي إلى A وينتمي أيضاً إلى B ، فيقال أن A محتواة في B ويكتب بالصورة $A \subset B$ (وعند التساوي نكتب $A \subseteq B$) ونقرأ : A محتواة في B أو A مجموعة جزئية في B .

— الفرق بين مجموعتين :

إن الفرق بين المجموعتين A و B هو مجموعة العناصر التي تنتمي إلى A ولا



تنتمي إلى B . ويرمز لذلك بالرمز $A \setminus B$

وهذا موضح في شكل فن المقابل أي :

حيث إن $A \cup B = \overline{A \cap B} = \overline{A} \cup \overline{B}$

$$\overline{A \cap B} = \overline{A} \cup \overline{B}$$

وبالمثل $\overline{A \cap B} = \overline{A} \cup \overline{B}$ وكذلك $\overline{A \cap B} = \overline{A} \cup \overline{B}$

$\overline{A \cap B} = \overline{A} \cup \overline{B}$

العمليات على المجموعات :

إذا كانت لدينا المجموعات الجزئية A, B, C من S ، فإن عمليات

المجموعات تحقق عدة خواص منها الآتي :

— قوانين التبديل :

$$A \cup B = B \cup A$$

$$A \cap B = B \cap A$$

— قوانين الدمج أو التجميع :

$$(A \cup B) \cup C = A \cup (B \cup C)$$

$$(A \cap B) \cap C = A \cap (B \cap C)$$

— قوانين التوزيع :

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

— قوانين دي مورغان :

$$(\bar{A} \cup \bar{B}) = 1 - [(A \cap B)]$$

$$(\bar{A} \cap \bar{B}) = 1 - [(A \cup B)]$$

وهذه القوانين تعطينا العلاقة بين عمليات الاتحاد والتقاطع.

3-3- خواص الاحتمال:

$$(1) \quad 0 \leq P(A) \leq 1 \quad \text{(قيمة تتراوح بين 0 و 1/5 والواحد الصحيح وهي}$$

موجبة دوماً)

$$(2) \quad P(\phi) = 0, P(S) = 1$$

$$(3) \quad \text{مجموع الحالات الممكنة تساوي الواحد الصحيح .}$$

وما يمكن ملاحظته من هذا التعريف هو أن الاحتمال P عدد ليس له وحدات ولا يمكن أن يكون سالب القيمة ولا يمكن أيضاً أن يتجاوز الواحد الصحيح .

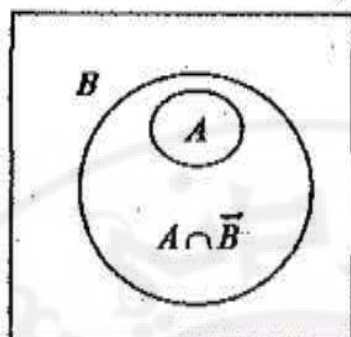
ملاحظات :

• إن احتمال الحادث المتم للحدث A يعطى :

$$P(\bar{A}) = 1 - P(A)$$

• إن احتمال الحادث المستحيل يساوي الصفر، أي: $P(\phi) = 0$:

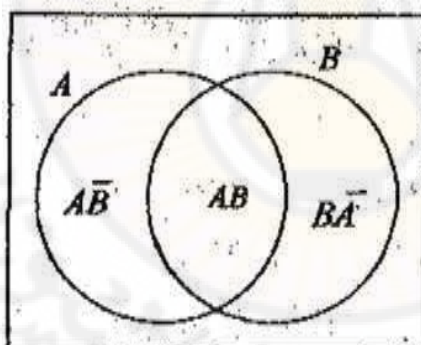
• إذا كانت لدينا الحوادث A, B وكان $A \subset B$ فإن $P(A) \leq P(B)$ ، كما في الشكل التالي (شكل فن) :



• إن احتمال اجتماع حادثين A, B يعطى :

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

انظر شكل فن المرفق :



مثال (3) :

لدينا قاعدتان لإطلاق صواريخ أرض/جو، وفجأة ظهرت طائرة معادية في مجال القاعدتين حيث أطلقت كل قاعدة صاروخاً واحداً على الطائرة، وكان:

$$P(A) = 0.93, P(B) = 0.89 \text{ وكذلك } P(A \cap B) = 0.87$$

المطلوب : أوجد احتمال إصابة الطائرة المعادية بصاروخ واحد على الأقل ؟

الحل :

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \\ = 0.93 + 0.89 - 0.87 = 0.95$$

مثال (4) :

صندوق به 5 كرات اثنتان بيضاوان وكرة واحدة سوداء وكرتان حمراوان، سحب كرة من هذا الصندوق بصورة عشوائية، فما احتمال أن تكون الكرة المسحوبة بيضاء أو سوداء ؟ .

الحل :

لتكن A حادثة الحصول على كرة بيضاء و B حادثة الحصول على كرة سوداء .

بما أن هذه الحوادث متنافية الحدوث أي بيضاء أو سوداء، فإننا نطبق قانون الجمع فنحصل على :

$$-P(A \cup B) = P(A) + P(B) - P(A \cap B) \\ = \frac{2}{5} + \frac{1}{5} - 0 = \frac{3}{5}$$

• ملاحظة : إذا كانت C و B و A ثلاثة حوادث فإن :

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - p(A \cap B)$$

$$P(A \cap C) - P(B \cap C) - P(A \cap B \cap C)$$

3-6- الاحتمال الشرطي Condition Probability :

يقوم هذا الاحتمال على إيجاد احتمالات معينة إذا علم تحقق حوادث أخرى ويُعبر عنه كالتالي: $P(A/B)$ ويقرأ احتمال حدوث (A) علماً بأن الحادث (B) قد وقع، ويعرف بالعلاقة التالية:

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

بحيث : $P(B) > 0$

$$P(A/B) \text{ أحياناً } P(B/A) = \frac{P(A \cap B)}{P(A)} \text{ فإن } P(A) > 0$$

(يقرأ " احتمال حدوث A شريطة أن B قد حدث .

مثال 5 :

لدينا 20 كرة متماثلة من حيث الشكل، وضعت في صندوقين، وكانت (13) منها بلون أحمر (R) و (7) بلون أصفر (Y) تبعاً للجدول الآتي:

اللون	الصندوق الأول (A)	الصندوق الثاني (B)	المجموع
R	5	8	13
Y	3	4	7
المجموع	8	12	20

اختر منها كرة عشوائياً :

- (1) أوجد احتمال أنها حمراء .
- (2) إذا كانت الكرة المختارة من اللون الأحمر، فما هو احتمال أنها من الصندوق الأول؟

الحل :

$$P(R) = (1/2)(5/8) + (1/2)(8/12) \quad (1)$$

$$= (5/16) + (8/24) = 0.64$$

(2)

$$P(A/R) = \frac{P(A \cap R)}{P(R)} = \frac{P(A) P(R/A)}{P(R)}$$

$$= \frac{(1/2)(5/8)}{0.64} = 0.48$$

مثال (6) :

إذا كان لدينا : $P(A) = 0.4$ ، $P(B) = 0.5$ ، $P(A \cap B) = 0.3$

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{0.3}{0.5} = 0.6 \quad \text{فإن :}$$

$$P(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{0.3}{0.4} = 0.75 \quad \text{وبالمثل نجد أن :}$$

خواص الاحتمال الشرطي :

$$P(B) \geq 0 \quad (1)$$

$$P(A/A) = 1 \quad (2)$$

(3) إذا كان $A_1, A_2, \dots$ حوادث متنافية:

$$A_i \cap A_j = \emptyset \quad \forall i \neq j \quad \text{فإن :}$$

$$P(A_1 \cup A_2 \cup \dots / B) = P(A_1/B) + P(A_2/B) + \dots$$

ملاحظات هامة :

(1) من خلال تعريف الاحتمال الشرطي يمكن كتابة :

$$P(A \cap B) = P(A)P(B/A)$$

أو :

$$P(A \cap B) = P(B)P(A/B)$$

4-6- قاعدة بايز Bayes' Theorem :

هي إحدى الطرائق المناسبة لحساب احتمال حادث (E) إذا كان هذا الحادث يأتي من قبل عدة حوادث متعددة، أي علينا تحديد كل مرحلة من مراحل التجربة الإحصائية ولتقريب فكرة هذه القاعدة إلى الأذهان نأخذ المثال الآتي :

مثال (7) :

ثلاثة صناديق تحتوي على كرات متساوية الحجم، الصندوق B_1 يحتوي على كرتين بلون أحمر (R) و 4 بيضاء (W)، والصندوق B_2 يحتوي على كرة

واحدة حمراء (R) ودائرتين بلون أبيض (W)، كما أن الصندوق B_3 يحتوي على 5 كرات حمراء (R) و4 بيضاء (W)، المطلوب: سحب كرة عشوائياً ما احتمال أن تكون حمراء؟
نجد من تعريف الاحتمال الشرطي أن:

$$P(R) = P(B_1)P(R/B_1) + P(B_2)P(R/B_2) + P(B_3)P(R/B_3)$$

$$= \left(\frac{1}{3}\right)\left(\frac{2}{6}\right) + \left(\frac{1}{6}\right)\left(\frac{1}{3}\right) + \left(\frac{1}{2}\right)\left(\frac{5}{9}\right) = \frac{4}{9}$$

سحبنا كرة عشوائياً فكانت حمراء المطلوب: أي صندوق تم اختيارها؟
وللإجابة على هذا التساؤل فإننا نحسب الاحتمالات الشرطية التالية:
 $P(B_3/R)$, $P(B_2/R)$, $P(B_1/R)$ وهي:

$$P(B_1/R) = \frac{P(B_1 \cap R)}{P(R)}$$

$$= \frac{P(B_1)P(R/B_1)}{P(B_1)P(R/B_1) + P(B_2)P(R/B_2) + P(B_3)P(R/B_3)}$$

$$= \frac{(1/3)(2/6)}{(1/3)(1/6) + (1/2)(4/9) + (1/2)(5/6)}$$

$$= \frac{2}{8} = \frac{1}{4}$$

$$P(B_2/R) = \frac{P(B_2)P(R/B_2)}{P(B_1)P(R/B_1) + P(B_2)P(R/B_2) + P(B_3)P(R/B_3)}$$

$$= 1/8$$

الاحتمال الأخير :

$$P(B_3 / R) = \frac{P(B_3)P(R / B_3)}{P(B_1)P(R / B_1) + P(B_2)P(R / B_2) + P(B_3)P(R / B_3)}$$
$$= \frac{(1/2)(5/9)}{\frac{4}{9}} = \frac{5}{8}$$

مثال (8):

تطبع ثلاث سكرتيرات جميع مراسلات مكتب إعلامي تسابع لإحدى وكالات الأنباء ، فإذا كانت السكرتيرة (A) تطبع (40%) و (B) (30%) و (C) تطبع (30%) الباقية، وكان احتمال أن تخطئ في الطباعة السكرتيرة (A) هو (0.02) و (B) هو (0.03) و (C) هو (0.04). سحبت ورقة من مراسلات ذلك المكتب فوجد فيها خطأ، فما احتمال أن تكون السكرتيرة (B) هي التي طبعتها.

الحل:

E : وجود خطأ في الورقة المطبوعة فيكون المطلوب $P(B/E)$

$$P(B / E) = \frac{P(B)P(E / B)}{P(A)P(E / A) + P(B)P(E / B) + P(C)P(E / C)}$$
$$= \frac{(0.30)(0.03)}{(0.40)(0.02) + (0.30)(0.03) + (0.30)(0.04)} = 0.31$$

تمارين غير محلولة

تمرين (1) :

يحتوي صندوق على (4) كرات حمراء مرقمة من (1) إلى (4)، و (3) كرات بيضاء مرقمة من (1) إلى (3)، وجميع هذه الكرات من الحجم نفسه .

والمطلوب :

- أ — سحب بصورة عشوائية كرة من الصندوق، فما احتمال أن تكون هذه الكرة حمراء أو تحمل رقم زوجي ؟
- ب — أعيدت الكرة المسحوبة سابقاً، ثم سحبت الكرات بالتتالي، ما احتمال تعاقب الكرات الحمراء والبيضاء في عملية السحب.

تمرين (2) :

أطلق رامي ثلاث طلقات متلاحقة على هدف ما، وكان احتمال إصابة الهدف (0.6) والمطلوب:

- ما هو احتمال إصابة الهدف بطلقة واحدة فقط .
- احسب عدد الطلقات الواجب رميها على الهدف ليكون احتمال إصابته (0.9) على الأقل.

تمرين (3) :

في أحد المختبرات (10) حاسبات، (6) منها تعمل بلغة البيسك فقط و (3) تعمل بلغة الفورتران فقط و (1) تعمل بكلتا اللغتين، فإذا اختبرت إحدى الحاسبات بصورة عشوائية، ما احتمال:

- أن تكون الحاسبة تعمل بلغة البيسك ولا تعمل بلغة الفورتران.
- أن تكون الحاسبة تعمل بلغة البيسك و الفورتران معاً.

- أن تكون الحاسبة تعمل بلغة البيسك .
- أن تكون الحاسبة تعمل بلغة الفورتران .
- ... أن تكون الحاسبة تعمل بلغة البيسك أو الفورتران .

تمرين (4) :

يحتفظ مشفى بسيارتي إسعاف للطوارئ واحتمال أن تكون السيارة جاهزة للتحرك عند الحاجة إليها هو (0.9)، وإذا علمت أن توفر أحد السيارتين مستقل عن الأخرى .

والمطلوب :

- 1 — أوجد احتمال أن لا تتوفر أيّ منهما .
- 2 — ما هو احتمال تلبية الطلب لسيارة إسعاف في حالة الطوارئ .

تمرين (5) :

إذا كان لدينا احتمال هطول مطر في يوم معين (0.1)، واحتمال وجود رياح نشطة في ذلك اليوم (0.05)، واحتمال وجود مطر ورياح نشطة (0.03).

والمطلوب :

- 1 — أوجد احتمال وجود مطر أو رياح نشطة في ذلك اليوم .
- 2 — أوجد احتمال هطول مطر في ذلك اليوم علماً أن الرياح نشطة .

تمرين (6) :

يتنافس أحمد، محمد، حسين على المرتبة الأولى في امتحان مقرر الإحصاء، فإذا كان احتمال فوز أحمد = $(2/3)$ من احتمال فوز محمد، واحتمال فوز حسين ضعف احتمال فوز محمد، فما هو احتمال فوز كل منهم ؟.

تمرين (9) :

إذا كانت (B, A) حادثتين ، حيث كان $P(A \cup B) = 3/4$ ، $P(B) = 2/5$ ،

$P(A) = x$. أوجد قيمة x في كل من الحالات التالية :

1 - إذا كان A و B متنافيتين .

2 - إذا كان A, B مستقلتين .



الفصل السابع

الاستدلال الإحصائي للعينات الكبيرة الحجم

- مقدمة
- توزيع المعاينة للأوساط الحسابية والنسب المئوية
- التقدير الإحصائي (خصائص التقدير: عدم التحيز، الكفاءة، الكفاية).
- تقدير المتوسط لمجتمع إحصائي، عندما يكون تباين المجتمع معلوماً.
- التقدير المجالي لنسبة المجتمع الإحصائي.
- تقدير الفرق بين متوسطي مجتمعين إحصائيين معلومي التباين.
- تقدير الفرق بين نسبتي مجتمعين إحصائيين.
- اختبار الفرضيات للعينات الكبيرة الحجم.
- اختيار الفرضيات لمتوسط مجتمع تباينه معلوم.
- مفهوم منطقة (الرفض، القبول، الحرجة).
- اختبار الفرضيات لنسبة مجتمع
- اختبار الفرضيات للفرق بين متوسطي مجتمعين
- تحديد حجم العينة
- تمارين غير محلولة

الاستدلال الإحصائي للعينات الكبيرة الحجم

1 - 7 - مقدمة :

يتعين على الباحث عند القيام بدراسة إحصائية أن يقرر ما هو الأسلوب الذي سيعتمده في جمع البيانات التي يحتاج إليها لهذه الدراسة، وما إذا كانت طريقة الجمع ستكون بطريقة الحصر الشامل، أم بطريقة العينة. لقد دلت التجارب أن معظم الأبحاث والدراسات تجري عادة بأسلوب العينات وليس بأسلوب الحصر الشامل لاعتبارات مادية وفنية، وعلى الرغم من تعرض العينة للأخطاء الاحتمالية وأخطاء التحيز فإنه غالباً ما يتم بالأخذ بهذا الأسلوب على اعتبار أن العينة هي جزء من المجتمع الإحصائي المدروس وهذا الجزء قد يعطي فرصة للتحكم والسيطرة على مختلف العوامل التي قد تنشأ عنها أخطاء التحيز، ونحن نستخدم أسلوب العينة من أجل الاستدلال منها على الصفات أو الميزات الرئيسية للمجتمع الإحصائي الذي أخذت منه.

يسمى المقياس الإحصائي المحسوب من العينة بـ (التابع الإحصائي) سواء أكان وسطاً حسابياً أو وسيطاً أو انحرافاً معيارياً، ويسمى نفس المقياس بالمعلمة أو (الثابت الإحصائي) فيما إذا كان محسوباً بدلالة المجتمع الإحصائي، وبفرض سحبنا عينة حجمها n_1 من مجتمع، وحسبنا من هذه العينة مقياساً معيناً، ثم سحبنا عينة ثانية من نفس الحجم ومن نفس المجتمع n_2 وحسبنا منها نفس المقياس وعينة ثالثة ورابعة وهكذا بالنسبة لجميع العينات المتساوية التي يمكن سحبها من هذا المجتمع، سنجد أمامنا عدداً من القيم لنفس المقياس، فإذا كان المقياس المحسوب هو الوسط الحسابي، سنجد أنه لدينا عدداً من الأوساط الحسابية بنفس عدد العينات المسحوبة وهذا بشكل توزيعاً تكرارياً نسميه بتوزيع معانية الأوساط الحسابية ، أما إذا كان المقياس هو الانحراف المعياري ، فإنه

سنجد لدينا عدداً من الانحرافات المعيارية بعدد العينات المحسوبة من المجتمع، وتتغير قيمة هذا التابع الإحصائي بتغير العينة، ولن تكون جميع القيم لهذا التابع الإحصائي التي حصلنا عليها من العينات متساوية، وإنما ستكون مختلفة عن بعضها البعض وتكون توزيعاً احتمالياً يدعى توزيع المعاينة، ويعرف بأنه التوزيع التكراري لأحد التوابع الإحصائية المحسوب في العينات العشوائية ذات الحجم الواحد والتي يمكن سحبها من مجتمع إحصائي واحد أو التوزيع الاحتمالي لجميع القيم الممكنة لإحصائه العينة. مع العلم أن هذا التوزيع الاحتمالي يخدم غرضين هما:

- أ - يساعد في الإجابة على الاحتمالات المتعلقة بإحصاء العينة.
- ب - يعطي الجانب النظري الذي يبرر صحة تطبيق أساليب الاستدلال الإحصائي، وحيث أن أغلب الإحصاءات نجد استخدامها في مجال الاستنتاج الإحصائي هي متوسط العينة وتباينها ونسبة عناصرها التي تحمل صفة معينة.

ملاحظات:

- (1): إن الفرق بين متوسط المجتمع (الثابت) ومتوسط العينة (التابع) يدعى بالخطأ المعياري للوسط الحسابي، مع العلم أن الانحراف المعياري لتوزيع معاينة الأوساط الحسابية، يعطى بالعلاقة التالية:

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$$

- (2): أخطاء المعاينة، عبارة عن الأخطاء الناتجة عن الفروق بين قيم المتوسطات للعينات والقيمة الحقيقية لمتوسط المجتمع.
- (3): إن التوزيع الطبيعي من وجهة النظر الواقعية مستحيل، غير أن نظرية الاحتمالات سمحت لنا بأن نستنتج، أن شكل توزيعات المعاينة للعينات

الكبيرة ($n \geq 30$) يكون لها شكل التوزيع الطبيعي اعتماداً على قانون الأعداد الكبيرة، ونظرية الحد المركزي للإحصاء.

— قانون الأعداد الكبيرة: يعتبر الأساس في نظرية الاحتمالات، و ينص: أنه كلما ازداد حجم العينة، فإن الاحتمال يقترب من اليقين، وأن الفرق المشاهد بين الثابت والتابع يصبح أصغر ما يمكن.

— نظرية الحد المركزي للإحصاء: تنص أنه كلما ازداد حجم العينة، فإن شكل توزيع المعاينة للعينات الكبيرة ($n \geq 30$) يكون لها شكل التوزيع الطبيعي، وإن برهان هذه النظرية خارج نطاق هذا الكتاب، وهي مهمة جداً من الناحيتين النظرية والتطبيقية وذلك لسببين، هما:

أولاً: الشرط في كون التباين محدوداً للتوزيع الذي سيتم منه اختيار العينة شرط محقق في معظم التوزيعات التي نواجهها في التطبيقات العملية.

ثانياً: إن حجم العينة (n) المطلوب لكي يكون التقريب جيد ليس كبيراً إلى حد ما، وكقاعدة عامة إذا كانت $n \geq 30$ فإن توزيع المعاينة للإحصاء ($\bar{X}$) سيكون تقريباً طبيعي إلى حد كبير، وفي الحقيقة إذا تم اختيار العينة من توزيع متصل ووحيد المنوال وملئوياً قليلاً، بأن التوزيع يمكن تقريبه بشكل جيد باستخدام التوزيع الطبيعي حتى إذا كان حجم العينة صغير إلى حد (5) أو (10).

إن أهمية هذه النظرية ستكون واضحة إلينا فيما بعد عندما نعلم أن التوزيع الطبيعي أداة هامة في الاستدلال الإحصائي، علاوة على ذلك نحن على ثقة على الأقل أن توزيع المعاينة لمتوسط العينة ($\bar{X}$) سيكون تقريباً توزيع طبيعي في أي حالة من الحالات الآتية:

1 — عندما تكون العينة من مجتمع إحصائي توزيعه طبيعي.

2 - عندما تكون العينة من مجتمع إحصائي توزيعه غير طبيعي ولكن حجم العينة كبير.

3 - عندما تكون العينة من مجتمع إحصائي صيغة دالته غير معروفة وأمكن حجمها كبير.

2-7- توزيع المعاينة للمتوسط والنسبة المئوية:

إذا كان لدينا مجتمع إحصائي توقعه (μ) وانحرافه المعياري (σ_x) وسحبنا منه جميع العينات العشوائية الممكنة ذات الحجم n ، فإن متوسطات هذه العينات تتوزع تقريباً توزيعاً طبيعياً بمتوسط $\mu_{\bar{x}} = \mu$ وانحراف معياري $\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$ ، أما إذا كانت الظاهرة تخضع لتوزيع ذي الحدين، فإن هذه النسب تتوزع طبيعياً بمتوسط $\mu_{\bar{x}} = P$ وانحراف معياري $\sigma_{\bar{x}} = \sqrt{\frac{Pq}{n}}$ ، وإن قيمة (Z) تعطى بالعلاقة:

$$Z = \frac{\bar{X}_i - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \quad \text{للمتوسط:}$$

$$Z = \frac{P - P'}{\sqrt{\frac{pq}{n}}} \quad \text{للمنسبة المئوية:}$$

وهذه المقادير تنتهي للتوزيع الطبيعي عندما $n \rightarrow \infty$ ، لذا لا يختلف شكل المجتمع كثيراً عن التوزيع الطبيعي عندما $n > 30$ ، أما إذا كان شكل المجتمع الأصلي طبيعياً، فإن توزيع العينة يتبع التوزيع الطبيعي مهما كان حجمها.

3-7 - توزيع معاينة الفرق بين متوسطي مجتمعين وبين نسبتي مجتمعين:

إذا كان لدينا ، مجتمعان إحصائيان متوسطهما (μ_1, μ_2) وانحرافهما δ_1 و δ_2 على الترتيب، وليكن ($\bar{x}_1$) متوسط للعينة (n_1) والمسحوبة من المجتمع الأول و ($\bar{x}_2$) متوسط للعينة (n_2) والمسحوبة من المجتمع الثاني، سيكون توزيع الفرق لـ $\bar{x}_1 - \bar{x}_2$ هو:

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\delta_1^2}{n_1} + \frac{\delta_2^2}{n_2}} \quad \text{(الخطأ المعياري لفرق الأوساط الحسابية):}$$

وبنفس الأسلوب سيكون توزيع الفرق لـ $P_1 - P_2$ هو:

$$\sigma_{P_1 - P_2} = \sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}} \quad \text{(الخطأ المعياري لفرق النسب المئوية):}$$

4-7 - التقدير الإحصائي : (Estimation):

نحن نعلم أن الاستدلال الإحصائي يعالج مسألتين هامتين هما اختبار الفروض والتقدير الإحصائي الذي يعرف بأنه استنتاج المعالم (الثوابت) والخصائص الأساسية للظواهر التي تخص المجتمع بدلالة المقاييس الإحصائية المحسوبة من العينة (الإحصاءات) مع الأخذ بعين الاعتبار للأخطاء الاحتمالية. والمقدر هو إحصاءة تحدد كيفية استخدام بيانات العينة كتقدير معلمة المجتمع الإحصائي المجهولة، وإن للمبدأ من وراء استخدام التقدير في المجالات العلمية المختلفة يستند على الافتراض بأن العاملين في هذه المجالات بحاجة لمعرفة بعض سمات المجتمع الإحصائي مدار البحث هذه السمات متمثلة في المعلمات، مثل: المتوسط والتباين ونسبة عناصر المجتمع الإحصائي التي تحمل صفة معينة، لذلك هناك سببين يبرران اعتماد الباحث أساليب التقدير للحصول

على معلومات تتعلق بهذه المعلومات وهما: أن العديد من المجتمعات ذات الاهتمام بالرغم من أنها محدودة ولكنها كبيرة وبالتالي تفحص جميع مفرداتها أمر غير ممكن من حيث التكاليف، ثانياً أن هناك مجتمعات لا نهائية أو غير محدودة وبالتالي تفحص جميع مفرداتها أمراً مستحيلاً، ونحن نميز:

— التقدير النقطي (Point estimation): عبارة عن تقدير قيمة الثابت الإحصائي برقم واحد بدلالة العينة المسحوبة عشوائياً من مجتمع ما.

— التقدير المجالي (Interval estimation): عبارة عن تقدير الثابت الإحصائي بعدد متناه من النقاط، أي يقع ضمن مجال محدد من القيم عند احتمال معين بدلالة التابع الإحصائي.

كأن نقول: أن متوسط دخل العامل في مصنع ما هو 15000 ليرة أو نقول: أن متوسط دخل العامل في مصنع ما يتراوح ما بين 10000 إلى 18000 ليرة، حيث استخدمنا في التعبير الأول التقدير النقطي، أي أننا قدرنا متوسط الدخل بقيمة واحدة، أما في التعبير الثاني، فقد قدرنا متوسط الدخل بمجال، نعتقد أن القيمة الحقيقية تقع داخل هذا المجال، مع العلم أن التقدير المجالي أكثر واقعية، حيث أنه يعطي سلسلة من القيم بدلاً من قيمة واحدة.

• خصائص التقدير الأمثل :

• عدم التحيز Unbiased estimators :

تعرف خاصية عدم التحيز على أنها: إذا كانت القيمة المتوقعة للمقدر مساوية للقيمة الحقيقية، فإن المقدر يسمى مقدراً غير متحيزاً.

• الكفاءة (الفعالية) Efficiency :

يمكن القول أن التقدير الأقل اختلافاً أو الأقل تشتتاً هو التقدير الأفضل، بمعنى كلما قل التباين زادت فعالية المقدر، وبالتالي زادت الثقة أكثر في الثابت الإحصائي.

• الكفاية : Sufficiency

تعني أن التقدير يحتوي على جميع المعلومات التي توفرت من العينة ولا يمكن لأي تقدير آخر أن يحتوي على معلومات ليست موجودة بالتقدير الأصلي، فيقال أن هذا التقدير كافٍ للمعلمة.

أخيراً تجدر الإشارة هنا، إلى أنه عادة ما يستخدم ما يسمى بالخطأ المعياري كمقياس لدقة التقدير، والذي عُرف سابقاً بأنه الانحراف المعياري للإحصاءة مثلاً الخطأ المعياري للمتوسط هو الانحراف المعياري لمتوسط العينة $\left(\sigma_x = \frac{\sigma}{\sqrt{n}}\right)$ وكلما زاد حجم العينة، كلما كان الخطأ المعياري صغيراً والعكس بالعكس، وبعبارة أخرى، كلما كانت قيمة الخطأ المعياري صغيرة كلما كانت العينات أكثر صدقاً في تمثيل المجتمع الإحصائي المأخوذة منها، ولهذا السبب عادة ما يتخذ الخطأ المعياري دليلاً على درجة تمثيل العينة لمجتمعها وعندما تكون (σ) مجهولة فإننا نستعوض عنها بقيمة الانحراف المعياري للعينة (S) عند حساب قيمة الخطأ المعياري.

7-5- تقدير المتوسط لمجتمع إحصائي (μ) عندما يكون تباين المجتمع σ^2 معلوماً:

إذا أخذت عينة عشوائية $X_1, X_2, \dots, X_n$ حجمها $n > 30$ من مجتمع طبيعي متوسطه غير معلوم μ وتباينه σ^2 معلوم، فيمكن إنشاء مجال احتمالي حول (x) باحتمال $(1-\alpha)$ على النحو التالي:

$$P\left(-Z_{\frac{\alpha}{2}} < \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} < Z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

$$-Z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right) \leq \bar{X} - \mu \leq Z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

إذا كان الاحتمال للأول هو $(1-\alpha)$ ، فإن الاحتمال للآخر يجب أن يكون أيضاً مساوياً لنفس الاحتمال أي $(1-\alpha)$ هذا يعني أن :

$$P\left[-Z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right) \leq \bar{X} - \mu \leq Z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)\right] = 1 - \alpha$$

ومن هنا فإن الاحتمال (عند درجة ثقة $(1-\alpha)\%$) هو :

$$\left[\bar{X} - Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \bar{X} + Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right] = \bar{X} \pm Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

سوف يتضمن المتوسط غير المعلوم (μ) حسب نظرية النهاية المركزية عندما $(n \geq 30)$.

تمرين :

سحبت عينة عشوائية حجمها $n=37$ مصباحاً من إنتاج أحد المصانع، فكان متوسط عمر إضاءة المصباح $\bar{X}=1478$ ساعة وبانحراف معياري قدره $S=36$ والمطلوب :

قدر باحتمال 95% متوسط عمر إضاءة المصابيح في المجتمع (مجتمع المصابيح) .

$$P\left[\bar{x} - z_{0.025}\left(\frac{\sigma}{\sqrt{n}}\right) \leq \mu \leq \bar{x} + z_{0.025}\left(\frac{\sigma}{\sqrt{n}}\right)\right] = 1 - \alpha$$

$$P\left[1478 - 1.96\left(\frac{36}{\sqrt{37}}\right) \leq x \leq 1478 + 1.96\left(\frac{36}{\sqrt{37}}\right)\right] = 0.95$$

$$P[1478 - 11.60 \leq \mu \leq 1478 + 11.60] = 0.95$$

$$P[1478 - 11.60 \leq \mu \leq 1478 + 11.60] = 0.95 \Rightarrow [1466.4, 1489.6]$$

وبالتالي يمكن القول أن متوسط عمر المصابيح في المصنع لن يقل عن (1466.4) ولن يزيد عن (1489.6) ساعة

ويمكن تفسير ذلك، أنه لو أخذنا (37) عينة عشوائية وحسبنا فترة الثقة من كل عينة لوجدنا أن 95 فترة تتضمن قيم الوسط الحسابي للمجتمع μ ، و 5 فترات لا تتضمن تلك القيم، مع العلم أنه يمكن إنشاء فترات ثقة أخرى عند مستويات مختلفة بإتباع نفس الأسلوب واستخدام جدول التوزيع الطبيعي المعياري للحصول على قيم Z ، إلا أن أكثر فترات الثقة شيوعاً واستخداماً هي:

• عند 90% مستوى ثقة يكون :

$$1 - \alpha = 0.90 \Rightarrow \alpha = 0.10 \Rightarrow Z_{0.10/2} = Z_{0.05} \Rightarrow \bar{x} \pm 1.645 \frac{\sigma}{\sqrt{n}}$$

• عند 95% مستوى ثقة يكون :

$$1 - \alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow Z_{0.05/2} = Z_{0.025} \Rightarrow \bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}$$

• عند 99% مستوى ثقة يكون :

$$1 - \alpha = 0.99 \Rightarrow \alpha = 0.01 \Rightarrow Z_{0.01/2} = Z_{0.005} \Rightarrow \bar{x} \pm 2.58 \frac{\sigma}{\sqrt{n}}$$

من خلال ذلك، نجد أن زيادة مستوى الثقة يصاحبه زيادة في قيم Z ، الأمر الذي يعني أن مجال الثقة يصبح أكثر وسعاً كلما زاد مستوى الثقة.

تمرين :

أخذت عينة عشوائية تتكون من (100) شخص أعمارهم (25) سنة يبدو أنهم طبيعيين، فوجد أن متوسط ضغط الدم الانقباضي لديهم (125) فإذا كان هناك اعتقاداً بأن الانحراف المعياري لضغط الدم الانقباضي لمجتمع هذه العينة يساوي (15).

المطلوب: أوجد (95%) فترة ثقة لتقدير متوسط ضغط الدم الانقباضي $\bar{X}$ بهذا المجتمع ؟

$$125 - (1.96) \left(\frac{15}{\sqrt{100}} \right) \leq \mu \leq 125 + (1.96) \left(\frac{15}{\sqrt{100}} \right) = 1 - \alpha$$

$$125 - 2.94 \leq \mu \leq 125 + 2.94$$

$$122.06 \leq \mu \leq 127.94$$

أي أن متوسط ضغط الدم الانقباضي للأشخاص الذين أعمارهم (25) سنة لا يقل عن (122.06) ولا يزيد عن (127.94) بالاعتماد على بيانات العينة.

ملاحظة :

إن حجم العينة دالة في طول فترة الثقة $(1-\sigma)$ والانحراف المعياري (σ) وذلك كما يلي:

أ — كلما زادت σ ، كلما زاد حجم العينة المطلوبة، وذلك بثبات حدود فترة والثقة المحددة.

ب — كلما زاد مستوى الثقة، كلما زاد حجم العينة، وذلك بثبات حدود فترة الثقة والانحراف المعياري.

ج — كلما قل طول فترة الثقة المطلوبة، كلما زادت حجم العينة، وذلك بثبات (σ) والثقة المحددة لها.

6 - 7 — التقدير المجالي لنسبة المجتمع الإحصائي (P) :

إذا سحبنا عشوائية $(n \geq 30)$ من مجتمع إحصائي وحسبنا النسبة المئوية $(P=x/n)$ ، فإنه بالإمكان تقدير الحدود التي تقسع ضمنها النسبة المئوية للمجتمع ، وذلك من خلال إنشاء مجال احتمالي حول P' باحتمال $(1-\alpha)$ على النحو التالي:

$$P \left[-Z_{\frac{\alpha}{2}} \leq \frac{(x/n) - P}{\sqrt{P(1-P)/n}} \leq Z_{\frac{\alpha}{2}} \right] = 1 - \alpha$$

$$P \left[\frac{x}{n} - Z_{\frac{\alpha}{2}} \sqrt{\frac{P(1-P)}{n}} \leq P \leq \frac{x}{n} + Z_{\frac{\alpha}{2}} \sqrt{\frac{P(1-P)}{n}} \right] = 1 - \alpha$$

حيث (P) غير معلوم. بفرض أن التقدير النقطي لـ P هو $P' = y/n$ فنكون فترة الثقة للمعلمة P هي :

$$P' \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{P'(1-P')}{n}}$$

يمكن كتابة العلاقة السابقة على النحو التالي:

$$P \left[\frac{P'}{n} + Z_{\frac{\alpha}{2}} \sqrt{\frac{P'(1-P')}{n}} \leq P \leq P' - Z_{\frac{\alpha}{2}} \sqrt{\frac{P'(1-P')}{n}} \right] = 1 - \alpha$$

تمرين :

في استطلاع للرأي ، تم سؤال عينة مكونة من (200) شخص عن أهم مصدر للمعلومات لديهم فأجاب (36) منهم بأنه التلفزيون، والمطلوب : قَدِّر باحتمال (95%) نسبة الأفراد في المجتمع الذين يعتبرون التلفزيون أهم مصدر للمعلومات لديهم؟

$$P' = \frac{x}{n} = \frac{36}{200} = 0.18$$

نحن نعلم أن :

$$1 - \alpha = 0.95 \Rightarrow Z_{\alpha/2} = Z_{0.025} = 1.96$$

$$0.18 \pm 1.96 \sqrt{\frac{(0.18)(0.82)}{200}} \Rightarrow [0.127, 0.233]$$

وبالتالي يمكن القول أن نسبة الأفراد في المجتمع الذين يعتبرون التلفزيون أهم مصدر للمعلومات لديهم لن تقل عن (0.127) ولن تزيد عن (0.233).

تمرين :

إذا علمت أن (28) شخص من بين (50) شخص تم اختيارهم بطريقة عشوائية يفضلون التدخين في الأماكن العامة، المطلوب: أوجد (90%) فترة ثقة حول نسبة الأفراد الذين يفضلون التدخين في العامة بالاعتماد على نتائج العينة.

الحل :

$$P = \frac{28}{50} = 0.56$$

$$0.56 - (1.65) \sqrt{\frac{(0.56)(0.44)}{50}} \leq P \leq 0.56 + 1.65 \sqrt{\frac{(0.56)(0.44)}{50}}$$

$$0.445 \leq P \leq 0.675$$

وبالتالي يمكن القول بأننا بثقة قدرها (90%) بأن نسبة الأفراد الذين يفضلون التدخين في الأماكن العامة بالمجتمع مدار البحث تقع ما بين (0.445 و 0.675) أي أن هناك اتجاه لعدم تفضيل التدخين بالأماكن العامة وذلك لأن النسبة أقل من (0.50).

7-7 - تقدير الفرق $(\mu_1 - \mu_2)$ بين متوسطي مجتمعين إحصائيين معلومي التباين:

لنفرض أن عينتين عشوائيتين بحجم n_1, n_2 قد سحبتا من مجتمعين مستقلين طبيعيين غير معلومة متوسطاتها μ_1, μ_2 . فإذا كان $\bar{X}_1$ هو متوسط العينة الأولى و $\bar{X}_2$ متوسط العينة الثانية. فإن تقدير الفرق $\mu_1 - \mu_2$ هو $\bar{X}_1 - \bar{X}_2$. وللحصول على تقدير فترة للفرق $\mu_1 - \mu_2$ ، فإننا سنتعامل مع الحالة التي يكون فيها تباين المجتمعين σ_1^2, σ_2^2 معلومين. وأن العينتين عشوائيتين مأخوذتين من مجتمع يتوزع طبيعياً وإن $\bar{X}_1 - \bar{X}_2$ تتوزع طبيعياً أيضاً وكذلك الفرق $\bar{X}_1 - \bar{X}_2$ له توزيع طبيعي بوسط $(\mu_1 - \mu_2)$ وتباين $\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$ وبناءً على ذلك يمكن كتابة العلاقة التالية فإن الإحصاء تعطى بالعلاقة التالية :

$$Z = \left[(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2) \right] / \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

تخضع للتوزيع الطبيعي $N(0,1)$. وعليه من أجل الاحتمال $(1-\alpha)$ يمكن كتابة العلاقة السابقة على النحو التالي :

$$P \left[-Z_{\frac{\alpha}{2}} \leq \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \leq Z_{\frac{\alpha}{2}} \right] = 1 - \alpha$$

$$P[(\bar{X}_1 - \bar{X}_2) - Z_{\frac{\alpha}{2}} \sigma_w \leq (\mu_1 - \mu_2) \leq (\bar{X}_1 - \bar{X}_2) + Z_{\frac{\alpha}{2}} \sigma_w] = 1 - \alpha$$

حيث أن : $\sigma_w = \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}$ وتكون فترة الثقة للفرق $(\mu_1 - \mu_2)$ وبمعلومية

(σ_1, σ_2) على الصورة التالية :

$$P[(\bar{X}_1 - \bar{X}_2) - Z_{\frac{\alpha}{2}} \sigma_w \leq (\mu_1 - \mu_2) \leq (\bar{X}_1 - \bar{X}_2) + Z_{\frac{\alpha}{2}} \sigma_w] = 1 - \alpha$$

ووفقاً لنظرية النهاية المركزية فإن النتيجة يمكن استخدامها لعينات عشوائية مستقلة من مجتمعات غير طبيعية معروفة تبايناتها σ_1^2, σ_2^2 شريطة أن يكون حجم العينات كبيراً كفاً $(n_1, n_2 \geq 30)$.

وعندما تكون σ_1^2, σ_2^2 غير معلومة ولكن أحجام العينات كبيرة $(n_1, n_2 \geq 30)$ فإنه يتم استبدال σ_1^2, σ_2^2 بتقديراتهم غير المتحيزة S_1^2, S_2^2 ، على الترتيب

ويصبح لدينا في هذه الحالة $S_w = \sqrt{S_1^2/n_1 + S_2^2/n_2}$ وعليه فإن فترة

الثقة للفرق $\mu_1 - \mu_2$ تكون :

$$[(\bar{X}_1 - \bar{X}_2) - Z_{\frac{\alpha}{2}} S_w, (\bar{X}_1 - \bar{X}_2) + Z_{\frac{\alpha}{2}} S_w] = (\bar{X}_1 - \bar{X}_2) \pm Z_{\frac{\alpha}{2}} S_w$$

تمرين :

سحبت عينتان عشوائيتان من طلاب إحدى الكليات لدراسة درجاتهم بمقرر الإحصاء فكانت النتائج الآتية:

والمطلوب: $n_1 = 150$, $n_2 = 80$, $\bar{x}_1 = 75,3$, $\bar{x}_2 = 70$, $S_1^2 = 60$, $S_2^2 = 40$

قدر باحتمال (90%) الفرق الحقيقي بين وسط درجات هؤلاء الطلاب.

الحل :

$$1.645\sigma_w = 1.645\sqrt{\frac{60}{150} + \frac{40}{80}}$$
$$= 1.560$$

$$[5.3 - 1.560, 5.3 + 1.560] = [3.74 + 6.86]$$

وبالتالي يمكن القول أنه باحتمال (90%) فإن الفرق بين متوسطي

المجتمعين لن يقل (3.74) ولن يزيد عن (6.86) درجة .

تمرين :

إذا علمت أنه تم اختيار عينة عشوائية تتكون من (22) تلميذ من تلاميذ الصف الأول الابتدائي في إحدى المدارس، كان متوسط الأطوال لهذه العينة (119.38) سم ثم اختيرت عينة عشوائية تتكون من (25) تلميذ من تلاميذ الصف الثاني الابتدائي وكان متوسط الأطوال (126) سم والانحراف المعياري لأطوال مجتمعي تلاميذ الصف الأول والثاني يساوي (4.5) سم و (5.13) سم على التوالي.

المطلوب :

أوجد (90%) فترة ثقة حول الفرق ما بين متوسطي أطوال مجتمعي

التلاميذ بالصف الأول والثاني الابتدائي.

الحل :

$$1.645\sigma_w = 1.645\sqrt{\frac{(4.5)^2}{22} + \frac{(5.13)^2}{25}}$$

$$= 1.645(1.405)$$

$$(119.38 - 126) - 2.311 \leq \mu_1 - \mu_2 \leq -6.62 + 2.311$$

$$-8.931 \leq \mu_1 - \mu_2 \leq -4.309$$

حيث أن حدي فترة الثقة سالبين، فإن ذلك يعني أن أطوال تلاميذ الصف الثاني أكبر من أطوال تلاميذ الصف الأول بصفة عامة.

8 - 7 - تقدير الفرق بين نسبتي مجتمعين إحصائيين:

في هذه الحالة يكون لدينا مجتمعان بحيث ينصب الاهتمام في استقراء وجود فروق بين النسب للمجتمعين بالاعتماد على عينتين عشوائيتين بسيطتين مستقلتين، بفرض أن: نسبة المجتمع الأول هي P_1 ونسبة المجتمع الثاني هي P_2 وأن $p'_1 = \frac{X_1}{n_1}$ هو تقدير لنسبة المجتمع الأول، وأن $p'_2 = \frac{Y_2}{n_2}$ هو أيضاً تقدير لنسبة المجتمع الثاني، حيث أن (n_1, n_2) عينتين عشوائيتين قد سحبنا من مجتمعين مستقلين طبيعيين، وأن تقدير الفرق $P_1 - P_2$ هو $(P'_1 - P'_2)$ ، وبناءً على ذلك، نوجد فترة الثقة للفرق $P_1 - P_2$ من خلال :

$$Z = \frac{(x_1 / n_1) - (x_2 / n_2) - (P_1 - P_2)}{\sqrt{P_1(1 - P_1) / n_1 + P_2(1 - P_2) / n_2}}$$

تخضع للتوزيع الطبيعي $N(0,1)$. وباحتمال $(1-\alpha)\%$ إنشاء المجال الاحتمالي حول $(P'_1 - P'_2)$ على النحو التالي:

$$P \left[(P'_1 - P'_2) \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{P'_1(1-P'_1)}{n_1} + \frac{P'_2(1-P'_2)}{n_2}} \right] = 1 - \alpha$$

تمرين:

لدراسة موقف العمال من الخصخصة، أخذت عينتان عشوائيتان
($n_1=100$) عامل ($n_2=120$) عاملة في إحدى الصناعات، فوجد أن (40) أيدوا
الموقف ، فيما أيدت (60) عاملة هذا الموقف.
والمطلوب : أوجد حدود الثقة (95%) (قدر باحتمال 95%) الفرق الحقيقي
لمواقف العمال والعاملات في هذه الصناعة من الخصخصة.

الحل :

$$\begin{array}{ll} \text{العمال} & P_1=40/100=40\% \quad q_1=60\% \\ \text{العاملات} & P_2=60/120=50\% \quad q_2=50\% \end{array}$$
$$P \left[(0.5 - 0.4) \pm 1.96 \sqrt{\frac{(0.4)(0.6)}{100} + \frac{(0.5)(0.5)}{120}} \right] = 0.95$$
$$P [0.04 , 0.24] = 0.95$$

وبالتالي يمكن القول أنه باحتمال (95%) فإن الفرق بين نسبيتي
المجتمعين لن يقل عن (0.04) ولن يزيد عن (0.24).

تمرين :

أجريت دراسة عن الاهتمام بصحة الأسنان، حيث اختيرت عينة
عشوائية من (500) شخص وجد من بينهم (220) مستواهم التعليمي شهادة
متوسطة و (280) مستواهم التعليمي شهادات جامعية وسئل كل منهم عن سبب
آخر زيارة لطبيب الأسنان، فكانت إجابة (44) شخص من المجموعة الأولى و
(150) من المجموعة الثانية أن السبب هو إجراء وقائي. المطلوب:
أوجد (95%) فترة ثقة لتقدير الفرق ما بين نسبيتي المجتمعين

الحل :

$$P_1 = \frac{44}{220} = 0.20$$

$$P_2 = \frac{150}{280} = 0.5$$

$$\sqrt{\frac{(0.2)(0.8)}{220} + \frac{(0.54)(0.46)}{280}}$$

$$= 0.0402 \quad \text{وعليه يكون:}$$

$$(0.2 - 0.54) - (1.96)(0.0402) \leq P_1 - P_2 \leq (0.2 - 0.536) + (1.95)(0.402)$$

$$-0.41 \leq P_1 - P_2 \leq -0.25$$

حيث أن مدى فترة الثقة سالبة فهذا يعني أن نسبة الأشخاص الذين يترددون على طبيب الأسنان لأسباب وقائية ومستواهم التعليمي جامعي أكثر من نسبة الذين مستواهم التعليمي متوسط.

7-9 - اختبار الفرضيات للعينات الكبيرة الحجم:

إن اختبار الفروض عبارة عن تحديد عما إذا كانت بيانات العينة تؤيد اعتقاداً معيناً في المجتمع أم لا، أي مدى تطابق الإحصاءات مع المعلومات والفروق هي أخطاء احتمالية، ومن ثم اتخاذ القرار المتعلق بالمجتمع الإحصائي من خلال بيانات العينة للتأكد من صحة أو عدم صحة تلك الفروض المتعلقة بالمجتمع من خلال الاختبار الإحصائي (اختبار المعنوية). والجديد بالملاحظة هنا، أن الفرضية دائماً جملة حول المجتمع أو التوزيع قيد الدراسة وليست جملة حول العينة.

ويقصد باختبار المعنوية، تحديد درجة الثقة بنتائج استخدام المقاييس الكمية للعينة المأخوذة من المجتمع ومعرفة مصداقية العينة في تمثيل المجتمع الذي سحبت منه، ولتور أخطاء الحظ والصدف في النتائج .

الثابت الإحصائي - التابع الإحصائي

$$Z = \frac{\text{التابع الإحصائي}}{\text{الخطأ المعياري للتابع الإحصائي}}$$

الخطأ المعياري للتابع الإحصائي

• أنواع الفروض:

الفرضية، عبارة عن مقولة تتعلق بالمجتمع الإحصائي تحمل الصواب أو الخطأ أو تخمينات تتعلق بالمجتمع وتنقسم إلى قسمين:

— فروض البحث: عبارة عن مزيج من التخمين والشعور حول شيء ما
مثال: افتراض المستوى التعليمي للفرد مرتبطاً ارتباطاً مباشراً بمستوى الأداء،
أي أن هناك شعور بأن الأفراد الأوفر تعليماً هم الأفضل في مستوى الأداء.

— الفروض الإحصائية: عبارة عن فرضيتين:

• الفرضية الصفرية أو الابتدائية (Null Hypothesis) ويرمز لها عادة بالرمز H_0 ، هي الفرضية الأساسية التي نرغب في اختبارها ويفترض صحتها.
• الفرضية البديلة (Alternative Hypothesis) ويرمز لها بالرمز H_1 ، وهي الفرضية التي تمثل البديل في حالة رفضت فرضية العدم (H_0) مع العلم أن الفرض الإحصائي قد يكون بسيطاً أو قد يكون مركباً أي لا يحدد توزيع المتغير بشكل كامل، وإن الفرضية H_0 سترفض لصالح H_1 فقط، إذا توفر دليل قاطع من خلال العينة على أن H_0 خاطئة، وإذا لم يتوفر ذلك الدليل فلا يمكن رفضها، ويمكن القول: أن النتيجتين المحتملتين هما إما رفض H_0 أو قبولها، وليبيان صحة أو خطأ فرضية ما، يمكن اختيار عينة عشوائية من المجتمع المراد دراسته وإجراء الاختبار عليها، لو عدنا للمثال السابق فإن فرضية العدم هي عدم وجود علاقة بين مستوى التعليم للفرد ومستوى الأداء ونحن علينا إثبات صحة أو عدم صحة فرضية العدم، ونحن بشكل عام نصيغ فرضية العدم غالباً بحسن النية.

والجدير بالملاحظة أيضاً، أن قيمة معلمة المجتمع الإحصائي المحددة في فرض العدم عادة ما يتم تحديدها بإحدى الطرائق الثلاثة الآتية:

أولاً : أن تكون، ناتجة من خبرة سابقة أو من تجربة مسبقة، وفي هذه الحالة يكون الهدف من اختبار الفرضية، هو تحديد فيما إذا كان هناك تغير في ظروف التجربة أم لا.

ثانياً: أن تكون محددة من بعض النظريات أو من نموذج يتعلق بالتجربة محل البحث، وفي هذه الحالة يكون الهدف من اختبار الفرضية، هو التحقق من صحة النظريات أو النموذج.

ثالثاً: تكون قيمة معلمة المجتمع الإحصائي ناتجة من عوامل خارجية مثل: المواصفات المطلوبة في إنتاج معين، وفي هذه الحالة يكون الهدف من اختبار الفرضية، هو التحقق من مطابقة الإنتاج للمواصفات المطلوبة أو عدم المطابقة.

إن ما يهمنا هنا هو الوصول إلى قرار حول صدق أو عدم صدق الفرضية، والأسلوب الذي يؤدي إلى ذلك القرار يسمى (اختبار الفرضية)، وإن هذا الأسلوب يعتمد على استخدام المعلومات التي تم الحصول عليها من العينة العشوائية المختارة من المجتمع الإحصائي محل البحث، فإذا كانت هذه المعلومات متناسقة مع الفرضية، فإن ذلك يؤدي إلى خلاصة مفادها أن: الفرضية صحيحة، ولكن إذا كانت غير متناسقة معها، فإن ذلك يؤدي إلى القول بأن الفرضية خاطئة.

بصفة عامة لاختبار أي فرضية يجب اختيار عينة عشوائية من مجتمع مدار البحث واستخدام المعلومات المتوفرة بها لحساب إحصاءة تساعد في اتخاذ القرار.

في اختبار الفرضيات ونتيجة لاستخدام معلومات عينة بدلاً من المجتمع كله، فإن الاستقرارات تكون عرضة للوقوع في نوعين من الخطأ هما:

- خطأ من النوع الأول يسمى الخطأ الناتج عن رفض H_0 عندما تكون صحيحة، (Type I Error) .
- خطأ من النوع الثاني يسمى الخطأ الناتج عن قبول H_0 عندما تكون خاطئة،

(Type II Error)

علماً أنه لا توجد أي طريقة معلومة تضمن عدم الوقوع في أي من الخطأين. ويمكن إعادة صياغة ما سبق ذكره في شكل جدول، كما هو مبين :

نوع القرار	طبيعة الفرضية	
	قبول H_0	رفض H_0
H_0 صحيحة	قرار صائب	خطأ من النوع الأول Type I error
H_0 خاطئة	خطأ من النوع الثاني Type II error	قرار صائب

ويسمى احتمال الوقوع في خطأ من النوع الأول (يرمز له عادة بالرمز α) بمستوى دلالة الاختبار أو المعنوية ويرمز لاحتمال وقوع خطأ من النوع الثاني بالرمز β .

7-10- اختبار الفرضيات لمتوسط مجتمع:

- تبين المجتمع σ^2 معلوم :

في هذه الحالة، فإن قيمة الاختبار Z تخضع للتوزيع الطبيعي المعياري، وتعطى بالعلاقة التالية:

$$Z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \sim N(0,1)$$

حيث أن $\bar{x}$ هو متوسط العينة العشوائية المأخوذة من المجتمع قيد الدراسة ، وأن μ_0 متوسط افتراضي (أو ابتدائي) للمجتمع المراد اختبار متوسطه

الحقيقي μ وأن σ^2 تبين هذا المجتمع كما أن $\sigma/\sqrt{n}$ هي الانحراف المعياري لمتوسط العينة $\bar{x}$ وتسمى الخطأ المعياري للمتوسط ويرمز له بالرمز $\sigma_{\bar{x}}$ ، وأن الفرضيات المراد اختبارها من أجل متوسط المجتمع μ يمكن أن تأتي في ثلاث إمكانات، إما أن تكون μ أكبر أو أصغر أو لا تساوي القيمة الابتدائية μ_0 ، ويمكن تلخيص اختبار الفرضيات حول متوسط مجتمع في الجدول التالي :

المنطقة التي ترفض فيها H_0	الفرضية البديلة	الفرضية الصفرية
منطقة الرفض أو المنطقة الحرجة	H_1	H_0
$\bar{x} \geq \mu_0 + Z_{\alpha} \frac{\sigma}{\sqrt{n}} \quad , \quad Z \geq Z_{\alpha}$	$\mu > \mu_0$	$\mu = \mu_0$
$\bar{x} \leq \mu_0 - Z_{\alpha} \frac{\sigma}{\sqrt{n}} \quad , \quad Z \leq -Z_{\alpha}$	$\mu < \mu_0$	$\mu = \mu_0$
$ \bar{x} - \mu_0 \geq Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \quad , \quad Z \geq Z_{\alpha/2}$	$\mu \neq \mu_0$	$\mu = \mu_0$

حيث أن $Z \geq Z_{\alpha}$ و $Z \leq -Z_{\alpha}$ و $|Z| \geq Z_{\alpha/2}$ هي على الترتيب اختبار من طرف أيمن أو من طرف أيسر أو من طرفين.

• الإجراءات المتبعة لاختبار فرضيات الدراسة:

لغرض التحقق من صحة فرضيات الدراسة يتم تحديد مستوى المعنوية المناسب من قبل الباحث عند مستوى 5% مثلاً، وهو المستوى الذي يمثل القيمة القصوى لاحتمال ارتكاب خطأ من النوع الأول كما ذكرنا سابقاً، أي: احتمال رفض الفرض الصفرية " فرض العدم" وهو في الواقع صحيح، واختبار الفرضية البحثية عادة يتم إتباع الإجراءات التالية:

1 - صياغة الفرضية البحثية في صورة فرضية إحصائية يراد اختبارها، وهي غالباً ما تأخذ شكل فرضية صفرية (H_0) تصاغ على أمل

رفضها، وفرضية بديلة (H_1) يتم الأخذ بها عند رفض الفرضية الصفرية.

2 - تحديد مستوى المعنوية أو الدلالة (α).

3 - تحديد قيمة مستوى المعنوية المشاهدة (P-value) وهي تعرف على أنها " أصغر قيمة لمستوى المعنوية (α) التي يمكن عندها رفض الفرضية الصفرية.

4 - المقارنة واتخاذ القرار الإحصائي، وذلك من خلال المقارنة بين مستوى المعنوية المشاهد بمستوى المعنوية أو الدلالة (α) وكقاعدة عامة، فإنه: إذا كانت (p-value) أقل من أو تساوي (α) فإنه يتم رفض الفرضية الإحصائية الصفرية، أما إذا كانت (p-value) أكبر من (α) فإن في هذه الحالة لا يتم رفض "أي قبول" الفرضية الإحصائية "الصفرية"، أو عندما تكون قيمة إحصاء الاختبار المحسوبة أكبر أو تساوي القيمة الجدولية، فإنه يتم رفض الفرضية الإحصائية "الصفرية"، أي أن قيمة إحصاء الاختبار وقعت في منطقة الرفض، أما إذا كانت قيمة إحصاء الاختبار المحسوبة أقل من القيمة الجدولية، فإنه لا يتم رفض الفرضية الإحصائية "الصفرية"، أي أن قيمة إحصاء الاختبار وقعت في منطقة عدم الرفض، أي منطقة القبول.

مثال : هناك فروق ذات دلالة إحصائية مابين اتجاهات الطلاب واتجاهات الطالبات نحو الإعلان عبر الرائي.

ولفرض اختبار هذه الفرضية من الضروري إعادة صياغتها في صورة فرضية إحصائية على النحو التالي :

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

• خطوات اختبار الفروض المتعلقة بالمتوسط عندما $(n \geq 30)$:

1 — القيام بتقدير افتراضي لقيمة متوسط مجتمع الدراسة أو من خلال دراسة شاملة لمجتمع الدراسة، ومن ثم صياغة فرضية العدم (H_0) وتصاغ بعكس الحالة التي نريد اختبارها ونحن دائماً نفترض حسن النية.

2 — سحب عينة عشوائية بحيث $(n \geq 30)$ لإمكانية استخدام التوزيع الطبيعي في اتخاذ القرار بالاعتماد على نص نظرية النهاية المركزية التي تنص أن توزيع المعاينة للمتوسط يتوزع توزيعاً طبيعياً.

3 — اختيار مستوى المعنوية أو الدلالة (α) : وهي عبارة عن الحد الأقصى للخطأ المسموح بارتكابه.

4 — إيجاد قيمة (Z) الدرجة المعيارية الجدولية

5 — حساب المتوسط الحسابي والانحراف المعياري (غير المتحيز) للعينة المسحوبة وتحديد الخطأ المعياري لتوزيع المعاينة للمتوسط من خلال :

$$\left(\sigma_x = \frac{\sigma}{\sqrt{n}} \right)$$

6 — القيام بحساب قيمة (Z) الدرجة المعيارية الفعلية، باستخدام معلومات العينة والمجتمع من خلال العلاقة التالية:

$$Z = \frac{|\bar{X} - \mu|}{\sigma_{\bar{x}}}$$

• اختبار المعنوية (Z) ، أي تحديد درجة الثقة بنتائج استخدام المقاييس المحسوبة من العينة والمأخوذة من المجتمع لمعرفة مدى مصداقية العينة في تمثيل المجتمع الذي سحبت منه.

8 — المقارنة واتخاذ القرار: من خلال المقارنة بين (Z) المحسوبة و (Z) الجدولية بقبول أو رفض فرضية العدم، فإذا وضعت الدرجة المعيارية (Z) داخل منطقة القبول نقبل فرضية العدم، أما إذا وقعت الدرجة المعيارية (Z)

داخل منطقة الرفض، فإننا نرفض فرضية العدم (H_0)، ونحن عادة نهتم بالقيمة الشهيرة لمستوى المعنوية المستمدة من جدول التوزيع الطبيعي الموجود في آخر الكتاب.

جدول مستوى المعنوية (α) للقيم الشهيرة وقيم (Z) المقابلة لها.

مستوى المعنوية (α)	0.01	0.05	0.10	4.5%	20%
الاحتمال ($1-\alpha$)	0.99	0.95	0.90	95.5%	80%
قيم (Z) اتجاهين	2.58	1.96	1.65	2	1.28
قيم (z) من اتجاه واحد	2.33	1.65	1.28		

مثال (1) :

بينت الدراسات التي أجرتها وزارة الصحة في إحدى الدول أن الكمية الموصى بها طبياً ويومياً من الزنك في غذاء الأفراد الذين أعمارهم أكثر من (50) سنة هي $\mu=15$ ملغ/يوم وبانحراف معياري $\sigma=6.43$ ، وللتأكد من ذلك، سحبت عينة من الأفراد حجمها (115) فرداً تراوحت أعمارهم بين (65-74) سنة، فوجد أن $(\bar{x}=11.3)$ وبانحراف معياري (8) ملغ/يوم والمطلوب: هل تبين هذه البيانات أن المتوسط اليومي μ من الزنك لهؤلاء المسنين هو أقل مما هو موصى به.

الحل :

في هذا المثال كون النقص في كمية الزنك هو موضع الشك بالنسبة للكمية الموصى بها فإن الفرضية الواجب اختبارها في مثل هذه الحالة هي :

$$H_0 : \mu = 15$$

$$H_1 : \mu < 15$$

وسنرفض الفرضية الصفرية إذا توفر لنا دليل قوي أن $\mu < 15$.
نلاحظ هنا أولاً أن $\bar{x}=11.3$ هي بالفعل قيمة أصغر من 15، ولكن هذا الفرق

قد يحدث بسبب تغير المعايير، لكن السؤال، هل الفرق بين 11.3 و 15 كبيراً إلى الدرجة التي يتوجب معرفة السبب الحقيقي وراء ذلك؟
للإجابة عن هذا السؤال نقوم بإجراء الاختبار :

— صياغة الفرضية :

$$H_0 : \mu = 15 \quad , \quad H_1 : \mu < 15 \quad (\text{الاختبار من اتجاه واحد})$$

— الاختبار الإحصائي :

$$\begin{aligned} Z &= \frac{|\bar{x} - \mu|}{\sigma / \sqrt{n}} \sim N(0,1) \\ &= \frac{|11.3 - 15|}{6.43 / \sqrt{115}} \\ &= \frac{|-3.7|}{0.6} = -6.16 \end{aligned}$$

— القرار :

نلاحظ أن القيمة تقع في أسفل ذيل منحنى Z وتعني أن $\bar{x}$ تبعد بأكثر من 6 انحرافات معيارية عن الوسط الحقيقي μ . هذه النتيجة تعني أن العينة قد أظهرت دليلاً قاطعاً لرفض H_0 وبالتالي نرفضها لصالح الفرضية البديلة، أي أن الزنك المستهلك في غذاء المسنين في الفترة العمرية 65-74 لا ترقى إلى المعدل المطلوب وهو 15 ملغ / اليوم.

تمرين: بفرض أن الدراسة قد أخذت بعين الاعتبار قياس معدل فيتامين E في غذاء نفس المجموعة من الأفراد المسنين، وأن الكمية الموصى بها من هذا الفيتامين هي 10 ملغ/ اليوم وأن $\sigma = 8.58$ أخذت عينة حجمها (115) فرد . أظهرت النتائج أن : $\bar{x} = 9.4$ وانحراف معياري $S = 6$ ملغ/يوم

المطلوب:

هل تبين هذه البيانات أن المتوسط اليومي (μ) من الفيتامين (E) لهؤلاء المسنين هو أقل من 10 ملغ/اليوم.

الحل:

— صياغة الفرضية:

$$H_0 : \mu = 10$$

$$H_1 : \mu < 10 \text{ (الاختبار في اتجاه واحد)}$$

— الاختبار الإحصائي:

$$Z = \frac{|\bar{x} - \mu|}{\sigma / \sqrt{n}} = \frac{|9.4 - 10|}{8.58 / \sqrt{115}} = 0.75$$

— القرار:

وهذا يعني أن $\bar{x}$ لا تبعد سوى 0.75 انحرافاً معيارياً عن الوسط الحقيقي μ الذي يجعل H_0 صحيحة. وبالتالي يمكن القول: أنه لا يوجد دليل قوي على أن معدل الاستهلاك للفيتامين E بين مجتمع المسنين في هذه الفئة العمرية أقل مما هو موصى به .

من خلال المثال السابق نجد أن: القيمتين $Z = 0.75$, $Z = 6.17$ مختلفتان إلى الحد الذي كان فيه قرار رفض أو قبول H_0 واضح المعالم تماماً. لكن في حالات أخرى قد لا يكون القرار واضحاً. لاسيما إذا أدخلنا مفهوم الاحتمال (درجة الثقة أو مفهوم مستوى المعنوية (α) لمعرفة وقوع القيمة (Z) ضمن منطقة الرفض أو منطقة القبول.

• منطقة الرفض، منطقة القبول، المنطقة الحرجة :

تعريف : تسمى مجموعة قيم الاختبار التي تقود إلى رفض الفرضية الصفرية بمنطقة الرفض، وتسمى مجموعة قيم الاختبار التي تقود إلى قبول الفرضية الصفرية بمنطقة القبول. وتسمى قيم الاختبار التي تفصل منطقة الرفض عن منطقة القبول بالقيم الحرجة. وتحدد القيم الحرجة عن طريق اختبار مستوى الثقة (α) أو الدلالة (α) وتحديد القيم الجدولية من جدول التوزيع الطبيعي المعياري التي تجعل المساحة في منطقة الرفض.

ونحن نميز ثلاثة خيارات في عملية اختبار الفرضية البديلة هي :

• اختبار من الطرفين : (Two-Tailed Test)

فإذا كان الاهتمام ينحصر على تقرير، ما إذا كان متوسط مجتمع μ يختلف عن قيمة محددة ولتكن μ_0 ، فإن الفرضية البديلة يجب أن تكون:

$$H_1 : \mu \neq \mu_0$$

اختبار فرضية لفرضية بديلة من هذا النوع يسمى باختبار من الطرفين.

• اختبار من الطرف الأيسر (left - Tailed)

إذا كان الاهتمام ينحصر على تقرير ما إذا كان متوسط مجتمع μ أقل من قيمة محددة ولتكن μ_0 ، فإن الفرضية البديلة يجب أن تكون :

$$H_1 : \mu < \mu_0$$

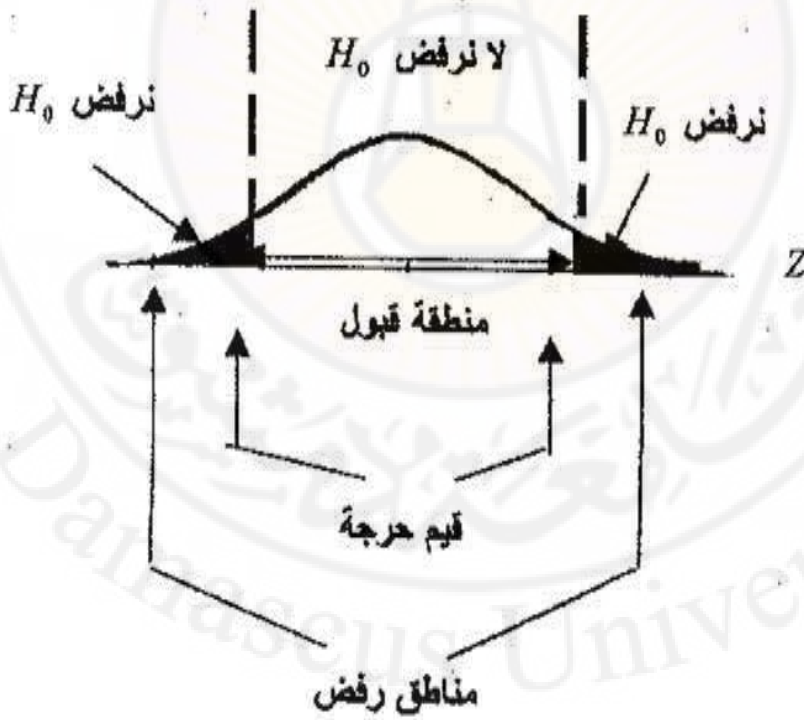
• اختبار من الطرف الأيمن (Right - tailed test) :

إذا كان الاهتمام ينحصر على تقرير، ما إذا كان متوسط مجتمع μ أكبر من قيمة ولتكن μ_0 ، فإن الفرضية البديلة يجب أن تكون :

$$H_1 : \mu > \mu_0$$

وبناءً على ذلك، فإن :

- الفرضية الصفرية في اختبار ذي الطرفين، سترفض كلما كانت قيمة الاختبار صغيرة جداً أو كبيرة جداً، أي أن منطقة الرفض في مثل هذا الاختبار ستحتوي جزأين الأول إلى اليسار والآخر إلى اليمين .
- الفرضية الصفرية في اختبار طرف أيسر، سترفض فقط إذا كانت قيمة الاختبار صغيرة جداً، أي أن منطقة الرفض في مثل هذا الاختبار ستحتوي جزءاً واحداً ويجب أن يكون إلى اليسار .
- الفرضية الصفرية في اختبار طرف أيمن، سترفض فقط إذا كانت قيمة إحصاء الاختبار كبيرة جداً، أي أن منطقة الرفض في مثل هذا الاختبار ستحتوي جزءاً واحداً إلى اليمين .



تمرين:

رغب أحد المستثمرين إنشاء مجمع سكني على هيئة شقق مفروشة فاخرة للإيجار، وأبلغته إحدى الشركات المتخصصة أن متوسط الإيجار اليومي لمثل هذه الشقق يبلغ 600 وحدة نقدية وبانحراف معياري (85) وحدة نقدية، وقبل أن يأخذ قراره، أراد التأكد من صحة معلومات تلك الشركة، فقام بسحب عينة عشوائية حجمها (36) شقة مماثلة لتلك الشقق فكان متوسط الإيجار اليومي (620) وحدة .

المطلوب: هل معلومات الشركة المتخصصة صحيحة، عند مستوى معنوية $\alpha=0.05$ ؟

— صياغة الفرضية : (معلومات الشركة صحيحة) $H_0 : \mu = 600$
(معلومات الشركة غير صحيحة) $H_1 : \mu \neq 600$

اختبار من الطرفين وبالتالي فالقيم الحرجة هي:

$$\pm Z_{\alpha/2} = \pm 1.96$$

— الاختبار الإحصائي :

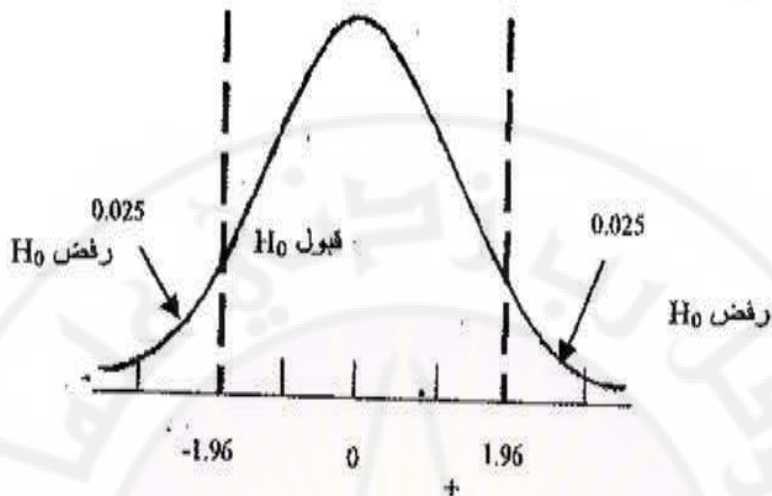
$$Z = \frac{|620 - 600|}{85/\sqrt{36}} =$$

$$\frac{|20|}{85/6} = |-1.41|$$

— اتخاذ القرار : من خلال المقارنة نجد أن :

$$Z = 1.14 < Z_{\alpha/2} = 1.96$$

وبالتالي فإن H_0 مقبولة، أي أن بيانات إجراءات الشقق التي حصل عليها المستثمر أكدت صحة ما ادعته الشركة المتخصصة.



7-11- اختبار الفرضيات لنسبة مجتمع :

يتبع اختبار الفرضيات للنسبة في حالة العينات الكبيرة نفس الخطوات التي تمت مع متوسط المجتمع تماماً، مع الأخذ بعين الاعتبار فروق المسميات والرموز التي ناقشناها سابقاً. وفيما يلي ملخص لخطوات إجراء مثل هذا الاختبار .

— صياغة الفرضية :

$$H_0 : P = P_0$$

$$H_1 : P \neq P_0$$

أو

$$H_1 : P > P_0$$

$$H_1 : P < P_0$$

الاختبار الإحصائي :

$$Z = \frac{|P' - P_0|}{\sqrt{P(1-P)/n}}$$

القرار : ترفض H_0 إذا وقعت قيمة إحصاءة الاختبار في منطقة الرفض، وتقبل إذا وقعت في منطقة القبول.

تمرين:

تنافس مرشحان الأول (x) والثاني (y) على مقعد مجلس النواب في إحدى الدوائر الانتخابية في إحدى المدن قبل إجراء الانتخابات بيوم واحد، وقد أراد فريق عمل المرشح (x) معرفة حجم شعبية مرشحهم وعما إذا كان سيفوز أم لا. أخذت عينة عشوائية من المسجلين في الدائرة الذين يحق لهم التصويت حجمها 3100 فرداً أجاب 1650 منهم بأنهم يعتزمون التصويت لصالح هذا المرشح (x). عند مستوى معنوية 5%، المطلوب : هل هذه المعلومات كافية لحصول المرشح على غالبية الأصوات؟.

الحل :

قبل أن تتم الحسابات لا بد لنا من التأكد أن شروط إجراء الاختبار متوفرة. فنجد:

$$nP = 0.5 * 3100 = 1550$$

$$n(1-P) = 3100 (1-0.5) = 1550$$

وكلاهما أكبر من 0.5 وتصبح بناءً على ذلك خطوات الاختبار قابلة للتطبيق:

الحل :

— صياغة الفرضية:

(لن يحصل المرشح (x) على غالبية الأصوات) $H_0 : P = 0.5$

(سيحصل المرشح (x) على غالبية الأصوات) $H_0 : P > 0.5$

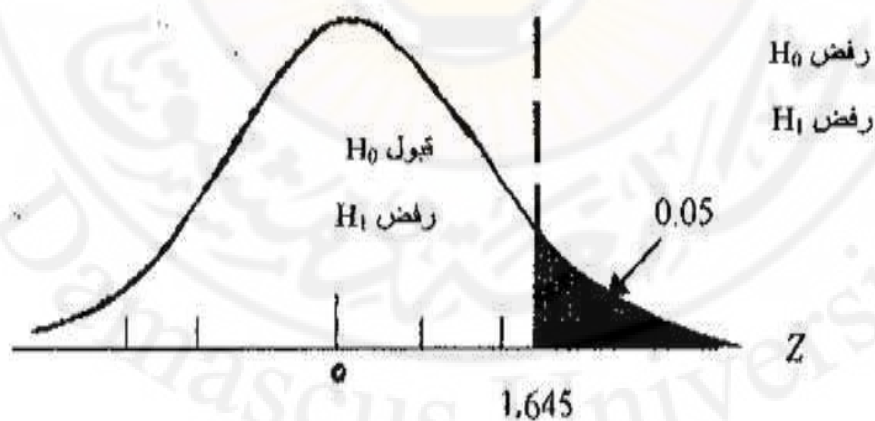
$$Z_{\alpha} = Z_{0.05} = 1.645 \Leftarrow \alpha = 0.05 \quad (\text{الاختبار من طرف واحد})$$

$$P' = \frac{x}{n} = \frac{1650}{3100} = 0.53$$

— الاختبار الإحصائي :

$$Z = \frac{|P' - P|}{\sqrt{P(1-P)/n}} = \frac{|0.53 - 0.5|}{\sqrt{\frac{(0.5)(0.5)}{3100}}} = 3.75$$

— اتخاذ القرار : من خلال المقارنة نجد أن: $Z = 3.75 > z_{\alpha} = 1.645$
وبالتالي H_0 مرفوضة، مما يعطي أن الاستبيان الذي أجراه فريق عمل المرشح (X) دليل على أن مرشحهم سيحصل على غالبية الأصوات.



7-12- اختبار الفرضيات للفرق بين نسبتين منويتين (P_2, P_1) :

كما هو الحال في اختبار الفرضيات فإننا نستخدم المتغير العشوائي الطبيعي المعياري لمقارنة قيمة إحصاء العينة مع القيمة الافتراضية لمعلمة المجتمع، ولاختبار الفرضيات المتعلقة بنسب مجتمعين، تأخذ Z الصورة التالية :

$$Z = \frac{(P_1' - P_2') - d_0}{\sigma_{P_1' - P_2'}}$$

حيث أن $d_0 = P_1 - P_2$ وأن $\sigma_{\hat{P}_1 - \hat{P}_2}$ هو الانحراف المعياري للفرق بين النسب المئوية $(\hat{P}_1 - \hat{P}_2)$ ، بناءً عليه نستخدم العلاقة :

$$S_{\hat{P}_1 - \hat{P}_2} = \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}$$

لتقدير $\sigma_{\hat{P}_1 - \hat{P}_2}$ ، ولكن عندما تكون الفرضية الصفرية لاختبار فرضية حول الفرق بين نسب مجتمعين هي :

$$\begin{aligned}\sigma_{\hat{P}_1 - \hat{P}_2} &= \sqrt{\frac{P(1-P)}{n_1} + \frac{P(1-P)}{n_2}} \\ &= \sqrt{P(1-P) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}\end{aligned}$$

أي أننا نحتاج فقط إلى تقدير واحد للقيمة P لتقدير $\sigma_{\hat{P}_1 - \hat{P}_2}$ ، ويمكن حسابه مباشرة من العلاقة التالية:

$$\hat{P} = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2}$$

ولتقدير الانحراف المعياري للفرق $\hat{P}_1 - \hat{P}_2$ فإننا نستخدم بدلاً من $\sigma_{\hat{P}_1 - \hat{P}_2}$ العلاقة التالية:

$$S_{\hat{P}_1 - \hat{P}_2} = \sqrt{\hat{P}(1 - \hat{P}) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

تمرين: في دراسة لنسبة الناخبين المؤيدين للمرشح (x) في إحدى المدن، سحبت عينتان عشوائيتان، عينة عشوائية في المنطقة (A) أظهرت أن (60) من (400) ناخب يؤيدون المرشح (x) وعينة عشوائية في المنطقة (B) أظهرت أن (40) من (300) ناخب يؤيدون هذا المرشح والمطلوب : اختبار الفرضية القائلة بأنه: ليس هناك اختلاف بين نسبة الذين يؤيدون هذا المرشح في المنطقتين عند مستوى معنوية $(\alpha = 0.05)$.

الحل :

P_1 = نسبة الناخبين المؤيدين للمرشح (x) في المنطقة (A)

P_2 = نسبة الناخبين المؤيدين للمرشح (x) في المنطقة (B).

— صياغة الفرضية :

$$H_0 : P_1 - P_2 = 0 \text{ أو } P_1 = P_2$$

$$H_1 : P_1 - P_2 \neq 0 \text{ أو } P_1 \neq P_2$$

عند $\alpha = 0.05$ فإن $Z_{\alpha/2} = 1.96$ والنسبة والاحتمال للمنطقتين (العينتين):

$$P_1' = \frac{60}{400} = 0.15 \text{ — للمنطقة A}$$

$$P_2' = \frac{40}{300} = 0.13 \text{ — للمنطقة B}$$

$$\hat{P} = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2} = \frac{400(0.15) + 300(0.13)}{400 + 300} = 0.14 \text{ فيكون لدينا، النسبة :}$$

الاختبار الإحصائي :

$$Z = \frac{(\hat{P}_1 - \hat{P}_2) - (P_1 - P_2)}{\sigma_{\hat{P}_1 - \hat{P}_2}} = \frac{(\hat{P}_1 - \hat{P}_2) - (P_1 - P_2)}{\sqrt{\hat{P}(1 - \hat{P})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$\frac{(0.15 - 0.13) - (0)}{\sqrt{(0.14)(1 - 0.14)\left(\frac{1}{400} + \frac{1}{300}\right)}} = 0.76$$

القرار : وبالتالي نقبل H_0 لأن $0.76 < 1.96$ بالتالي ليس هناك فرق معنوي بين نسب المؤيدين للمرشح (x) في المنطقتين (A) و (B) .

7-12- اختبار الفرضيات للفرق بين متوسطي مجتمعين (μ_1, μ_2) :

يعتمد اختبار الفرضيات في هذه الحالة على قيم توابع العينة $(x_1 - x_2)$ بدلاً من x في حالة عينة من المجتمع الواحد، وبناءً على ذلك، فإننا نستخدم معادلة الفرق $x_1 - x_2$ لإجراء اختبار الفرضية. وبذلك يكون الاختبار المطلوب لإجراء اختبار الفرضيات في مثل هذه الحالة على النحو التالي:

$$Z = \frac{(x_1 - x_2) - (d_0)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

حيث أن: $d_0 = \mu_1 - \mu_2$ وأن: $x_1 - x_2$ القيمة الملاحظة لـ $(\mu_1 - \mu_2)$ للفرق بين الوسطين الحقيقيين، هي :

$$\sigma_{x_1 - x_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$\sigma_{\bar{x}_1 - \bar{x}_2}$ هو الخطأ المعياري للفرق $x_1 - x_2$ وإذا كانت σ_1^2, σ_2^2 غير معلومتين، فإنه بالإمكان الاستعاضة عنهما بتباينات العينات S_1^2, S_2^2 مادامت أن الأحجام لهذه العينات كبيرة، أي $n_1, n_2 \geq 30$.
تمرين:

في دراسة حول متوسط عمر المرأة عند الزواج الأول في إحدى الدول، سحبت من مدينتين عينتين عشوائيتين، فكانت النتائج في الجدول المرفق .

العينة	العينة الأولى	العينة الثانية
المؤشر		
حجم العينة	$n_1 = 248$	$n_2 = 225$
متوسط عمر المرأة عند الزواج الأول	$\bar{x}_1 = 24$	$\bar{x}_2 = 26$
الانحراف المعياري حول عمر المرأة عند الزواج (بالسنوات)	$S_1 = 3$	$S_2 = 4$

المطلوب :

باستخدام هذه المعلومات، هل هناك فرق جوهري بين متوسط أعمار النساء المتزوجات في هاتين المدينتين؟
صياغة الفرضية :

$$H_0 : \mu_1 - \mu_2 = 0$$

$$H_1 : \mu_1 \neq \mu_2 \neq 0$$

باستخدام مستوى دلالة $\alpha = 0.05$ واختبار من طرفين، فإن قيم Z الحرجة كما نعلم هي $Z_{\alpha/2} = \pm 1.96$

– الاختبار الإحصائي :

$$Z = \frac{(x_1 - x_2) - (d_0)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{(24 - 26) - 0}{\sqrt{\frac{(3)^2}{248} + \frac{(4)^2}{225}}} = |6.45|$$

القرار : وبالتالي نرفض H_0 لأن $6.45 > 1.96$ ونستنتج أن هناك فرقاً معنوياً بين متوسطي أعمار النساء المتزوجات في هاتين المدينتين.

7-13- تحديد حجم العينة:

عند اختيار أسلوب العينات لإجراء الدراسة الإحصائية، يتبادر إلى ذهن السؤال التالي، ما هو حجم العينة اللازم لإجراء الدراسة ؟ وبالرغم من أن هذا السؤال لا توجد له إجابة قاطعة، أن حجم العينة يتناسب عكساً مع خطأ المعاينة، وكلما زدنا حجم العينة، كلما قللنا من مقدار الخطأ الاحتمالي أو خطأ المعاينة ، لكننا بنفس الوقت لا نريد أن نزيد من حجم العينة إلى الحد الذي نفقد معه مزايا استخدام أسلوب العينات في الدراسة. وبعبارة أخرى، أنه إذا زدنا حجم العينة إلى الحد الذي نصل به إلى حجم المجتمع، نكون قد ضحينا بأسلوب العينات، وإذا صغرنا من حجم العينة لن تكون صادقة في تمثيلها للمجتمع المسحوبة منه، لذلك عند تحديد حجم العينة لا بد من مراعاة ما يلي:

1 - الاعتبارات المادية.

2 - الاعتبارات العلمية.

بالنسبة للاعتبارات المادية : حجم العينة يتحدد في ضوء الميزانية المتاحة للدراسة من كواادر بشرية، ومالية، ووقت، الأمر الذي يقودنا إلى تصغير أخطاء المعاينة.

أما بالنسبة للاعتبارات العلمية : يتوقف تحديد حجم العينة علمياً على مدى الدقة المطلوب توافرها في بيانات العينة أولاً، ونشتت القيم في المجتمع ثانياً.

إن الدقة معيار أساسي في تحديد حجم العينة، ويقصد بها: ما هو حجم الخطأ الذي نستطيع أن نقبله، في تقدير معالم المجتمع باستخدام بيانات العينة، هذا يتطلب منا معرفة الانحراف المعياري للمجتمع، وهو أمر يتعذر علينا معرفته في أغلب الأحيان، لذلك نعتبر الانحراف المعياري للعينة العشوائية تقديراً غير متحيزاً للانحراف المعياري للمجتمع.

نستنتج من خلال الاستعراض السابق أنه لتحديد حجم العينة نتبع

الخطوات التالية:

1 - نحدد الحد الأقصى للخطأ المسموح به، ولنفرض أنه مساوٍ (E)، أي أننا نأمل في الحصول على عينة يكون الفرق بين وسطها الحسابي والوسط الحسابي للمجتمع، محل الدراسة، في حدود مقدار معين من الخطأ هو (E).

فإذا أردنا تحديد نسبة البطالة في بلد ما، وكنا نرغب في أن لا تزيد قيمة الحد الأقصى للخطأ المسموح عن (3%)، وكانت النسبة الحقيقية للبطالة في المجتمع مساوية (20%)، فإن النسبة المثوية للبطالة في العينة يجب أن لا تقل عن (17%) وأن لا تزيد عن (23%).
إن الخطأ الأعظمي المرتكب هو :

$$E = Z_{\frac{\alpha}{2}} \sqrt{\frac{p'q}{n}}$$

وذلك باحتمال قدره :

$$(1 - \alpha)\%$$

$$E^2 = Z_{\frac{\alpha}{2}}^2 \frac{p'q}{n} \quad \text{ومن العلاقة السابقة نجد :}$$

$$n = \frac{Z_{\frac{\alpha}{2}}^2 p'q}{E^2} \quad \text{وبالإصلاح فإن :}$$

وتفسير هذه العلاقة أننا واثقون باحتمال $(1 - \alpha)\%$ أن الخطأ المرتكب لن يزيد

$$n = \frac{Z_{\frac{\alpha}{2}}^2 p'q}{E^2} \quad \text{على القيمة (E) إذا كان حجم العينة :}$$

وبما أن $p(1 - p)$ غير معلومة، فإننا نعتبر أن $p = \frac{1}{2}$ ، $q = \frac{1}{2}$

وإن $p(1 - p)$ يأخذ قيمته العظمى عندما $\bar{p} = \frac{1}{2}$.

إذاً يمكن القول أن الخطأ المرتكب لن يتجاوز (E) إذا كان حجم العينة :

$$n = \frac{Z_{\frac{\alpha}{2}}^2}{4(E^2)}$$

تمرين:

في استفتاء سريع حول أحد المرشحين وجد أن (960) من أصل (1500) تم اختيارهم عشوائياً، كانوا يؤيدون ذلك المرشح .

والمطلوب :

- 1 — أوجد مجال ثقة 98% للنسبة الحقيقية لمؤيدي هذا المرشح .
- 2 — أوجد الخطأ الأعظمي المرتكب في قولنا أن نسبة المؤيدين الحقيقية تساوي نسبة المؤيدين في هذه العينة بثقة 98% .
- 3 — ما هو حجم العينة الواجب سحبها عشوائياً حتى لا يتجاوز الخطأ 0.02 زيادة أو نقصاناً وذلك بثقة 98% .

الحل :

$$P = \frac{x}{n} = \frac{960}{1500} = 0.64$$

من جدول (Z) نجد أن:

$$Z_{\alpha/2} = Z_{0.49} = 2.33$$

$$\bar{p} - Z_{\alpha/2} \sqrt{\frac{\bar{p} \cdot \bar{q}}{n}} \leq p \leq \bar{p} + Z_{\alpha/2} \sqrt{\frac{\bar{p} \cdot \bar{q}}{n}} \quad - 1$$

$$0.64 - 2.33 \sqrt{\frac{0.64(0.36)}{1500}} \leq p \leq 0.64 + 2.33 \sqrt{\frac{0.64(0.36)}{1500}}$$

$$(0.61, 0.67)$$

يمكن القول أنه باحتمال 98% لن تقل نسبة المؤيدين لهذا المرشح فسي المجتمع الذي سحبت منه العينة عن 61% ولن تزيد عن 67% .

2 - نحن نعلم أن $p = 0.64$:

$$E = Z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}}$$

$$E = 2.33 \sqrt{\frac{0.64(0.36)}{1500}} = 0.03$$

3 - حتى لا يتجاوز الخطأ 0.02 زيادة أو نقصاناً بثقة 98% فإن حجم العينة المطلوب هو :

$$n = \frac{Z_{\alpha/2}^2}{4E^2} = \frac{(2.33)^2}{4(0.03)^2} = 15.08$$

وفي حال تقدير متوسط المجتمع (μ) من متوسط العينة $\bar{X}$ فإن الخطأ الأعظمي المرنكب، يعطى بالعلاقة :

$$E = Z_{\alpha/2} \frac{\delta}{\sqrt{n}}$$

وذلك باحتمال $(1-\alpha)\%$.

وبالتالي فإن حجم العينة الواجب سحبها حتى لا يتجاوز الخطأ في التقدير مقداراً (E) وذلك باحتمال $(1-\alpha)\%$ هو :

$$n = \frac{Z_{\alpha/2}^2 \delta^2}{E^2}$$

تمرين : في عينة من 36 عاملاً كان متوسط النقاط التي حاز كل منهم عليها هو (2,6) نقطة وانحرافها المعياري $S = 0.3$ ، والمطلوب :

1 - أوجد 99% حد الثقة لمتوسط المجتمع .

2 - كم يجب أن يكون حجم العينة الواجب سحبها عشوائياً بحيث لا يتجاوز الخطأ في تقدير الوسط الحسابي للمجتمع 0.05 وذلك باحتمال 99% من الثقة.

الحل :

بما أن $n > 30$ يمكن اعتبار : $S = \delta = 0.3$

ومن جدول (Z) نجد أن :

$$Z_{\alpha/2} = Z_{0.945} = 2.58$$

ويكون مجال الثقة هو :

$$\bar{x} - Z_{0.945} \frac{S}{\sqrt{n}} \leq \mu \leq \bar{x} + Z_{0.945} \frac{S}{\sqrt{n}}$$

$$2.6 - 2.58 \frac{0.3}{\sqrt{36}} \leq \mu \leq 2.6 + 2.58 \frac{0.3}{\sqrt{36}}$$

$$2.47 \leq \mu \leq 2.73$$

والتفسير أنه باحتمال 99% فإن متوسط النقاط في مجتمع العمال (μ) الذي

سُحِبَت منه العينة لن يقل عن 2.47 ولن يزيد عن 2.73 .

$$n = \left[\frac{(2.58)(0.3)}{0.05} \right]^2 = 239.63 \text{ عاملاً } 2$$

والتفسير :

إننا نثقون باحتمال 99% بأنه لن يختلف μ بأكثر من 0.05 إذا كان حجم

العينة مساوياً لـ 239 عاملاً .

ملاحظة (1) :

في الحالات العملية في حال عدم إعطاء احتمال الدقة (درجة الثقة) فإننا

نأخذ أكبر احتمال ممكن وهو (99.73%) ويقابل في جدول (Z) القيمة (3).

ملاحظة (2) :

عند تحديد حجم العينة للنوعيات (النسبة المئوية) وكانست (p) غير معلومة كنا نفترض أن ($p = 1/2$) و ($q = 1/2$) ، لكن ليس هذا هو الحال في أغلب الحالات العملية، فقد تتوفر لدينا معلومات عن قيمة تحقق الظاهرة المدروسة في المجتمع وأنهما تتراوح بين قيمتين دنيا وعليا مثلاً: ($10 \leq P \leq 20$) ، فعند حساب حجم العينة نأخذ القيمة الأقرب إلى الـ (50%) ومن ثم نقوم بتطبيق العلاقة التالية لحساب حجم العينة:

$$n = \frac{Z^2 \cdot P \cdot q}{E^2}$$

تمرين :

لقد كان من المرغوب به تحديد نسبة المرشحين للانتخابات من حملة الإجازة الجامعية، وقد قدرت هذه النسبة بين (12-22%) المطلوب: ماهو حجم العينة الواجب سحبها عشوائياً بغية تقدير النسبة الحقيقية لحملة الإجازة الجامعية لهؤلاء المرشحين على ألا يزيد الخطأ المرتكب في التقدير عن (0.05) وعند احتمال الدقة لا يقل عن (99.5%)

الحل :

$$P = 22\%$$

$$n = \frac{Z^2 \cdot p \cdot q}{E^2} = \frac{(2)^2 (0.22)(0.78)}{(0.05)^2} = 275 \text{ مرشحاً}$$

تمارين غير محلولة

– تمرين (1) :

رغب أحد الباحثين في تقدير نسبة الأسر التي لا يزيد عدد أفرادها عن (3) أشخاص، وكان الاعتقاد بأن نسبة هذه الأسر تتراوح بين (60%) و(70%)

والمطلوب:

ما هو حجم العينة المراد سحبها عشوائياً لتقدير النسبة الحقيقية للأسر التي لا يزيد عدد أفرادها عن (3) أشخاص، وألا يزيد الخطأ عن (4%) واحتمال الدقة لا يقل عن (95.5%)

تمرين (2) :

تدعي مجلة في علم الإدارة أن (25%) من قرائها هم طلبة جامعيون. أخذت عينة عشوائية من (200) قارئ لها، ووجد أن (42) فقط هم طلبة جامعيون.

والمطلوب:

- 1 – اختبر الفرض القائل بأن نسبة القراء من الطلبة الجامعيين وما زالت (25%) عند مستوى معنوية $\alpha = 0.05$
- 2 – باستخدام بيانات العينة، أنشئ فترة (95%) ثقة لنسبة القراء من المجتمع الذين هم طلبة جامعيون .

تمرين (3):

ترغب إحدى القنوات التلفزيونية في معرفة، ما إذا كان درجة شعبية برنامجين تلفزيونيين متساوية، سحبت عينة عشوائية من (400) من المشاهدين المتوقعين للبرنامج الأول (A) فوجد أن (220) منهم يشاهدون البرنامج بانتظام،

كما سحبت عينة أخرى مستقلة من حجم مماثل للمشاهدين المتوقعين للبرنامج الثاني (B) فوجد أن (260) منهم يشاهدون البرنامج بانتظام. المطلوب: اختبر الفرض القائل بتساوي النسبتين عند مستوى معنوية ($\alpha=0.05$).

تمرين (4):

يعتقد أن أكثر من (50%) من مشاهدي التلفزيون يشاهدون نشرة الأخبار المسائية في القناة الثانية. سحبت عينة عشوائية من (1600) مشاهد وجد أن (840) منهم يشاهدون نشرة الأخبار المسائية في القناة الثانية. المطلوب: اختبر عند مستوى معنوية ($\alpha=0.01$) صحة هذا الاعتقاد.

تمرين (5):

ترغب وزارة العمل في إحدى الدول تقدير متوسط عمر عاطلين عن العمل باستخدام عينة عشوائية بحيث لا تزيد الفرق بين متوسط العينة والمتوسط الحقيقي عن (2) وذلك بدرجة ثقة (0.99).

المطلوب: حدد حجم العينة المناسب، فيما إذا علمت أن الانحراف المعياري لأعمار العاطلين عن العمل هو (5) سنوات.

تمرين (6):

قررت إدارة شركة تجارية لها فرعين أن تقدر الفرق في متوسط حجم المبيعات للفرعين، أخذت عينتين مستقلتين من مبيعات يوم ما في الفرعين، ووجدت النتائج الآتية:

الفرع الأول

$$n_1=32$$

$$\bar{X}_1=5000$$

$$s_1=1500$$

الفرع الثاني

$$n_2=36$$

$$\bar{X}_2=3750$$

$$s_2=1300$$

والمطلوب: أنشئ تقدير للفرق بين متوسط المبيعات لدى الفرعين.

تمرين (7) - أثناء أزمة مياه في إحدى المدن رأت الشركة المختصة إجراء معاينة عشوائية على عدادات مياه منطقة سكنية بغية مراقبة الاستهلاك اليومي في المياه، في يوم معين أظهرت عينة من (50) عداداً أن متوسط الاستهلاك $x = 25$ بانحراف معياري $S = 5m^3$ أنشئ تقدير فترة ثقة (95%) لمتوسط استهلاك مجتمع تلك المدينة من المياه.

تمرين (8) :

في عينة عشوائية مكونة من (900) شخص من مشاهدي التلفزيون تبين أن (180) من بينهم يشاهدون برنامجاً تلفزيونياً معيناً. أوجد فترة الثقة (0.99) للنسبة الحقيقية لمشاهدي التلفزيون الذين يشاهدون هذا البرنامج .

تمرين (9) :

تدعي عمادة كلية الإعلام في إحدى الجامعات، أن نسبة الطلاب في الكلية هي (48%) والطالبات (52%)، وللتأكد من ذلك، سحبت عينة عشوائية من (100) طالب وطالبة، فأظهرت النتائج أن نسبتهم هي (40%).
والمطلوب: هل كان الادعاء صحيحاً، عند مستوى معنوية $(\alpha = 0.05)$.

تمرين (10) :

في مصنع للنسيج ادخل أحد المهندسين تحسينات عليه مدعياً تخفيف نسبة التلف إلى (4%)، للتأكد من ذلك أخذت عينة عشوائية من (200) قطعة فظهرت (12) قطعة تالفة.
المطلوب: هل كان هذا الادعاء صحيحاً عند مستوى معنوية $(\alpha = 0.05)$.

تمرين (11) :

في عينة عشوائية تتكون من (100) شخص، وجد من بينهم (25) مرض .
المطلوب: أوجد 95% وفترة ثقة حول نسبة المدخنين في هذا المجتمع بصفة عامة.

الفصل الثامن

الاستدلال الإحصائي للعينات الصغيرة الحجم

- مقدمة
- توزيع (t) ستودنت
- التقدير المجالي لمعاملات المجتمع في العينات الصغيرة الحجم
- تقدير الفرق بين متوسطي مجتمعين إحصائيين غير معلومي الثباين.
- اختبار الفرضيات للعينات الصغيرة الحجم (حالة مجتمع واحد)
- اختبار معنوية الفروق للأوساط الحسابية (حالة مجتمعين)
- توزيع كاي مربع (χ^2)
- التقدير المجالي لمعلمة مجتمع
- توزيع فيشر (F)
- اختبار المقارنة بين تبايني مجتمعين



الاستدلال الإحصائي للعينات الصغيرة الحجم

1-8- مقدمة :

سبق أن درسنا توزيعات المعاينة لكل من المتوسط والنسب المئوية والفريق بين متوسطي مجتمعين.. الخ، وبيناً أنه عند تطبيق نظرية النهاية المركزية نحتاج لمتوسط وتباين المجتمع μ, σ^2 . وقد لاحظنا أن توزيع المعاينة لـ (μ) يتبع التوزيع الطبيعي، إذا كان المجتمع المدروس يخضع للتوزيع الطبيعي والانحراف المعياري للمجتمع معلوماً، لكن في كثير من الحالات لا يمكن حساب تباين المجتمع σ^2 ، لذا نقبل من أجل عينات $n \geq 30$ تباين S^2 كتقدير جيد للتباين σ^2 وبالتالي نظرية النهاية المركزية محققة هنا وتصبح الدرجة المعيارية (Z) بالشكل:

$$Z = \frac{\bar{x} - \mu}{S/\sqrt{n}}$$

أما إذا كانت $n < 30$ فإن (Z) في هذه الحالة تتبع توزيعاً جديداً يدعى توزيع (t) ستودنت.

2-8- توزيع (t) ستودنت :

لقد بينا أنه عندما يكون حجم العينة $n \leq 30$ والانحراف المعياري للمجتمع المدروس (δ) غير معلوم، في هذه الحالة، فإن البيانات تتبع توزيعاً آخر يدعى توزيع (t ستودنت)، ويمكن القول: أن هذا التوزيع يشبه التوزيع الطبيعي من حيث التماثل، ولكنه أكثر انبساطاً منه، ويقترب من التوزيع الطبيعي المعياري، عندما تزداد (n) حجم العينة، بمعنى آخر، أن توزيع ستودنت: هو عبارة عن مجموعة من المنحنيات المتماثلة أو المتناظرة حول المعلمة، ويتغير شكل هذه المنحنيات تبعاً لتغير عدد درجات الحرية، التي تعرف بأنها عدد المشاهدات المستقلة التي تتكون منها العينة أو حجم العينة ناقصاً عدد المعلمات

الإحصائية، التي يجب تقديرها منها، ويرمز لها بالرمز (v) ، حيث $V = n-1$ ، وعندما تكون درجات الحرية قليلة، فإن المنحني يصبح أكثر تفلطحاً وتشتتاً بالمقارنة مع التوزيع الطبيعي المعياري، ويتطابق توزيع ستودنت مع التوزيع الطبيعي المعياري $N(0,1)$ عندما يصبح درجات الحرية $n \geq 30$ ، إن أهم تطبيقات توزيع ستودنت هي إمكانية استخدامه كتوزيع للمعينة، للعينات الصغيرة الحجم. وتعطى (t) بالعلاقة التالية:

$$t = \frac{\bar{x} - \mu}{S / \sqrt{n}}$$

من خلال هذه العلاقة يمكن ملاحظة أن توزيع (t) يشابه التوزيع الطبيعي المعياري من حيث كون متوسطه $\mu=0$ (إلا أن تباينه لا يساوي الصفر).

— خصائص التوزيع t :

- 1 — القيمة المتوقعة للدرجة المعيارية (t) هي: $E(t) = 0$
- 2 — التباين لتوزيع t عندما $n \geq 30$: $Var(t) = \frac{v}{v-2}$
- 3 — يستخدم توزيع (t) بصورة خاصة عند عدم معلومية تباين المجتمع الإحصائي المدروس والذي سحبت منه العينة العشوائية.
- 4 — تتقارب قيم (t) من بعضها البعض عندما تكون أحجام العينات كبيرة.
- 5 — يعتمد توزيع (t) على شرط التوزيع الطبيعي للمجتمع الإحصائي المدروس، وبخلاف ذلك قد لا تكون نتائجه دقيقة.

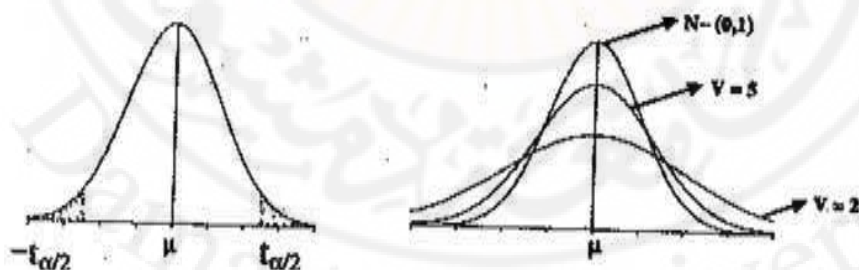
3-8- التقدير المجالي لمعلمات المجتمع في العينات صغيرة الحجم (تقدير متوسط المجتمع μ) عندما يكون التباين σ^2 غير معلوم:
إذا رغبنا في تقدير متوسط مجتمع إحصائي (μ) وكان $n < 30$ يمكن كتابة المجال الاحتمالي، لـ (μ) وفق الصيغة التالية :

$$P(-t_{\alpha/2} < t < t_{\alpha/2}) = 1 - \alpha$$

$$P\left(-t_{\alpha/2} < \frac{\bar{x} - \mu}{S/\sqrt{n}} < t_{\alpha/2}\right) = 1 - \alpha$$

$$P\left(\bar{x} - t_{\alpha/2}\left(\frac{S}{\sqrt{n}}\right) < \mu < \bar{x} + t_{\alpha/2}\left(\frac{S}{\sqrt{n}}\right)\right) = 1 - \alpha$$

وبالتالي يمكن القول، بأن (μ) تقع ضمن المجال السابق باحتمال قدره $(1 - \alpha)$ وتعطى قيم $t_{\alpha/2}$ من جداول (t) في نهاية الكتاب، حيث تعطى قيم t لكل درجة من درجات الحرية وذلك من أجل مستويات دلالة محددة .



تمرين (1) :

في إنتخابات مجلس إدارة إحدى الشركات تم اختيار (16) مرشحاً عشوائياً، فوجد أن متوسط عمر المرشح $x = 30$ سنة وبانحراف معياري (5) سنوات.

والمطلوب :

أوجد مجال الثقة لمتوسط عمر المرشحين باحتمال 95% (أو قدر باحتمال 95% متوسط عمر المرشحين لمجلس إدارة هذه الشركة.

الحل :

$$P\left(\bar{x} - t_{\alpha/2}\left(\frac{S}{\sqrt{n}}\right) < \mu < \bar{x} + t_{\alpha/2}\left(\frac{S}{\sqrt{n}}\right)\right) = 1 - \alpha$$

وبالتعويض :

$$P\left(30 - 1.753\left(\frac{5}{\sqrt{16}}\right) < \mu < 30 + 1.753\left(\frac{5}{\sqrt{16}}\right)\right) = 0.95$$

إن قيمة (t) الجدولية من أجل :

$$t(1.735)$$

$$\alpha = 0.05$$

$$v = 16 - 1 = 15$$

وعليه يكون :

$$[27.80 < \mu < 32.19]$$

وبالتالي يمكن القول أنه باحتمال 95% إن متوسط عمر المرشح لن يقل عن 27.80 سنة ولن يزيد عن 32.19 سنة.

تمرين (2) :

لقياس السرعة القصوى لسيارة جديدة، أجرينا (16) تجربة فحصلنا على متوسط للسرعة القصوى يساوي $x = 200$ كم / ساعة وانحراف معياري (10) كم / ساعة .

والمطلوب :

إيجاد مجال الثقة لمتوسط السرعة القصوى لهذا النوع الجديد من السيارات باحتمال 99%.

الحل :

$$P\left(x - t_{\alpha/2}\left(\frac{S}{\sqrt{n}}\right) < \mu < x + t_{\alpha/2}\left(\frac{S}{\sqrt{n}}\right)\right) = 1 - \alpha$$

وبالتعويض :

$$P\left(200 - 2.602\left(\frac{10}{\sqrt{16}}\right) < \mu < 200 + 2.602\left(\frac{10}{\sqrt{16}}\right)\right) = 0.99 \Rightarrow [193.495, 206.505]$$

وبالتالي يمكن القول أنه باحتمال 99% فإن متوسط السرعة لن يقل عن 193.495 كم / ساعة، ولن يزيد عن 206.5 لهذا النوع الجديد من السيارات.

4-8- تقدير الفرق بين متوسطي مجتمعين إحصائيين $(\mu_1 - \mu_2)$ غير معلومي التباين وحجم العينات صغيراً :

إذا كان $(n_1, n_2) \leq 30$ و (σ_1^2, σ_2^2) غير معلومين، في هذه الحالة نقوم بحساب الخطأ المعياري وفقاً لتوزيع (t) ستودنت، بشرط $\sigma_1^2 = \sigma_2^2$ والتباين المشترك الذي يعطى بالعلاقة :

$$S_1^2 = \frac{S_1^2 \cdot (n_1 - 1) + S_2^2 \cdot (n_2 - 1)}{n_1 + n_2 - 2}$$

والخطأ المعياري يعطى بالعلاقة :

$$\delta_{x_1 - x_2} = S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

ونحن نعلم أنه من أجل احتمال $(1 - \alpha)$ فإن (t) تتوزع وفق العلاقة :

$$P[-t_{\alpha/2} < t < t_{\alpha/2}] = 1 - \alpha$$

وإن الاختبار (t) يأخذ الشكل :

$$t = \frac{(x_1 - x_2) - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

وبالتعويض نجد أن :

$$P\left[-t_{\alpha/2} < \frac{(x_1 - x_2) - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} < t_{\alpha/2}\right] = 1 - \alpha$$

وبالتالي تصبح العلاقة :

$$P\left[(x_1 - x_2) - t_{\alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu - \mu < (x_1 - x_2) + t_{\alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right] = 1 - \alpha$$

تمرين (3):

بفرض أننا سحبنا عينتين عشوائيتين مستقلتين عن متوسط أسعار الكتب في مجال الاقتصاد والعلوم الإدارية (بالدولار) على النحو التالي:

$$\text{العلوم الاقتصادية } n_1=10, \bar{x}_1 = 9, S_1^2 = 25$$

$$\text{العلوم الإدارية } n_2=10, \bar{x}_2 = 3, S_2^2 = 22$$

والمطلوب : أوجد باحتمال قدرة 95% التقدير المجالي للفرق، بين متوسطي أسعار الكتب في هذين المجتمعين، بافتراض تساوي التباينات في المجتمع وأسعار الكتب تتبع التوزيع الطبيعي.

الحل :

$$S_p = \sqrt{\frac{S_1^2 \cdot (n_1 - 1) + S_2^2 \cdot (n_2 - 1)}{n_1 + n_2 - 2}}$$

$$= \sqrt{\frac{25(10-1) + 22(10-1)}{10+10-2}} = 4.847$$

$$P\left[(9-3) - 1.734(4.847)\sqrt{\frac{1}{10} + \frac{1}{10}} < \mu_1 - \mu_2 < (9-3) + 1.734(4.847)\sqrt{\frac{1}{10} + \frac{1}{10}}\right] = 0.95$$

إن قيمة (t) الجدولية من أجل :

$$t(1.734)$$

$$\alpha=0.05$$

$$v=n_1 + n_2 - 2 = 10 + 10 - 2 = 18$$

$$[1.15 < \mu_1 - \mu_2 < 4.84]$$

وبالتالي يمكن القول أنه باحتمال 95% فإن المجال [1.55, 4.84] يحتوي القيمة الحقيقية للفرق بين المتوسطين.

تمرين (4):

بفرض أنه تم اختيار عينتين عشوائيتين مستقلتين من مجتمعين طبيعيين مجهولي التباين ولكن تباينهما متساوي، وكانت نتائج الدراسة كما يلي:

$$n_1 = 4 \quad x_1 = 0.22 \quad S_1^2 = 0.001$$

$$n_2 = 5 \quad x_2 = 0.17 \quad S_2^2 = 0.002$$

المطلوب: أوجد 95% فترة ثقة لتقدير الفرق ما بين متوسطي المجتمعين .

الحل:

$$S_p^2 = \frac{3(0.001) + 4(0.002)}{7} = 0.0016$$

$$SP = 0.04$$

$$SP \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = (2.365)(0.04) \sqrt{\frac{1}{4} + \frac{1}{5}} = 0.06$$

$$(0.22 - 0.17) - 0.06 \leq \mu_1 - \mu_2 \leq (0.22 - 0.17) + 0.06$$

$$-0.01 \leq \mu_1 - \mu_2 \leq 0.11$$

وبالتالي، أن فترة الثقة تحتوي على الصفر، وعليه يمكننا القول: بأنه لا يوجد دليل واضح على وجود فرق ما بين متوسطي المجتمعين.

5-8- اختبار الفرضيات في حالة العينات الصغيرة :

اختبار الفروض للمتوسط (μ) (حالة مجتمع واحد) :

عندما يكون لدينا عينة صغيرة الحجم $n \leq 30$ وأن (δ^2) غير معلوم، في هذه الحالة نبني قرارنا اعتماداً على توزيع (t) ستوننت من أجل عدد من درجات الحرية $v = n - 1$ وتكون صياغة الفرضية (H_0 و H_1) مشابهاً تماماً

$$H_1: \mu \neq \mu_0 \quad H_0: \mu = \mu_0$$

ويتم الاختبار الإحصائي يتم من خلال العلاقة :

$$t = \frac{x - \mu}{S / \sqrt{n}}$$

تمرين (5):

تدعي إحدى المكتبات أن متوسط سعر الكتب المرجعية هذه السنة هو (20) دولاراً وبغية التأكد من ذلك سحبت عينة من (16) كتاباً فوجد أن $\bar{x} = 24$ دولاراً وبانحراف معياري قدره (3) دولارات، بفرض أن أسعار الكتب تتبع التوزيع الطبيعي. المطلوب: هل تدل هذه البيانات على أن متوسط سعر الكتاب المرجعي مازال $x = 20$ عند مستوى معنوية $\alpha = 0.05$

الحل :

$$1 - \text{صياغة الفرض } H_1: \mu \neq 20, H_0: \mu = 20$$

2 - إجراء الاختبار الإحصائي :

$$t = \frac{\bar{x} - \mu}{S / \sqrt{n}} = \frac{|24 - 20|}{3 / \sqrt{16}} = 5.33$$

3 - اتخاذ القرار :

من خلال المقارنة بين $t = 5.33$ وقيمة t الجدولية من أجل ($v=15$) و ($\alpha = 5\%$) فإن $t = 1.753$ وبالتالي ($t > t_{\alpha/2}$)، نرفض الفرضية (H_0) ونقول أن متوسط سعر الكتاب لم يعد يساوي ($x = 20$) دولاراً.

تمرين (6):

قامت إحدى شركات الطيران بدراسة حول عدد المقاعد التي تبقى شاغرة أثناء الرحلات الداخلية، فوجدت أن متوسط عدد المقاعد التي تبقى أثناء الرحلات يساوي (30.2) مقعد، وبالتالي قررت الشركة تخفيض سعر التذاكر حتى يكون متوسط عدد المقاعد الشاغرة (20) مقعداً فأقل، وبعد ثلاثة أشهر أجريت دراسة على عينة من (41) رحلة، فوجدت أن متوسط عدد المقاعد الشاغرة يساوي (17.3) وبتباين (30.25) المطلوب: ما هي الخلاصة التي يمكن أن تصل إليها الإدارة عند مستوى معنوية (5%)

الحل: الفرضية :

$$H_0: \mu = 20$$

$$H_1: \mu < 20$$

الاختبار الإحصائي :

$$t = \frac{|\bar{x} - \mu|}{S / \sqrt{n}} \\ = \frac{17.3 - 20}{\sqrt{30.25} / \sqrt{41}} = 3.14$$

المقارنة والقرار:

من جدول (t) نجد أن :

$$\alpha = 0.05$$

$$v = 40$$

$$t = 1.684$$

من خلال المقارنة نجد أن $3.14 > 1.689$ وبالتالي نرفض (H_0) ، أي أن تخفيض سعر التذاكر قد كان عاملاً أساسياً في تخفيض عدد المقاعد الشاغرة أثناء الرحلات الداخلية.

8-6- اختبار معنوية الفروق للأوساط الحسابية $(\mu_1 - \mu_2)$ (حالة مجتمعين) :
 في هذه الحالة إن قيم (δ_1^2, δ_2^2) غير معلومتين وإن $(n_1, n_2) \leq 30$ والمجتمعين يتوزعان توزيعاً طبيعياً، وإن رفض أو قبول الفرض (H_0) يعتمد على قيمة (t) توزيع ستودنت من أجل عدد درجات حرية $(V=n_1+n_2-2)$ ضمن شرط تساوي التباينات في المجتمعين، أي أن :

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

حيث : $d_0 = \mu_1 - \mu_2$

$$S_p = \sqrt{\frac{S_1^2(n-1) + S_2^2(n-1)}{n_1 + n_2 - 2}}$$

تمرين (7) :

رغب مدير الأفراد في إحدى الشركات معرفة إذا كان لطريقتين مختلفتين في التدريب نفس التأثير، سحبت عيّنتان مستقلتان من عمال هذه الشركة عشوائياً، وتم تدريب كل منهما بطريقة مختلفة، وبعد انتهاء مدة التدريب أجرى لهما امتحاناً موحداً لقياس مستوى العمال، فحصلنا على البيانات الآتية

$$n_1 = 18, \quad x_1 = 80, \quad S_1^2 = 36$$

$$n_2 = 12, \quad x_2 = 85, \quad S_2^2 = 34$$

المطلوب : اختبر الفرض القائل: بأن هاتين الطريقتين بالتدريب متساويتان عند مستوى معنوية $\alpha=0.05$ ، علماً أن التوزيع التكراري للمجتمعين طبيعياً والتباينات متساوية.

الحل :

1 - صياغة الفرضية :

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

2 - الاختبار الإحصائي :

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{S_p \sqrt{1/n_1 + 1/n_2}}$$

$$S_p = \sqrt{\frac{S_1^2(n-1) + S_2^2(n-1)}{n_1 + n_2 - 2}}$$

$$S_p = \sqrt{\frac{36(18-1) + 34(12-1)}{18+12-2}} = 5.93$$

$$t = \frac{(85-80)}{5.93 \sqrt{1/18 + 1/12}} = 2.27$$

3 - اتخاذ القرار :

من خلال المقارنة بين $t = 2.27$ وقيمة t الجدولية $t = 2.048$ من أجل $V = 28$ و $\alpha = 5\%$ والاختبار من اتجاهين نجد أن: $|t| > t$ عليه نرفض الفرضية H_0 ونقبل الفرضية البديلة H_1 ونقول بأن هناك فرقاً حقيقياً بين الطريقتين.

تمرين (8) :

يرغب مدير إحدى المؤسسات قياس قدرة العامل على تعلم لغة جديدة، وللتأكد من ذلك، سحبت عينتان مستقلتان من إحدى المؤسسات وتم قياس قدرة كل منهما على تعلم اللغة الجديدة وبعد الانتهاء، أجري لهما امتحان موحد لقياس قدرة عمال العينتين، فحصلنا على البيانات التالية :

العينة الأولى	العينة الثانية
$n_1 = 10$	$n_2 = 8$
$\bar{x}_1 = 95$	$\bar{x}_2 = 97$
$S_1^2 = 40$	$S_2^2 = 36$

وبافتراض أن درجات اختبار جميع العمال تتبع للتوزيع الطبيعي والتباينات متساوية.

المطلوب: اختبر الفرض القائل بتساوي قدرة عمال المينتين على تعلم اللغة الجديدة، عند مستوى دلالة $\alpha = 5\%$.

الحل :

1 - صياغة الفرض :

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

2 - إجراء الاختبار الإحصائي :

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{S_p \sqrt{1/n_1 + 1/n_2}}$$

$$S_p = \sqrt{\frac{S_1^2 \cdot (n_1 - 1) + S_2^2 \cdot (n_2 - 1)}{n_1 + n_2 - 2}}$$

$$S_p = \sqrt{\frac{40(10-1) + 36(8-1)}{10+8-2}} = 6.36$$

$$t = \frac{(85 - 80) - 0}{6.36 \sqrt{1/10 + 1/8}} = 0.66$$

3 - اتخاذ القرار :

من خلال المقارنة بين $t=6.36$ وقيمة t الجدولية المساوية لـ 2.119 من أجل $v=16$ و $\alpha=5\%$ والاختبار من اتجاهين وبالتالي $t < t$ ونقبل الفرضية H_0 ونقول ليس هناك فرقاً بين قدرة عمال العينتين على تعلم الكلمات الجيدة.

٣- اختبار (t) للبيانات المزدوجة (the Paired t - Test) :

لقد تعرضنا في البند السابق لطريقة اختبار تساوي متوسطي مجتمعين عندما تكون المعاينة من مجتمعين طبيعيين مجهولي التباين، وافترضنا أن العينتين مستقلتين وأن هذه الطريقة قد تؤدي في بعض الأحيان إلى وجود فروق جوهرية بين متوسطي المجتمعين، ولكن الحقيقة خلاف ذلك، والسبب راجع إلى وجود فروق جوهرية بين متوسطي المجتمعين بالظاهرة قيد الدراسة، فمثلاً عند مقارنة نوعين من السماد، فإذا كانت قطعة الأرض التي استخدم فيها النوع الأول من السماد أعطت محصولاً أفضل من القطعة التي استخدم فيها النوع الثاني من السماد، فهذا لا يعني أن النوع الأول أفضل من الثاني، وذلك لأنه من الممكن أن يكون لنوعية التربة أو غير ذلك من العوامل تأثير أكبر من تأثير نوعية السماد.

وعليه للتخلص من تأثيرات العوامل الخارجية، فإننا نلجأ إلى اختيار مفردات العينتين على شكل أزواج متناظرة بحيث يكون لكل مفردة من مفردات العينة الأولى مفردة متناظرة لها من مفردات العينة الثانية ومساوية لها تقريباً من جميع الأوجه التي لها علاقة بالمتغير قيد الدراسة، ففي مثالنا السابق نختار أزواجاً من قطع الأرض التي يتشابه مفردات كل زوج منها من حيث نوع التربة والظروف الأخرى المحيطة بها، مع مراعاة استخدام نفس النوع من البذور في كل حالة،

وبالتالي إذا ظهرت فروق بين المحصولين، فإن ذلك سيكون راجعاً لنوعية السماد.

إن تحليل البيانات المزدوجة تتحول من مسألة عينتين إلى مسألة عينة واحدة وذلك من خلال تحليل الفروق ما بين الأزواج، فإذا كان متوسط هذه الفروق يساوي الصفر، فهذا يعني أنه لا يوجد فرق بينهما، ولتوضيح ذلك، نطرح لدينا التمرين التالي:

تمرين :

أراد مركز البحوث الزراعية التابع لوزارة الزراعة بإحدى الدول المقارنة بين نوعين من القمح، فحدد (9) مناطق وقام بزرع (5) هكتارات في كل منطقة فأعطت التجربة النتائج التالية:

X	38	23	35	41	44	29	37	31	39
Y	45	25	31	38	50	33	36	40	43

وبفرض أن كمية المحصول من النوعين تتبع التوزيع الطبيعي، اختبر الفرض القائل بأن متوسط كمية المحصول من النوعين متساوية عند مستوى معنوية ($\alpha=0.05$)

الحل :

الفرضية :

$$H_0 : \mu_x = \mu_y$$

$$H_1 : \mu_x \neq \mu_y$$

الاختبار الإحصائي :

X	y	d=x-y	d ²
38	45	-7	49
23	25	-2	4
35	31	4	16
41	38	3	9
44	50	-6	36
29	33	-4	16
37	36	-1	1
31	40	-9	81
39	43	-5	25

$$\sum d = -27, \sum d^2 = 237, \bar{d} = \frac{-27}{9} = -3 \quad S_d^2 = \left(\frac{\sum d^2}{n} \right) - \left(\frac{\sum d}{n} \right)^2 = 4.41$$

$$t = \frac{\bar{d}}{S_d / \sqrt{n}} = \frac{-3}{4.41 / \sqrt{9}} = -2.03$$

المقارنة والقرار : من جدول (t) نجد أن :

$$df = 9 - 1 = 8$$

$$\alpha = 0.05$$

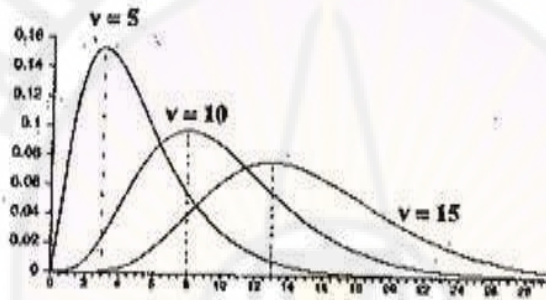
$$t = 2.306$$

من خلال المقارنة نجد أن :

(الجدولية) $-2.03 < 2.30$ (المحسوبة) وبالتالي نقبل فرضية (H_0) .

7-8- توزيع كاي مربع (χ^2) :

يتألف توزيع كاي مربع (χ^2) مثل توزيع (t) ستودنت، من مجموعة من المنحنيات، التي يتغير شكلها بتغير عدد درجات الحرية (v)، فعندما يكون ($v \leq 2$)، فإن منحنى التوزيع يأخذ شكل حرف الراء المقلوبة، وبزيادة عدد درجات الحرية يصبح شكل المنحنى ملتوياً نحو اليمين، ويقترب شكل المنحنى من منحنى التوزيع الطبيعي عندما يصبح عدد درجات الحرية ($v > 30$) وله تطبيقات عديدة في اختبارات جودة المطابقة والتجانس والاستقلالية.



نعلم بأن الأوساط الحسابية للعينات الصغيرة تتوزع وفق التوزيع (t)، أما بالنسبة للانحرافات المعيارية والمحسوبة من عينات صغيرة، فإن شكل توزيعها الاحتمالي ملتوٍ نحو اليمين (موجب الالتواء)، وتتوزع وفق (χ^2) الذي يعتمد على معلمتين أساسيتين هما، تباين المجتمع الإحصائي (δ^2) وحجم العينة المدروسة (n) وتعطى قيمة الاختبار بالعلاقة :

$$\chi^2 = \frac{(n-1)S^2}{\delta^2}$$

— خصائص توزيع كاي مربع :

- 1 — $\chi^2 \geq 0$
- 2 — توزيع كاي مربع يعتمد على متغير واحد فقط هو درجة الحرية ($V=n-1$)
- 3 — $E(\chi^2) = v$
- 4 — $\text{var}(\chi^2) = 2v$

8-8- التقدير المجالي لمعلمة المجتمع (δ^2) :

لتقدير قيمة تباين المجتمع الإحصائي δ^2 والذي يتميز بتوزيعه الطبيعي، يمكن الاستفادة من تعريف المتغير العشوائي χ^2_α الذي يفترض أن تكون العينة العشوائية مأخوذة من مجتمع إحصائي ذي توزيع طبيعي.

$$\chi^2 \leq \frac{(n-1)S^2}{\delta^2}$$

وبالتالي يمكن تقدير قيمة المعلمة (δ^2) عندما تكون مجهولة، بدلالة قيمة (S^2) والمحسوب من عينة عشوائية صغيرة الحجم، بدرجة احتمال ($1-\alpha$) وعدد من درجات الحرية ($V=n-1$). على النحو التالي :

$$P\left(a \leq \frac{(n-1)S^2}{\delta^2} \leq b\right) = 1 - \alpha$$

وبحل المتراجحات السابقة نحصل على :

$$P\left(\frac{(n-1)S^2}{b} \leq \delta^2 \leq \frac{(n-1)S^2}{a}\right) = 1 - \alpha$$

تمرين (9) :

إذا فرضنا أن كميات التفاح التي يتم تعبئتها في صناديق باستخدام آلة خاصة بها تتبع التوزيع الطبيعي، وإن هذه الآلة سوف تعتبر دقيقة بشكل كافي، إذا كان الانحراف المعياري (σ) لكميات التفاح المعبأة في كل صندوق أقل من (0.25) ملغم. فإذا تم اختيار عينة عشوائية من (6) صناديق من إنتاج هذه الآلة فكان متوسط كميات التفاح بها (16) وانحراف معياري $S=0.87$. المطلوب : أوجد 95% فترة ثقة لتقدير الانحراف المعياري (σ) لكميات التفاح المعبأة باستخدام هذه الآلة.

الحل:

$$\alpha=0.05 \quad 1-\alpha=0.95 \quad 1-\alpha=0.975$$

من الجدول نجد أن :

$$\chi^2_{0.025,5} = 12.832$$

$$\chi^2_{0.975,5} = 0.831$$

وبالتالي فإن فترة الثقة حول σ^2 ستكون كالآتي :

$$\frac{(n-1)S^2}{\chi^2_{\alpha/2, n-1}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-\alpha/2, n-1}}$$

$$\frac{0.876}{12.832} \leq \sigma^2 \leq \frac{0.876}{0.831}$$

$$0.261 \leq \sigma^2 \leq 1.02$$

وبالتالي إن الحد الأدنى لفترة الثقة أكبر من (0.25) وعليه لا يمكن اعتبار هذه الآلة تعمل بشكل دقيق.

9-8- اختبار الفروض للتباين σ^2 :

لإجراء الاختبار بين (S^2) و (δ^2) لمعنوية الفرق عند مستوى دلالة (α) وعدد من درجات الحرية $(V=n-1)$ يأخذ المتغير χ^2 الصيغة التالية:

$$\chi^2 = \frac{(n-1)S^2}{\delta^2}$$

وبمقارنة χ^2 المحسوبة مع χ^2_{α} الجدولية عند مستوى دلالة محدد (α) وعدد درجات حرية $(V=n-1)$ يتم اتخاذ القرار.

تعرين (10) :

بفرض أن أعمار العمال في الصناعة الغذائية تتبع التوزيع الطبيعي بانحراف معياري قدره (6) سنوات، سحبت عينة عشوائية من (25) عامل من إحدى الصناعات الغذائية، فوجد أن الانحراف المعياري للأعمار فسي العينة يساوي (5).

المطلوب : هل تدل هذه البيانات على أن تباين أعمار العمال بهذه الصناعة يختلف عن تباين أعمار العمال في الصناعات الغذائية عند مستوى دلالة $(\alpha = 0.05)$.

الحل :

1 - صياغة الفرضية :

$$H_0 : S^2 = \delta^2$$

$$H_1 : S^2 \neq \delta^2$$

2 - الاختبار الإحصائي :

$$\chi^2 = \frac{(n-1)S^2}{\delta^2} = \frac{24(5)^2}{(6)^2} = 16.66$$

3 - اتخاذ القرار :

نبحث عن قيمة $\chi^2_{0.05}$ من أجل $V = 24$ في جداول (χ^2) فنجد أن :

$$\chi^2_{0.05} = 36.415$$

من خلال المقارنة نجد أن : $\chi^2 = (16.66) < \chi^2 = (36.415)$ ، وبالتالي نقبل الفرضية H_0 ونقول ليس هناك من فرق معنوي بين تباين أعمار العمال في هذه الصناعة وتباين أعمار العمال في الصناعات الغذائية .

8-10- توزيع فيشر (F) :

عند مناقشتنا للاختبارات الخاصة بـ $(\mu_1 - \mu_2)$ في توزيع (t) ستودنت، افترضنا دوماً تجانس التباينات في المجتمعات المسحوبة منها العينات، أي أن التباين في العينة الأولى مساوٍ للتباين في العينة الثانية، لذلك وقبل إجراء اختبار معنوية الفروق للأوساط الحسابية للعينات الصغيرة الحجم لا بد من إجراء اختبار تجانس التباين في المجتمعات الإحصائية المسحوبة منها العينات، من خلال النسبة $\frac{S_1^2}{S_2^2}$ التي تمثل توزيعاً تكرارياً غير متماثلاً وهو ملتوٍ نحو اليمين، فكلما اقتربت النسبة من الواحد الصحيح اقترب تبايني المجتمعين

الإحصائين المدروسين من بعضهما البعض، وبالعكس كلما ابتعدت عن الواحد ابتعدت تبائني المجتمعين عن بعضهما البعض.

إن توزيع فيشر (F) يستخدم لاختبار تساوي تبائني مجتمعين وقيسة الاختبار تعطى بالعلاقة التالية :

$$F = \frac{S_1^2}{S_2^2}$$
 وهو عبارة عن نسبة التباين الأكبر إلى التباين الأصغر، وإن قيمة الاختبار (F) تأخذ قيمة موجبة دوماً.

وعند المقارنة واتخاذ القرار نقارن قيمة (F) المحسوبة من العلاقة السابقة مع قيمة (F) الجدولية، عند مستوى دلالة محدد وعدد مسن درجات الحرية، للتباين في المقام ($V=n_1-1$) ونحصل عليه من سطر جداول (F) وعدد من درجات الحرية، للتباين البسط ($V=n_2-1$) ونحصل عليه من عمود جداول (F)، بمعنى أننا نحصل على قيمة (F) الجدولية من نقطة تقاطع السطر مع العمود، مع أنه في بعض الحالات تؤخذ أقرب قيمة لدرجات الحرية فسي حال عدم توافرها في الجداول.

ويمكن القول، أنه كلما اقترب حاصل القسمة في علاقة (F) من الواحد الصحيح، ازداد الاحتمال بأن يكون الفرق غير معنوي بين التباينات، وبالعكس كلما ازدادت قيمة (F) ازداد الاحتمال بأن يكون الفرق معنوياً. إن توزيع (F) مشابه لتوزيع (χ^2) من حيث توزع المنحنيات التي يتغير شكلها بتغير عدد درجات الحرية وإن التوزيعان (t) و (χ^2) يعتمدان كلياً على معلمة واحدة فقط هي درجات الحرية بينما التوزيع (F) على معلمتين (V_1) و (V_2).

11-8- اختبار المقارنة بين تبائني مجتمعين :

نفرض أننا سحبنا عينتين عشوائيتين من مجتمعين إحصائيين مستقلين، وحصلنا على البيانات التالية :

العينة الأولى

$$n_1 = 12$$

$$\bar{x}_1 = 85$$

$$S_1 = 4$$

العينة الثانية

$$n_2 = 10$$

$$\bar{x}_2 = 81$$

$$S_2 = 5$$

والمطلوب :

اختبر الفرض القائل بأن تبايني المجتمعين متساويين $\sigma_1^2 = \sigma_2^2$ عند

$$\alpha=0.05$$
 مستوى دلالة

الحل :

1- صياغة الفرضية :

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

2 - الاختبار الإحصائي :

$$F = \frac{S_2^2}{S_1^2} = \frac{(5)^2}{(4)^2} = \frac{25}{16} = 1.5625$$

3 - المقارنة والقرار :

$$F(3,11)$$

من جداول F نجد أن :

$$V = (11,9)$$

$$\alpha=0.05$$

من خلال المقارنة نجد أن : $F < F_{\alpha}$ وبالتالي نقبل فرضية العدم (H_0)

والمجتمعين متجانسين أي أن $\sigma_1^2 = \sigma_2^2$

تعمرين:

في دراسة لنوعين من الدواء لعلاج مرض معين تمت مقارنتهما من حيث الفترة الزمنية التي يستغرقها كل منها لتسكين الألم، حيث أعطى الدواء الأول لعينة تتكون من (16) شخص والدواء الثاني لعينة تتكون من (13) شخص أيضاً، فكان تبين العينة الأولى يساوي (64) والثانية (16).
المطلوب: من هذه المعلومات هل يمكن القول بوجود فرق معنوي مابين تبايني المجتمعين عند مستوى معنوية (5%) .

الفرضية :

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

الاختبار الإحصائي :

$$F = \frac{S_1^2}{S_2^2} = \frac{64}{16} = 4$$

المقارنة والقرار :

$$n-1=12$$

$$m-1=15$$

$$F_{0.025,15,12} = 3.18$$

من جدول F نجد أن :

وبالتالي نجد (الجدولية) $F > F$ (المحسوبة) ونرفض (H_0)

تمارين غير محلولة

تمرين (1) :

ترغب إحدى الشركات الكبرى في معرفة ما إذا كان متوسط ربحها في السيارة الواحدة أقل من 3500 وحدة نقدية، سحبت عينة عشوائية من (16) فاتورة من الفواتير السابقة، فوجد أن متوسط الربح في العينة هو (3200) وحدة نقدية وبتباينها معياري 125 وحدة نقدية، المطلوب :

1 - هل تؤيد هذه البيانات الفرض القائل بأن متوسط الربح يقل عن (3500) وحدة نقدية عند $\alpha = 0.05$ ، وذلك بافتراض أن الأرباح تتبع التوزيع الطبيعي.

2 - استخدم بيانات العينة لإيجاد فترة ثقة 0.95 للمتوسط الحقيقي للأرباح في المجتمع.

تمرين (2) :

ترغب إحدى الكليات في معرفة متوسط المسافة التي يقطعها الطلبة من منازلهم إلى الكلية، فأخذت عينة عشوائية من (16) طالباً فوجد أن $\bar{x} = 12$ كيلو متراً وبتباينها $S^2 = 36$ والمطلوب :

1 - أوجد فترة ثقة 0.99 للمتوسط الحقيقي في المجتمع للمسافة المقطوعة .

تمرين (3) :

تدعي إحدى الكليات بأن متوسط عدد ساعات الدراسة الأسبوعية للطلاب في الأسابيع الأخيرة قبل بدء الامتحان هو (20) ساعة. أخذت عينة عشوائية من (16) طالباً، فوجد أن $\bar{x} = 18$ ساعة و $S = 4$ ساعات، المطلوب : هل تؤدي هذه البيانات إلى قبول الإدعاء السابق عند مستوى معنوية ($\alpha = 0.01$) بفرض أن ساعات الدراسة تتبع التوزيع الطبيعي .

تمرين (5):

تقوم إحدى المؤسسات بتدريب خريجي قسم الإعلام الجدد باستخدام برنامجين مختلفين للتدريب، حيث يلتحق كل خريج جديد بأحد البرنامجين عشوائياً. سحبت عينة من (25) من الخريجين الذين تم تدريبهم باستخدام البرنامج (A) كما سحبت عينة أخرى مستقلة من (16) مراقب الذين تم تدريبهم باستخدام برنامج التدريب (B)، والبيانات التالية تبين النتائج خلال (30) يوم من انتهاء التدريب .

(A)	(B)
$\bar{x}_2 = 80$	$\bar{x}_1 = 90$
$S_2^2 = 36$	$S_1^2 = 25$

والمطلوب :

هل تدل هذه البيانات على أن برنامج التدريب (A) يكافئ برنامج التدريب (B) عند مستوى معنوية $\alpha = 0.05$. بافتراض أن درجات الاختبار تتبع التوزيع الطبيعي والتباينات متساوية $\delta_1^2 = \delta_2^2$.

تمرين (5) :

سحبت عينة عشوائية من (30) طالباً من إحدى الجامعات السورية، فوجد أن تباين إنفاقهم السنوي بمئات الوحدات النقدية مساوٍ $S^2=100$ ، وإذا علمت أن الإنفاق السنوي لجميع طلبة الجامعات يتبع التوزيع الطبيعي بتباين قدره $S^2=90$ بمئات الوحدات النقدية .

المطلوب : هل تدل هذه البيانات على أن تباين إنفاق طلبة هذه الجامعة أكبر معنوياً من تباين إنفاق طلبة كل الجامعات، وذلك عند $(\alpha = 0.05)$.

تمرين (6) :

بفرض لدينا عينتين عشوائيتين مستقلتين لأجور العمال والعاملات في إحدى الشركات، فكانت النتائج الآتية:

العمال	العاملات
$n_1=5$	$n_2=8$
$\bar{x}_1=10750$	$\bar{x}_2=10491$
$S_1=150$	$S_2=200$

وبافتراض أن أجور العمال والعاملات تتبع التوزيع الطبيعي وإن التباينات متساوية.

المطلوب :

اختبر الفرض القائل: بأنه ليس هناك من فرق معنوي بين أجور العمال والعاملات هذه الشركة وذلك عند مستوى معنوية $(\alpha=0.05)$.

تمرين (7):

رغب أحد الباحثين في تقدير نسبة الأسر التي لا يزيد عدد أفرادها عن (3) أشخاص، وكان الاعتقاد بأن نسبة هذه الأسر تتراوح بين (60%) و (70%)

والمطلوب:

ما هو حجم العينة المراد سحبها عشوائياً لتقدير النسبة الحقيقية للأسر التي لا يزيد عدد أفرادها عن (3) أشخاص، وألا يزيد الخطأ عن (4%) واحتمال الدقة لا يقل عن (95.5%)

تمرين (8) :

تدعي مجلة في علم الإدارة أن (25%) من قرائها هم طلبة جامعيون. أخذت عينة عشوائية من (200) قارئ لها، ووجد أن (42) فقط هم طلبة جامعيون.

والمطلوب:

- 1 - اختبر الفرض القائل بأن نسبة القراء من الطلبة الجامعيين وما زالت (25%) عند مستوى معنوية ($\alpha=0.05$)
- 2 - باستخدام بيانات العينة، أنشئ فترة (95%) ثقة لنسبة القراء من المجتمع الذين هم طلبة جامعيون.

تمرين (9) :

ترغب إحدى القنوات التلفزيونية في معرفة، ما إذا كان درجة شعبية برنامجين تلفزيونيين متساوية، سحبت عينة عشوائية من (400) من المشاهدين المتوقعين للبرنامج الأول (A) فوجد أن (220) منهم يشاهدون البرنامج بانتظام، كما سحبت عينة أخرى مستقلة من حجم مماثل للمشاهدين المتوقعين للبرنامج الثاني (B) فوجد أن (260) منهم يشاهدون البرنامج بانتظام. المطلوب: اختبر الفرض القائل بتساوي النسبتين عند مستوى معنوية ($\alpha=0.05$).

تمرين (10):

يعتقد أن أكثر من (50%) من مشاهدي التلفزيون يشاهدون نشرة الأخبار المسائية في القناة الثانية. سحبت عينة عشوائية من (1600) مشاهد وجد أن (840) منهم يشاهدون نشرة الأخبار المسائية في القناة الثانية. المطلوب: اختبر عند مستوى معنوية ($\alpha=0.01$) صحة هذا الاعتقاد.

تمرين (11) :

ترغب وزارة العمل في إحدى الدول تقدير متوسط عمر العاطلين عن العمل باستخدام عينة عشوائية، بحيث لا تزيد الفرق بين متوسط العينة والمتوسط الحقيقي عن (2) وذلك بدرجة ثقة (0.99).
المطلوب: حدد حجم العينة المناسب سحبه عشوائياً، إذا علمت أن الانحراف المعياري لأعمار العاطلين عن العمل هو (5) سنوات.

تمرين (12) :

قررت إدارة شركة تجارية لها فرعان أن تقدر الفرق في متوسط حجم المبيعات لفرعين أخذت عيّنتان مستقلتان من مبيعات يوم ما في الفرعين فوجدت النتائج الآتية :

الفرع الأول	الفرع الثاني
$n_1=32$	$n_2=36$
$\bar{X}_1=5000$	$\bar{X}_2=3750$
$S_1=1500$	$S_2=1300$

والمطلوب: أنشئ تقديراً للفرق بين متوسط المبيعات لدى الفرعين.

تمرين (13) :

أثناء أزمة مياه في إحدى المدن رأت الشركة المختصة إجراء معاينة عشوائية على عدادات مياه منطقة سكنية بغية مراقبة الاستهلاك اليومي في المياه، في يوم معين فأظهرت عينة من (50) عداداً، أن متوسط الاستهلاك $\bar{x}=25$ بانحراف معياري $S=5m^3$ أنشئ تقدير فترة (95%) ثقة لمتوسط استهلاك مجتمع تلك المدينة من المياه.

تمرين (14):

في عينة عشوائية مكونة من (900) شخص من مشاهدي التلفزيون تبين أن (180) من بينهم يشاهدون برنامجاً تلفزيونياً معيناً. أوجد فترة الثقة (0.99) للنسبة الحقيقية لمشاهدي التلفزيون الذين يشاهدون هذا البرنامج.

تمرين (15):

تدعي عمادة كلية الإعلام في إحدى الجامعات أن نسبة الطلاب في الكلية هي (48%) والطلابات (52%)، وللتأكد من ذلك، سحبت عينة عشوائية من (100) طالب وطالبة أظهرت النتائج أن نسبة الطلاب هي (40%).
والمطلوب: هل كان الإدعاء صحيحاً، عند مستوى معنوية ($\alpha=0.05$) .

تمرين (16):

أدخل أحد المهندسين في مصنع للنسيج أدخل أحد المهندسين تحسينات عليه مدعياً تخفيف نسبة التلف إلى (4%)، للتأكد من ذلك أخذت عينة عشوائية من (200) قطعة، أظهرت (12) قطعة تالفة.
والمطلوب: هل كان هذا الإدعاء صحيحاً عند مستوى معنوية ($\alpha=0.05$) .



الفصل التاسع

الإحصاء الالاعلمي

— مقدمة

— اختبار حسن المطابقة

— اختبار الاستقلال لجداول التوافق $(2*2)$ ، $(3*2)$ ، $(3*3)$

— تصحيح ياتس

— الاختبارات الإحصائية الالاعلمية للمقارنة بين عينتين

• اختبار الإشارة

• اختبار مان ويتني

• اختبار ولكوكسون

— الاختبارات الإحصائية الالاعلمية بين عدة عينات

• اختبار كروسكال — واليس

— الاختبارات الإحصائية الالاعلمية للمقارنة بين عينتين للبيانات الاسمية

والترتبية

• اختبار كولموجروف — سمير نوف — ثنائي الحدين — فشر —

الوسيط

— نماين غير محلولة

المجلد الثاني

العدد الثاني



الإحصاء اللامعلمي

1-9 مقدمة :

إن معظم أساليب اختبار الفروض التي ناقشناها سابقاً تعرف بالطرائق المعلمية Parametric methods، وذلك لأنها تهتم بمعالم (ثوابت) المجتمع مثل $(\bar{x}, \sigma^2, P)$ ، وعلى الرغم من ذلك يوجد العديد من المسائل التي لا تستطيع فيها تطبيق الطرق المعلمية، ولا سيما عندما تؤخذ المقاييس أو المشاهدات في صور رتب يصبح من الأفضل استخدام الطرائق اللامعلمية Non parametric أضف إلى ذلك أن الطرائق المعلمية تعتمد على افتراض معرفة التوزيع الاحتمالي للمجتمع، فعند إجراء اختبارات الفروض الإحصائية للمتوسط والتباين أو للفرق بين وسطين حسابيين والنسبة بين تباينين، افترضنا أن التوزيع توزيعاً طبيعياً أو قريب من الطبيعي للمجتمعات الأصلية التي سحبت منها العينات، أما بالنسبة للطرائق اللامعلمية فإنها لا تفترض شكلاً معيناً لتوزيع المجتمع الأصلي، أي أنها حرة التوزيع free distribution، وإن العبارتين لامعلمي وحررة التوزيع ليس لهما نفس المعنى تماماً، فعبارة لامعلمي تستخدم لوصف اختبارات الفروض التي لا تتضمن أية معلومة من معالم المجتمع، وحررة التوزيع (توزع حر) تستخدم لوصف اختبارات الفروض التي لا تفترض شكلاً معيناً لتوزيع المجتمع الأصلي.

لكن في حال تحقق شرط تبعية البيانات للتوزيع الطبيعي ينصح باستخدام الطرائق المعلمية والتي تؤدي في هذه الحالة إلى نتائج أدق، أما الطرائق اللامعلمية تستخدم فقط عندما يتأكد الباحث من عدم إمكانية استخدام الطرق المعلمية، علماً أن الطرائق اللامعلمية تتطلب عدداً أقل من الفروض وسهولة الشرح والتطبيق، ومن هذه الاختبارات المتعلقة بالإحصاء اللامعلمي نذكر:

9-2- اختبار حسن المطابقة:

لقد بات معروفاً أن الدراسات الإحصائية تعتمد على فرض نموذج إحصائي تخضع له البيانات المتوفرة. فعلى سبيل المثال (درجات الطلاب تخضع للتوزيع الطبيعي، وأعداد حوادث السير تخضع لتوزيع بواسون وعدد المصابيح غير الصالحة في مصنع تخضع لتوزيع ثنائي الحدين وهكذا ..).

لكن السؤال الذي يطرح نفسه، كيف يمكن التأكد من أن البيانات المتوفرة تخضع لهذا النموذج الإحصائي المعطى؟ إن الإجابة على هذا السؤال يتم من خلال اختبار كاي مربع χ^2 لحسن المطابقة، وتقوم فلسفة هذا الاختبار على حساب مقياس يعبر عن مدى التطابق بين المشاهدات الحقيقية (التكرارات الحقيقية) مع المشاهدات المتوقعة (التكرارات المتوقعة)، فيما إذا كان النموذج الإحصائي صحيحاً، بمعنى آخر، إن اختبار حسن المطابقة هو عبارة عن اختبار فيما إذا كان توزع المفردات للعينة يوافق أو يطابق ماهو مفروض أو متوقع أن يكون توزعها في المجتمع. فإذا رمزنا بـ E للتكرارات المتوقعة و Q للتكرارات الحقيقية المعروفة من خلال العينة، فإنه يمكن حساب الاختلاف أو مدى التطابق بين هذه التكرارات من خلال العلاقة الآتية:

$$\chi^2 = \sum_{i=1}^r \left(\frac{(Q-E)^2}{E} \right)$$

وخطوات الاختبار هي نفسها الخطوات المعروفة لدينا.

تمرين (1) :

رغبت إحدى المؤسسات الاجتماعية معرفة رأي طلبة قسم الإعلام بالتعليم المفتوح ، فقامت المؤسسة بأخذ عينة عشوائية من (50) طالب ، فوجد أن (32) منهم يؤيدون الفكرة و (18) منهم يعارضون ، هل يمكن القول بأن

الآراء تنقسم بشكل متساوٍ حول هذا الموضوع، وذلك عند مستوى معنوية $\alpha=0.05$.

الحل:

1 - صياغة الفرضية:

$$H_0 : P_1 = P_2 = 1/2$$

$$H_1 : P_1 \neq P_2 \neq 1/2$$

2 - الاختبار الإحصائي :

$$\chi^2 = \sum \left(\frac{(Q - E)^2}{E} \right)$$

لحساب المجموع السابق نضع الجدول التالي :

الآراء \ التكرارات	Q	E	(Q-E)	(Q-E) ²	$\left(\frac{(Q - E)^2}{E} \right)$
مويد	32	25	7	49	1.96
معارض	18	25	-7	49	1.96
Σ	5	50			$\chi^2=3.92$

$$\chi^2 = 1.96 + 1.96 = 3.92$$

3 - اتخاذ القرار: من جداول χ^2 وعند $\alpha=0.05$ و $V=2-1=1$ نجد أن :

$$\chi^2(\alpha, v) = \chi^2(0.05, 1) = 3.84$$

ومن خلال المقارنة نجد أن :

$$\chi^2 = 3.92 > \chi^2(\alpha, v) = 3.841 \text{ (المحسوبة)}$$

وبالتالي نرفض فرضية العدم H_0 ونقول أن الآراء في هذا المجتمع لا تنقسم بالتساوي.

تمرين (2):

يهتم الباحثون في إحدى الدول بمعرفة الحالة التعليمية لجميع من لهم الحق في الانتخاب، وقد تم من خلال خبرة سابقة تصنيف جميع من لهم حق الانتخاب حسب حالتهم التعليمية إلى خمسة أقسام: ابتدائي، إعدادي، ثانوي، جامعي، دراسات عليا، وإن النسب المقابلة لهذه الأقسام على النحو التالي:

دراسات عليا	جامعي	ثانوي	إعدادي	ابتدائي	الحالة التعليمية
5	10	20	35	30	النسبة المئوية %

أخذت عينة عشوائية من (1000) ناخب فوجد أن (270) ابتدائي، (315) إعدادي، (230) ثانوي، (125) جامعي، (60) دراسات عليا.

والمطلوب:

اختبار فيما إذا كان التوزيع الحالي للحالة التعليمية لمن لهم حق الانتخاب في هذه الدولة مازال هو نفس التوزيع السابق الذي تم الحصول عليه من خلال خبرة سابقة، وذلك عند مستوى معنوية $\alpha=0.05$.

الحل:

1 - صياغة الفرضية :

$$H_0: P_1 = P_2 = P_3 = P_4 = P_5$$

علماً أن الحالة التعليمية تتوزع حسب النسب الآتية:

$$(5\%, 10\%, 20\%, 35\%, 30\%)$$

$$H_1: P_1 \neq P_2 \neq P_3 \neq P_4 \neq P_5$$

2 - الاختبار الإحصائي :

$$\chi^2 = \sum \left(\frac{(Q-E)^2}{E} \right)$$

نحسب التكرارات المتوقعة:

$$E_1 = \frac{(30)(1000)}{100} = 300 \quad (إعدادي) \quad E_2 = \frac{(35)(1000)}{100} = 350$$

$$E_3 = 200, \quad E_4 = 100, \quad E_5 = 50$$

جدول تحديد التكرارات الحقيقية والمتوقعة :

الحالة التعليمية (الخلايا)	Q (المتوقعة)	E (المتوقعة)	(Q-E)	(Q-E) ²	$\left(\frac{(Q-E)^2}{E} \right)$
ابتدائي	270	300	-30	900	3.00
إعدادي	315	350	-35	1225	3.50
ثانوي	230	200	300	900	4.50
جامعي	125	100	25	625	6.25
دراسات عليا	60	50	10	100	2.00
Σ	1000	1000			$\chi^2=19.25$

3 - اتخاذ القرار: من جداول χ^2 نجد أن :

$$\chi^2_{(0.05,4)} = 9.48$$

ومن خلال المقارنة نجد :

$$\chi^2 = 19.25 > \chi^2(\alpha, v) = 9.48$$

وبالتالي نرفض فرضية العدم H_0 ، ونقول: بأن التوزيع للحالة التعليمية قد تغير عما كان عليه في السابق.

تمرين (3) :

رغب أحد الباحثين معرفة آراء المواطنين بشأن الاتجاه نحو عمل المرأة، وقد تم من خلال خبرة سابقة تصنيف جميع الآراء والنسب المقابلة لها على النحو التالي :

الآراء	مؤيد بشدة	مؤيد	محايد	معارض	معارض بشدة	Σ
النسب المئوية %	30	20	10	25	15	100

قام الباحث باختيار عينة عشوائية تتكون من (200) مواطن لاستطلاع آرائهم بشأن الاتجاه حول عمل المرأة من خلال اختيار إجابة واحدة للسؤال الموجه إليهم، وهو هل تؤيد عمل المرأة خارج البيت في بعض الأعمال الموجودة في المجتمع؟ فكانت النتائج على النحو التالي:

الآراء	مؤيد بشدة	مؤيد	محايد	معارض	معارض بشدة	Σ
عدد المواطنين	75	55	15	45	10	100

المطلوب: اختبر الفرض القائل: بأن هناك تشابه بين نسب إجابات العينة مع النسب في المجتمع، من خلال الخبرة السابقة عند مستوى ($\alpha=0.05$)

الحل:

I - صياغة الفرضية :

$$H_0: P_1=P_2=P_3=P_4=P_5$$

تتوزع الحالة (الآراء) على النحو التالي :

(15%, 25%, 10%, 20%, 30%)

$$H: P_1 \neq P_2 \neq P_3 \neq P_4 \neq P_5$$

(أي لا تتوافق النسب في هذه الحالة مع النسب السابقة).

2 - الاختبار الإحصائي :

$$\chi^2 = \sum \left(\frac{(Q - E)^2}{E} \right)$$

نحسب التكرارات المتوقعة:

$$E_1 = \frac{(30)(200)}{100} = 60 \quad \text{مؤيد بشدة}$$

$$E_2 = \frac{(20)(200)}{100} = 40 \quad \text{(مؤيد)}$$

وبنفس الأسلوب لبقية التكرارات المتوقعة

جدول تحديد التكرارات الحقيقية والمتوقعة :

الحالة التعليمية (الخلايا)	Q (الفعلية)	E (المتوقعة)	(Q-E)	(Q-E) ²	$\left(\frac{(Q-E)^2}{E} \right)$
1	75	60	15	225	3.750
2	55	40	15	225	5.625
3	15	20	-5	25	1.250
4	45	50	-5	25	0.5
5	10	30	-20	400	13.333
Σ	200	200			$\chi^2 = 24.548$

3 - اتخاذ القرار: من جداول χ^2 نجد أن :

$$\chi^2_{(\alpha, v)} = \chi^2_{(0.05, 4)} = 9.487$$

ومن خلال المقارنة نجد :

$$\chi^2 = 24.548 > \chi^2_{(\alpha, v)} = 9.487$$

وبالتالي نرفض فرضية العدم H_0 ، أي أن النسب لم تعد كالسابق وليس هناك تشابه بين إجابات العينة والمجتمع.

ملاحظات حول اختبار حسن المطابقة:

- 1 - شاهدنا في بعض التطبيقات أننا افترضنا أن نسبة الإجابات في المجتمع الأصلي كانت متساوية لكل الاختبارات (التوزيع بالتساوي)، لكن هذا ليس هو الحال في جميع التجارب أو التطبيقات أو البحوث، فقد تكون التكرارات المتوقعة E تتوزع وفقاً لقواعد أو نسب معينة معروفة في ضوء خبرة سابقة؛
- 2 - عند استخدام اختبار χ^2 لجودة المطابقة، ينبغي ألا تقل القيمة المتوقعة (التكرار المتوقع) عن (5) أي عن 20% من العدد الكلي، ويمكن في بعض الأحيان اللجوء إلى ضم الخلايا (أي التكرارات المتوقعة) التي أقل من (5) إلى بعضها البعض، أما في حالة كانت هذه الخلية الوحيدة التي تقل عن (5) فلا تؤخذ بعين الاعتبار وتبقى كما هي.
- 3 - هناك إمكانية دمج عدة قيم محسوبة لـ χ^2 في قيمة واحدة، بمعنى إذا كان لدينا قيم كاي مربع لعدة عينات خاصة بنفس الموضوع المدروس ومسحوبة في نفس المجتمع، فإن هذه القيم يمكن إضافتها إلى بعضها لاتخاذ قراراً أفضل من القرار الذي اتخذ على أساس البيانات الخاصة بعينة واحدة فقط، ويجب أن تكون درجات الحرية في العينة متساوية في

كل التجارب وتكون درجة الحرية لاختبار χ^2 الكلية تساوي مجموع جميع درجات الحرية بالتجارب، وبعبارة أخرى، فإن شروط دمج قيمة χ^2 هي:

1 - أن تكون الجداول في التجارب من نفس النمط.

2 - أن تكون التجارب تحمل نفس درجات الحرية:

9-3- اختبار الاستقلال لجداول التوافق (2*2) أو (2*3) أو (3*3):

يقوم هذا الاختبار على اختبار التطابق النسبي بين التكرارات الحقيقية والتكرارات المتوقعة لظاهرتين، ومن ثم اتخاذ قرار بشأن استقلال متغير ما إحصائياً عن متغير آخر، أي تحديد فيما إذا كان هناك علاقة ما بين مجموعتين من الصفات في المجتمع، من خلال بيانات العينة. بمعنى آخر، اختبار العلاقة بين متغيرات الدراسة موضع الاهتمام غير القائم على الارتباط الخطي، وإنما ينصب اهتمامنا على التأكد من جودة العلاقة موضع البحث وهل يعملان باستقلالية عن بعضهما البعض، ويتم إعداد جدول التوافق الذي يبحث في مدى التطابق بين (Q) و (E) من خلال بيانات العينة على اعتبار كل ظاهرة منقسمة إلى قسمين أو أكثر، ويتم الاختبار بنفس الأسلوب السابق مع ملاحظة أن : حساب التكرارات المتوقعة E يتم على النحو التالي :

$$E = \frac{(\text{مجموع العمود}) (\text{مجموع السطر})}{(\text{المجموع الكلي})}$$

تمرين (4):

قام أحد الباحثين باختيار عينة عشوائية من مشاهدي البرامج التلفزيونية واستطلع آرائهم لمعرفة رأيهم في البرامج التلفزيونية المقترحة على القناتين الأولى والثانية، فكانت النتائج على النحو التالي :

القناة النوع	الأولى	الثانية	Σ
ذكور	110	90	200
إناث	40	60	100
Σ	150	150	300

والمطلوب:

اختبر الفرض القائل باستقلال كل نوع (ذكور، إناث) واختيار القناة عند مستوى معنوية $\alpha=0.05$.

الحل:

1 - صياغة الفرضية: (فرضية الاستقلال) لا توجد علاقة بين النوع (ذكور، إناث) وبين اختيار القناة، وكل ظاهرة مستقلة عن الأخرى.

2 - الاختبار الإحصائي : (اختبار التوافق) :

$$\chi^2 = \sum \left(\frac{(O - E)^2}{E} \right)$$

تحديد التكرارات المتوقعة E :

$$E_1 = \frac{(200)(150)}{300} = 100$$

$$E_3 = \frac{(100)(150)}{300} = 50$$

$$E_2 = 100, \quad E_4 = 50$$

نضع جدول تحديد التكرارات الحقيقية والتكرارات المتوقعة:

(الخلايا)	Q	E	(Q-E)	(Q-E) ²	$\left(\frac{(Q-E)^2}{E}\right)$
ذكور قناة (1)	110	100	10	100	1
ذكور قناة (2)	90	100	-10	100	1
إناث قناة (1)	40	50	-10	100	2
إناث قناة (2)	60	50	10	100	2
Σ	300	300			$\chi^2=6$

3 - اتخاذ القرار:

من جداول χ^2 عند مستوى دلالة $\alpha=0.01$ وعدد من درجات الحرية

$$V=(2-1)(2-1)$$

أي $V=1$ نجد أن : $\chi^2_{(0.01,1)}=6.634$

ومن خلال المقارنة نجد $\chi^2_{(\alpha,v)}=6.634 < \chi^2=6$ وبالتالي نقبل فرضية

الاستقلال ، ونقول ليس هناك علاقة بين الجنس واختيارهم للقناة.

– جداول (2 * 3):

تمرين (5):

البيانات التالية توضح نتائج دراسة حول تسلم المرأة لمناصب سياسية:

الآراء النوع	مؤيد	معارض	محايد	Σ
ذكور	80	40	40	160
إناث	100	80	60	240
Σ	180	120	100	400

والمطلوب:

اختبار فيما إذا كان هناك علاقة بين الرأي والنوع، عند مستوى معنوية

$$\alpha=0.05$$

الحل:

1 – صياغة الفرضية : (فرضية الاستقلال) لا توجد علاقة بين الآراء والنوع

وكل ظاهرة مستقلة عن الأخرى.

2 – الاختبار الإحصائي (اختبار التوافق):

$$\chi^2 = \sum \left(\frac{(Q-E)^2}{E} \right)$$

نحسب التكرارات المتوقعة :

$$E_1 = \frac{(180)(160)}{400} = 72$$

$$E_3 = 48$$

$$E_5 = 40$$

وهكذا لبقية التكرارات ...

جدول تحديد التكرارات الحقيقية والمتوقعة:

جدول تحديد التكرارات الحقيقية والمتوقعة :

الخلايا	Q	E	(Q-E)	(Q-E) ²	$\left(\frac{(Q-E)^2}{E}\right)$
1	80	72	8	64	0.89
2	100	108	8	64	0.59
3	40	48	8	64	1.33
4	80	72	8	64	0.89
5	40	40	0	0	0
6	60	60	0	0	0
Σ	400	400			$\chi^2=3.7$

3 - اتخاذ القرار : من جداول χ^2 وعند $\alpha = 0.05$ ودرجات حرية $V=(3-1)(2-1)=2$

نجد $\chi^2_{(0.05,2)} = 5.99$ وبالتالي $\chi^2_{(\alpha,v)} = 3.7 < \chi^2$ وبالتالي نقبل فرضية الاستقلال ونقول بأنه لا توجد علاقة بين الآراء والنوع .

هـ جداول (3×3) :

تمرين (6) :

قام باحث بإجراء دراسة لمعرفة ما إذا كان هناك علاقة بين الأفراد الذين يحملون مؤهلات علمية والاتجاه نحو عادات شرائهم للصحف والمجلات وكانت النتائج كما يلي:

عادات شرائهم المستوى التعليمي	غالباً	أحياناً	لا	Σ
إعدادي	56	71	12	139
ثانوي	47	163	38	248
جامعي	14	42	85	141
Σ	117	276	135	528

والمطلوب :

اختبر الفرض القائل بعدم وجود علاقة بين عادات الشراء والمستوى التعليمي، عند مستوى معنوية $\alpha = 0.05$.

الحل :

1 - صياغة الفرضية : (فرضية الاستقلال) لا توجد علاقة بين عادات الشراء والمستوى التعليمي، وكل ظاهرة مستقلة عن الأخرى .

2 - الاختبار الإحصائي : (اختبار التوافق):

$$\chi^2 = \sum \left(\frac{(Q - E)^2}{E} \right)$$

نحسب التكرارات المتوقعة E :

$$E_1 = \frac{(139)(117)}{528} = 30.80$$

$$E_4 = 54.95$$

$$E_3 = 35.54$$

$$E_2 = 72.66$$

وكذلك : E_8, E_7, E_6, E_5 بنفس الأسلوب ...

جدول تحديد التكرارات الحقيقية والمتوقعة التالي :

الخلايا	Q	E	(Q-E)	(Q-E) ²	$\left(\frac{(Q-E)^2}{E} \right)$
1	56	30.80	25.2	635.04	20.62
2	71	72.66	-1.66	2.75	0.04
3	12	35.54	-23.54	554.13	15.59
4	47	54.95	-7.95	63.20	1.15
5	163	129.64	33.36	1112.89	8.58
6	38	63.41	-25.41	645.67	10.18
7	14	31.23	-17.24	297.21	9.51
8	42	73.70	-31.7	1004.89	13.63
9	85	36.05	48.95	2396.10	66.47
Σ	528	528			$\chi^2 = 145.77$

3- اتخاذ القرار : من جداول χ^2 عند $\alpha = 0.05$ ودرجات حرية $V=3$
 $(3-1)=4$ نجد $\chi^2_{(0.05,4)} = 9.487$ ، ومن خلال المقارنة $\chi^2 = 145.77 > \chi^2_{(0.05,4)} = 9.487$
 $\chi^2_{(\alpha, v)} = 9.487$
 وبالتالي نرفض فرضية الاستقلال ونقول أن هناك علاقة بين الإنجاز العلمي
 والحالة الزوجية لدكاترة الجامعات.

3-9- تصحيح ياتس :

يستخدم هذا التصحيح في اختبارات الاستقلال لجداول التوافق
 (2×2) و (2×3) إذا كان أحد التكرارات المتوقعة أقل من (10) في
 إحدى الخلايا، وتقوم على طرح $(1/2)$ من كل تكرار حقيقي أكبر من التكرار
 المتوقع E المقابل له، وإضافة $(1/2)$ إلى كل تكرار حقيقي Q إذا كان أقل من
 التكرار المتوقع E المقابل له (وذلك لجميع التكرارات الحقيقية)، وينتج عن هذا
 التصحيح تقليل الفرق بين التكرارات الحقيقية والتكرارات المتوقعة، باعتبار أن
 التكرارات الحقيقية تعتمد على قيم صحيحة، مع العلم أن التكرارات الحقيقية
 تختلف وتقفز قفزات ذات درجات منفصلة، بينما جدول χ^2 الذي يمثل توزيع
 درجات χ^2 يعتبر ذا قيم متصلة، لذلك يستخدم من أجل تخفيف هذه القفزات، مع
 ذلك عندما تكون التكرارات (في العينة) كبيرة الحجم لا داع لاستخدام هذا
 التصحيح ويصبح دون معنى.

تمرين (7):

أجرى أحد الباحثين الاجتماعيين بحثاً عن دور الاتصال عن طريق
 معرفة اتجاهات الناس إزاء برامج الإذاعة، فسأل عينة من الأفراد عددها (43)
 شخصاً، السؤال التالي:

هل من السهل الاستماع إلى للأخبار بدلاً من قراءتها، فأجاب (10) أشخاص (بنعم) من أصل (19) شخص من المدراء، وأيضاً أجابوا (20) شخصاً (بنعم) من أصل (24) شخص من العمال، فإذا علمت أن الإجابات تلتخصر بـ نعم أو لا، والمطلوب:

اختبر فيما إذا كان هناك علاقة بين نوع الإجابة والمهنة عند مستوى معنوية $\alpha=0.05$ ؟

الحل:

1 - صياغة الفرضية : (فرضية الاستقلال) لا توجد علاقة بين نوع الإجابة (نعم، لا) والمهنة، وكلأ منهما مستقل عن الآخر.

2 - الاختبار الإحصائي:

$$\chi^2 = \sum \left(\frac{(O - E)^2}{E} \right)$$

نضع جدول التكرارات النظرية :

المهنة \ الإجابة	المدراء	عمال	Σ
نعم	10	20	30
لا	9	4	13
Σ	19	24	43

ثم نحسب التكرارات المتوقعة E

نحسب التكرارات المتوقعة E :

$$E_1 = \frac{(30)(19)}{43} = 13.26$$

$$E_3 = 5.74$$

$$E_2 = \frac{(30)(24)}{43} = 16.74$$

$$E_4 = 7.26$$

نلاحظ أن بعض التكرارات المتوقعة أقل من 10، أي $E < 10$ والعينة صغيرة لذلك لابد من إجراء تصحيح ياتس.

— فإذا كان $E > Q$ نضيف $1/2$ لـ Q

— وإذا كان $E < Q$ نطرح $1/2$ من Q

كما هو واضح من خلال الجدول التالي :

الخلايا	Q	E	Q_1	$(Q_1 - E)$	$(Q_1 - E)^2$	$\left(\frac{(Q_1 - E)^2}{E} \right)$
1	10	13.26	10.5	-2.76	7.62	0.57
2	20	16.74	19.5	+2.76	7.62	0.45
3	9	5.74	8.5	+2.76	7.62	1.33
4	4	7.26	4.5	-2.76	7.62	1.05
Σ	43					$\chi^2 = 4.3$

3 — اتخاذ القرار: من جداول χ^2 عند $\alpha = 0.05$ و $v = 1$ نجد $\chi^2 = 3.84$ وبالمقارنة $\chi^2_{(a,v)} = 3.84 < \chi^2 = 3.4$ وبالتالي نقبل فرضية الاستقلال، ونقول بأنه لا يوجد علاقة بين نوع الإجابة والمهنة.

ملاحظات:

1 — في بعض الأحيان قد تعطى البيانات على شكل نسب مئوية ففي هذه الحالة، لا بد من تحويل النسب المئوية إلى أرقام مطلقة ومن ثم إجراء الاختبار. مثلاً: إذا كان لدينا مجتمع الدراسة ($n = 150$) حيث أن (80%) منهم نساء و (20%) رجال فمن الضروري تحويل هذه البيانات إلى (120) نساء و (30) رجال.

2 — نحن نعلم أن فكرة جداول الاستقلال (2×2) قائمة على فكرة الاستقلال بين الظاهرتين المدروستين أو عدم الاستقلال، لذلك في بعض الأحيان بعد أن نقوم بإجراء الاختبار للجدول (2×2) نرغب معرفة النسبة المئوية التي تحدد مدى الارتباط أو الاستقلال بين الظاهرتين عن طريق

استخدام معامل الاقتران الذي هو عبارة عن مقياس يقيس شدة العلاقة الارتباطية بين ظاهرتين في الجداول (2 × 2) وتعطى قيمة بالعلاقة :

$$r_c = \frac{AD - BC}{AD + BC}$$

وكلما اقتربت قيمة هذا المعامل من الواحد كانت العلاقة قوية، بمعنى إذا كانت $r_c > 50\%$ دل ذلك على وجود علاقة بين المتغيرات وبالعكس، يستخدم هذا المعامل أحياناً لدعم القرار المتخذ في اختبار (2 × 2)، فيما إذا كان قرارنا صحيحاً أم لا. فعلى سبيل المثال، لو كان قرارنا قبول فرضية الاستقلال، وقمنا بحساب معامل الاقتران وكان $r_c = 0.29$ ، أي أن $r_c < 50\%$ أي أن وهذا يدل على أن فالعلاقة ضعيفة بين الظاهرتين، وهذا يؤكد قرار الاختبار الإحصائي بقبول فرضية الاستقلال، أي لا علاقة بين الظاهرتين المدروستين.

3 — بنفس الأسلوب السابق عندما يكون لدينا جداول (3 × 3) نستخدم معامل التوافق :

$$r_A = \sqrt{\frac{G-1}{G}} ; G = \sum \left(\frac{f_{ij}^2}{f_i \cdot f_j} \right)$$

وكلما اقتربت قيمة r_A من الواحد كانت العلاقة قوية ومتوافقة (أي توافق قوي بين البيانات)، وأيضاً يستخدم هذا المعامل لدعم القرار المتخذ في اختبار (3 × 3)، فيما إذا كان قرارنا صحيحاً أم لا، فإذا كان $r_A > 0.50$ فالقرار المتعلق بالاختبار هو رفض فرضية الاستقلال، وبالعكس إذا كان $r_A < 0.05$ كان القرار هو قبول فرضية الاستقلال، وسيوضح تطبيق هذه العلاقة لاحقاً.

4-9- الاختبارات الإحصائية اللاحتمالية للمقارنة بين عيّنتين (اختبار الإشارة،

اختبار مان ويتني، اختبار ولكوكسون):

• اختبار الإشارة Sign Test :

لقد استخدمنا في السابق اختبار (t) لتحديد ما إذا كان متوسط المجتمعين متساويين، وبفرض أن كل مجتمع من المجتمعين له توزيع طبيعي وأن التباينين مجهولين ومتساويين، لكنه وفي كثير من الأحيان لا يتحقق أحد الشرطين أو كلاهما، وبالتالي لا نستطيع استخدام اختبار (t)، ويصبح البديل هو اختبار الإشارة القائم على أن المجتمعين لهما نفس التوزيع ويقوم هذا الاختبار على اختبار صفة واحدة من صفتين مثلاً (الصح والخطأ) أو (النجاح والفشل) أو (اختيار الصفة الأكثر انتشاراً)، وسوف نميز ثلاث حالات هي :

أ - حالة عينة واحدة : يتم هذا الاختبار على النحو التالي : نختار رمز الإشارة الموجبة (+) للصفة التي تهمننا والإشارة السالبة (-) للفشل، ويمكن أن يكون بالعكس، إذا كان الأمر الذي يثير اهتمامنا هو مقدار الفشل، وهكذا بالنسبة لأية صفة وعكسها.

وفرضية العدم (H_0) ستكون ليس هناك أي فرق بين الصفة وعكسها أو عدد الإشارات الموجبة (+) تعادل عدد الإشارات السالبة (-) ، والاختبار الإحصائي يتم بتحديد مستوى المعنوية (α)، ثم نحصى عدد الإشارات الموجبة وبعدها ننظر في الجدول الخاص باختبار الإشارة (في آخر الكتاب) عن القيمة المقابلة لعدد الإشارات الموجبة أو بالعكس إذا كانت الإشارة السالبة هي التي تثير اهتمامنا، ومن ثم اتخاذ القرار من خلال المقارنة بين P و P_{α} و (P) الجدولية لقبول أو رفض فرضية العدم، مع العلم أنه يمكن الاستعاضة عن جداول اختبار الإشارة بجداول ثاني الحدين. ونستوضح ما تم شرحه من خلال التمرين التالي :

تمرين (8) :

قام أحد الباحثين باختيار عينة عشوائية من (10) مشاهدين واستطلع آرائهم حول مشاهدة إحدى القنوات العربية، فوجد من بينهم اثنين فقط يشاهدان هذه القناة.

المطلوب: اختبر الفرض القائل: بأن مشاهدي هذه القناة أقل من مشاهدي القنوات الأخرى، عند مستوى معنوية $\alpha=0.10$ ؟

— صياغة الفرضية : $H_0: P = 1/2$

$H_1: P < 1/2$

— الاختبار الإحصائي :

$$n = 10 \Rightarrow (-8), (+2)$$

$$n=10$$

— اتخاذ القرار : من جداول الإشارة نجد أن : $P_{10} (+2, -8)$

$$(-8), (+2)$$

$$P = \alpha = 0.10$$

$$P_{\alpha} = (0.05469)$$

من خلال المقارنة نجد أن : $P_{\alpha} = (0.05469) < P = (0.10)$ وبالتالي نرفض فرضية العدم (H_0) ونقبل الفرضية البديلة (H_1) ونقول أن مشاهدي هذه القناة أقل من مشاهدي القنوات الأخرى .

ب — حالة وسيط عينة :

إذا كان لدينا مجتمعاً مستمراً (أي لا توجد فيه مفردتان لهما نفس القيمة) وأردنا اختبار أن وسيطه (Med) يزيد عن قيمة معينة أو ينقص عنها، فإننا نختار عينة عشوائية حجمها (n) منه وننظر كم من مفردات هذه العينة

يتجاوز القيمة المعينة [أي بإشارة موجبة (+)] وكم يقل عن تلك القيمة [أي بإشارة سالبة (-)] أو العكس، نوضح ذلك بالتمرين التالي:

تمرين (9) :

تم وضع امتحان لقسم الإعلام في إحدى الجامعات لأحد المقررات، بحيث أن وسيط توزيع الزمن الذي ينتهي فيه الامتحان والخروج من القاعة يزيد عن ساعتين، وللتأكد من ذلك أخذت عينة عشوائية من الطلاب الذين دخلوا ذلك الامتحان حجمها (25) طالباً، فوجد أن (16) طالباً منهم انتهوا من ذلك الامتحان وخرجوا بعد مرور أكثر من ساعتين على بدء الامتحان.

والمطلوب:

اختبر الفرض القائل: بأن الزمن اللازم الذي استغرقه الطلاب في هذا الامتحان يتوزع حول توزيع وسيطه (M) الذي يتجاوز ساعتين من الزمن اللازم، عند مستوى معنوية $\alpha = 0.10$

— صياغة الفرضية : $H_0: Med = 2$

$H_1: Med > 2$

— الاختبار الإحصائي :

$$n = 25 \Rightarrow (-9), (+16)$$

— اتخاذ القرار : من جداول الإشارة نجد أن :

$$n=25 \Rightarrow (-9), (+16)$$

$$P = 0.11476$$

من خلال المقارنة نجد أن : $P = (0.11476) > P = (0.10)$ وبالتالي نرفض فرضية العدم (H_0) ونقبل الفرضية البديلة (H_1) أي أن الزمن اللازم يتجاوز الساعتين .

جـ - حالة عينتين غير مستقلتين :

يقصد بالعينتين الزوجيتين (عينتان غير مستقلتين)، أي أن أحدهما مرتبطة بالأخرى (صورة أزواج من القيم) كأن تكون عينة واحدة أخذ لها القياس مرتين مرة قبل التجربة، ثم مرة أخرى بعد التجربة، مثل إنتاج العمال في الساعة قبل الإضراب وإنتاجهم لساعة بعد الإضراب، ويتم القياس للبيانات التي في صورة أزواج من خلال الفرق بين مفردة معينة والمفردة المقابلة لها، وإذا كان الفرق سالباً يعطي إشارة (-) وإذا كان الفرق موجباً يعطي إشارة (+) ويعطي الفرق عند التساوي الصفر وفي هذه الحالة، تستقط القيمة من الحساب ويعتبر (n) عدد الفروق التي لها إشارة موجبة أو سالبة. ونوضح ذلك من خلال التمرين التالي :

تمرين (10) :

تم اختيار (11) شخصاً بطريقة عشوائية، لمعرفة فيما إذا كان لنظام تغذية معين أثر على تخفيض الوزن قبل وبعد استخدام نظام التغذية عند مستوى معنوية $\alpha=0.10$ ، فأعطت النتائج الآتية:

رقم الشخص	1	2	3	4	5	6	7	8	9	10	11
الوزن قبل (x)	75	63	40	80	75	56	60	75	81	85	61
الوزن بعد (y)	77	60	41	82	73	56	69	79	86	93	63

والمطلوب : اختبر الفرض القائل بعدم فعالية نظام التغذية مقابل الفرض القائل بأن نظام التغذية أثر على تخفيض الوزن بعد استخدام النظام .
الحل :

— صياغة الفرضية : $H_1: P < 1/2$ ، $H_0: P = 1/2$

– الاختبار الإحصائي :

رقم الشخص	1	2	3	4	5	6	7	8	9	10	11
الوزن قبل (x)	75	63	40	80	75	56	60	75	81	85	61
الوزن بعد (y)	77	60	41	82	73	56	69	79	86	93	63
الإشارة	+	-	+	+	-	0	+	+	+	+	+

نضع إشارة (+) في حالة زيادة الوزن ونضع إشارة (-) في حالة انخفاض الوزن ونضع (0) في حالة التساوي قبل وبعد، وتستبعد القراءة رقم (6) من الدراسة ونتعامل فقط مع (10) حالات .

اتخاذ القرار :

$$(+8), (-2) \Rightarrow n=10$$

$$P = 0.05469$$

من خلال المقارنة نجد أن : $P = (0.10) < P = (0.05469)$

وبالتالي نرفض فرضية العدم (H_0) ونقبل الفرضية البديلة (H_1) أي يوجد فعالية تأثير لهذا النظام في التغذية على تخفيض الوزن .

• حالة خاصة : تقريب اختبار الإشارة من التوزيع الطبيعي المعياري :

عندما يكون حجم العينة المدروسة كبيراً فإنه يمكن الاستغناء عن القيمة الفعلية (P) المستمدة من جداول اختبار الإشارة ونستخدم جداول (Z) المستمدة من التوزيع الطبيعي المعياري وفي الأغلب يعتبر حجم العينة (n) في هذه الحالة كبيراً إذا لم تقل عن (10) وبالتالي توزيع عدد الإشارات الموجبة تابعاً للتوزيع الطبيعي بمتوسط ($n/2$) وبتباين ($n/4$)، وفكرة التقريب من التوزيع الطبيعي

جاءت من أن اختبار الإشارة قائم على مفهوم النجاح والفشل وعدد مرات النجاح والفشل (P), (1-P) لذلك يتبع التوزيع الثنائي $C_n^k P^k q^{n-k}$ وبالتالي :

$$E(x) = np = \mu$$

$$\sigma = \sqrt{npq}$$

ونحن نعلم بالإمكان تقريب توزيع ثنائي الحدين (التوزيع المنقطع) إلى التوزيع الطبيعي (Z) (التوزيع المستمر) عند زيادة عدد مرات التجربة، وعليه يمكن الحصول على (Z) بطرح المتوسط والقسم على الانحراف المعياري ثم إجراء المقارنة مع $(Z_{\alpha/2})$ من جدول التوزيع الطبيعي المعياري لتحديد قبول أو رفض فرضية العدم (H_0)، فإذا رمزنا لعدد الإشارات الموجبة في العينة بالرمز (T) فيمكن إيجاد (Z) بافتراض أن عدد (n) كبيراً على النحو التالي:

$$Z = \frac{T - n/2}{\sqrt{n/4}}$$

تمرين (11) :

نعود إلى التمرين (10) السابق والمتعلق بنظام التغذية والأثر على تخفيض الوزن ، ونستخدم التوزيع الطبيعي كتوزيع تقريبي لاختبار الإشارة لاختبار تأثير نظام التغذية على تخفيض الوزن عند مستوى معنوية ($\alpha=0.10$) ؟

الاختبار الإحصائي : $(+8), (-2) \Rightarrow n=10$

$$Z = \frac{8 - n/2}{\sqrt{n/4}} = \frac{8 - 5}{\sqrt{2.5}} = 1.897$$

اتخاذ القرار : من خلال جداول (Z) نجد أن :

$$Z_{\alpha=0.10} = 1.28$$

من خلال المقارنة نجد أن $Z(1.897) > (1.28)$ وبالتالي نرفض فرضية العدم (H_0) ونقول ليس لهذا النوع من نظام التغذية أي تأثير على تخفيض الوزن.

ملاحظة: في بعض الأحيان لزيادة الدقة في التقريب يستخدم ما يسمى بمعامل الاتصال (عند الانتقال من التوزيع الثنائي إلى التوزيع الطبيعي) وتصبح العلاقة:

$$Z = \frac{T - \frac{n}{2} - \frac{1}{2}}{\sqrt{\frac{n}{4}}} \quad \text{عندما } T > \frac{n}{2}$$

$$Z = \frac{T - \frac{n}{2} + \frac{1}{2}}{\sqrt{\frac{n}{4}}} \quad \text{عندما } T < \frac{n}{2}$$

لوعدنا للتمرين (10) السابق، المطلوب: اختبر باستخدام معامل الاتصال.

الاختبار الإحصائي:

$$Z = \frac{T - \frac{n}{2} - \frac{1}{2}}{\sqrt{\frac{n}{4}}} = \frac{8 - 5 - \frac{1}{2}}{\sqrt{2.5}} = 1.581$$

$$\text{لأن } T = 8 > \frac{n}{2} = 5$$

اتخاذ القرار: $Z(1.581) > Z(1.28)$ وبالتالي نفس القرار السابق نرفض (H_0)

تمرين (12):

تم استخدام عقار جديد لعلاج مرض السكر، لمعرفة فيما إذا كان له تأثير على ضربات القلب، فأعطى العلاج لـ (12) مريضاً وكان عدد ضربات القلب في الدقيقة لكل منهم على النحو التالي:

رقم المريض	1	2	3	4	5	6	7	8	9	10	11	12
قبل العلاج	76	80	91	75	81	77	79	82	88	81	78	85
بعد العلاج	78	81	92	74	84	77	78	83	83	80	79	85

والمطلوب :

اختبر الفرض القائل بأن للعقار تأثير على ضربات القلب وذلك باستخدام

التوزيع الطبيعي كتوزيع تقريبي لاختبار الإشارة $\alpha = 0.01$

.... صياغة الفرضية : $H_0 : P_1 = P_2 = 1/2$

$H_1 : P_2 > P_1$

الاختبار الإحصائي :

رقم المريض	1	2	3	4	5	6	7	8	9	10	11	12
قبل العلاج	76	80	91	75	81	77	79	82	88	81	78	85
بعد العلاج	78	81	92	74	84	77	78	83	83	80	79	85
الإشارة	+	+	+	-	+	0	-	+	-	-	+	0

تم استبعاد قرانتين : $n=10$ وعدد الإشارات الموجبة 6 أي :

$$n=10 \Rightarrow (-4), (+6)$$

$$Z = \frac{T - n/2}{\sqrt{n/4}} = \frac{6 - 5}{\sqrt{10/4}} = 0.632$$

اتخاذ القرار :

من جداول (Z) نجد أن : $Z(2.33)$ (اتجاه واحد)

$$\alpha = 0.01$$

من خلال المقارنة نجد أن : $Z(2.33) < Z(0.63)$ وبالتالي نقبل فرضية العدم (H_0) وليس للعقار تأثير .

وبناءً على ما سبق شرحه، نشير إلى الملاحظات الآتية :

1 - أكثر استخدامات اختبار الإشارة يكون في اختبار :

... الفرض القائل بعدم فعالية نظام معين أو فعالية نظام معين (التخذية مثلاً)؛

— الفرض القائل بعدم تأثير أو تغير نظام معين (إنتاجية، دواء) أو تأثير وتغير نظام معين؛

2 - معظم اختبارات الإشارة قائمة على عينتين مرتبطتين وغير مستقلتين (قبل وبعد) أي البيانات في صورة أزواج للقيم؛

3 - في حالة عدم توفر جداول الإشارة يمكن استخدام جداول ثنائي الحدين (في نهاية الكتاب).

• اختبار مان ويتني Mann Whitney :

يستخدم هذا الاختبار عند الرغبة في فحص الفرق بين عينتين مختارتين ومستقلتين عن بعضهما. ويعتبر هذا الاختبار البديل الآخر لاختبار (t) في حالة عدم معرفة التوزيع الاحتمالي الذي تتبعه الظاهرة المدروسة، ويقوم هذا الاختبار على تحديد الرتبة المناظرة لكل قراءة بنفس أسلوب الرتب عند حساب معامل ارتباط الرتب (سبيرمان)، ومن ثم المقارنة بين مجموعات الرتب والهدف من هذا الاختبار هو إجراء المقارنة لمعرفة فيما إذا كانت إحدى مجموعات الرتب تقل بصورة ملموسة عن المجموعة الأخرى، وسنتناول هنا، حالة عينتين مستقلتين:

— حالة عينتين مستقلتين :

لقد تناولنا اختبار مجتمع معين وأيضاً مجتمعين في صورة أزواج للقيم، أي مرتبطين وغير مستقلين، أما الآن فسوف نتناول كيفية اختبار مجتمعين اثنين مستقلين (أي لا يؤثر أحدهما على الآخر) ويشترط في المجتمعين محل الدراسة، أن يكونا متطابقين من حيث شكل التوزيع وحجمه، وبالتالي إمكانية اختبار النزعة المركزية للمجتمعين، أي فيما إذا كان لهما نفس التمرکز وفي الغالب يؤخذ الوسيط وفي حالة التماثل ممكن استخدام الوسيط والوسط الحسابي وإجراء الاختبار يتم على النحو التالي :

إذا كان لدينا مجتمعان متماثلان أحدهما (x) ووسيطه (M_x) والآخر (y) ووسيطه (M_y) فإنه يمكن اختبار وسيطي المجتمعين (M_x) و (M_y) فيما إذا كان $(M_x > M_y)$ أو $(M_x < M_y)$ أي $(M_x \neq M_y)$ ، لذا تؤخذ عينة عشوائية من كل مجتمع من المجتمعين، نختار الرمز (x) لأحد المجتمعين ويفضل أن يكون المجتمع الذي سحبت منه العينة الأصغر (n) و (y) المجتمع الآخر الذي سحبت منه العينة الأكبر (m) ، والحجم الكلي للعينتين هو $(N=n+m)$.

تؤخذ عينتا المجتمعين وتمزجان معاً مع الاحتفاظ بالمقدرة على معرفة مصدر كل مفردة في العينة الكبيرة التي حجمها (N) ومن ثم يتم ترتيب هذه العينة المشتركة. ثم نجمع الرتب الخاصة، أي نجمع الرتب لمفردات العينة القادمة من المجتمع (x) ويرمز لها (T_x) وأيضاً بنفس الأسلوب (T_y) أو من خلال : $(T_y = 1/2 N(N+1) - T_x)$ بعد الحصول على قيمة T_x يتم الكشف عن قيمة (P) المقابلة لها في جداول مان ويتني من خلال (m, n) بحيث (T_x) على يمينها و (T_y) على يسارها ومن ثم المقارنة بين (P) ومستوى المعنوية (α)

لقبول أو رفض الفرضية H_0 . مع ملاحظة أنه يتم استخدام جداول مان ويتني، عندما يكون حجم العينة $(n, m) \leq 10$ و لتوضيح ذلك لدينا التمرين التالي:

تمرين (13) :

في دراسة لمعرفة التأثير الجانبي لدواء جديد مطور في رفع ضربات القلب عند المرضى الذين يعالجون به عن الدواء القديم الذي يستخدم في العلاج لحالات المرضى نفسها، أخذت عينة عشوائية متجانسة لهؤلاء المرضى حجمها (10) مريض ووزعت عشوائياً إلى مجموعتين في كل مجموعة (5) مريض، عولجت كل واحدة من هاتين المجموعتين عشوائياً بأحد النوعين من هذا الدواء . حيث عولجت إحدى المجموعتين بالدواء القديم و الأخرى بالدواء المطور، ثم قيسَت الزيادة في ضربات القلب الناتجة عن العلاج بأحد الدوائين فكانت النتيجة كما يلي :

رقم المفردة	1	2	3	4	5
الزيادة في ضربات القلب (الدواء القديم)	15	22	23	11	27
الزيادة في ضربات القلب (الدواء الجديد)	26	25	28	18	29

المطلوب :

اختبر الفرض القائل بأن الدواء الجديد يؤدي إلى ارتفاع ضربات القلب لدى المرضى المعالجين به، عند مستوى معنوية $(\alpha = 0.10)$.

الحل :

— صياغة الفرضية :

$$H_0 : M_x = M_y$$

$$H_1 : M_x < M_y$$

– الاختبار الإحصائي :

رقم المفردة	1	2	3	4	5
الدواء القديم (x)	15	22	23	11	27
Tx (الرتب)	2	4	5	1	8
الدواء الجديد (y)	26	25	28	18	29
Ty (الرتب)	7	6	9	3	10

$$\sum T_x = 2 + 4 + 5 + 1 + 8 = 20 \quad \text{عينة المجتمع (x)}$$

اتخاذ القرار :

من خلال جداول مان ويتني نجد أن :

$$N: n=5, \quad m=5, \quad P=0.0754$$

من خلال المقارنة نجد أن : $P = (0.0754) < P = (0.10)$ وبالتالي
نقبل الفرضية البديلة (H_1) ونرفض فرضية العدم (H_0) أي الدواء الجديد يسبب
ارتفاع في ضربات القلب.

– تقريب اختبار مان ويتني بالتوزيع الطبيعي المعياري :

ويتم ذلك عندما تكون العينات كبيرة ($n, m=15$)، فإنه يمكن تقريب
هذا الاختبار باستخدام التوزيع الطبيعي المعياري للحصول على القيمة المعيارية
(Z). وذلك على النحو التالي:

$$Z = \frac{U - \frac{(n_1)(n_2)}{2}}{\sqrt{\frac{(n_1)(n_2)(n_1 + n_2 + 1)}{12}}}$$

حيث أن :

$$U = (n_1)(n_2) + \frac{n_1(n_2 + 1)}{2} - \sum T_x$$

ولتوضيح كيفية إجراء هذا الاختبار لدينا التمرين (14) التالي :

تمرين (14):

رغب أحد الباحثين دراسة أثر مشاهدة البرامج التلفزيونية على كمية المعلومات العامة لدى طلبة الجامعة فقام باختيار مجموعتين عشوائيتين من طلبة الجامعة، حيث أخضعت المجموعة الأولى (A) لمشاهدة برامج تلفزيونية لمدة شهر، والمجموعة الثانية (B) لم تخضع لمشاهدة أي برنامج تلفزيوني، وبعد الانتهاء طبق اختبار معلومات عامة على المجموعتين فكانت النتائج في الجدول الآتي :

رقم الطالب	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
المجموعة (A) (x)	60	62	47	54	60	48	64	50	54	49	-	-	-	-	-
المجموعة (B) (y)	65	62	59	62	67	51	56	68	69	45	66	64	55	60	70

المطلوب :

اختبر الفرض القائل: بعدم وجود تأثير للبرامج التلفزيونية على كمية المعلومات العامة لدى الطلبة، عند مستوى معنوية $\alpha=0.05$ ؟

الحل :

— صياغة الفرضية : $H_0 : \sum x = \sum y$

$H_1 : \sum x \neq \sum y$

— الاختبار الإحصائي :

$$Z = \frac{U - \frac{(n_1)(n_2)}{2}}{\sqrt{\frac{(n_1)(n_2)(n_1 + n_2 + 1)}{12}}}$$

رقم الطالب	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Σ
المجموعة (A) (x)	60	62	47	54	60	48	64	50	54	49	-	-	-	-	-	89.6
الفرق T_x	13	16	2	7.5	13	3	18.5	5	7.5	4	-	-	-	-	-	-
المجموعة (B) (y)	65	62	59	62	67	51	56	68	69	45	66	64	55	60	70	-
الفرق T_y	20	16	11	16	22	6	10	23	24	1	21	18.5	9	13	25	235.5

$$U = (n_1)(n_2) + \frac{n_1(n_1+1)}{2} - \sum T_x$$

$$\sum T_x = 89.5 \quad , \quad \sum T_y = 235.5$$

$$U = (10)(15) + \frac{10(15+1)}{2} - 89.5 = 115.5$$

$$Z = \frac{115.5 - \frac{(10)(15)}{2}}{\sqrt{\frac{(10)(15)(10+15+1)}{12}}} = 0.80$$

اتخاذ القرار :

من جداول (Z) نجد أن : $Z_{0.05} = 1.96$

من خلال المقارنة نجد أن : $Z = (0.80) < Z_{0.05} = (1.96)$ وبالتالي نقبل فرضية العدم (H_0) ونقول لا يوجد تأثير للبرامج التلفزيونية على كمية المعلومات للمجموعتين .

• اختبار ولوكسون Wilcoxon :

إن القصور الأساسي في اختبار الإشارة يتركز في تجاهله التام لمقدار الفرق في كل زوج من القيم، وقد توصل ولوكسون إلى اختبار يأخذ بعين الاعتبار مقدار الفرق في كل زوج من القيم (اختبار الإشارة والرتب) ، حيث يتطلب هذا الاختبار ترتيب القيم المطلقة للفروق ترتيباً تصاعدياً وتهمل أزواج القيم ذات الفروق الصفرية، مع العلم أن الفروق ترتب دون النظر لإشارتها، وبعد ذلك يتم حساب مجموع الرتب السالبة ومجموع الرتب الموجبة، إذ يطبق هذا الاختبار على هذين المجموعتين، ليبيان فيما إذا كان يوجد فرق بين مجموعتين مرتبطتين مرتبطين من القراءات .

هذا ويمكن استخدام اختبار ولكوكسون لاختبار عينتين زوجيتين لمعرفة فيما إذا كان هناك بينهما ارتباط أم لا (البيانات بصورة أزواج من القيم)، وهذا الاختبار مشابه لاختبار الإشارة ولكن يأخذ الرتب بعين الاعتبار، وبناءً عليه يمكن تقريب اختبار ولكوكسون بالتوزيع الطبيعي المعياري حيث يتم ذلك، عندما تكون العينات كبيرة [$(n, n) > 10$ أزواج] فإنه يمكن تقريب هذا الاختبار باستخدام التوزيع الطبيعي المعياري للحصول على القيمة المعيارية (Z) على النحو التالي:

$$Z = \frac{w - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

حيث أن : w عبارة عن مجموع الرتب للإشارة الأقل تكراراً
n : عدد الفروق غير الصفرية.

ونوضح ذلك من خلال التمرين التالي :

تمرين (15) :

رغب أحد الباحثين دراسة مدى تأثير استخدام طريقة جديدة لتدريس الإحصاء لطلاب السنة الأولى، فأخذ عينة من (10) أزواج من الطلبة المتكافئين، وبعد ذلك قام باستخدام الطريقة المعتادة لتدريس الإحصاء على المجموعة (A) والطريقة الجديدة على المجموعة (B)، وبعدها أجرى امتحاناً لهؤلاء الطلبة، فكانت النتائج كما في الجدول التالي :

رقم الزوج	1	2	3	4	5	6	7	8	9	10
الدرجة بالطريقة القديمة (y)	53	61	59	63	41	64	55	54	60	45
الدرجة بالطريقة الجديدة (x)	75	65	72	58	48	62	59	59	69	68

المطلوب : اختبار الفرض القائل: بعدم وجود فرق معنوي بين الطريقتين لتدريس الإحصاء باستخدام اختبار الإشارة والرتب (ولكوكسون) ، وذلك عند مستوى معنوية $\alpha=0.01$ ؟

الحل :

$H_0: \sum T(x-y) = \sum T(x-y)$: صياغة الفرضية :

$H_1: \sum T(x-y) \neq \sum T(x-y)$

== الاختبار الإحصائي :

$$Z = \frac{w - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

رقم الزوج	1	2	3	4	5	6	7	8	9	10
الدرجة بالطريقة القديمة (y)	53	61	59	63	41	64	55	54	60	45
الدرجة بالطريقة الجديدة (x)	75	65	72	58	48	62	59	59	69	68
(x-y)	+22	+4	+13	-5*	+7	-2*	+4	+20	+9	+23
الفرق بالقيمة المطلقة	22	4	13	5	7	2	4	20	9	23
رتب الفرق T(x-y)	9	2.5	7	4*	5	1*	2.5	8	6	10

الموجبة $\sum T(x-y) = +9 + 2.5 + 7 + 5 + 2 + 2.5 + 8 + 6 + 10 = 50$

السالبة $\sum T(x-y) = 4 + 1 = 5$

نختار الإشارة الأقل تكراراً نجد أنها السالبة ، فيكون :

$$Z = \frac{5 - \frac{10(11)}{4}}{\sqrt{\frac{10(11)(21)}{24}}} = -2.29$$

اتخاذ القرار : من جداول (Z) نجد أن : $Z_{\alpha} = Z_{0.01} = 2.58$

من خلال المقارنة نجد أن : $Z = 2.29 < Z_{0.01} = 2.58$

وبالتالي نقبل فرضية العدم (H_0) ونقول لا يوجد فرق معنوي بين الطريقتين .

5-9- الاختبارات الإحصائية اللاحتمالية بين عدة عينات (اختبار كروسكال - واليس)

• اختبار كروسكال - واليس Kruska - Wallis

عند إجراء اختبارات الفروض المتعلقة بالفروق بين متوسطات ثلاثة مجتمعات أو أكثر، نفترض أن المجتمعات الأصلية تتبع التوزيع الطبيعي، وأن تبايناتها مجهولة ومتساوية، فإذا تحققت هذه الافتراضات، فإنه يمكننا استخدام اختبار (F)، أما إذا كانت الفروق المشاهدة لمعرفة بين متوسطات العينات راجعة للصدفة، فإنها تشير إلى فروق معنوية بين المجتمعات المتناظرة، أما إذا لم تتحقق هذه الافتراضات، فيمكن استخدام طريقة للاحتمالية ملائمة تسمى اختبار مجموع الرتب (اختبار كروسكال - واليس) .

— حالة عينات مستقلة : يستخدم هذا لاختبار مدى تطابق توزيعات عدة مجتمعات ويرمز له بالرمز (H)، ويتطلب هذا الاختبار افتراض أن توزيعات المجتمعات مستمرة ولا يتطلب أي افتراض آخر، وهذا الاختبار يتبع توزيع (χ^2) . والاختبار الإحصائي يتم من خلال :

$$H = \frac{12}{N(N+1)} \sum_{k=1}^K \frac{R_k^2}{nk} - 3(N+1)$$

حيث أن :

K : عدد العينات (المجموعات) (k=1,2, ..., k)

R_k^2 : مجموع رتب كل عينة (مجموعة)

nk : حجم كل عينة (مجموعة)

N : عدد جميع المشاهدات في العينات .

ولتوضيح ذلك لدينا التمرين التالي :

تمرين (16) :

قامت إحدى شركات الإعلان بتصنيف العاملين فيها إلى ثلاث مجموعات وفقاً لمدة خبرتهم فيها، وقد رغبت إدارة الشركة في معرفة مدى اختلاف الدافع المعنوي بين هذه المجموعات الثلاث، فقامت باختيار عينة عشوائية من كل مجموعة مكونة من (10) موظفين، حيث قسمت دوافعهم المعنوية بمقياس ذي (40) نقطة، فحصلت الإدارة على البيانات الواردة في الجدول التالي:

المجموعة A	21	23	25	27	31	33	30	35	13	39
المجموعة B	20	22	26	28	30	32	19	19	36	38
المجموعة C	15	16	17	24	18	29	30	30	37	24

المطلوب : هل يمكننا استنتاج تساوي الدافع المعنوي بين هذه المجموعات الثلاث عند $(\alpha=0.01)$.

الحل :

— صياغة الفرضية :

$$H_0 : \sum R_1 = \sum R_2 = \sum R_3$$

$$H_1 : \sum R_1 \neq \sum R_2 \neq \sum R_3$$

— الاختبار

الإحصائي :

$$H = \frac{12}{N(N+1)} \sum_{k=1}^K \frac{R_k^2}{nk} - 3(N+1)$$

الرقم	1	2	3	4	5	6	7	8	9	10	Σ
المجموعة A	21	23	25	27	31	33	30	35	13	39	-
الرتب	9	11	14	16	22	24	20	26	1	30	173
المجموعة B	20	22	26	28	30	32	19	34	36	38	-
الرتب	8	10	15	17	20	23	7	25	27	29	181
المجموعة C	15	16	17	24	18	29	30	14	37	24	-
الرتب	3	4	5	12.5	6	18	20	2	28	12.5	111

$$R_3=111, \quad R_2=181, \quad R_1=173$$

$$H = \frac{12}{30(31)} \left(\frac{(173)^2}{10} + \frac{(181)^2}{10} + \frac{(111)^2}{10} \right) - 3(30+1) = 3.78$$

— اتخاذ القرار : من جداول χ^2 نجد أن :

$$\chi^2_{(0.01, 2)} = 9.210$$

$$V = k - 1 = 3 - 1 = 2$$

من خلال المقارنة نجد أن $\chi^2 = (3.78) < \chi^2_{(0.01, 2)} = (9.210)$

وبالتالي نقبل فرضية العدم (H_0) وهذا يعني تساوي الدافع المعنوي بين هذه المجموعات الثلاث .

ملاحظات :

1 — اختبار مجموع الرتب (كروسكال — واليس) هو دائماً اختبار من اتجاه واحد ومن اليمين ؛

2 — بشكل عام نكون أمام حالتين :

أ — تركز الرتب الكبيرة في مجموعة واحدة أو في عدد قليل جداً

من المجموعات، وبالتالي تكون قيمة H المحسوبة صغيرة

نسبياً، وهذا يدل على عدم اختلاف توزيعات المجموعات؛

ب - تركب الرتب الصغيرة في المجموعات الأخرى، فيكون عندها قيمة H كبيرة نسبياً، مما يعكس هذا اختلاف في توزيعات المجتمعات.

10-6- الاختبارات الإحصائية اللامعلمية للمقارنة بين عينتين للبيانات الاسمية والرتبية

• اختبار كولموجوروف - سميرنوف Kolmogorov - Smirnov

— حالة عينة واحدة : يستعان بهذا الأسلوب في حالة البيانات الاسمية للتحقق من صحة فرضية العدم (H_0) القائلة: بأن الفروق بين (Q و E) كانت نتيجة الحظ والصدف، ويعتمد هذا الاختبار على جداول (χ^2)، علماً بأنه أكثر دقة من اختبار χ^2 عند التعامل مع حجم عينة ($n \leq 30$) ويعطى الاختبار بالعلاقة التالية:

$$K.S = \left| \frac{f_Q}{n} - \frac{f_E}{n} \right|$$

حيث إن :

n : حجم العينة

f_Q : التكرار المتجمع الحقيقي

f_E : التكرار المتجمع المتوقع

E : التكرار الحقيقي

وستكون ($K.S$) عبارة عن أكبر فرق مطلق بين النسب المتجمعة الحقيقية والمتوقعة، والتمرين التالي يوضح لنا الاختبار:

تمرين (17) :

قام أحد الباحثين بدراسة لمعرفة رغبة أطفال إحدى رياض الأطفال في اختيارهم للعبة معينة، وبغية تحقيق ذلك وضعت اللعبة بألوان مختلفة، وكانت النتائج التي حصل عليها في الجدول التالي :

اللون	أبيض	أصفر	أحمر	أزرق	أخضر	Σ
العدد	3	9	16	1	1	30

والمطلوب :

اختبر الفرض القائل: بأن اختيار الطفل للعبة لا علاقة له باللون عند مستوى معنوية $(\alpha=0.05)$ ؟

الحل :

صياغة الفرضية : فرضية الاستقلال — ليس هناك علاقة بين اختيار اللعبة ولون اللعبة.

— الاختبار الإحصائي :

$$K.S = \left| \frac{f_o}{n} - \frac{f_e}{n} \right|$$

حساب التكرارات المتوقعة (E) : $30/5 = 6$

الجدول المساعد :

الألوان (الخلايا)	Q	ت.م — f _Q	ت.م.ن — Q	E	ت.م — E	ت.م.ن — E	$\left \frac{f_Q}{n} - \frac{f_E}{n} \right $
الأبيض	3	3	3/30	6	6	6/30	3/30
الأصفر	9	12	12/30	6	12	12/30	0
الأحمر	16	28	28/30	6	18	18/30	10/30
الأزرق	1	29	29/30	6	24	24/30	5/30
الأخضر	1	30	30/30	6	30	30/30	0

من خلال الجدول نجد أن أكبر فرق مطلق هو (10/30=0.33) أي أن:
(k.s = 0.33)

— اتخاذ القرار : من جداول (k.S) نجد أن : $K.S_{(0.05, 30)} = 0.24$ وبالتالي $K.s = 0.33 > K.s = 0.24$ من خلال المقارنة نجد أن نرفض فرضية الاستقلال وهناك فروق جوهرية (معنوية) في اختيار الأطفال للعب باختلاف اللون، بمعنى أن هناك علاقة بين اختيار الطفل للعبة ولونها.
— حالة عينتين مستقلتين : يستعان بهذا الأسلوب لاختبار الفروق بين عينتين، عندما تكون البيانات الخاصة لأحد المتغيرين اسمية والآخر رتيبة، ويمكن استخدامه في حالة البيانات الإسمية لكلا المتغيرين، ويتم الاختبار على النحو التالي :

$$K = \left[\frac{f_{G1}}{n_1} - \frac{f_{G2}}{n_2} \right] \sqrt{\frac{(n_1)(n_2)}{n_1 + n_2}}$$

$$K = F \sqrt{\frac{(n_1)(n_2)}{n_1 + n_2}}$$

حيث أن :

f_{G1} : التكرار التجميعي الصاعد للمجموعة الأولى

f_{G2} : التكرار التجميعي الصاعد للمجموعة الثانية

n_1, n_2 : حجم العينات

F : أكبر فرق مطلق بين التكرارين المتجمعين النسبيين .

نوضح بالتمرين التالي :

تمرين (18) :

بفرض أن كلية الإعلام بإحدى الجامعات تقبل الحاصلين على الثانوية العامة (علمي، أدبي)، وبهدف معرفة معنوية الفروق بين نتائج الطلبة حملة الثانوية العامة العلمية وحملة الثانوية العامة الأدبية في مقرر الإحصاء، كانت النتائج على النحو التالي :

التقدير التخصص	ضعيف جداً	ضعيف	مقبول	جيد	جيد جداً	ممتاز	Σ
أدبي	12	19	15	7	6	2	61
علمي	4	4	9	11	16	20	64

المطلوب :

اختبر الفرض القائل: بأنه ليس هناك فروقاً معنوية في تقديرات مقرر الإحصاء بين الطلاب (العلمي والأدبي) عند مستوى معنوية $(\alpha = 0.01)$ ؟
الحل :

— صياغة الفرضية : $H_1: G_1 \neq G_2$, $H_0 : G_1 = G_2$

— الاختبار الإحصائي :

$$K = \left[\frac{f_{G1}}{n_1} - \frac{f_{G2}}{n_2} \right] \sqrt{\frac{(n_1)(n_2)}{n_1 + n_2}}$$

الجدول المساعد :

الفرق المطلق (F)	ت.م.ن f_{G2}/n_2	ت.م. f_{G2}	العلمي f_{G2}	ت.م.ن f_{G1}/n_1	ت.م. f_{G1}	الأدبي Q_1	التقدير
0.13	4/64	4	4	12/61	12	12	ضعيف جداً
0.38	8/64	8	4	31/61	31	19	ضعيف
0.49	17/64	17	9	46/61	46	15	مقبول
0.43	28/64	28	11	53/61	53	7	جيد
0.28	44/64	44	16	59/61	59	6	جيد جداً
0	64/64	64	20	61/61	61	2	ممتاز

$$K = F \sqrt{\frac{(n_1)(n_2)}{n_1 + n_2}}$$

— اتخاذ القرار : من خلال الجدول نجد أن أكبر فرق مطلق هو (0.49)

أي أن : $F = 0.49$

وبالتالي نجد أن :

$$K = 0.49 \sqrt{\frac{(61)(64)}{61 + 64}} = 2.74$$

من جداول (k.S) نجد أن : $K_{0.01} = 1.22$ ، ومن اتجاهين :

من خلال المقارنة نجد أن : $K = 2.74 > K_{0.01} = 1.22$ ، المجموعة وبالتالي

نرفض فرضية العدم، ونقول: أن هناك فروقاً معنوية بين الطلاب العلمي

والأدبي في تقديراتهم في مقرر الإحصاء .

• اختبار ثنائي الحدين Binomiale :

حالة عينة واحدة (بيانات إسمية) : يستعان بهذا الاختبار عندما يكون لدينا عينة واحدة وقد اختيرت عشوائياً وتم الحصول منها على بيانات بصورة إسمية وثنائية التقسيم، مثل مجموعة جاءت إجاباتها على صورة (نعم، لا)، (موافق، غير موافق)، (حاضر، لم يحضر)، (صح، خطأ)، ويتم الاختبار من خلال العلاقة التالية :

$$K = \frac{n_1 - NP_1}{\sqrt{NP_1P_2}}$$

n_1 : عدد الإجابات للبديل الأول

N : حجم العينة الكلية

P_1 : احتمال ظهور إجابات البديل الأول = 0.5

P_2 : احتمال ظهور إجابات البديل الثاني = 0.5، وهنا نميز حالتين:

الحالة (أ): عندما يكون $n > 20$ (في هذه الحالة نستخدم جداول Z)، ونوضح ذلك بالتمرين التالي :

تمرين (19):

في دراسة على عينة من الأطفال ($n=30$) من عمر أربع سنوات، بهدف التفريق بين شكل المكعب وشكل الكرة، أظهرت التجربة أن (20) طفلاً استطاعوا التفريق بين الشكلين في حين فشل الباقي، المطلوب : اختبر الفرض القائل: بأن هؤلاء الأطفال تمكنوا من التمييز بين الشكلين عند مستوى معنوية $\alpha = 0.05$

الحل :

— صياغة الفرضية : $H_0 : P_1 = P_2 = 1/2$ ، $H_1 : P_1 \neq P_2 \neq 1/2$

— الاختبار الإحصائي :

$$K = \frac{n_1 - NP_1}{\sqrt{NP_1P_2}}$$

$$K = \frac{20 - (30)(0.5)}{\sqrt{(30)(0.5)(0.5)}} = 1.82$$

— اتخاذ القرار : من جداول (Z) نجد أن :

$$Z_{(0.05)} = 1.65 \text{ (اتجاه واحد)}$$

من خلال المقارنة نجد أن :

$$Z(1.82) > Z(1.65)$$

وبالتالي يمكننا القول أن هناك فرقاً معنوياً بين عدد من استطاع التمييز ومن لم يستطع، أي أن الأطفال في هذا العمر يمكنهم التفريق بين هذين الشكلين. الحالة (ب): عندما يكون $n < 20$ (في هذه الحالة نستخدم جداول ثنائي الحدين):

n_1 : عدد مفردات البديل الأول

N : حجم العينة

ولتوضيح كيفية إجراء الاختبار في هذه الحالة، نعطي التمرينين التاليين :

تمرين (20):

تم توجيه سؤال لعينة عشوائية حجمها ($n=16$) طالباً من طلاب كلية الإعلام في إحدى الجامعات حول المواعيد المفضلة لديهم لحضور محاضرات الإحصاء (8-10) صباحاً أم (11-1) ظهراً، فجاءت النتائج تشير إلى تفضيل (9) طلاب للفترة (8-10). المطلوب : اختبر الفرضية القائلة: بأن الطلاب لا يفضلون حضور محاضرات الإحصاء من (8-10) على الفترة (11-1) ظهراً عند مستوى $(\alpha=0.05)$ ؟

الحل :

$$n=9 \quad N=16$$
$$1/2 = \text{احتمال البديل} , \quad 2 = \text{عدد البدائل}$$

— صياغة الفرضية :

$$H_0: P_1 = P_2 = 1/2$$
$$H_1: P_1 \neq P_2 \neq 1/2$$

— الاختبار الإحصائي :

$$P = \alpha = 0.05$$

— إتخاذ القرار : من خلال جداول ثنائيي الحدين عندما : $n_1=9$ $N=16$
 $P=0.5$

نجد أن القيمة المقابلة (0.1746)

من خلال المقارنة نجد أن : $P=0.1746 > P=0.05$ وبالتالي نرفض
فرضية العدم (H_0)، ونقول أن الطلاب لا يفضلون وقتاً على الآخر لدراسة
الإحصاء في ضوء بيانات العينة.

تمرين (21) :

في دراسة بإحدى المدارس الخاصة عن لون الزي (الملبس) المرغوب
الارتداء كانت العينة ($n=19$) طالباً عرضت عليهم الألوان (بنّي، أبيض،
أزرق) وأشارت النتائج إلى تفضيل (11) طالباً للون الأبيض، والمطلوب: اختبر
الفرض القائل: بأن إجابات الطلاب تمثل إجابات المجتمع الذي تم اختيارهم منه،
وذلك عند مستوى معنوية ($\alpha=0.01$) ؟

الحل :

$$N = 19 \quad n_1 = 11 \quad 3 = \text{عدد البدائل}$$

$$0.33 = 1/3 \quad \text{احتمالي البديل}$$

— صياغة الفرضية :

$$H_0 : P_1 = P_2 = P_3 = 1/3$$

$$H_1 : R \neq P_2 \neq P_3 \neq 1/3$$

— الاختبار الإحصائي :

— اتخاذ القرار : من خلال جداول ثنائي الحدين عندما :

$$n_1=11, \quad N=19$$

احتمالات البدائل هي : 0.20 , 0.25 , 0.50

والاحتمال البديل الأقرب هو (0.25)، ومن خلال الجدول، نجد أن الاحتمال المقابل لها (0.0018) .

من خلال المقارنة نجد أن : $P(0.01) < P(0.0018)$ وبالتالي نرفض فرضية العدم (H_0) ونقبل (H_1)، أي البدائل غير متساوية، أي أن نسبة المفضلين للون الأبيض هم الأغلبية

• اختبار فيشر Fisher وحالة مجموعتين مستقلتين :

يستخدم هذا الاختبار للمقارنة بين عينتين مستقلتين (ذكور، إناث) مثلاً، و متغير اسمي ينقسم إنقساماً ثنائياً (ناجح، راسب)، بمعنى أن البيانات الخاصة بالمتغيرين المستقل (النوع) والتابع (التحصيل) إسمية ثنائية التقسيم، ويتم الاختبار من خلال العلاقة التالية :

$$F = \frac{[(a.d - b.c) \pm n/2]}{\sqrt{\frac{(a+b)(c+d)(a+c)(b+d)}{n-1}}}$$

مع العلم أن : $(+n/2)$ عندما يكون $(a.d - b.c > 0)$ و $(-n/2)$ عندما يكون $(a.d - b.c < 0)$

ولتوضيح هذا الاختبار، نعطي التمرين التالي :

تمرين (22) :

صممت تجربة على مجموعتين، استخدم في المجموعة الأولى نظام الحوافز ولم يستخدم في المجموعة الثانية، وذلك بهدف معرفة رأي العمال في تقديرهم لروؤسائهم نتيجة الحوافز، فكانت النتائج الآتية :

الرأي	تقدير	عدم تقدير	Σ
A	18	6	24
B	10	11	21
Σ	28	17	45

المطلوب :

اختبر الفرض القائل: بأن للحوافز أثراً على رأي العاملين أي تقدير رؤسائهم عند مستوى معنوية $(\alpha=0.05)$ ؟

— الاختبار الإحصائي :

$$d=11, \quad c=10, \quad b=6, \quad a=18$$

نعلم أن

$$a.d - b.c > 0 \quad \text{نطبق :}$$

$$(18)(11) - (6)(10) > 0 \quad \text{نجد أن :}$$

أي أن :

$$F = \frac{[(a.d - b.c) + n/2]}{\sqrt{\frac{(a+b)(c+d)(a+c)(b+d)}{n-1}}} = \frac{[(18)(11) - (6)(10) + 45/2]}{\sqrt{\frac{(28)(17)(24)(21)}{45-1}}} = 1.56$$

ـ اتخاذ القرار :

من جداول (Z)، نجد أن قيمة F (اتجاه واحد) عند مستوى دلالة $\alpha=0.05$ هي $1.06 \leq$

من خلال المقارنة نجد أن : $F_{0.05} = (1.65) < F = (1.56)$ وبالتالي لا يوجد فروق معنوية بين تقدير الرؤساء وعدم تقديرهم عند استخدام نظام الحوافز أو عدم استخدامه أي أن نظام الحوافز ليس له تأثير على الرأي في تقدير الرؤساء .

• اختبار الوسيط Median :

ـ حالة عينتين مستقلتين : يستخدم هذا الاختبار عندما يرغب الباحث التحقق من معنوية الفروق بين وسيطي عينتين، ويعتمد هذا الاختبار على إيجاد الوسيط للعينتين موضع الدراسة ثم يحدد عدد المفردات التي تقع قيمتهم فوق الوسيط، وكذلك يحدد عدد المفردات التي تقع قيمتهم مساوية أو أقل. وتصنف المفردات في جدول مزدوج (2×2) كما هو موضح بالشكل :

العينة	أعلى من الوسيط	أقل أو يساوي الوسيط	Σ
I	a	b	a+b
II	c	d	C+d
Σ	a+c	b+d	

ويتم الاختبار وفق العلاقة التالية :

$$\chi^2 = \frac{n[a.d - b.c]^2}{(a+b)(c+d)(a+c)(b+d)}$$

وبدرجة حرية واحدة .

نوضح كيفية إجراء هذا الاختبار التمرين التالي :

تمرين (23) :

رغب باحث إجراء بحث يتعلق بأوزان الطلاب، فقام باختيار عينة عشوائية من طلاب كلية الإعلام وعينة أخرى عشوائية من بقية طلاب الكليات الأخرى، وكانت البيانات على النحو التالي :

طلاب كلية الإعلام:

60	65	62	70	75	67	61	80	66	65	63	70	70	69	64	61
----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

طلاب الكليات الأخرى :

70	71	75	59	60	67	62	79	62	68	74	71	70	65	82
----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

المطلوب :

اختبر الفرض القائل: بعدم وجود فروق معنوية بين أوزان طلاب كلية الإعلام وطلاب بقية الكليات عند مستوى معنوية $(\alpha = 0.01)$ ؟

الحل :

— الطريقة الأولى : نقوم بدمج المجموعتين معاً فتصبح العينة الكلية (31) طالباً ثم نرتب هذه المفردات من الأصغر إلى الأكبر أو بالعكس، ثم يتم تحديد المفردة أو المفردتين التي يسبقها عدد من المفردات يساوي عدد المفردات التي تليها، وإذا كان هناك درجتين يجب جمعهما والقسمة على (2) .

— الطريقة الثانية : يتم استخدامها إذا كانت مسألة الترتيب تحتاج إلى وقت وذلك بإنشاء جدول تكراري .

الفئات	f	ت.ت.ص
58-62	8	8
63-67	8	16
68-72	9	25
73-77	3	28
78-82	3	31

$$\sum f/2 = \frac{31}{2} = 15.5$$

$$Med = L + \left(\frac{\sum f/2 - \sum fe}{f_{up}} \right) \cdot C$$

$$= 63 + \left(\frac{15.5 - 8}{8} \right) \cdot 5 = 67.7$$

نقوم بترتيب الجدول (2 × 2)

العينة	أكبر من الوسيط	الوسيط فأقل	Σ
كلية الإعلام	a 6	b 10	16
الكلية الأخرى	c 10	d 5	15
Σ	16	15	31

— الاختبار الإحصائي :

$$\chi^2 = \frac{n[a.d - b.c]^2}{(a+b)(c+d)(a+c)(b+d)}$$

$$\chi^2 = \frac{31[(6)(5) - (10)(10)]^2}{(16)(15)(16)(15)} = 2.64$$

— اتخاذ القرار : من جداول χ^2 نجد أن :

$$\chi^2(3.841)$$

$$\alpha = 0.01$$

$$V=1$$

من خلال المقارنة نجد أن : $\chi^2 = (3.841) > \chi^2 = (2.64)$ وبالتالي نقبل فرضية

$$\alpha=0.01$$

العدم (H_0)

$$V=1$$

ونقول لا توجد فروقاً معنوية بين أوزان طلاب كلية الإعلام وبقية

الكليات في الجامعة.

— حالة عدة عينات مستقلة : لقد بينا في السابق استخدام الوسيط للمقارنة بين عينتين مستقلتين، وبنفس الأسلوب يمكن إجراء المقارنة لعدة عينات مستقلة، وذلك وفق الخطوات الآتية:

- 1 — يتم دمج مفردات العينات موضع الدراسة وكأنها مفردات عينة واحدة.
- 2 — يستخرج وسيط مفردات هذه العينة الواحدة، إما بوضع المفردات من الأصغر إلى الأكبر أو بالعكس، ثم استخراج الوسيط على اعتبار أن تلك المفردة التي يسبقها عدد من المفردات يساوي عدد المفردات التي تليها، وذلك بعد الترتيب لهذه المفردات تصاعدياً أو تنازلياً، ومن ثم تطبيق القانون لحساب قيمة الوسيط .
- 3 — تصنيف مفردات كل عينة كما كانت قبل الدمج في ضوء قيمة الوسيط بحيث نضع عوضاً عن المفردة إشارة (+) إذا كانت قيمة هذه الدرجة أكبر من الوسيط ونضع عوضاً عن المفردة إشارة (-) إذا كانت قيمة هذه الدرجة أصغر أو تساوي قيمة الوسيط.

4 - يحدد عدد تكرارات الإشارة (الموجبة) وعدد التكرارات للإشارة (السالبة) في كل عينة من العينات موضوع المقارنة قبل الدمج وتصنف التكرارات في جدول كما هو موضح :

العينات	+	-	Σ
I	a	b	a+b
II	c	d	c+d
III	e	f	e+f
	a+c+e	b+d+f	N

ولبيان كيفية إجراء هذا الاختبار، نحل التمرين التالي :

تمرين (24):

قامت إحدى شركات الإعلان الطرقي بتصنيف العاملين فيها إلى ثلاث مجموعات وفقاً لمدة سنوات الخدمة فيها، وطبقت اختباراً معيناً على هذه المجموعات، بهدف التعرف على مدى الاختلاف بين هذه المجموعات، فكانت النتائج الآتية :

المجموعة A	7	9	15	4	3	2
المجموعة B	8	5	9	3	11	8
المجموعة C	6	9	14	8	13	12

المطلوب : اختبر الفرض القائل: بأنه ليس هناك فروقاً معنوية بين المجموعات الثلاث بالنسبة للاختبار ؟

الحل :

– الاختبار الإحصائي :

نقوم بدمج العينات الثلاث للحصول على الوسيط كالآتي :

رقم المجموعة	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
المفردات	2	3	3	4	5	6	7	8	8	8	9	9	9	11	12	13	14	15

ترتيب الوسيط

$$k = \frac{n+1}{2} = \frac{18+1}{2} = 9.5 \Rightarrow$$

$$\text{الوسيط } Med = \frac{8+8}{2} = 8$$

تصنيف البيانات يتم على النحو التالي :

العينة	+	-	Σ
	أكبر من الوسيط	أقل أو يساوي الوسيط	
A	a 2	b 4	6
B	c 2	d 4	6
C	e 4	f 2	6
Σ	8	10	18=N

حساب التكرارات المتوقعة (E) :

$$a = \frac{(6)(8)}{18} = 2.67 \quad b = \frac{(6)(10)}{18} = 3.33$$

وهكذا لبقية الخلايا

$$\chi^2 = \left(\sum \frac{(Q-E)^2}{E} \right)$$

$$\chi^2 = \frac{(2-2.67)^2}{2.67} + \frac{(4-3.33)^2}{3.33} + \dots + \frac{(2-3.33)^2}{3.33} = 1.79$$

اتخاذ القرار : من خلال جداول (χ^2) نجد أن :

$$\chi^2 (9.21)$$

$$\alpha = 0.01$$

$$V = (2-1)(3-1)=2$$

من خلال المقارنة نجد أن : $\chi^2 = (1.79) < \chi^2 = (9.21)$ وبالتالي نقبل

الفرضية القائلة بأنه لا يوجد فروق معنوية بين المجموعات بالنسبة للاختبار .

تمارين غير محلولة

تمرين (1) :

تهتم إدارة إحدى الكليات بمعرفة تقديرات النجاح لجميع المقررات ، وقد تم من خلال خبرة سابقة تصنيف هذه التقديرات لأحد المقررات على النحو التالي :

ممتاز	جيد جداً	جيد	مقبول	ضعيف	التقديرات
5%	10%	25%	30%	30%	النسبة

أخذت عينة عشوائية من (1000) طالب، فوجد أن : (250) طالب تقدير ضعيف، (300) طالب تقدير مقبول، (200) طالب تقدير جيد، (150) طالب تقدير جيد جداً، (100) طالب تقدير ممتاز .

والمطلوب : اختبار فيما إذا كان التوزيع الحالي للتقديرات مازال نفس التوزيع السابق الذي تم الحصول عليه من خبرة سابقة .

تمرين (2) :

رغب أحد الباحثين معرفة رغبات طلبة الجامعات فسي نسوع الكتب المفضلة لديهم، فاختار عينة عشوائية من (56) طالباً وطلب من كل واحد منهم أن يختار الكتب التي يفضلها أكثر من غيرها من بين أربعة أنواع منها هي: كتب تربوية، اجتماعية، اقتصادية وسياسية، وطلب منهم عدم اختيار أكثر من نوع واحد فوجد أن : (18) منهم يفضلون الكتب التربوية، (8) منهم يفضلون الكتب الاجتماعية، (19) منهم يفضلون الكتب الاقتصادية، و (11) منهم يفضلون الكتب السياسية

والمطلوب: هل رغبات هؤلاء الطلاب تمثل رغبات جميع الطلاب المتساوية في الرغبات لكل نوع من أنواع الكتب.

تمرين (3) :

يعتقد أن مجتمع الموظفين في القطاع العام في إحدى الدول العربية بحسب خبرة سابقة يتوزع حسب الفئات التالية :

الفئات	1	2	3	4	5
النسب المئوية %	10	15	25	30	20

أخذت عينة عشوائية تمثل هذا المجتمع حجمها (210) موظفاً فوجد أنها تتوزع على هذه الفئات على النحو التالي :

الفئات	1	2	3	4	5
عدد الموظفين	15	35	60	50	50

والمطلوب:

اختبر صحة هذا الاعتقاد عند مستوى معنوية $\alpha=0.05$

تمرين (4) :

وجه أحد الباحثين سؤال لعينة عشوائية من الطلاب حجمها (500) طالب حول إجراء الامتحانات بطريقة الأتمتة (الخيارات المتعددة)، وجاءت التكرارات لبدائل خمسة على النحو:

الآراء	مؤيد بشدة	مؤيد	محايد	معارض	معارض بشدة	Σ
التكرارات	200	60	100	120	20	500

فإذا كانت النسب المتوقعة طبقاً للدراسة هي :

الآراء	مؤيد بشدة	مؤيد	محايد	معارض	معارض بشدة	Σ
النسب المئوية %	50	12	18	17	3	100

والمطلوب : اختبر الفرض القائل: بأن هناك تشابه بين نسب إجابات طلاب العينة والنسب المتوقعة طبقاً للدراسات السابقة عند مستوى معنوية $\alpha = 0.05$ ؟

تمرين (5):

تتوزع فئات أحد المجتمعات على النحو التالي :

الفئات	A	B	C	D	E	Σ
النسبة %	15	30	12	33	10	100

أخذت عينة عشوائية من هذا المجتمع فجاء توزيع الفئات على النحو التالي :

الفئات	A	B	C	D	E	Σ
التكرارات	24	40	20	44	12	140

والمطلوب:

اختبر الفرض القائل: بأن هناك تشابه بين نسب إجابات أفراد العينة والنسب المتوقعة عند مستوى معنوية $\alpha=0.05$ ؟

تمرين (6) :

في عينة عشوائية من (100) عامل من عمال إحدى الشركات، وجد أن من بين (70) عامل متعلم (15) عامل مدخن ومن بين غير المتعلمين وجد (14) عامل لا يدخنون والمطلوب: هل يمكن إرجاء صفة التدخين أصلاً إلى صفة التعلم عند مستوى معنوية $\alpha=0.05$ ؟

تمرين (7):

البيانات التالية عبارة عن نتائج دراسة تتعلق بالتعليم المفتوح في إحدى

الجامعات:

الرأي النوع	مؤيد	معارض	محايد	Σ
ذكور	70	15	35	120
إناث	55	10	15	80
Σ	125	25	50	200

والمطلوب:

اختبر فيما إذا كان هناك علاقة بين الرأي والنوع عند مستوى معنوية $\alpha=0.05$

تمرين (8):

في عينة عشوائية من (100) طالب وطالبة كانت أعداد الناجحين

والراسبين في امتحان مادة الإحصاء، كما يلي:

النتيجة النوع	ناجح	راسب	Σ
طالب	40	15	55
طالبة	35	10	45
Σ	75	25	100

والمطلوب:

اختبر فيما إذا كان هناك علاقة بين النوع الطلبة ونتائجهم في الامتحان

عند مستوى معنوية $\alpha=0.05$ ؟

تمرين (9) :

سحبت عينة عشوائية مكونة من (330) من الناخبين في إحدى الدول لمعرفة رأيهم في قانون الأحزاب الجديد، حيث تم طبقاً لانتمائاتهم الحزبية، والجدول التالي يمثل النتائج التي حصلنا عليها :

الرأي الانتماء الحزبي	مؤيد	معارض	محايد	Σ
الحزب (A)	62	52	24	138
الحزب (B)	98	68	26	192
Σ	160	120	50	330

والمطلوب:

اختبر الفرض القائل: باستقلال كلاً من الانتماء الحزبي والرأي في قانون الأحزاب الجديد عند مستوى معنوية $\alpha=0.05$ ؟

تمرين (10):

في دراسة حول الزواج المدني في إحدى الدول، أخذت عينة عشوائية من طلبة الجامعات فكانت النتائج كالآتي:

الرأي النوع	مؤيد	معارض	محايد	Σ
ذكور	40	32	18	90
إناث	60	40	20	120
Σ	100	72	38	210

المطلوب:

اختبار فيما إذا كان هناك علاقة بين النوع والرأي عند مستوى معنوية $\alpha=0.05$.

تمرين (11):

يرغب البعض في معرفة فيما إذا لمرتبة عضو هيئة التدريس لها علاقة باستخدام المكتبة الجامعية، سحبت عينة عشوائية مكونة من (1000) عضو هيئة تدريس من إحدى الجامعات. حيث تم تصنيفهم حسب المرتبة العلمية إلى : أستاذ، أستاذ مساعد، مدرس وحسب استخدامهم للمكتبة الجامعية، فكانت النتائج في الجدول التالي :

المرتبة الاستخدام	أستاذ	أستاذ مساعد	مدرس	Σ
مستفيدون	90	150	160	400
غير مستفيدين	110	250	240	600
Σ	200	400	400	1000

والمطلوب:

اختبر الفرض القائل: بعدم وجود علاقة بين مرتبة عضو هيئة التدريس وبين استخدام المكتبة الجامعية عند مستوى معنوية $\alpha=0.05$ ؟

تمرين (12) :

بفرض أننا نرغب في معرفة ما إذا كان ذكاء الأشخاص ومعدل أوزانهم مستقلان عن بعضهما، سحبت عينة عشوائية مكونة من (90) طالب من طلاب إحدى الكليات، فحصلنا على التكرارات التالية :

معدل الوزن الذكاء	ضعيف	متوسط	ثقيل	Σ
ذكاء طبيعي	10	20	10	40
ذكاء ضعيف	25	10	15	50
Σ	35	30	25	90

المطلوب :

اختبر الفرض القائل باستقلال المتغيرين، عند مستوى معينة $\alpha = 0.01$ ؟

تمرين (13) :

يعتقد البعض في عدم استقلال إنتاجية العمال الأسبوعية قبل وبعد الإجازة، وللتحقق من ذلك، سحبت عينة عشوائية من (10) عمال وسجل إنتاجهم قبل وبعد الإجازة في الجدول التالي:

رقم العامل	1	2	3	4	5	6	7	8	9	10
قبل الإجازة	61	74	95	47	95	36	39	18	70	85
بعد الإجازة	69	77	93	48	99	38	33	20	73	89

والمطلوب :

اختبر الفرض القائل: بعدم تغير إنتاجية العمال في مقابل الفرض البديل بأن الإنتاجية قد زادت بعد الإجازة، عند مستوى معنوية $\alpha=0.01$ ؟

تمرين (14):

قام أحد الباحثين باختيار عينة عشوائية (15) طالباً من إحدى الكليات في الجامعة ووجه إليهم سؤالاً بشأن دور الإعلام في خدمة المجتمع وبعد إجراء التجربة وجد أن (8) من طلاب العينة كانت إجاباتهم صحيحة. والمطلوب: هل إجابات الطلبة تمثل إجابات المجتمع الذي تم اختيارهم منه أم لا، عند مستوى معنوية $\alpha=0.01$.

تمرين (15) :

اختبر (10) خبراء للتذوق للمفاضلة بين الشراب (A) والشراب (B) فكان تفضيلهم كما يلي:

رقم الخبير	1	2	3	4	5	6	7	8	9	10
التفضيل	B	A	A	B	B	B	A	B	A	B

المطلوب:

اختبر الفرض القائل بأن: الشراب (B) أفضل من الشراب (A) عند مستوى معنوية $\alpha=0.05$

تمرين (16):

الجدول التالي يمثل الدرجات التي حصل عليها خمسة طُلاب في الاختبار النهائي لكل من الحاسوب، الإحصاء، الرياضيات:

الحاسوب	98	86	78	84	87
الإحصاء	90	91	95	97	93
الرياضيات	99	89	80	88	96

المطلوب:

اختبر الفرض القائل بتساوي مستوى الطلبة في الموضوعات الثلاث عند مستوى معنوية $\alpha=0.05$.

تمرين (17) :

لدراسة أثر التحصيل العلمي في الاتجاه نحو عمل المرأة قسام أحد الباحثين باختيار ثلاث عينات، العينة (A) من حملة الشهادة الإعدادية وحجمها (5) والعينة الثانية (B) من حملة الشهادة الثانوية وحجمها (4)، والعينة الثالثة (C) من حملة الشهادة الجامعية وحجمها (4).
طبق مقياس الاتجاه نحو عمل المرأة وهو عبارة عن قائمة تحتوي (50) سؤالاً يتطلب الإجابة عليها بـ نعم أو لا، وبعد تطبيق هذا المقياس دلت النتائج، كما في الجدول التالي :

الرقم	1	2	3	4	5
العينة (A)	40	35	32	23	20
العينة (B)	42	25	24	18	-
العينة (C)	48	46	41	-	

المطلوب:

اختبر الفرض القائل بتساوي: مستوى الإجابات في مقياس الاتجاه نحو عمل المرأة، عند مستوى معنوية $\alpha=0.05$ ؟



الفصل العاشر

تحليل التباين

Analysis of Variance

– مقدمة

– تحليل التباين الأحادي التصنيف

– تمارين غير محلولة



تحليل التباين Analysis of Variance

1-10 مقدمة:

إن تحليل التباين، عبارة عن أسلوب إحصائي يأخذ بعين الاعتبار جميع العينات لإجراء اختبارات الفروض المتعلقة بالفروق بين متوسطات ثلاث مجتمعات أو أكثر، ويفترض أن تكون المجتمعات الأصلية تتبع التوزيع الطبيعي، وإن تبايناتها مجهولة ومتساوية، وعند تحقق هذه الشروط يمكن استخدام اختبار (F) لمعرفة معنوية الفروق المشاهدة بين متوسطات العينات العشوائية المأخوذة من هذه المجتمعات.

وبشكل عام، يتضمن التحليل الإحصائي تشتت متوسطات العينات المأخوذة من المجتمعات المطلوب دراستها، ويسمى هذا التحليل بتحليل التباين أحادي التصنيف باعتباره يتضمن عاملاً واحداً فقط، وهناك تحليل التباين ثنائي التصنيف لأنه يتضمن عاملين، ونحن سوف نتناول التباين الأحادي التصنيف فقط في دراستنا.

10-2- تحليل التباين أحادي التصنيف (One – way analysis of varince):

إن الهدف من تحليل التباين أحادي التصنيف معرفة ما إذا كانت الفروق المشاهدة بين متوسطات ثلاث عينات أو أكثر تعكس فروقاً حقيقية بين متوسطات المجتمعات التي سحبت منها هذه العينات أم أنها فروقاً تعود للصدفة والحظ فقط، فمثلاً قد يرغب باحث ما معرفة ما إذا كان هناك فرقاً معنوياً في مقارنة كفاءة طرق عديدة لتعليم الإحصاء لطلاب الكليات في الجامعة أو رغبة أحد الباحثين في معرفة ما إذا كان هناك فرقاً معنوياً في معدل استهلاك الوقود لأنواع مختلفة من السيارات. وعليه يتم تحليل التباين الأحادي التصنيف وفق المراحل التالية :

1 - صياغة الفرضية :

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

$$H_1 : \mu_1 \neq \mu_2 \neq \mu_3$$

2 - الاختبار الإحصائي :

$$F = \frac{MSB}{MSE}$$

حيث إن :

MSB هي متوسط التباين بين المجموعات

MSE هي متوسط التباين داخل المجموعات.

ونحن نعلم بأن التباين داخل المجموعات + التباين بين المجموعات = التباين الكلي

حيث إن :

$$SST = SSB + SSE$$

$$SSB = n_1(x_1 - \bar{X})^2 + n_2(x_2 - \bar{X})^2 + n_3(x_3 - \bar{X})^2$$

حيث أن : (x_i) متوسط العينة.

$(\bar{X})$ المتوسط العام .

$$\bar{X} = \frac{\sum \bar{x}}{K} \quad (\text{في حالة أحجام العينات متساوية}) \quad (\text{المتوسط العام})$$

$$\bar{X} = \frac{n_1x_1 + n_2x_2 + n_3x_3}{n_1 + n_2 + n_3} \quad (\text{في حالة أحجام العينات غير متساوية}) \quad (\text{المتوسط العام})$$

حيث أن : (K) عدد العينات.

$$SSE = S_1^2(n_1 - 1) + S_2^2(n_2 - 1) + S_3^2(n_3 - 1) \quad (\text{التباين داخل المجموعات}).$$

$$MSE = \frac{SSE}{N - 1} \quad (\text{متوسط التباين داخل المجموعات})$$

$$MSB = \frac{SSE}{K - 1} \quad (\text{متوسط التباين بين المجموعات}).$$

N : الحجم الكلي للعينات.

ولبيان كيفية تطبيق تحليل التباين نعرض التمرين التالي :

تمرين (1) :

بفرض استخدام ثلاث مجموعات من الطلبة لتعليمهم الإحصاء بثلاث طرائق تعليم مختلفة وذلك لمعرفة ما إذا كان هناك فرقاً معنوياً بين نتائج هذه الطرائق. أعطي اختبار موحد في الإحصاء للمجموعات الثلاث، فحصلنا على النتائج التالية:

الطريقة A	5	5	6	7	7
الطريقة B	1	2	5	5	7
الطريقة C	3	5	5	6	6

المطلوب:

بفرض أن درجات الاختبار تتبع التوزيع الطبيعي، وأن العينات مستقلة، استخدم ($\alpha=0.05$) لاختبار فرضية العدم القائلة بأن المتوسطات في المجموعات الثلاث متساوية.

الحل — صياغة الفرضية:

1 — صياغة الفرضية :

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

$$H_1 : \mu_1 \neq \mu_2 \neq \mu_3$$

2 - الاختبار الإحصائي :

$$F = \frac{MSB}{MSE}$$

الجدول المساعد :

المجموعة المقياس	A	B	C
n	5	5	5
$\sum x$	30	20	25
$\bar{X}$	6	4	5
S^2	1.2	2.18	1.2

$$SSB = n_1(x_1 - \bar{X})^2 + n_2(\bar{x}_2 - \bar{X})^2 + n_3(\bar{x}_3 - \bar{X})^2$$

$$\bar{X} = \frac{x_1 + x_2 + x_3}{k} = \frac{6 + 4 + 5}{3}$$

$$SSB = 5(1)^2 + 5(-1)^2 + 5(0)^2 = 10$$

$$MSB = \frac{SSB}{K-1} = \frac{10}{3-1} = 5$$

$$SSE = S_1^2(n_1 - 1) + S_2^2(n_2 - 1) + S_3^2(n_3 - 1)$$

$$= 1.2(4) + 2.18(4) + 1.2(4) = 18.32$$

$$= \frac{18.32}{15-3} = 1.53$$

$$MSB = \frac{SSB}{N-K}$$

$$F = \frac{MSB}{MSE} = \frac{5}{1.53} = 3.27$$

— اتخاذ القرار: من جداول (F) نجد أن :

$$F(3.885)$$

$$\alpha = 0.05$$

$$V = K - 1 = 2$$

$$V = N - K = 12$$

من خلال المقارنة نجد أن $F = 3.27 < F_{(0.05, 2, 12)}$ وبالتالي نقبل فرضية العدم ونقول أن متوسطات الطرق الثلاث متساوية، وعليه يمكننا وضع نتائج تحليل التباين (ANOVA) في الجدول الآتي:

النسبة F	التباين	درجة الحرية	مجموع المربعات	مصدر التباين
$\frac{5}{1.53} = 3.27$	MSB=5	2	SSB=10	بين المجموعات
	MSE=1.53	12	SSE=18.32	داخل المجموعات
		14	SST=28.32	الكل

تمرين :

أعطي اختبار موحد لثلاث مجموعات من المتقدمين لشغل ثلاث وظائف، بحيث تتقدم كل مجموعة منها لشغل وظيفة معينة، فيما يلي الدرجات التي حصل عليها المتقدمون في كل مجموعة من هذه المجموعات.

A	5	4	3	2	1
B	6	4	2	1	-
C	3	2	1	-	-

المطلوب:

بفرض أن الدرجات تتبع التوزيع الطبيعي، وأن العينات مستقلة، اختبر
الفرض القائل بتساوي متوسطات الدرجات عند مستوى معنوية ($\alpha=0.01$)

1 - صياغة الفرضية :

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

$$H_1 : \mu_1 \neq \mu_2 \neq \mu_3$$

2 - الاختبار الإحصائي :

$$F = \frac{MSB}{MSE}$$

الجدول المساعد :

المجموعة المقياس	A	B	C
n	5	4	3
$\sum x$	15	13	6
$\bar{X}$	3	3.25	2
S^2	2	3.69	0.66

$$SSB = n_1(x_1 - \bar{X})^2 + n_2(x_2 - \bar{X})^2 + n_3(x_3 - \bar{X})^2$$

$$\bar{X} = \frac{n_1x_1 + n_2x_2 + n_3x_3}{n_1 + n_2 + n_3}$$

العينات غير متساوية

$$\bar{X} = \frac{(5)(3) + (4)(3.25) + (3)(2)}{5 + 4 + 3} = 2.83$$

$$\begin{aligned} SSB &= 5(3-2.83)^2 + 4(3.25-2.83)^2 + 3(2-2.83)^2 \\ &= (0.14) + (0.70) + (2.06) = 2.90 \end{aligned}$$

$$MSB = \frac{SSB}{K-1} = \frac{2.90}{3-1} = 1.45$$

$$\begin{aligned} SSE &= S_1^2(n_1 - 1) + S_2^2(n_2 - 1) + S_3^2(n_3 - 1) \\ &= 2(5-1) + 3.69(4-1) + 0.66(3-1) = 20.39 \end{aligned}$$

$$MSE = \frac{SSE}{N-K} = \frac{20.39}{12-3} = 2.26$$

$$F = \frac{MSB}{MSE} = \frac{1.45}{2.26} = 0.64$$

— اتخاذ القرار: من جداول (F) نجد أن : $F(8.022)$

$$\alpha = 0.01$$

$$V = K-1 = 2$$

$$V = N - K = 12 - 3$$

من خلال المقارنة نجد أن : $F = 0.64 < F_{(0.05, 2, 9)}$ وبالتالي نقبل
فرضية العدم (H_0) ونقول ليس هناك فرقاً معنوياً بين متوسطات الدرجات لشغل
الوظائف، وعليه يكون جدول تحليل التباين (ANOVA) الآتي :

النسبة F	التباين	درجة الحرية	مجموع المربعات	مصدر التباين
$\frac{1.45}{2.26} = 0.64$	MSB=1.45	2	SSB=2.90	بين المجموعات
	MSE=2.26	9	SSE=27.2	داخل المجموعات
		11	SST=29.92	الكل

تمارين غير محلولة

تمرين (1) :

تم تخصيص ثلاثة مخرجين لإنجاز عمل تلفزيوني معين حيث سجل إنتاجهم اليومي لمدة (5) أيام، يبين الجدول التالي الإنتاجية اليومية لهؤلاء المخرجين.

اليوم	1	2	3	4	5
المخرج (A)	44	46	55	55	60
المخرج (B)	50	56	60	64	65
المخرج (C)	46	48	48	52	61

المطلوب:

بفرض أن الإنتاج اليومي يتبع التوزيع الطبيعي وأن العينات مستقلة
اختبر الفرض القائل: بتساوي متوسطات إنتاجية المخرجين الثلاثة عند مستوى معنوية $\alpha=0.01$ ؟

تمرين (2) :

البيانات التالية تمثل عدد حوادث السيارات اليومية التي حدثت على أربعة طرق سريعة لها نفس الطول وذلك خلال (5) أيام :

اليوم	1	2	3	4	5
A	3	5	5	6	6
B	1	2	3	4	5
C	5	5	6	7	7
D	1	2	5	5	7

المطلوب:

بفرض أن أعداد حوادث السيارات اليومية تتبع التوزيع الطبيعي وأن العينات مستقلة، اختبر الفرض القائل: بتساوي متوسطات أعداد الحوادث اليومية على هذه الطرق الأربع وذلك عند مستوى معنوية $(\alpha=0.05)$ ؟

تمرين :

بفرض وجود ثلاثة أحياء يقطن في كل منها عدد متساوٍ تقريباً من السكان.

الجدول التالي يمثل عدد الذين تم القبض عليهم شهرياً لارتكاب جرائم معينة في كل حي من هذه الأحياء وذلك خلال فترة (6) أشهر:

الشهر	1	2	3	4	5	6
X	3	4	4	5	5	7
Y	2	3	4	5	-	-
Z	4	5	6	-	-	-

المطلوب:

بفرض أن أعداد المقبوض عليهم في الأحياء الثلاث تتبع التوزيع الطبيعي وأن العينات مستقلة، اختبر الفرض القائل: بتساوي متوسطات أعداد المقبوض عليهم عند مستوى معنوية $(\alpha=0.05)$ ؟

تمرين (3) :

استخدمت ثلاثة أنواع من برامج التدريب المختلفة (A,B,C) في ثلاث مجموعات في العمال المتمثلين في سنوات الخدمة والكفاءة، فحصلنا على جدول يبين الدرجات التي حصل عليها هؤلاء العمال في كل مجموعة من هذه المجموعات كما يلي:

المجموعة (A)	78	84	86	88	94	96	92	83	75
المجموعة (B)	64	68	73	76	78	78	80	60	93
المجموعة (C)	50	73	77	79	79	79	60	63	90

المطلوب:

افرض أن الدرجات تتبع الطبيعي وأن المجموعات مستقلة، اختبر
الفرض القائل: بتساوي متوسطات الدرجات، عند مستوى معنوية $(\alpha=0.01)$ ؟



الملاحق

الرموز والأحرف الأبجدية



الرموز والأحرف الأبجدية المستخدمة في هذا الكتاب:

الأحرف الأبجدية: أولاً :

الأبجدية العربية	الأحرف الأبجدية اللاتينية			الأحرف الأبجدية اليونانية		
	حرف صلب	حرف كبير	اللفظ	حرف صغير	حرف كبير	اللفظ
أ	a	A	ا	α	A	ألفا
ب	b	B	بي	β	B	بيتا
ج	c	C	سي	γ	Γ	قاما
د	d	D	دي	δ	Δ	ديلتا
هـ	e	E	إي	ϵ		إيسيلون
و	f	F	إف	ζ	Z	زيتا
ز	g	G	جي	η	H	إيتا
ح	h	H	اتش	θ	θ	تيتا
ط	i	I	أي	i	+	يوتا
ي	j	J	جي	k	K	كبا
ك	k	K	كي	λ	Λ	لمدا
ل	l	L	ال	μ	M	ميو
م	m	M	ام	v	N	نيو
ن	n	N	ان	ξ	Ξ	كسي
س	o	O	او	o	O	أوميكرون
ع	p	P	بي	π	Π	بي
ف	q	Q	كيو	ρ	P	رو

ص	r	R	أر	σ	Σ	سيكما
ق	s	S	اس	τ	T	تاو
ر	t	T	تي	ν	Y	يوسيلون
ش	u	U	يو	φ	ϑ	فاي
ت	v	V	في	χ	X	كاي
ث	w	W	دبل يو	ψ	Ψ	بسمي
خ	x	X	إكس	ω	Ω	أوميغا
ذ	y	Y	واي	$\wp$	ς	اس
ض، ظ، غ	z	Z	زد			

ثانياً : الأرقام المستخدمة:

الأرقام العربية	0	1	2	3	4	5	6	7	8	9	10
الأرقام الهندية	0	1	2	3	4	5	6	7	8	9	10
الأرقام الرومانية		I	II	III	IV	V	VI	VII	VIII	IX	X

ثالثاً : الرموز والإشارات المستخدمة:

العملية	الرمز	العملية	الرمز	العملية	الرمز
ضرب	* أو ×	أكبر أو يساوي	$\leq$	اللوغاريتم العشري	$\log(x)$
تقسيم	/ أو ÷	أكبر من	$<$	اللوغاريتم الطبيعي	$\ln(x)$
الجمع	+	أصغر أو يساوي	$\geq$	التوقع الرياضي	$E(x)$
الطرح	-	أصغر من	$>$	التباين	σ^2
يؤدي إلى	$\Rightarrow$	يساوي	$=$	الانحراف المعياري	σ
واو الملف	&	للتقريب	#	الانحراف المعياري للعبء	S
المضارب		يتطابق	$\equiv$	العلمي	!
المجاميع	Σ	يساوي تقريباً	$\approx$	المحدد	Δ
مجموع المربعات	ΣX^2	لا نهاية	∞	النسبة المئوية	%
الاجتماع	U	النقاطع	$\cap$	الوسط الحسابي للمجتمع	μ
التوافيق	C_n^k	الجذر	$\sqrt{X}$	الوسط الحسابي للعينة	$\bar{X}$
حجم المجتمع	N	حجم العينة	n	الاحتمال التراكمي	$F(Xs)$



قائمة المراجع بالعربية

- أبو صالح، محمد صبحي؛ عوض، عدنان محمد ، 1990 — مقدمة في الإحصاء — مركز الكتب الأردني.
- أبو عمة، عبد الرحمن، راضي، الحسني عبد البر، الهادي، محمود إبراهيم، 1995، الإحصاء والاحتمالات، الرياض، جامعة الملك سعود.
- الجاعولي، فريد؛ إسماعيل، حبيب؛ غانم، عدنان، 1999، مبادئ الإحصاء، جامعة صنعاء.
- أوسليغان، جورج، بانكروفت، جوردن، 1983، الرياضيات والإحصاء لدراسات المحاسبة والأعمال، دار ماكجرو هيل للنشر، ترجمة مقدسي، جمال سامي، ومراجعة السيد محمد الغزي.
- درويش، رمضان، الإحصاء الوصفي، جامعة دمشق، محاضرات أُلقيت على طلبة كلية التربية، للعام الدراسي (2007, 2008).
- درويش، رمضان، مبادئ الإحصاء، جامعة دمشق، محاضرات أُلقيت على طلبة كلية الآداب، قسم الإعلامي للعام الدراسي (2007, 2008).
- حميدان، عدنان؛ الجاعولي، فريد؛ ناصر آغا، عمار؛ العواد، منذر، 2003، مبادئ الإحصاء، منشورات جامعة دمشق.
- طيوب، محمود؛ غانم، عدنان، 2002، مبادئ الإحصاء، جامعة صنعاء.
- طيوب، محمود، الإحصاء اللامعلمي، (جامعة صنعاء، محاضرات أُلقيت على طلبة كلية التجارة، للعام الدراسي (2001, 2000).
- مخول، مطانيوس، غانم، عدنان، 2006 مبادئ الإحصاء، كلية الآداب والعلوم الإنسانية، قسم المكتبات، جامعة دمشق.

المراجع باللغة الروسية

- 1 - جرسيموفيتش، أ. ن، الإحصاء الرياضي، الطبعة الثانية، موسكو 1983.
- 2 - ايفتشينكو، ج، 1، مينفيديف، يو، ان، نظرية الاحتمالات والإحصاء الرياضي، موسكو، 1984.
- 3 - بوروفكوف، أ، أ، الإحصاء الرياضي، تقدير المعالم واختيار الفرضيات، موسكو، 1988.
- 4 - كينينكا، ب، ف، نظرية الاحتمالات، موسكو، 1994.

المراجع الأجنبية

- 1- Kolmogrove A.N, Theory Of Probability and Mathematical Statistics. Moscow "Seince", 1967.
- 2- Mitropolskyi. A.K., Statistical Calculations techises. Moscow, "Seince", 1978.
- 3- Druzhynin n.k , Sampling and Experiment. Gernal Logical Principles of Organization. Moscow "Statistica" 1977.
- 4- Fisher R.A., Statistical Methods for an analyst. Moscow 1958.
- 5- Cohen J. , Statistical Power Analysis. Hillsdale, New gersey, 1988.
- 6- Collyer c. and Enns J., analysis of Variance: The Basic Designs. Chicogo, 1976.
- 7- Kirk R. Experimental Deizgn: Procedures for the Behavioral Sciences. California, 1982.

مواقع الانترنت

- 1- <http://www.stutsft.com/textbook/stathome.html>
- 2- <http://www.bmj.com/collections/statsbk/11.shtml>
- 3- <http://davidmlance.com/hyperstat/intro.html>
- 4- <http://www.psychstat.missouristate.edu/introbook/sbko?.htm>
- 5- <http://www.penhall.com/groebner>





الجداول الإحصائية



ملحق الجداول الإحصائية

Appendix Tables

Table 1 Binomial Coefficients معاملات التوزيع الثنائي

$$C_x^n = \binom{n}{x} = \frac{n!}{x!(n-x)!}, x = 0, 1, 2, \dots, 13$$

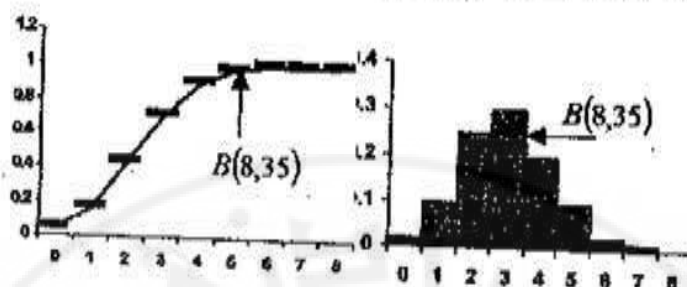
n	0	1	2	3	4	5	6	7	8
0	1								
1	1	1							
2	1	2	1						
3	1	3	3	1					
4	1	4	6	4	1				
5	1	5	10	10	5	1			
6	1	6	15	20	15	6	1		
7	1	7	21	35	35	21	7	1	
8	1	8	28	56	70	56	28	8	1
9	1	9	36	84	126	126	84	36	9
10	1	10	45	120	210	252	210	120	45
11	1	11	55	165	330	462	462	330	165
12	1	12	66	220	495	792	924	792	495
13	1	13	78	286	715	1287	1716	1716	1287
14	1	14	91	364	1001	2002	3003	3432	3003
15	1	15	105	455	1365	3003	5005	6435	6435
16	1	16	120	560	1820	4368	8008	11440	12870
17	1	17	136	680	2380	6188	12376	19448	24310
18	1	18	153	816	3060	8568	18564	31824	43758
19	1	19	171	969	3876	11628	27132	50388	75582
20	1	20	190	1140	4845	15504	38760	77520	125970
21	1	21	210	1330	5985	20349	54264	116280	203490
22	1	22	231	1540	7315	26334	74613	170544	319770
23	1	23	253	1771	8855	33649	100947	245157	490314
24	1	24	276	2024	10626	42504	134596	346104	735471
25	1	25	300	2300	12650	53130	177100	480700	1081575
26	1	26	325	2600	14950	65780	230230	657800	1562275

تابع الجدول (I) ..

$$C_x^n = \binom{n}{x} = \frac{n!}{x!(n-x)!}, x = 0, 1, 2, \dots, 13$$

n	9	10	11	12	13
0					
1					
2					
3					
4					
5					
6					
7					
8					
9	1				
10	10	1			
11	55	11	1		
12	220	66	12	1	
13	715	286	78	13	1
14	2002	1001	364	91	14
15	5005	3003	1365	455	105
16	11440	8008	4368	1820	560
17	24310	19448	12376	6188	2380
18	48620	43758	31824	18564	8568
19	92378	92378	75582	50388	27132
20	167960	184756	167960	125970	77520
21	293930	352716	352716	293930	203490
22	497420	646646	705432	646646	497420
23	817190	1144066	1352078	1352078	1144066
24	1307504	1961256	2496144	2704156	2496144
25	2042975	3268760	4457400	5200300	5200300
26	3124550	5311735	7726160	9657700	10400600

Table II The Binomial Distribution جدول التوزيع الثنائي

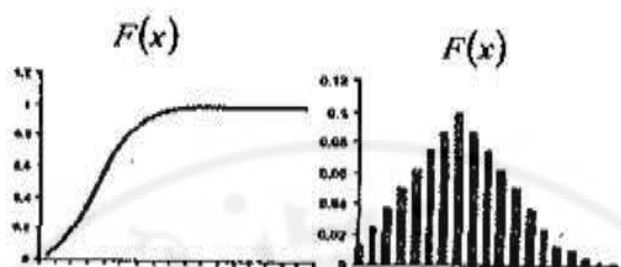


$$F(x) = P(X \leq x) = \sum_{k=0}^x \binom{n}{k} p^k (1-p)^{n-k}$$

		p								
n	x	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45
2	0	0.9025	0.8100	0.7225	0.6400	0.5625	0.4900	0.4225	0.3600	0.3025
	1	0.9975	0.9900	0.9775	0.9600	0.9375	0.9100	0.8775	0.8400	0.7975
	2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
3	0	0.8574	0.7290	0.6141	0.5120	0.4219	0.3430	0.2746	0.2160	0.1664
	1	0.9928	0.9720	0.9393	0.8960	0.8438	0.7840	0.7183	0.6480	0.5748
	2	0.9999	0.9990	0.9966	0.9920	0.9844	0.9730	0.9571	0.9360	0.9089
	3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4	0	0.8145	0.6561	0.5220	0.4096	0.3164	0.2401	0.1785	0.1296	0.0915
	1	0.9860	0.9477	0.8905	0.8192	0.7383	0.6517	0.5630	0.4752	0.3910
	2	0.9995	0.9963	0.9880	0.9728	0.9492	0.9163	0.8735	0.8208	0.7585
	3	1.0000	0.9999	0.9995	0.9984	0.9961	0.9919	0.9850	0.9744	0.9590
	4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
5	0	0.7738	0.5905	0.4437	0.3277	0.2373	0.1681	0.1160	0.0778	0.0503
	1	0.9774	0.9185	0.8352	0.7373	0.6328	0.5282	0.4284	0.3370	0.2562
	2	0.9988	0.9914	0.9734	0.9421	0.8965	0.8369	0.7648	0.6826	0.5931
	3	1.0000	0.9995	0.9978	0.9933	0.9844	0.9692	0.9460	0.9130	0.8688
	4	1.0000	1.0000	0.9999	0.9997	0.9990	0.9976	0.9947	0.9898	0.9815
	5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
6	0	0.7351	0.5314	0.3771	0.2621	0.1780	0.1176	0.0754	0.0467	0.0277
	1	0.9672	0.8857	0.7765	0.6554	0.5339	0.4202	0.3191	0.2333	0.1636
	2	0.9978	0.9842	0.9527	0.9011	0.8306	0.7443	0.6471	0.5443	0.4415
	3	0.9999	0.9987	0.9941	0.9830	0.9624	0.9295	0.8826	0.8208	0.7447
	4	1.0000	0.9999	0.9996	0.9984	0.9954	0.9891	0.9777	0.9590	0.9308
	5	1.0000	1.0000	1.0000	0.9999	0.9998	0.9993	0.9982	0.9959	0.9917

6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n	x								
7	0	0.6983	0.4783	0.3206	0.2097	0.1335	0.0824	0.0490	0.0280
	1	0.9556	0.8503	0.7166	0.5767	0.4449	0.3294	0.2338	0.1586
	2	0.9962	0.9743	0.9262	0.8520	0.7564	0.6471	0.5323	0.4199
	3	0.9998	0.9973	0.9879	0.9667	0.9294	0.8740	0.8002	0.7102
	4	1.0000	0.9998	0.9988	0.9953	0.9871	0.9712	0.9444	0.9037
	5	1.0000	1.0000	0.9999	0.9996	0.9987	0.9962	0.9910	0.9812
	6	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9994	0.9984
	7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n	x								
8	0	0.6634	0.4305	0.2725	0.1678	0.1001	0.0576	0.0319	0.0168
	1	0.9428	0.8131	0.6572	0.5033	0.3671	0.2553	0.1691	0.1064
	2	0.9942	0.9619	0.8948	0.7969	0.6785	0.5518	0.4278	0.3154
	3	0.9996	0.9950	0.9786	0.9437	0.8862	0.8059	0.7064	0.5941
	4	1.0000	0.9996	0.9971	0.9896	0.9727	0.9420	0.8939	0.8263
	5	1.0000	1.0000	0.9998	0.9988	0.9958	0.9887	0.9747	0.9502
	6	1.0000	1.0000	1.0000	0.9999	0.9996	0.9987	0.9964	0.9915
	7	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9993
	8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n	x								
9	0	0.6302	0.3874	0.2316	0.1342	0.0751	0.0404	0.0207	0.0101
	1	0.9288	0.7748	0.5995	0.4362	0.3003	0.1960	0.1211	0.0705
	2	0.9916	0.9470	0.8591	0.7382	0.6007	0.4628	0.3373	0.2318
	3	0.9994	0.9917	0.9661	0.9144	0.8343	0.7297	0.6089	0.4826
	4	1.0000	0.9991	0.9944	0.9804	0.9511	0.9012	0.8283	0.7334
	5	1.0000	0.9999	0.9994	0.9969	0.9900	0.9747	0.9464	0.9006
	6	1.0000	1.0000	1.0000	0.9997	0.9987	0.9957	0.9888	0.9750
	7	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996	0.9986	0.9962
	8	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9997
	9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n	x								
10	0	0.5987	0.3487	0.1969	0.1074	0.0563	0.0282	0.0135	0.0060
	1	0.9139	0.7361	0.5443	0.3758	0.2440	0.1493	0.0860	0.0464
	2	0.9885	0.9298	0.8202	0.6778	0.5256	0.3828	0.2616	0.1673
	3	0.9990	0.9872	0.9500	0.8791	0.7759	0.6496	0.5138	0.3823
	4	0.9999	0.9984	0.9901	0.9672	0.9219	0.8497	0.7515	0.6331
	5	1.0000	0.9999	0.9986	0.9936	0.9803	0.9527	0.9051	0.8338
	6	1.0000	1.0000	0.9999	0.9991	0.9965	0.9894	0.9740	0.9452
	7	1.0000	1.0000	1.0000	0.9999	0.9996	0.9984	0.9952	0.9877
	8	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9983
	9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997

Table III The Poisson Distribution جدول توزيع بواسون



$$F(x) = P(X \leq x) = \sum_{k=0}^x \frac{\lambda^k e^{-\lambda}}{k!}$$

$\lambda = E(X)$										
x	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
0	0.905	0.819	0.741	0.670	0.607	0.549	0.497	0.449	0.407	0.368
1	0.995	0.982	0.963	0.938	0.910	0.878	0.844	0.809	0.772	0.736
2	1.000	0.999	0.996	0.992	0.986	0.977	0.966	0.953	0.937	0.920
3	1.000	1.000	1.000	0.999	0.998	0.997	0.994	0.991	0.987	0.981
4	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999	0.998	0.996
5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999
6	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

$\lambda = E(X)$										
x	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2
0	0.333	0.301	0.273	0.247	0.223	0.202	0.183	0.165	0.150	0.135
1	0.699	0.663	0.627	0.592	0.558	0.525	0.493	0.463	0.434	0.406
2	0.900	0.879	0.857	0.833	0.809	0.783	0.757	0.731	0.704	0.677
3	0.974	0.966	0.957	0.946	0.934	0.921	0.907	0.891	0.875	0.857
4	0.995	0.992	0.989	0.986	0.981	0.976	0.970	0.964	0.956	0.947
5	0.999	0.998	0.998	0.997	0.996	0.994	0.992	0.990	0.987	0.983
6	1.000	1.000	1.000	0.999	0.999	0.999	0.998	0.997	0.997	0.995
7	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999	0.999
8	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

$\lambda = E(X)$										
x	2.2	2.4	2.6	2.8	3.0	3.2	3.4	3.6	3.8	4.0
0	0.111	0.091	0.074	0.061	0.050	0.041	0.033	0.027	0.022	0.018
1	0.355	0.308	0.267	0.231	0.199	0.171	0.147	0.126	0.107	0.092
2	0.623	0.570	0.518	0.469	0.423	0.380	0.340	0.303	0.269	0.238

3	0.819	0.779	0.736	0.692	0.647	0.603	0.558	0.513	0.473	0.433
4	0.928	0.904	0.877	0.848	0.815	0.781	0.744	0.706	0.668	0.629
5	0.975	0.964	0.951	0.935	0.916	0.895	0.871	0.844	0.816	0.785
6	0.993	0.988	0.983	0.976	0.966	0.955	0.942	0.927	0.909	0.889
7	0.998	0.997	0.995	0.992	0.988	0.983	0.977	0.969	0.960	0.949
8	1.000	0.999	0.999	0.998	0.996	0.994	0.992	0.988	0.984	0.979
9	1.000	1.000	1.000	0.999	0.999	0.998	0.997	0.996	0.994	0.992
10	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999	0.998	0.997
11	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999
12	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

$$\lambda = E(X)$$

x	4.2	4.4	4.6	4.8	5.0	5.2	5.4	5.6	5.8	6.0
0	0.015	0.012	0.010	0.008	0.007	0.006	0.005	0.004	0.003	0.002
1	0.078	0.066	0.056	0.048	0.040	0.034	0.029	0.024	0.021	0.019
2	0.210	0.185	0.163	0.143	0.125	0.109	0.095	0.082	0.072	0.238
3	0.395	0.359	0.326	0.294	0.265	0.238	0.213	0.191	0.170	0.433
4	0.590	0.551	0.513	0.476	0.440	0.406	0.373	0.342	0.313	0.629
5	0.753	0.720	0.686	0.651	0.616	0.581	0.546	0.512	0.478	0.785
6	0.867	0.844	0.818	0.791	0.762	0.732	0.702	0.670	0.638	0.889
7	0.936	0.921	0.905	0.887	0.867	0.845	0.822	0.797	0.771	0.949
8	0.972	0.964	0.955	0.944	0.932	0.918	0.903	0.886	0.867	0.979
9	0.989	0.985	0.980	0.975	0.968	0.960	0.951	0.941	0.929	0.992
10	0.996	0.994	0.992	0.990	0.986	0.982	0.977	0.972	0.965	0.997
11	0.999	0.998	0.997	0.996	0.995	0.993	0.990	0.988	0.984	0.999
12	1.000	0.999	0.999	0.999	0.998	0.997	0.996	0.995	0.993	1.000
13	1.000	1.000	1.000	1.000	0.999	0.999	0.999	0.998	0.997	1.000
14	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999	1.000
15	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
16	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

$$\lambda = E(X)$$

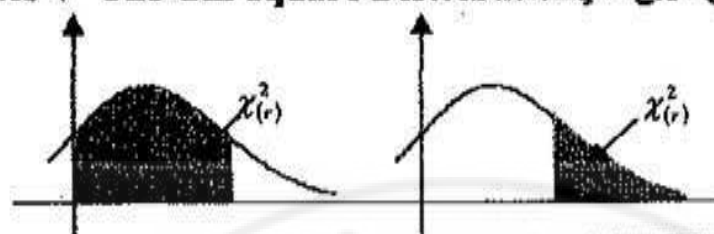
x	6.5	7	7.5	8	8.5	9	9.5	10	10.5	11.0
0	0.002	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1	0.011	0.007	0.005	0.003	0.002	0.001	0.001	0.000	0.000	0.000
2	0.043	0.030	0.020	0.014	0.009	0.006	0.004	0.003	0.002	0.001
3	0.112	0.082	0.059	0.042	0.030	0.021	0.015	0.010	0.007	0.005
4	0.224	0.173	0.132	0.100	0.074	0.055	0.040	0.029	0.021	0.015
5	0.369	0.301	0.241	0.191	0.150	0.116	0.089	0.067	0.050	0.038
6	0.527	0.450	0.378	0.313	0.256	0.207	0.165	0.130	0.102	0.079
7	0.673	0.599	0.525	0.453	0.386	0.324	0.269	0.220	0.179	0.143
8	0.792	0.729	0.662	0.593	0.523	0.456	0.392	0.333	0.279	0.232
9	0.877	0.830	0.776	0.717	0.653	0.587	0.522	0.458	0.397	0.341
10	0.933	0.901	0.862	0.816	0.763	0.706	0.645	0.583	0.521	0.460
11	0.966	0.947	0.921	0.888	0.849	0.803	0.752	0.697	0.639	0.579
12	0.984	0.973	0.957	0.936	0.909	0.876	0.836	0.792	0.742	0.689
13	0.993	0.987	0.978	0.966	0.949	0.926	0.898	0.864	0.825	0.781
14	0.997	0.994	0.990	0.983	0.973	0.959	0.940	0.917	0.888	0.854

15	0.999	0.998	0.995	0.992	0.986	0.978	0.967	0.951	0.932	0.907
16	1.000	0.999	0.998	0.996	0.993	0.989	0.982	0.973	0.960	0.944
17	1.000	1.000	0.999	0.998	0.997	0.995	0.991	0.986	0.978	0.968
18	1.000	1.000	1.000	0.999	0.999	0.998	0.996	0.993	0.988	0.982
19	1.000	1.000	1.000	1.000	0.999	0.999	0.998	0.997	0.994	0.991
20	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.998	0.997	0.995
21	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999	0.998
22	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999
23	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

$$\lambda = E(X)$$

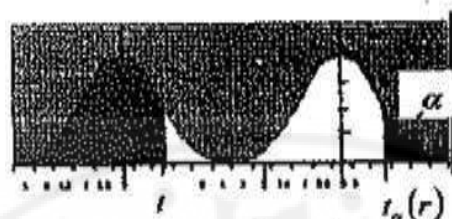
x	11.5	12	12.5	13	13.5	14	14.5	15	15.5	16
0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.003	0.002	0.002	0.001	0.001	0.000	0.000	0.000	0.000	0.000
4	0.011	0.008	0.005	0.004	0.003	0.002	0.001	0.001	0.001	0.000
5	0.028	0.020	0.015	0.011	0.008	0.006	0.004	0.003	0.002	0.001
6	0.060	0.046	0.035	0.026	0.019	0.014	0.010	0.008	0.006	0.004
7	0.114	0.090	0.070	0.054	0.041	0.032	0.024	0.018	0.013	0.010
8	0.191	0.155	0.125	0.100	0.079	0.062	0.048	0.037	0.029	0.022
9	0.289	0.242	0.201	0.166	0.135	0.109	0.088	0.070	0.055	0.043
10	0.402	0.347	0.297	0.252	0.211	0.176	0.145	0.118	0.096	0.077
11	0.520	0.462	0.406	0.353	0.304	0.260	0.220	0.185	0.154	0.127
12	0.633	0.576	0.519	0.463	0.409	0.358	0.311	0.268	0.228	0.193
13	0.733	0.682	0.628	0.573	0.518	0.464	0.413	0.363	0.317	0.275
14	0.815	0.772	0.725	0.675	0.623	0.570	0.518	0.466	0.415	0.368
15	0.878	0.844	0.806	0.764	0.718	0.669	0.619	0.568	0.517	0.467
16	0.924	0.899	0.869	0.835	0.798	0.756	0.711	0.664	0.615	0.566
17	0.954	0.937	0.916	0.890	0.861	0.827	0.790	0.749	0.705	0.659
18	0.974	0.963	0.948	0.930	0.908	0.883	0.853	0.819	0.782	0.742
19	0.986	0.979	0.969	0.957	0.942	0.923	0.901	0.875	0.846	0.812
20	0.992	0.988	0.983	0.975	0.965	0.952	0.936	0.917	0.894	0.868
21	0.996	0.994	0.991	0.986	0.980	0.971	0.960	0.947	0.930	0.911
22	0.998	0.997	0.995	0.992	0.989	0.983	0.976	0.967	0.956	0.942
23	0.999	0.999	0.998	0.996	0.994	0.991	0.986	0.981	0.973	0.963
24	1.000	0.999	0.999	0.998	0.997	0.995	0.992	0.989	0.984	0.978
25	1.000	1.000	0.999	0.999	0.998	0.997	0.996	0.994	0.991	0.987
26	1.000	1.000	1.000	1.000	0.999	0.999	0.998	0.997	0.995	0.993
27	1.000	1.000	1.000	1.000	1.000	0.999	0.999	0.998	0.997	0.996
28	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999	0.999	0.998
29	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999
30	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999
31	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
32	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
33	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
34	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table V The Chi-square Distribution توزیع مربع کای



	0.01	0.025	0.05	0.100	0.900	0.950	0.975	0.990	0.995
r	$\chi^2_{0.99(r)}$	$\chi^2_{0.975(r)}$	$\chi^2_{0.95(r)}$	$\chi^2_{0.90(r)}$	$\chi^2_{0.10(r)}$	$\chi^2_{0.05(r)}$	$\chi^2_{0.025(r)}$	$\chi^2_{0.01(r)}$	$\chi^2_{0.005(r)}$
1	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.554	0.831	1.145	1.610	9.236	11.070	12.832	15.086	16.750
6	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.647	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	4.107	5.009	5.892	7.041	19.812	22.362	24.736	27.688	29.819
14	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.558
25	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	12.878	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.994
29	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.335
30	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
60	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321

Table VI The t Distribution توزیع استیودنت

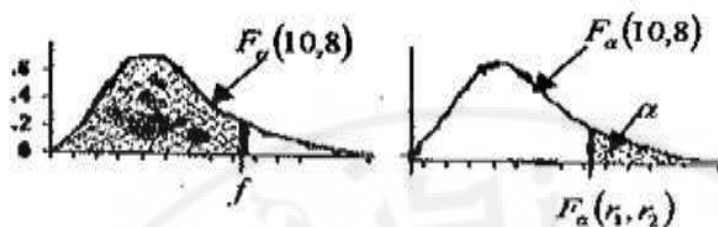


$$P(T \leq t) = \int_{-\infty}^t \frac{\Gamma[(r+1)/2]}{\sqrt{r\pi} \Gamma(r/2) (1+w^2/r)^{(r+1)/2}} dw$$

$$P[(T \leq -t)] = 1 - P(T \leq t)$$

	0.60	0.75	0.90	0.95	0.975	0.99	0.995
<i>r</i>	<i>t</i> _{0.40} (<i>r</i>)	<i>t</i> _{0.25} (<i>r</i>)	<i>t</i> _{0.10} (<i>r</i>)	<i>t</i> _{0.05} (<i>r</i>)	<i>t</i> _{0.025} (<i>r</i>)	<i>t</i> _{0.01} (<i>r</i>)	<i>t</i> _{0.005} (<i>r</i>)
1	0.325	1.000	3.078	6.314	12.700	31.821	63.656
2	0.289	0.816	1.886	2.920	4.303	6.965	9.825
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841
4	0.271	0.741	1.533	2.132	2.778	3.747	4.604
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707
7	0.263	0.711	1.415	1.885	2.365	2.998	3.499
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169
11	0.260	0.697	1.363	1.798	2.201	2.718	3.108
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807
24	0.256	0.685	1.318	1.711	2.064	2.492	2.797
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787
26	0.256	0.684	1.315	1.706	2.056	2.479	2.779
27	0.256	0.684	1.314	1.703	2.052	2.473	2.771
29	0.256	0.683	1.311	1.699	2.045	2.462	2.758
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750
∞	0.253	0.674	1.282	1.645	1.960	2.326	2.576

Table VII The F Distribution توزيع فيشر



$$P(F \leq f) = \int_0^f \frac{\Gamma((r_1 + r_2)/2) (r_1/r_2)^{r_1/2} w^{r_1/2-1}}{\Gamma(r_1/2) \Gamma(r_2/2) (1 + r_1 w/2)^{(r_1+r_2)/2}} dw$$

$\alpha\%$	r_2	Numerator Degrees of Freedom r_1 درجات الحرية للمسط								
		1	2	3	4	5	6	7	8	9
0.05	1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5
0.025	1	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.6	963.3
0.01	1	4052.2	4999.3	5403.5	5624.3	5764.0	5859.0	5928.3	5981.0	6022.4
0.05	2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
0.025	2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39
0.01	2	98.50	99.00	99.16	99.25	99.30	99.33	99.36	99.38	99.39
0.05	3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
0.025	3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47
0.01	3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.34
0.05	4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
0.025	4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90
0.01	4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66
0.05	5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
0.025	5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68
0.01	5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16
0.05	6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
0.025	6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52
0.01	6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98
0.05	7	5.39	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
0.025	7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82
0.01	7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72
0.05	8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
0.025	8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36
0.01	8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91
0.05	9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
0.025	9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03
0.01	9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35
0.05	10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
0.025	10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78
0.01	10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94
0.05	12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
0.025	12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44
0.01	12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39
0.05		4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59

0.025	15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12
0.01		8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89
0.05	20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
0.025		5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84
0.01		8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46
0.05	24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
0.025		5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70
0.01		7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26
0.05	30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
0.025		5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57
0.01		7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07
0.05	40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
0.025		5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45
0.01		7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89
0.05	60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
0.025		5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33
0.01		7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72
0.05	120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96
0.025		5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22
0.01		6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56
0.05	∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88
0.025		5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11
0.01		6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41

Table VII The F Distribution (Fisher) توزيع فيشر (يتبع)

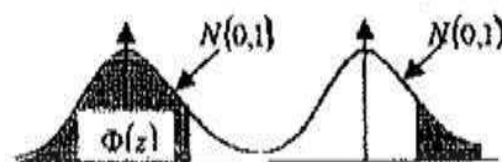
α	r_2	Numerator Degrees of Freedom r_1 درجات الحرية البسط							
		12	15	20	24	30	40	60	120
0.05	1	243.90	245.95	248.02	249.1	250.10	251.14	252.20	253.25
0.025		976.72	984.87	993.08	997.27	1001.4	1003.6	1009.8	1014.0
0.01		6107	6157	6209	6234	6260.4	6286.4	6313	6340
0.05	2	19.41	19.43	19.45	19.5	19.46	19.47	19.48	19.49
0.025		39.41	39.43	39.45	39.46	39.46	39.47	39.48	39.49
0.01		99.42	99.43	99.45	99.46	99.47	99.48	99.48	99.49
0.05	3	8.74	8.70	8.66	8.6	8.62	8.59	8.57	8.55
0.025		14.34	14.25	14.17	14.12	14.08	14.04	13.99	13.95
0.01		27.03	26.87	26.69	26.60	26.50	26.41	26.32	26.22
0.05	4	5.91	5.86	5.80	5.8	5.75	5.72	5.69	5.66
0.025		8.75	8.66	8.56	8.51	8.46	8.41	8.36	8.31
0.01		14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56

Table VIII

جدول الأرقام العشوائية في الفترة (0,1)

9588	0198	3820	0894	6705	2742	4569
4900	7466	7496	2909	4978	3722	3079
7521	2725	3902	4063	0949	8802	7023
5843	8325	0528	9800	2068	9029	9665
3949	5233	7119	8639	8131	2475	7942
8715	3299	6692	9803	4528	3506	4945
8401	7484	8384	0224	1242	3068	3077
9207	0149	8058	6870	8529	7588	6798
7736	7574	8588	1319	1872	8386	2689
6520	7841	7373	6333	0333	7685	6023
4165	7946	3533	1031	3257	7900	7313
4838	1276	1246	7801	6225	0457	4350
7947	9142	2410	4292	9920	7496	3669
0253	3400	6426	1920	4342	1146	5190
8348	7762	5936	2109	1036	0439	5598
6720	7203	2655	2512	6737	4896	0476
0628	4188	5390	4364	9341	5278	5372
3661	2808	6778	9171	3486	0070	8395
9784	8036	3352	4960	3833	1634	2902
6119	7492	6927	0885	5729	5241	3936
7482	6518	9499	2966	7333	3457	5776
5938	9499	8154	6110	2970	2117	8560
4725	7083	3201	7851	0122	0435	5401
4298	3395	5982	9762	3821	6014	4061
8745	3018	6243	7117	6795	7009	4831
2084	5778	1475	9227	6156	7372	7152
2578	7162	3083	1147	2882	2377	9183
0045	7556	2820	1923	6061	6569	1805
8138	7994	0808	5852	3319	8838	7975
4187	5360	9691	0736	2104	5497	3998
9239	6149	2790	8068	0064	5110	2736
6640	0233	9391	4707	6740	8089	4086
5314	1778	0038	4353	3054	3854	7375
1982	4704	4771	5948	5567	8559	6118
6178	7492	1538	8617	8481	2830	0787
2817	6883	8082	0589	6475	0674	8508
3270	8611	9218	7240	6695	9635	9953
2826	4639	7201	6681	2385	4960	2165
1734	1450	5790	0361	7877	7867	2580
4870	1693	8312	6607	5553	3707	1932
8207	3300	7158	1835	8846	8869	1195
8248	1060	7053	8547	9854	4010	3488
7596	8039	2547	3006	3395	4834	9408
4959	4082	2318	5127	578	4345	2857
3329	7550	9481	2988	4642	3842	4167

Table 9 IX Iard Normal Distribution التوزيع الطبيعي المعياري



$$P(Z \leq z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-w^2/2} dw, \quad [\Phi(-z) = 1 - \Phi(z)]$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714
0.2	0.5793	0.5832	0.5871	0.5911	0.5948	0.5987	0.6026	0.6064	0.6103
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823
0.8	0.7881	0.7910	0.7939	0.7968	0.7995	0.8023	0.8051	0.8078	0.8106
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810
1.2	0.8849	0.8869	0.8888	0.8906	0.8925	0.8944	0.8962	0.8980	0.8997
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162
1.4	0.9192	0.9207	0.9222	0.9237	0.9251	0.9265	0.9279	0.9292	0.9306
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913
2.4	0.9918	0.9920	0.9922	0.9924	0.9927	0.9929	0.9931	0.9932	0.9934
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973
2.8	0.9974	0.9975	0.9976	0.9977	0.9978	0.9979	0.9980	0.9981	0.9982
2.9	0.9984	0.9985	0.9986	0.9987	0.9988	0.9989	0.9990	0.9991	0.9992
3.0	0.9993	0.9994	0.9995	0.9996	0.9997	0.9998	0.9999	0.9999	0.9999
α	0.400	0.300	0.200	0.100	0.050	0.025	0.010	0.005	0.001
z _α	0.253	0.524	0.842	1.282	1.645	1.960	2.326	2.576	3.090
z _{α/2}	0.842	1.036	1.282	1.645	1.960	2.240	2.576	2.807	3.291

جدول رة X (10)

جدول إختبار الإشارة

n	p	n	p	n	p	n	p
1 0 .50000	1 10 0 .00098	10 15 0 .00003	15 18 0 .00000	18			
2 0 .25000	2 1 .01074	9 1 .00049	14 1 .00007	17			
1 .75000	1 2 .05469	8 2 .00369	13 2 .00066	16			
3 0 .12500	3 3 .17188	7 3 .01758	12 3 .00377	15			
1 .50000	2 4 .37695	6 4 .05923	11 4 .01544	14			
4 0 .06250	4 5 .62305	5 5 .15088	10 5 .04813	13			
1 .31250	3 11 0 .00049	11 6 .30362	9 6 .11894	12			
2 .68750	2 1 .00586	10 7 .50000	8 7 .24034	11			
5 0 .03125	5 2 .03271	9 16 0 .00002	16 8 .40726	10			
1 .18750	4 3 .11328	8 1 .00026	15 9 .59274	9			
2 .50000	3 4 .27441	7 2 .00209	14 19 0 .00000	19			
6 0 .01563	6 5 .50000	6 3 .01064	13 1 .00004	18			
1 .10938	5 12 0 .00024	12 4 .03841	12 2 .00036	17			
2 .34375	4 1 .00317	11 5 .10506	11 3 .00221	16			
3 .65625	3 2 .01929	10 6 .22725	10 4 .00961	15			
7 0 .00781	7 3 .07300	9 7 .40181	9 5 .03178	14			
1 .06250	6 4 .19385	8 8 .59819	8 6 .08353	13			
2 .22656	5 5 .38721	7 17 0 .00001	17 7 .17964	12			
3 .50000	4 6 .61279	6 1 .00014	16 8 .32380	11			
8 0 .00391	8 13 0 .00012	13 2 .00117	15 9 .50000	10			
1 .03516	7 1 .00171	12 3 .00636	14 20 0 .00000	20			
2 .14453	6 2 .01123	11 4 .02452	13 1 .00002	19			
3 .36328	5 3 .04614	10 5 .07173	12 2 .00020	18			
4 .63672	4 4 .13342	9 6 .16615	11 3 .00129	17			
9 0 .00195	9 5 .29053	8 7 .31453	10 4 .00591	16			
1 .01953	8 6 .50000	7 8 .50000	9 5 .02069	15			
2 .08984	7 14 0 .00006	14 6 .05766	14 6 .05766	14			
3 .25391	6 1 .00092	13 7 .13159	13 7 .13159	13			
4 .50000	5 2 .00647	12 8 .25172	12 8 .25172	12			
	3 .02869	11 9 .41190	11 9 .41190	11			
	4 .08978	10 10 .58810	10 10 .58810	10			
		9 5 .21198	9 5 .21198	9			
		8 6 .39526	8 6 .39526	8			
		7 7 .60474	7 7 .60474	7			

جدول رقم (11)

جدول ويلكسون (مان-ويتني) Wilcoxon (Mann-Whitney)

$$n = 1(1)5, m = 1(1)10$$

m	n=1	m	n=1	m	n=2	m	n=2
1	1 0.50000	2	10 1 0.09091	11	7 3 0.02778	17	10 10 0.30303
			2 0.18182	10	4 0.05556	16	11 0.37879
2	1 0.33333	3	3 0.27273	9	5 0.11111	15	12 0.45455
		2 0.66667	2	4 0.36364	8	6 0.16667	14
			5 0.45455	7	7 0.25000	13	m
3	1 0.25000	4	6 0.54545	6	8 0.33333	12	n=3
	2 0.50000	3			9 0.44444	11	3 6 0.05
		m	n=2		10 0.55556	10	7 0.10
4	1 0.20000	5	2 3 0.16667	7			8 0.20
	2 0.40000	4	4 0.33333	6	8 3 0.02222	19	9 0.35
		3 0.60000	3	5 0.66667	5	4 0.04444	12
					5 0.08889	17	10 0.50
					4 6 0.02857	18	
5	1 0.16667	6	3 3 0.10000	9	6 0.13333	16	4 6 0.02857
	2 0.33333	5	4 0.20000	8	7 0.20000	15	7 0.05714
	3 0.50000	4	5 0.40000	7	8 0.26667	14	8 0.11429
			6 0.60000	6	9 0.35556	13	9 0.20000
					10 0.44444	12	10 0.31429
6	1 0.14286	7			11 0.55556	11	11 0.42857
	2 0.28571	6	4 3 0.06667	11			12 0.77143
	3 0.42857	5	4 0.13333	10			13 0.68571
		4 0.57143	4	5 0.26667	9	9 3 0.01818	21
			6 0.40000	8	4 0.03636	20	5 6 0.01786
7	1 0.12500	8	7 0.60000	7	5 0.07273	19	7 0.03571
	2 0.25000	7			6 0.10909	18	8 0.07143
	3 0.37500	6	5 3 0.04762	13	7 0.16364	17	9 0.12500
	4 0.50000	5	4 0.09524	12	8 0.21818	16	10 0.19643
			5 0.19048	11	9 0.29091	15	11 0.28571
8	1 0.11111	9	6 0.28571	10	10 0.36364	14	12 0.39286
	2 0.22222	8	7 0.42857	9	11 0.45455	13	13 0.50000
		3 0.33333	7	8 0.57143	8	12 0.54545	12
		4 0.44444	6				6 6 0.01190
		5 0.55555	5	6 3 0.03571	15	10 3 0.01515	23
			4 0.07143	14	4 0.03030	21	7 0.02381
9	1 0.10000	10	5 0.14286	13	5 0.06061	20	8 0.04762
	2 0.20000	9	6 0.21429	12	6 0.09091	19	9 0.08333
	3 0.30000	8	7 0.32143	11	7 0.13636	18	10 0.13095
	4 0.40000	7	8 0.42857	10	8 0.18182	17	11 0.19048
	5 0.50000	6	9 0.57143	9	9 0.24242	16	12 0.27381
							13 0.35714

جدول رقم (12)

جدول ویلیکسون (مان-ویتنی) (Wilcoxon (Mann-Whitney)

$n = 1 (1) 5, m = 1 (1) 10$

m	n=3	m	n=3	m	n=4	m	n=4
6	14 0.45238 16	9	15 0.24091 24	5	15 0.14286 25	8	12 0.00808 40
	15 0.54762 15		16 0.30000 23		16 0.20635 24		13 0.01414 39
			17 0.36364 22		17 0.27778 23		14 0.02424 38
7	6 0.00833 27		18 0.43182 21		18 0.36508 22		15 0.03636 37
	7 0.01667 26		19 0.50000 20		19 0.45238 21		16 0.05455 36
	8 0.03333 25				20 0.54762 20		17 0.07677 35
	9 0.05833 24	10	6 0.00350 36				18 0.10707 34
	10 0.09167 23		7 0.00699 35	6	10 0.00476 34		19 0.14141 33
	11 0.13333 22		8 0.01399 34		11 0.00952 33		20 0.18384 32
	12 0.19167 21		9 0.02448 33		12 0.01905 32		21 0.23030 31
	13 0.25833 20		10 0.03846 32		13 0.03333 31		22 0.28485 30
	14 0.33333 19		11 0.05594 31		14 0.05714 30		23 0.34141 29
	13 0.41667 18		12 0.08042 30		15 0.08571 29		24 0.40404 28
	14 0.50000 17		13 0.10839 29		16 0.12857 28		25 0.46667 27
	15 0.58333 16		14 0.14336 28		17 0.17619 27		26 0.53333 26
			15 0.18531 27		18 0.23810 26		
8	6 0.00606 30		16 0.23427 26		19 0.30476 25	9	10 0.00140 46
	7 0.01212 29		17 0.28671 25		20 0.38095 24		11 0.00280 45
	8 0.02424 28		18 0.34615 24		21 0.45714 23		12 0.00559 44
	9 0.04242 27		19 0.40559 23		22 0.54286 22		13 0.00979 43
	10 0.06667 26		20 0.46853 22				14 0.01678 42
	11 0.09697 25		21 0.53147 21	7	10 0.00303 38		15 0.02517 41
	12 0.13939 24				11 0.00606 37		16 0.03776 40
	13 0.18788 23	m	n=4		12 0.01212 36		17 0.05315 39
	14 0.24848 22	4	10 0.01429 26		13 0.02121 35		18 0.07413 38
	15 0.31515 21		11 0.02857 25		14 0.03636 34		19 0.09930 37
	16 0.38788 20		12 0.05714 24		15 0.05455 33		20 0.13007 36
	17 0.46061 19		13 0.10000 23		16 0.08182 32		21 0.16503 35
	18 0.53939 18		14 0.17143 22		17 0.11515 31		22 0.20699 34
			15 0.24286 21		18 0.15758 30		23 0.25175 33
9	6 0.00455 33		16 0.34286 20		19 0.20606 29		24 0.30210 32
	7 0.00909 32		17 0.44286 19		20 0.26364 28		25 0.35524 31
	8 0.01818 31		18 0.55714 18		21 0.32424 27		26 0.41259 30
	9 0.03182 30				22 0.39394 26		27 0.46993 29
	10 0.05000 29	5	10 0.00794 30		23 0.46364 25		28 0.53007 28
			11 0.07273 28				
	12 0.10455 27		12 0.03175 28		11 0.01587 29		24 0.53636 24
	13 0.14091 26		13 0.05556 27	8	10 0.00202 42	10	10 0.00100 50
	14 0.18636 25		14 0.09524 26		11 0.00404 41		11 0.00200 49
							12 0.00400 48

XIV - جدول معامل ارتباط الرتب

n	مستوى المعنوية (اختبار الجانب الواحد)	
	.05	.01
4	1.000	
5	.900	1.000
6	.829	.943
7	.714	.893
8	.643	.833
9	.600	.783
10	.564	.746
12	.504	.701
14	.456	.645
16	.425	.601
18	.399	.564
20	.377	.534
22	.359	.508
24	.343	.485
26	.329	.465
28	.317	.448
30	.306	.432

T

جدول الأرقام العشوائية في الفترة (0,1)

9588	0198	3820	0894	6705	2742	4569
4900	7466	7496	2909	4978	3722	3079
7521	2725	3902	4063	0949	8802	7023
5843	8325	0528	9800	2068	9029	9665
3949	5233	7119	8639	8131	2475	7942
8715	3299	6692	9803	4528	3506	4945
8401	7484	8384	0224	1242	3068	3077
9207	0149	8058	6870	8529	7588	6798
7736	7574	8588	1319	1872	8386	2689
6520	7841	7373	6333	0333	7685	6023
4165	7946	3533	1031	3257	7900	7313
4838	1276	1246	7801	6225	0457	4350
7947	9142	2410	4292	9920	7496	3669
0253	3400	6426	1920	4342	1146	5190
8348	7762	5936	2109	1036	0439	5598
6720	7203	2655	2512	6737	4896	0476
0628	4188	5390	4364	9341	5278	5372
3661	2808	6778	9171	3486	0070	8395
9784	8036	3352	4960	3833	1634	2902
6119	7492	6927	0885	5729	5241	3936
7482	6518	9499	2966	7333	3457	5776
5938	9499	8154	6110	2970	2117	8560
4725	7083	3201	7851	0122	0435	5401
4298	3395	5982	9762	3821	6014	4061
8745	3018	6243	7117	6795	7009	4831
2084	5778	1475	9227	6156	7372	7152
2578	7162	3083	1147	2882	2377	9183
0045	7556	2820	1923	6061	6569	1805
8138	7994	0808	5852	3319	8838	7975
4187	5360	9691	0736	2104	5497	3998
9239	6149	2790	8068	0064	5110	2736
6640	0233	9391	4707	6740	8089	4086
5314	1778	0038	4353	3054	3854	7375
1982	4704	4771	5948	5567	8559	6118
6178	7492	1538	8617	8481	2830	0787
2817	6883	8082	0589	6475	0674	8508
3270	8611	9218	7240	6695	9635	9953
2826	4639	7201	6681	2385	4960	2165
1734	1450	5790	0361	7877	7867	2580
4870	1693	8312	6607	5553	3707	1932
8207	3300	7158	1835	8846	8869	1195
8248	1060	7053	8547	9854	4010	3488
7596	8039	2547	3006	3395	4834	9408
4959	4082	2318	5127	578	4345	2857
3329	7550	9481	2988	4642	3842	4167

T:

Standard Normal Distribution التوزيع الطبيعي المعياري



$$P(Z \leq z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-w^2/2} dw, \quad [\Phi(-z) = 1 - \Phi(z)]$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714
0.2	0.5793	0.5832	0.5871	0.5911	0.5948	0.5987	0.6026	0.6064	0.6103
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480
0.4	0.6554	0.6591	0.6628	0.6665	0.6700	0.6736	0.6772	0.6808	0.6844
0.5	0.6915	0.6950	0.6985	0.6995	0.7054	0.7088	0.7123	0.7157	0.7190
0.6	0.7257	0.7291	0.7324	0.7354	0.7389	0.7422	0.7454	0.7486	0.7517
0.7	0.7580	0.7611	0.7642	0.7672	0.7704	0.7734	0.7764	0.7794	0.7823
0.8	0.7881	0.7910	0.7939	0.7968	0.7995	0.8023	0.8051	0.8078	0.8106
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810
1.2	0.8849	0.8869	0.8888	0.8906	0.8925	0.8944	0.8962	0.8980	0.8997
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162
1.4	0.9192	0.9208	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306
1.5	0.9332	0.9345	0.9357	0.9368	0.9378	0.9388	0.9398	0.9406	0.9419
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887
2.3	0.9893	0.9896	0.9898	0.9900	0.9904	0.9906	0.9909	0.9911	0.9913
2.4	0.9918	0.9920	0.9922	0.9924	0.9927	0.9929	0.9931	0.9932	0.9934
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973
2.8	0.9974	0.9975	0.9976	0.9977	0.9978	0.9979	0.9980	0.9981	0.9982
2.9	0.9984	0.9985	0.9986	0.9987	0.9988	0.9989	0.9990	0.9991	0.9992
3.0	0.9993	0.9994	0.9995	0.9996	0.9997	0.9998	0.9999	0.9999	0.9999
منحني التوزيع الطبيعي المعياري									
α	0.400	0.300	0.200	0.100	0.050	0.025	0.010	0.005	0.001
z _α	0.253	0.524	0.842	1.282	1.645	1.960	2.326	2.576	3.090
z _{α/2}	0.842	1.036	1.282	1.645	1.960	2.240	2.576	2.807	3.291

المصطلحات الإحصائية
Statistical Terminology
إنكليزي - عربي

A

Absolute	مطلق
Absolute value	القيمة المطلقة
Accuracy	دقة
Adjustment data	تعديل البيانات
Aggregation	تجميع
Allowance Value	حدود السماح
Aggregate	كلى، شامل
Analysis of variance	تحليل التباين
Applied Statistics	الإحصاء التطبيقي
Arbitrary constant	ثابت اختياري
Area	مساحة
Area under the normal curve	مساحة تحت المنحنى الطبيعي
Array	ترتيب
Association	اقتران
Alternative hypothesis	الفرض البديل
Asymmetry	عدم التماثل
Asymmetrical	غير متماثل

Asymptotic	تقريبي
Attribute	صفة
Average	متوسط
Axis	محور
Axis of abscissas	المحور السيني أو الأفقي
Axis of ordinates	المحور العمودي

B

Base	أساس أو قاعدة
Bayes theorem	نظرية بايز
Basic statistics	إحصاءات أساسية
Battery of test	مجموعة الاختبارات
Best fit	التوفيق الأفضل
Between Groups	بين المجموعات
Bias in surveys	التحيز في المسوح
Binominal Distribution	توزيع ثنائي الحدين
Bivariate	ذو متغيرين
Breakdown	تقسيم
Business Cycles	دورات اقتصادية
Business Indicators	مؤشرات اقتصادية

C

Calculation	حساب
Calculus of probability	حساب الاحتمالات
Census	تعداد
Census enumerator	عداد
Census schedule	استمارة التعداد
Causation	سببية
Central tendency	النزعة المركزية
Central limit theorem	نظرية الحد المركزي
Change	تغير
Chi-Squared test	اختبار كاي مربع
Classification	تصنيف
Class frequency	تكرار الفئة
Code	رمز
Coefficient	معامل
Coefficient of Correlation	معامل الارتباط
Coefficient of Detemination	معامل التحديد
Column	عمود
Confidence Intervals	فترات الثقة
Confidence limits	حدود الثقة
Consistency	اتساق
Constant	ثابت

Contingency	توافقي
Contingency table	جدول التوافقي
Continuous	متصل، مستمر
Control	مراقبة
Coordinate	إحداثي
Coordinate axis	محور الإحداثيات
Corrected rate	معدل مصحح
Correction factor	عامل التصحيح
Correlation	ارتباط
Correlation table	جدول ارتباطي
Correlation ratio	نسبة الارتباط
Classical probability	احتمال كلاسيكي
Conditional probability	احتمال شرطي
Continuous distribution	توزيع متصل (مستمر)
Critical region	منطقة حرجة
Curves	منحنيات

D

Data	بيانات
Deduction	استنباط
Degrees of freedom	درجات الحرية
Dependent variable	متغير تابع

Derivative function	دالة مشتقة
Design of experiments	تصميم التجارب
Determinant	محددة
Deviation	انحراف
Diagram (figure)	شكل بياني
Differential	تفاضلي
Discrete distribution	توزيع منقطع
Dispersion	تشتت
Distributions of probability	توزيعات احتمالية
Distribution function	دالة التوزيع
Double sampling	معاينة مزدوجة
Downward bias	تحيز نحو الأدنى
Downward trend	اتجاه هابط

E

Effect	تأثير
Efficiency of estimates	فعالية التقدير
Elimination	حذف
Enumerator	عداد
Equation	معادلة
Error	خطأ
Error of Estimate	خطأ التقدير

Error of Variance	تباين الأخطاء
Estimate / Estimation	تقدير
Estimation equation	معادلة التقدير
Event	واقعة / حادثة
Expected value	القيمة المتوقعة
Experimental error	الخطأ التجريبي
Experimental units	وحدات تجريبية
Exponential Curve	منحنى آسي
Exponential equation	معادلة آسية

F

" F " Table	جدول " ف "
Factor	عامل
Factor analysis	تحليل عوامل
Factorial analysis	تحليل عاملي
Family budget	ميزانية الأسرة
Fluctuations	تقلبات
Forecasting	تنبؤ
Forms	استمارات
Frame	إطار
Frequency	تكرار
Frequency Curve	منحنى التكرار
Frequency density	كثافة التكرار
Frequency distribution	توزيع التكرار
Frequency polygon	مضلع التكرار
Frequency table	جدول التكرار

Function
Function relationship

دالة
علاقة دالة

G

Gaussian curve
Geometric mean
Graduation
Graph
Graphic analysis
Graphic presentation
Goodness of fit test
Grouping
Grouping error
Grouped data

منحنى غوس
وسط هندسي
تدرج
رسم بياني
تحليل بياني
عرض بياني
اختبار جودة التوفيق
تجميع / تصنيف
خطا التجميع
بيانات مبنوية

H

Harmonic equation
Harmonic mean
Heteroscedasticity
Histogram
Homogeneous
Hypothesis testing

دالة توافقية
وسط توافقي
اختلاف التباين
المدرج التكراري
متجانس
اختبار الفروض

I

Ideal number

الرقم الأمثل

Identity	متطابقة
Independent variable	متغير مستقل
Index	دليل أو مقياس
Index number	الرقم القياسي
Infinite	لا نهائي
Inferential statistics	احصاء استدلالي
Interclass correlation	ارتباط بين الفترات
Inersection of sets	تقاطع المجموعات
Interval estimation	فترة، مجال التقدير
Interview	مقابلة / استجواب
Inverse correlation	ارتباط عكسي
Inverse function	دالة عكسية

J

Joint distribution	توزيع مشترك
Joint regression	انحدار مشترك

K

Kurtosis	تفرطح
-----------------	-------

L

Large sample	عينات كبيرة
Law of large numbers	قانون الأعداد الكبيرة

Least squares	المربعات الصغرى
Least squares method	طريقة المربعات الصغرى
Level of significane	مستوى المعنوية
Likelihood	إمكان
Limit	نهاية أو حد
Linear	خطي / مستقيم
Linear correlation	ارتباط خطي
Linear equation	معادلة خطية
Linear interpolation	استكمال خطي
Linear regression	الحدار خطي
Linear trend	اتجاه مستقيم

M

Master Sample	عينة رئيسة
Matrix	مصفوفة
Miximum variation	أقصى اختلاف
Mean	وسط
Method of least squanes	طريقة المربعات الصغرى
Median	الوسيط
Model	نموذج
Moment	عزم
Multiple correlation	ارتباط متعدد

Multiple factor analysis

تحليل عاملي متعدد

Multipleication of probabilities

ضرب الاحتمالات

Mutually exclusive events

احداث متنافية

N

Negative correlation

ارتباط سالب

Non-Linear correlation

ارتباط غير خطي

Non-Statistics

إحصاء لا معلمي

Normal curve

منحنى طبيعي

Normal distribution

توزيع طبيعي

Normal equation

معادلة طبيعية

Null hupothesis

فرضية العدم

O

Observation

مشاهدة

One – tailed test

اختبار من طرف واحد

P

Parabola

قطع مكافئ

Parameter

معطمة، ثابت

Partial correlation

ارتباط جزئي

Partial regression

انحدار جزلي

Percentage	نسبة مئوية
Population (statistical)	مجتمع (إحصائي)
Positive correlation	ارتباط موجب
Prediction	تنبؤ
Point estimation	تقدير بنقطة
Poisson distribution	توزيع بواسون
Power of a test	قوة الاختبار
Probability	احتمال
Prior probability	الاحتمال المسبق
Probability density	كثافة الاحتمال
Probablity sampling	المعاينة الاحتمالية
Q	
Qualitative analysis	تحليل وصفي
Quality control	مراقبة الجودة
Quantitative analysis	تحليل كمي
Quanity Index	الرقم القياسي للكميات
Questionnaire	استبيان / استمارة
Quota sampling	عينة الحصص
R	
Random distribution	توزيع عشوائي
Random error	خطأ عشوائي

Random numbers	أرقام عشوائية
Random sample	عينة عشوائية
Random selection	اختيار عشوائي
Random variable	متغير عشوائي
Range	مدى
Rank correlation	ارتباط الرتب
Rank correlation coefficient	معامل ارتباط الرتب
Rate	معدل
Ratio	نسبة
Ratio estimate	تقدير نسبي
Ratio test	اختبار النسب
Regression	انحدار
Regression coefficient	معامل الانحدار
Regression equation	معادلة الانحدار
Regression estimate	تقدير الانحدار
Regression line	خط الانحدار
Rejection region	منطقة الرفض

S

Sample	عينة
Sample census	مسح بالعينة
Sample design	تصميم العينة
Sampling	معاينة

Sampling error	خطأ المعاينة
Sampling method	طريقة المعاينة
Sampling unit	وحدة المعاينة
Sample size	حجم العينة
Sample space	فضاء العينة
Scatter diagram	شكل الانتشار
Second degree curve	منحنى من الدرجة الثانية
Set theory	نظرية المجموعات
Sensitivity	حساسية
Sequential analysis	تحليل متتابعي
Significance level	مستوى المعنوية
Simple correlation	ارتباط بسيط
Small samples	عينات صغيرة
Sources of error	مصدر الخطأ
Spurious correlation	ارتباط وهمي
Square	مربع
Standard normal distribution	توزيع طبيعي معياري
Standard deviation	انحراف معياري
Standard error	خطأ معياري
Standard error of estimate	خطأ معياري للتقدير
Standard error of the difference between means	خطأ معياري للفرق بين الوسطين
Statistical induction	استنتاج إحصائي

Statistical inference	استدلال إحصائي
Statistical method	الطريقة الإحصائية
Statistical probability	احتمال إحصائي
Statistical processing	التجهيز الإحصائي
Statistics	إحصاء
Survey	مسح (استقصاء)
Student's – distribution	توزيع (t) ستودنت

T

' T ' Table	جدول " ت " ستودنت
Table	جدول
Tabulation	جدول أو تبويب
Test	اختبار
Test of goodness of fit	اختبار جودة التوفيق
Test of homogeneity	اختبار التجانس
Test of significance	اختبار المعنوية
Third degree curve	منحنى من الدرجة الثالثة
Time series	سلسلة زمنية
Two – tailed test	اختبار من طرفين
Type I error	خطأ من النوع الأول
Type II error	خطأ من النوع الثاني

U

Union of sets	اتحاد المجموعات
---------------	-----------------

Unbiased estimate	تقدير غير متحيز
Unit of measurement	وحدة القياس
Universe (statistical)	مجتمع (إحصائي)
Upward tend	اتجاه صاعد

V

Variable	متغير
Variance	تباين
Variance analysis	تحليل التباين
Variance ratio	نسبة التباين
Variation	اختلاف
Variation coefficient	معامل اختلاف
Venn diagram	شكل فن

W

Weight average	تثقيلات، أوزان
Weight average	متوسط مائل أو مرجح

X

X - axis	محور السينات
----------	--------------

Y

Y-axis	محور الصادات
Yates continuity correction	تصحیح یاتس

Z

Z - test	الختبار (ز)
----------	-------------

اللجنة العلمية:

أ.د : امطانيوس مخول

د : هاني عمران

د : نبيل علي

المدقق اللغوي:

أ. د : محمد موعد

حقوق الطبع والترجمة والنشر محفوظة لمديرية الكتب والمطبوعات الجامعية