



الجمهورية العربية السورية
جامعة دمشق - كلية الزراعة
قسم الاقتصاد الزراعي



التحليل الإحصائي

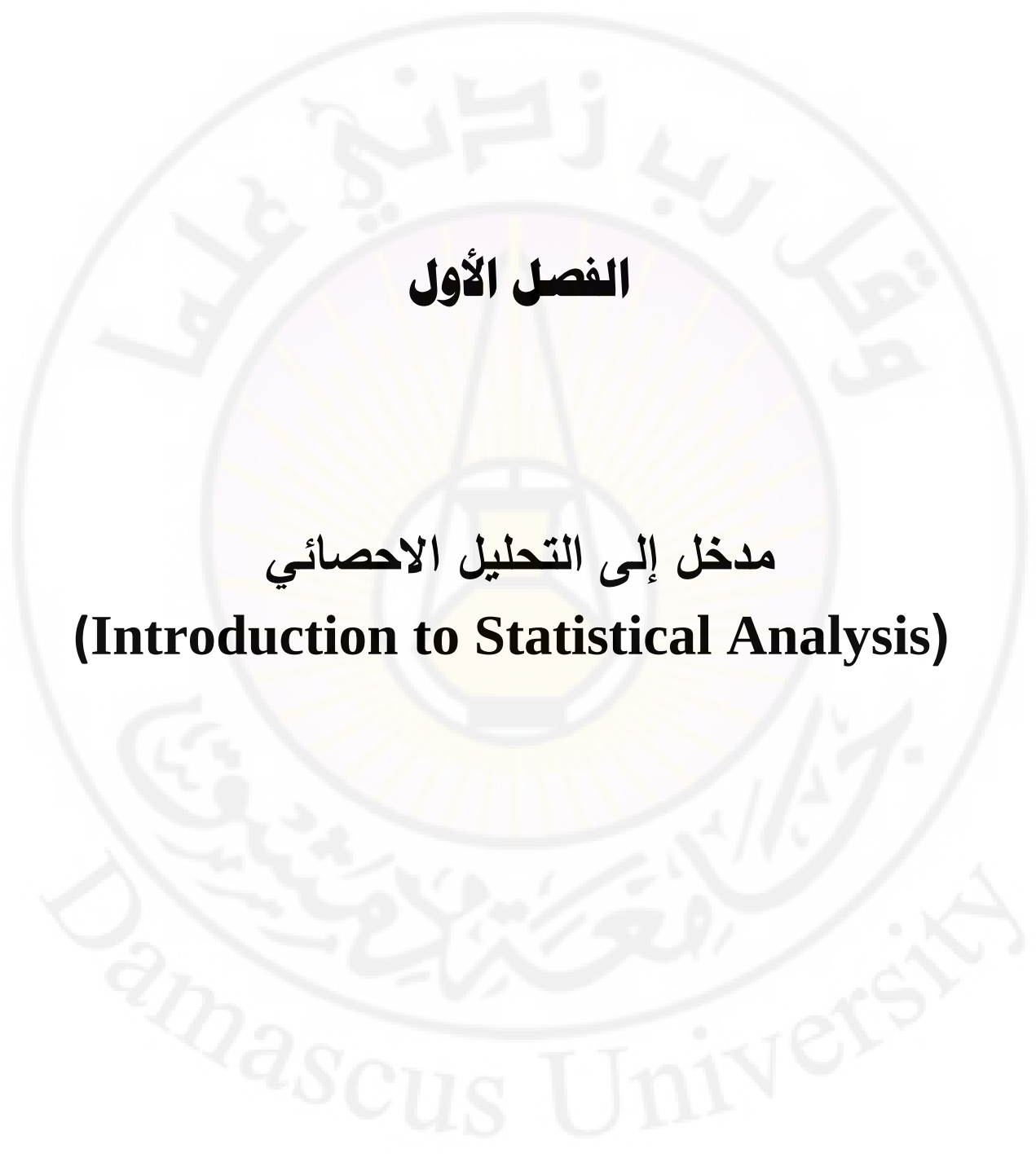
(السنة الرابعة – إرشاد + اقتصاد زراعي)

الفصل الأول

د. خالد السلطان

2023/2022

Damascus University

The background of the page features a large, faint watermark of the Damascus University logo. The logo is circular, with the university's name in Arabic 'جامعة دمشق' at the top and 'Damascus University' at the bottom. In the center is a shield with a sunburst design.

الفصل الأول

مدخل إلى التحليل الإحصائي (Introduction to Statistical Analysis)

مدخل إلى التحليل الإحصائي

Introduction to Statistical Analysis

لا يمكن تعلم المبادئ الأساسية لكيفية التحليل الإحصائي للبيانات إلا إذا كان الطالب أو الباحث ملماً بالمفاهيم الأساسية في الإحصاء الاستدلالي والتي على أساسها يختار الاختبارات الإحصائية المناسبة لمعالجة بياناته واختبار فرضيات بحثه، حتى تمكنه من قراءة وتفسير نتائجه ومخرجاته التي يتحصل عليها من تجاربه.

لذلك في هذا الفصل سوف يتم إلقاء الضوء على المبادئ الأساسية في أنواع الإحصاء الاستدلالي وأنواع ومقاييس المتغيرات وأنواع الاختبارات الإحصائية المناسبة.

تعريف الإحصاء: هو العلم الذي يهتم بجمع البيانات أو المعلومات الوصفية أو الكمية، وعرضها في جدول ورسوم بيانية، وتحليلها بطريقة مختصرة تيسر الإدراك والفهم.

أنواع الإحصاء:

إن التمييز بين أنواع الإحصاء يتم وفق أساسين الأول يتعلق بطبيعة المشكلة التي يهتم الباحث بدراستها والغرض الذي من أجله تستخدم البيانات، أما الأساس الثاني فيتعلق بنوع البيانات المراد تحليلها ومستوى قياسها، فاستخدام الأسلوب الإحصائي المناسب يعتمد على طبيعة البيانات (وصفية أو كمية) ومستوى قياس المتغير موضوع البحث (أسمية أو رتبية أو فترية أو نسبية) وهنا يتم التمييز بين الإحصاء المعلمي (البارامتري) والإحصاء اللا معلمي (اللا بارامتري).

أولاً: تصنيف علم الإحصاء من حيث طبيعة المشكلة:

يُصنف الإحصاء من حيث طبيعة المشكلة التي يعالجها إلى:

● الإحصاء الوصفي (Descriptive Statistics):

هو ذلك الفرع الذي يهتم بالأساليب العلمية لجمع وتنظيم وتيويب و طرق عرض البيانات سواء كان جدولياً، أو بيانياً، وتلخيصها في رقم أو مجموعة أرقام تعبر عن وصف العينة أو المجتمع محل الدراسة، ويشمل كل من مقاييس النزعة المركزية (المتوسط، الوسيط، و المنوال) ومقاييس التشتت (المدى، الانحراف المعياري، التباين، الربيعيات) والمقاييس الخاصة بالعلاقات الإحصائية (الارتباط، الانحدار).

وظائف الإحصاء الوصفي:

يمكن تلخيص وظائف الإحصاء الوصفي بالنقاط التالية:

- 1- جمع وتنظيم وتبويب البيانات.
- 2- عرض البيانات جدولياً (جداول تكرارية) أو بيانياً (مدرج تكراري، منحني تكراري، دائرة بيانية).
- 3- تلخيص البيانات بحساب بعض المقاييس الإحصائية الأساسية مثل
 - مقاييس النزعة المركزية (المتوسط-الوسيط-المنوال....).
 - مقاييس التشتت (المدى-التباين-الانحراف المعياري.....).
 - مقاييس الالتواء والتفلطح.
 - مقاييس الارتباط والانحدار.

● **الإحصاء استدلالي (Inferential Statistics):** هو ذلك الفرع من الإحصاء الذي يهتم بالاستدلال على معالم المجتمع المجهولة عن طريق بيانات عينة مسحوبة من ذلك المجتمع. ويبدأ الإحصاء الاستدلالي حيث ينتهي الإحصاء الوصفي.

وظائف الإحصاء الاستدلالي

- 1) يهتم بأساليب التحليل التي تستدعي المقارنة والاستنتاج والحكم على ظواهر المجتمع من خلال بيانات العينة.
- 2) يهتم باستخلاص نتائج عن المجتمع عن طريق عينة مسحوبة من ذلك المجتمع معتمداً على طريقة التقدير الإحصائي (Statistical Estimation) حيث يتم حساب قيمة مقدرة تسمى إحصاءه (وهي قيمة مسحوبة من بيانات العينة) كمقدر للقيمة الفعلية في المجتمع والتي تسمى (معلمة Parameter) وهي القيمة الفعلية لصفة من صفات المجتمع غالباً تكون مجهولة ومطلوب تقديرها، كذلك يمكن تطبيق اختبار الفروض الإحصائية (Hypothesis Testing) كطريقة من طرق الإحصاء الاستدلالي للحكم على فرضيات حول معالم المجتمع.

ثانياً: تصنيف علم الإحصاء من حيث نوع البيانات:

يُصنف علم الإحصاء من حيث نوع البيانات التي يعالجها إلى إحصاء معلمي (بارامتري) وإحصاء لا معلمي (لا بارامتري)، وهذان المصطلحان ليسا مترادفين بل يشيران إلى جانبين مختلفين في عملية الاستدلال الإحصائي، أي أن الإحصاء المعلمي واللا معلمي هما فرعان للإحصاء الاستدلالي.

الإحصاء المعلمي (البارامتري) والإحصاء اللامعلمي (اللابارامتري):

المعلم (parameter): هو مصطلح إحصائي يمثل مقياس أو قيمة تصف المجتمع الأصلي الذي اشتقت منه العينة. وهي قيمة غير معروفة بالنسبة للباحث و إنما يقدرها من بيانات العينة.

الإحصاء المعلمي أو البارامتري (Parametric Statistics):

الإحصاء المعلمي: هو مجموعة الأساليب الإحصائية التي تتطلب تحقق شروط وافتراسات محددة حول المجتمع الذي تسحب منه العينة، أهمها اعتدالية توزيع بيانات المجتمع وطبيعة ونوع البيانات (فئوية أو نسبية).

الإحصاء اللامعلمي أو اللابارامتري (Non Parametric Statistics):

الإحصاء اللامعلمي: هو مجموعة الطرق والأساليب الإحصائية البديلة التي تستخدم في حالة عدم تحقيق الافتراضات والشروط السابقة حول المجتمع التي تسحب منه العينة، وكذلك تطبق الاختبارات اللامعلمية عندما تكون البيانات لا معلمية (وصفية: اسمية أو ترتيبية).

وهذا النوع يسمى بإحصاء التوزيعات الحرة.

أي أن الاختبارات الإحصائية اللامعلمية يمكن تطبيقها على مدى واسع دون ان تشترط اعتدالية التوزيع لبيانات المجتمع الذي سحبت منه العينة.

الاختبارات الإحصائية المعلمية

تستعمل الاختبارات المعلمية في حالة العينات الكبيرة ومع مستويات القياس العليا النسبية والفترية (Interval, Ratio) للمتغيرات.

مزايا الاختبارات المعلمية:

- الاختبارات المعلمية هي أقوى و أدق و أكثر ثقة في اختبار الفرضيات الإحصائية من الاختبارات اللامعلمية وذلك لأنها تعمل على إيجاد الفروق الموجودة في البيانات نفسها ومن ثم فهي أكفأ وأقدر على رفض فرضية العدم من الاختبارات اللامعلمية.
- عند اختلال احد شروط الاختبارات المعلمية لا يكون امامنا إلا اللجوء الى الاختبارات اللامعلمية اذ حينها ستكون نتائجها أوثق وأدق من الطرائق المعلمية.

عيوب الاختبارات اللامعلمية:

من جانب آخر يؤخذ على الاختبارات المعلمية العديد من المآخذ منها:

- تُعد الاختبارات المعلمية أكثر صعوبة من الاختبارات اللامعلمية بسبب محدودية نوع البيانات التي تستعمل في الاختبارات المعلمية.

الاختبارات الإحصائية اللامعلمية:

يطلق مصطلح اختبارات لا معلمية (Nonparametric Tests) على الاختبارات التي تستعمل في الحالات التي لا يستطيع فيها الباحث معرفة أو حتى افتراض التوزيع الاحتمالي للمجتمع أو المجتمعات التي جاءت منها البيانات المراد اختبارها. وتسمى هذه الاختبارات أيضاً باسم الاختبارات حرة التوزيع (Distribution Free Testing) أي غير المقيدة بالتوزيع الطبيعي.

وتعرف الاختبارات اللامعلمية على أنها نوع من الطرائق الإحصائية التي يمكن استعمالها في التوصل إلى نتائج بخصوص المجتمع بغض النظر عن نوع التوزيع الاحتمالي لمجتمع العينة أو نوع البيانات التي يمكن الحصول عليها .

كذلك يتم استعمال الاختبارات اللامعلمية في حالة انتهاك أو فقد أحد شروط الاختبارات المعلمية إذ لا يكون أمامنا إلا اللجوء إلى الاختبارات اللامعلمية لأن حينها ستكون نتائجها أوثق وأدق من الاختبارات المعلمية.

مزايا الاختبارات اللامعلمية:

تمتاز الاختبارات اللامعلمية بعدة مزايا مقارنة بالاختبارات المعلمية أهمها:

- تُعد من الطرائق المبسطة وسهلة التطبيق والفهم، وتتميز بقلة الافتراضات المطلوبة، أي أنها لا تحتاج إلى افتراضات كثيرة عن توزيع المجتمع الذي سحبت منه العينة.
- معظم الطرائق اللامعلمية لا تتطلب أن يكون حجم العينة كبيراً أو إجراء حسابات معقدة لذلك يمكن جمع البيانات وتصنيفها وتحليلها بسرعة أكبر.
- البيانات المستعملة في الاختبارات اللامعلمية تكون عادة مقاسة بمقاييس دنيا مثل المقياس (الاسمي أو الترتيبي)

عيوب الاختبارات اللامعلمية:

من الناحية الأخرى تعاني الاختبارات اللامعلمية من مأخذ ينبغي التنبيه لها أهمه:

- الاختبارات اللامعلمية بصفة عامة أقل قوة من الاختبارات المعلمية، وهذا يعرف بالكفاءة النسبية (Relative Efficiency) للاختبارات اللامعلمية مقارنة بالاختبارات المعلمية
- عند استيفاء الاختبارات المعلمية لمتطلباتها وافترضاها وشروطها يجب عدم استعمال الاختبارات اللامعلمية.

متى يتم استخدام الاختبارات اللامعلمية ؟

- (1) إذا كان توزيع المجتمع الذي سحبت منه العينة غير معروف ويرغب الباحث في إجراء استدلال إحصائي حول معالم المجتمع، ففي مثل هذه الحالة يعتمد على التوزيع الفعلي للعينة (Empirical Distribution) ولذلك تسمى هذه الاختبارات أيضا باسم الاختبارات أو الطرق حرة التوزيع (Distribution Free Methods)
- (2) تستخدم هذه الاختبارات في حالة التعامل مع عينات صغيرة الحجم (أقل من 30).
- (3) في حالة أن تكون المتغيرات مقاسة بمقياس اسمي (Nominal Scale) أو بمقياس رتبي (Ordinal Scale) حيث لا يمكن في هذه الحالة استخدام الاختبارات المعلمية.
- (4) إذا لم تتوافر شروط أو افتراضات استخدام الاختبارات البارامترية في اختبارات الفروض وأهمها شرط التوزيع المعتدل للمجتمع الذي تسحب منه العينة وكذلك تجانس التباين.

الفرق بين الإحصاء البارامترية والإحصاء اللابارامترية

- ترتبط المعلمية واللامعلمية في الإحصاء الاستدلالي بجملة من العوامل المترابطة أهمها نوع البيانات وشكل توزيعها وحجم العينة.
- الاختبارات المعلمية تشترط حجماً كبيراً للعينة (أكبر من 30) مع بيانات كمية (فتوية أو نسبية) بشرط أن يكون توزيعها طبيعياً.
- بينما الاختبارات اللامعلمية لا يُشترط استخدامها ما سلف من شروط وضوابط ويمكن تلخيص الفروق بين الأساليب المعلمية واللامعلمية في الجدول التالي:

وجه المقارنة	الإحصاء البارامترية Parametric Tests	الإحصاء اللابارامترية Non Parametric Test
التعريف	هي الأساليب الإحصائية التي تستخدم في التحقق من صحة الفروض المتعلقة بمجموعات قيم بارامترات محددة، أي يعتمد على معالم المجتمع.	هي الأساليب الإحصائية التي تستخدم في التحقق من صحة الفروض المتعلقة بمجموعات قيم بارامترات غير محددة، أي لا يعتمد على معالم المجتمع. ويسمى بإحصاء التوزيعات الحرة أو الفرضيات الضعيفة.
الشروط	يشترط اعتدالية التوزيع لبيانات المجتمع الذي تسحب منه العينة (طبيعية البيانات). ان يكون حجم العينة كبير وتم اختياره عشوائيا. يستخدم في حالة البيانات الكمية (القياس الفكري والنسبي).	- لا يشترط اعتدالية التوزيع. - حجم العينة صغير. - يستخدم في حالة البيانات الوصفية (القياس الاسمي والترتيبي).
الاختبارات	اختبارات (t) - معامل ارتباط بيرسون - تحليل التباين (F).	التكرارات - النسب المئوية اختبار ولكسون - اختبار كروسكال واليس - مربع كاي - مان ويتي، معامل ارتباط الرتب لسبيرمان .

البيانات المعلمية:

تكون البيانات معلمية إذا كانت تحقق الشرطين التاليين:

1- إذا كانت مقاييس البيانات كمية (نسبي أو فئوي).

2- ان تتوزع البيانات وفق التوزيع الطبيعي.

البيانات اللا معلمية:

تكون البيانات غير معلمية إذا كانت مقاسة بالمقياس الاسمي أو الترتيبي.

طبيعة وأنواع البيانات (Nature and Types of Data)

يقصد بطبيعة البيانات في هذه الحالة أنواع المتغيرات ومستويات قياسها، من الضروري للباحث أن يحدد نوع المتغير الذي يقوم بدراسته قبل عمل أي تحليل احصائي، فنوع المتغير له علاقة بنوع التحليل الاحصائي الذي يريد استخدامه الباحث، وإن أي إخلال بذلك يؤثر على الافتراضات الأساسية الخاصة بكل تحليل احصائي ويؤدي إلى خلل كبير في نتائج التحليل الاحصائي.

وتنقسم المتغيرات عند قياسها إلى أربعة أنواع رئيسية وهي: المتغيرات الأسمية، الترتيبية، الفئوية، والنسبية، وقبل الخوض في تفاصيل أنواع المتغيرات لابد من توضيح مفهوم القياس.

القياس و مستويات القياس (Levels of Measurement):

مستويات القياس تعني كيفية قياس المتغيرات أو كيف نقاس المتغيرات.

القياس: هو إعطاء قيم عددية أو رمزية لمتغير عن طريق القياس.

هناك متغيرات يمكن قياسها بدقة باستخدام معدات وأجهزة مثل الطول والوزن ودرجة الحرارة، وهناك متغيرات لا يمكن قياسها بمقاييس دقيقة بدقة مثل الحب والكراهية ودرجة الرضا الوظيفي.

ما هو الهدف من دراسة مستويات القياس؟

الإجابة: إن الهدف الأساسي من دراسة مستويات القياس للمتغيرات، هو تحديد الاختبار الاحصائي المناسب لكل نوع من المتغيرات من أجل عملية التحليل الاحصائي لملاحظات المتغيرات، حيث أن لكل مقياس أساليب واختبارات خاصة به وتناسب طبيعته.

وعموماً يوجد أربعة أنواع من المقاييس أو مستويات للقياس (Scales of Measurement) مرتبة تصاعدياً من البسيط إلى الأكثر وضوحاً وهي:

- المقياس الاسمي (Nominal Scale)
- بالمقياس الترتيبي (Ordinal Scale)
- المقياس الفتري (Interval Scale)
- المقياس النسبي (Ratio Scale)

مقاييس قياس المتغيرات :

في هذا القسم سوف نتناول بالتفصيل أنواع المتغيرات ومقاييس هذه المتغيرات بالأمثلة التوضيحية:

المتغير (Variable): إن مصطلح "متغير" يعبر عن شيئاً يتغير ويأخذ قيماً مختلفة، فهو أي ظاهرة تأخذ قيماً مختلفة بين مفرداتها مثل الاسم، الجنس، السن، الطول، التحصيل العلمي.

فالمتغير مصطلح يدل على صفة محددة، تأخذ عدداً من الحالات أو القيم أو الخصائص، وقد يشير المتغير إلى مفهوم معين يجري تعريفه إجرائياً في البحث ويتم قياسه كمياً أو وصفه كيفياً، فالذكاء مثلاً صفة عقلية لدى الأفراد بدرجات متفاوتة وهو متغير، لأنه ليس بنفس القيمة أو الدرجة أو المستوى عند جميع الأفراد، وبالتالي يمكن للباحث أن يقدر درجات هذا المتغير بمقياس رتبي (غبي، متوسط الذكاء، ذكي)

ويمكن تصنيف المتغيرات وفقاً لعدة معايير أما لطبيعة البيانات التي تمثلها أو لطبيعة العلاقة بينها.

● المتغيرات الأسمية (Nominal Variables)

إن القيم الخاصة بالمتغير الاسمي تختلف عن بعضها في النوعية لا في الكمية، ويسمى مستوى القياس لهذا النوع بالمقياس الاسمي (Nominal Scale)

المقياس الاسمي يصنف المشاهدات فقط.

تستخدم الأعداد للتمييز بين المفردات، وفي هذه الحالة لا يكون للعدد أي مدلول كمي، فمثلاً يمكننا استعمال الرقمين (0 و 1) ليدلا على متغير الجنس، حيث يدل الرقم (0) على الذكر، والرقم (1) يدل على الأنثى. نلاحظ أن الرقمين (0، 1) لا يدلان على القيم العددية لذلك لا نجري عليهم أي عملية حسابية (الجمع الطرح والضرب والتقسيم). كذلك لا تستخدم فيها مقاييس النزعة المركزية على هذه المتغيرات باستثناء المنوال. **العمليات الحسابية:** لا يتم إجراء أي عملية حسابية على هذا النوع من المتغيرات.

الأعمال الإحصائية أو المقاييس الإحصائية:

الجدول التكرارية، التكرارات، النسب، والمنوال.

الاختبارات الإحصائية: في هذا النوع من المتغيرات يتم تطبيق الاختبارات الإحصائية غير المعلمية فقط.

أمثلة على المتغيرات الأسمية: الجنسية (سوري=1، لبناني=2، مصري=3) الحالة العائلية (متزوج=1، عاذب=2، مطلق=3، أرمل=4)

بيانات المقياس الاسمي والتحليل الإحصائي:

توجد طريقتان يمكن من خلالهما جمع بيانات المقياس الاسمي:

1. من خلال طرح سؤال مفتوح، يمكن ترميز اجاباته بأرقام يحددها الباحث (عاذب=1، متزوج=2، مطلق=3).
2. البديل الآخر هو طرح سؤال متعدد الخيارات.

في كلتا الحالتين، سيتم تحليل البيانات المجمعة باستخدام التكرارات والنسب المئوية أو المنوال،

٢ المتغيرات الترتيبية (Ordinal Variables)

المتغيرات الترتيبية تقع في مستوى أعلى من المتغيرات الاسمية وتقاس بالمقياس الترتيبي (Ordinal Scale) وهذا المقياس يسمح بترتيب القيم او الفئات حسب سلم معين.

المقياس الترتيبي يصنف المشاهدات ويرتبها ولا يبين الفرق بينها.

فمثلاً: لو أخذنا مستويات الدخل (منخفض، متوسط، وعال) فيمكننا القول بأن مستوى الدخل العالي أكبر من مستوى الدخل المتوسط ولكن لا نستطيع تحديد الفرق أي كم يزيد الدخل العالي عن المتوسط، هذه البيانات على الرغم أنها غير عددية تم ترتيبها وفق ترتيب هرمي.

الأعمال أو المقاييس الإحصائية:

يتم استخدام وحساب الوسيط بالإضافة إلى الأعمال السابقة الخاصة بالمقياس الاسمي (الجدول التكرارية، التكرارات، النسب، والمنوال)

العمليات الحسابية: لا يتم إجراء أي عملية حسابية على هذا النوع من المتغيرات.

الاختبارات الإحصائية: في هذا النوع من المتغيرات يتم تطبيق الاختبارات الإحصائية غير المعلمية فقط.

أمثلة على المتغيرات الترتيبية: المستوى التعليمي (أمي، ابتدائي، ثانوي، جامعي)، مستوى الذكاء (متدني، متوسط، مرتفع)، الرتبة الأكاديمية (مدرس، أستاذ مساعد، استاذ). السعادة، ومستوى الألم، درجات الرضا (غير راضي تماماً غير راضين محايد، راضي، راضي جداً)

البيانات الترتيبية والتحليل الاحصائي: يمكن تنظيم بيانات المقياس الترتيبي في شكل جداول تكرارية ليتم حساب التكرارات والنسب والوسيط والمنوال.

أيضاً، يمكن استخدام طرق مثل اختبار مان وتني واختبار كروسكال ووالس لتحليل البيانات الترتيبية. ويتم تنفيذ هذه الطرق عموماً للمقارنة بين مجموعتين ترتيبيتين أو أكثر.

في اختبار مان وتني، يستطيع الباحث استنتاج أي المتغيرات لمجموعة واحدة أكبر أو أصغر من المتغير الآخر لمجموعة مختارة عشوائياً. بينما في اختبار كروسكال ووالس، يستطيع الباحث تحليل ما إذا كان لمجموعتين ترتيبيتين أو أكثر نفس الوسيط أم لا.

● المتغيرات الفترية (Interval Variables)

المتغيرات الفترية تقع في مستوى أعلى من المتغيرات في المستوى الرتبي وتقاس بالمقياس الفترى (Interval Scale) وهذا المقياس يتضمن خاصية تساوي المسافات بين الرتب، والمسافات المتساوية تدل على مقادير متساوية من الخاصية التي يتم قياسها، ولذلك يسمى هذا المقياس في بعض الأحيان بمقياس "المسافة" والأرقام التي يتم استخدامها لفئات المتغير تدل على نوع المعدود ترتيبه وكمه.

المقياس الفترى يصنف المشاهدات ويرتبها ويبين الفرق ولا يوجد صفر مطلق.

ويلاحظ في المتغيرات الفترية أن الصفر لا يشير إلى انعدام وجود الظاهرة المدروسة، فدرجة الحرارة إذا كانت صفراً لا تعني عدم وجود حرارة، كذلك حصول الطالب على صفر في الإحصاء لا يعني أنه لا يعرف شيئاً في هذا المقرر، وهذا الصفر يسمى بالصفر النسبي أو الافتراضي وليس صفراً مطلقاً. وللعلم يستخدم هذا المقياس بشكل كبير في الدراسات التربوية والاجتماعية.

العيب الوحيد لهذا المقياس هو عدم وجود نقطة إنطلاق محددة مسبقاً أو قيمة للصفر الحقيقي.

العمليات الحسابية: يتم تطبيق عمليات (الجمع والطرح) فقط على هذا النوع من المتغيرات.

الأعمال أو المقاييس الإحصائية: يتم حساب واستخدام مقاييس النزعة المركزية والتشتت المختلفة بالإضافة إلى الأعمال السابقة الخاصة بالمقياس الاسمي والترتيبي (الجدول التكرارية، التكرارات، النسب، والمنوال) ويتم تطبيق تحليل الارتباط والانحدار الذي يستخدم على نطاق واسع في تحليل بيانات الفترات.

الاختبارات الإحصائية: في هذا النوع من المتغيرات يتم تطبيق الاختبارات الإحصائية المعلمية بشرط أن يكون توزيع البيانات طبيعياً أو حجم العينات كبير، كذلك يمكن تطبيق الاختبارات غير المعلمية.

أمثلة على المتغيرات الفترية:

- درجة الحرارة،
- الزمن يعد مثلاً شائعاً جداً لمقياس الفترات حيث تكون القيم ثابتة وقابلة للقياس.
- تقويم السنوات والوقت يندرج أيضاً ضمن هذه الفئة من مقاييس القياس
- مقياس ليكرت، يعتبر أكثر مقاييس الفترات استخداماً.

● المتغيرات النسبية (Ratio Variables)

تأخذ المتغيرات النسبية مكاناً أعلى من كل المتغيرات السابقة أو في قمتها وتقاس بالمقياس النسبي (Ratio Scale) وهذا المقياس يتضمن فضلاً عن خصائص المستويات السابقة (تصنيف وترتيب ومسافات متساوية) خاصية النسبة وهي تنسب الأرقام أو العناصر إلى بعضها إضافة إلى وجود الصفر الحقيقي " المطلق " .

المقياس النسبي يصنف المشاهدات ويرتبها ويبين الفرق ويوجد صفر مطلق.

فمثلاً متغير السرعة يقاس بالمقياس النسبي حيث أن قيمة السرعة (10، 20، 30 ...الخ) فيها تصنيف وترتيب والمسافات بينها متساوية إضافة إلى وجود الصفر الحقيقي المطلق الذي يشير إلى انعدام وجود الظاهرة المدروسة، فسرعة السيارة عندما تكون صفراً يعني إنها واقفة. وللعلم يستخدم هذا المقياس بشكل كبير في الدراسات الطبية.

العمليات الحسابية: يتم تطبيق العمليات الحسابية الأربع على هذا النوع من المتغيرات.

الأعمال أو المقاييس الإحصائية: يتم استخدام وحساب كافة الأعمال والمقاييس الإحصائية أي يمكن تطبيق التحليل الوصفي والاستدلالي على هذا النوع من المتغيرات.

الاختبارات الإحصائية: في هذا النوع من المتغيرات يتم تطبيق الاختبارات الإحصائية المعلمية بشرط أن يكون توزيع البيانات طبيعياً أو حجم العينات كبير، كذلك يمكن تطبيق الاختبارات غير المعلمية.

أمثلة على المتغيرات النسبية: الطول، الوزن، العمر، السرعة، عدد أفراد الأسرة، الدخل، عدد الحوادث في مكان معين. فمثلاً إذا كان هناك شخص وزنه (80 كغ) وشخص آخر وزنه (40 كغ) فإننا نقول أن وزن الشخص الأول يساوي ضعف وزن الشخص الثاني، لكن عندما نقول أن درجة الحرارة (30) مئوية ودرجة حرارة (15) مئوية فهذا لا يعني أن درجة الحرارة الأولى ضعف الثانية في الأثر ولكن أكبر منها. لا يوجد قيم سالبة لهذا النوع.

• بسبب وجود قيمة للصفر الحقيقي، فإن المقياس النسبي لا يمتلك قيم سالبة

ويمكن تلخيص مستويات القياس الأربعة للمتغيرات بالجدول الآتي:

المستوى	العمليات الرياضية	الخصائص القياسية	أمثلة
الاسمي (Nominal)	العد	- عدد لا يدل على كم أو مقدار - الأرقام تحل محل الأسماء - الأرقام قد تمثل فئات	الجنس - الجنسية - الحالة العائلية

		- لا يمكن اجراء العمليات الحسابية على الأرقام	
الترتيبي (Ordinal)	الترتيب	- كم لا يشار إليه بعدد - الأرقام مرتبة تصاعدياً أو تنازلياً - المسافات بين الرتب غير متساوية	درجة الرضا الوظيفي - المؤهل العلمي - الرتبة العسكرية، درجة الذكاء.
الفترتي (Interval)	الجمع والطرح	- عدد يدل على كم أو مقدار (قيم متصلة) - يتكون من مسافات متساوية - له صفر اعتباري " ليس مطلق" - يمكن مقارنة المسافات بين الدرجات	درجة الحرارة - العلامات في الاختبارات والقياسات النفسية (الذكاء، الانفعال)
النسبي (Ratio)	جميع العمليات الرياضية	- عدد يدل على كم أو مقدار (قيم متصلة) - يتكون من وحدات متساوية وله صفر مطلق - يمكن مقارنة المسافات بين الدرجات	السرعة - الطول - الوزن - الدخل - التكاليف

ويمكن تلخيص الاختبارات الاحصائية المعلمية واللا معلمية بالجدول التالي:

نوع الاختبار	الاختبارات المعلمية	الاختبارات غير المعلمية
مقارنة متوسط عينة بمتوسط مجتمع	- اختبار (Z) للعينة الواحدة - اختبار (t) للعينة الواحدة One Sample T-Test	اختبار الإشارة (Sign Test) للعينة الواحدة
الفرق بين عينتين مستقلتين	اختبار (T) لعينتين مستقلتين (Independent sample T-Test)	اختبار (مان وتتي) لعينتين مستقلتين Mann-Whitney Test
الفرق بين عينتين مرتبطتين	اختبار (T) لعينتين مرتبطتين (Paired sample T-Test)	اختبار (ويلكوكسن) لعينتين مرتبطتين Wilcoxon Tes اختبار الإشارة لعينتين مرتبطتين (Sign Test)
الفرق بين متوسطات أكثر من عينتين مستقلة	اختبار تحليل التباين (ONE-WAY) (ANOVA)	اختبار (كروسكال واليس) Kruskal-WallisTest
الفرق بين متوسطات أكثر من عينتين مرتبطة	اختبار تحليل التباين ذات القياسات المتكررة	اختبار (فريدمان) لأكثر من عينتين مرتبطة Friedman Test

العلاقة بين مستويات القياس والاختبارات الإحصائية:

إن معرفة مستويات قياس المتغيرات المعتمدة في الدراسة تؤثر بشكل مباشر على نوع الاختبارات الإحصائية المناسبة لتحليل البيانات الميدانية، فكل اختبار يصلح لنوع معين من البيانات (أو لمستوى قياسها) وكلما تغيرت نوع البيانات تغير معها الاختبار المناسب.

ومن الجدير بالذكر أن هناك علاقة بين مستويات القياس للمتغيرات بالأساليب الإحصائية المناسبة للبيانات والتي يمكن تلخيصه بالجدول الآتي:

الاحصاء	الاسمي (Nominal)	الترتيبي (Ordinal)	الفتري (Interval)	النسبي (Ratio)
الوصفي	- التكرارات - النسب لمئوية - الأعمدة البيانية - المنوال	- التكرارات - النسب المئوية - الأعمدة البيانية - المنوال - الوسيط - ارتباط سبيرمان	- التكرارات - النسب المئوية - المدرج/ المضلع التكراري - المنوال - الوسيط - المتوسط - مقاييس التشتت (التباين - المدى - الانحراف المعياري) - ارتباط سبيرمان	- التكرارات - النسب المئوية - المدرج/ المضلع التكراري - المنوال - الوسيط - المتوسط - مقاييس التشتت (التباين - المدى - الانحراف المعياري) - ارتباط سبيرمان
الاستدلالي	اختبار كاي تربيع	- مانويتتي / فريدمان - ولكوكسون - كروسكال واليس	- اختبار (t) - تحليل التباين (F)	- اختبار (t) - تحليل التباين (F)

أسس تحديد الاختبارات الإحصائية المناسبة للدراسة:

وفي الأخير نشير إلى أن أصعب مرحلة في البحث هي تحديد الاختبارات الإحصائية المناسبة لقياس الفرضيات، لذلك عند اختيار الاختبار الإحصائي يجب أن يتبع عدة معايير من خلال الإجابة على هذه الأسئلة الخمسة الآتية:

س1: ماهي نوع الفرضية التي تريد قياسها (متوسط، تباين، نسبة) ؟

س2: مانوع التصميم التجريبي الذي يستخدمه الباحث؟

س3: ماعدد العينات المستخدمة في البحث؟ وهل هي مستقلة أم مرتبطة ؟

س4: مانوع البيانات الخاصة بمتغيرات البحث ؟

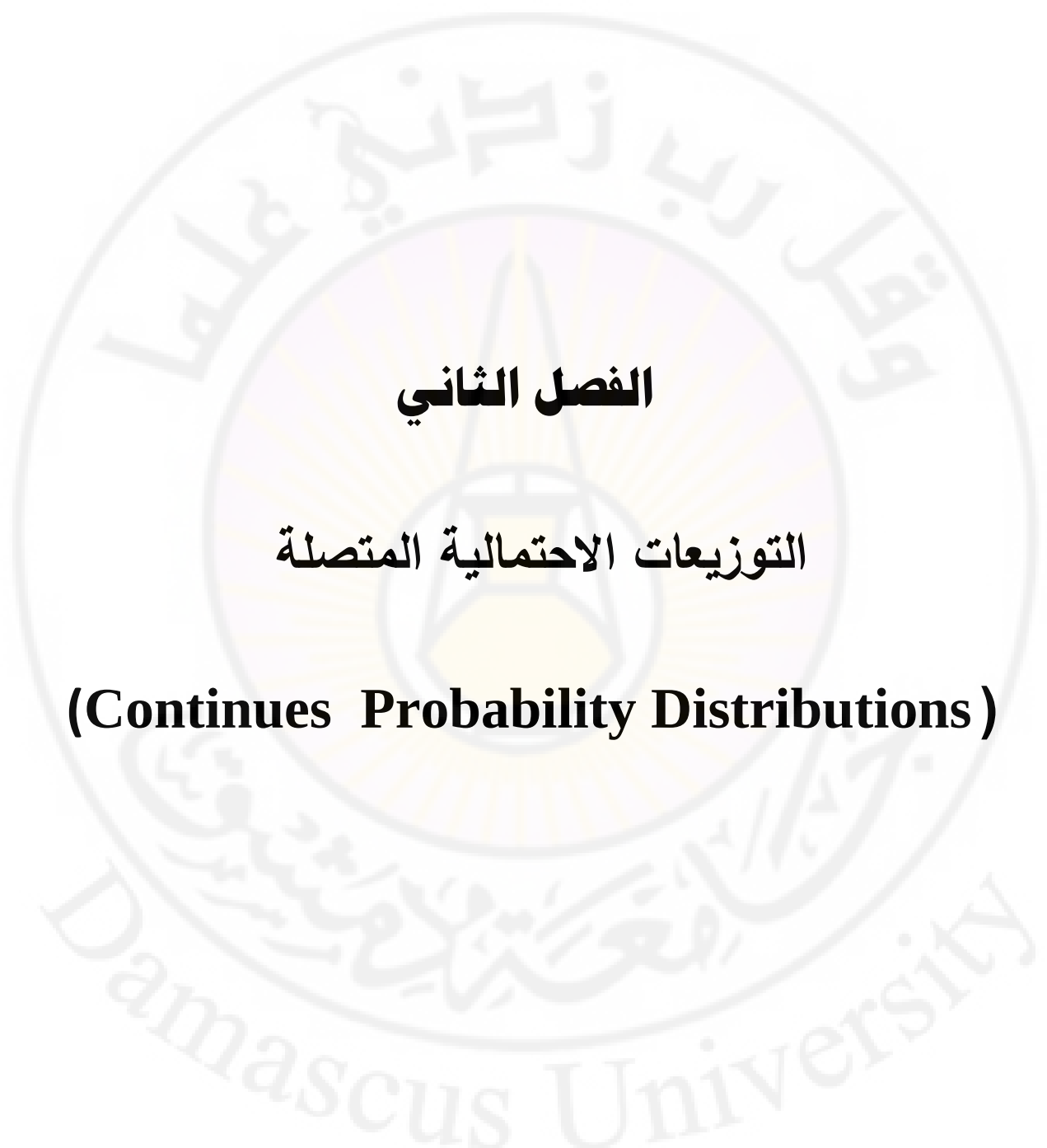
س5: ما طبيعة توزيع البيانات (بارامترية أم لا بارامترية)؟

ويمكن وضع الإجابات على الأسئلة السابقة في الجدول الاتي:

العينات	الفرض	التصميم التجريبي	نوع البيانات	الاختبار الاحصائي
عينة واحدة	التحقق من جودة المطابقة (وصفي)	مجموعة واحدة ذات الاختبار الواحد	اسمية	ذوي الحدين - اختبار كا ² - سميرنوف
			رتبية	سميرنوف - الإشارة
			كمية	اختبار (Z) - اختبار (t) لعينة واحدة
عينتان مستقلتان	الفروق بين المتوسطات (فروقي)	مجموعتان تجريبية	اسمية	اختبار كا ² - فشر - سميرنوف
			رتبية	الوسيط - مان ويتني - التتابع
			كمية	اختبار (t) لعينتين مستقلتين
عينتان مترابطتان	الفروق بين القياسات (فروقي)	مجموعة واحدة (ذات اختبارين قبلي وبعدي)	اسمية	اختبار ماكنمار
			رتبية	ولكوكسن - الإشارة
			كمية	اختبار (t) لعينتين مترابطتين
عدة عينات مستقلة	الفروق بين المجموعات (فروقي)	المجموعات المتعددة	اسمية	اختبار كا ²
			رتبية	الوسيط - كروسكال واليس
			كمية	تحليل التباين - تحليل التباين
عدة عينات مترابطة	الفروق بين القياسات (فروقي)	مجموعة واحدة ذات الاختبارات المتعددة	اسمية	كوجران
			رتبية	فريدمان
			كمية	تحليل التباين ذي القياسات المتكررة

جدول لاختبارات العلاقات الاحصائية:

عدد العينات	الفرض	التصميم التجريبي	نوع البيانات	الاختبار الاحصائي
عينة واحدة أو عينتان أو عدة عينات	الارتباط بين القياسات أو العلاقة بين المتغيرات (علائقي)	مجموعة واحدة ذات اختبار قبلي أو بعدي أو عدة اختبارات	اسمية	معامل ارتباط فاي - معامل التوافق - معامل الارتباط الرباعي
			رتبية	معامل ارتباط سبيرمان - معامل ارتباط كندال
			كمية	ارتباط بيرسون - الارتباط القانوني - الارتباط المتعدد
عينة واحدة أو عينتان أو عدة عينات	دراسة تنبؤية للمتغيرات أو عضوية الجماعة (تأثري)	مجموعة واحدة أو عدة مجموعات مع عدة اختبارات	كمية	تحليل السلاسل الزمنية - تحليل الانحدار بأنواعه المختلفة
عينة واحدة أو عينتان أو عدة عينات	دراسات عاملية "البناء العملي" (تأثري)	مجموعة واحدة أو عدة مجموعات مع عدة اختبارات	كمية	التحليل العاملي الاستكشافي - التحليل العاملي التوكيدي



الفصل الثاني

التوزيعات الاحتمالية المتصلة

(Continues Probability Distributions)

الفصل الثاني

التوزيعات الاحتمالية المتصلة

Continues Probability Distributions

المتغيرات العشوائية والتوزيعات الاحتمالية

المتغير (Variable): المتغير يعبر عن شيئاً يتغير فهو أي ظاهرة تأخذ قيماً مختلفة بين مفرداتها مثل الجنس، السن، الطول، التحصيل العلمي.

المتغيرات العشوائية (Random Variables): وهي المتغيرات المعبرة عن الظواهر العشوائية أو التي تعبر عن نتائج التجربة العشوائية، حيث يتوقع العديد من القيم المتغيرة وفق نظام احتمالي معين يختلف من ظاهرة إلى أخرى.

مثل نتيجة رمي حجر النرد، فنتيجة الرمية تكون متغيراً عشوائياً ونتائجها هي $\{1:2:3:4:5:6\}$ لأنها تتغير من محاولة إلى أخرى (فهي ليست شيئاً ثابتاً)، وهذا التغير يكون عشوائياً لأنه يخضع في تغيره لعوامل العشوائية أو الصدفة البحتة، حيث لا يستطيع أحد أن يحدد نتيجة الرمية مسبقاً.

كذلك دخل دولة ما هو متغير عشوائي، لأنه يتغير من سنة لأخرى طبقاً لعوامل كثيرة مثل الطلب على الصادرات، وقيمة العملة، وإمكانية وجود بدائل لصادراتها... وهكذا، وتنقسم المتغيرات العشوائية إلى قسمين.

1- المتغيرات العشوائية المنفصلة

2- المتغيرات العشوائية المتصلة (المستمرة)

(1) المتغيرات العشوائية المتقطعة (Discrete Random Variables)

إذا كان المتغير العشوائي يأخذ عدداً من القيم المنفصلة عن بعضها أو المتميزة distinct أو المختلفة فيقال أن المتغير العشوائي متقطع، وذلك كما في مثال (نتيجة حجر النرد) نجد أن المتغير يأخذ ست قيم فقط وهي 1, 2, 3, 4, 5, 6. وفي مثال قطعة النقود نجد أن المتغير يأخذ ثلاث قيم فقط هي 0, 1, 2. بمعنى أنه يمكن حساب احتمال أن يأخذ المتغير العشوائي المتقطع أي قيمة.

(2) المتغيرات العشوائية المتصلة (Continuous Random Variables)

أما إذا كان المتغير العشوائي يأخذ عدداً غير محدود أو لا نهائي من القيم داخل نطاق تغيره فيكون المتغير العشوائي متصلاً. فمثلاً عند قياس أطوال الطلاب بالجامعة نقول أن أطوالهم تتراوح بين 150 – 200 سنتيمتر وبنفس الطريقة يمكن القول بأن دخل الفرد السنوي في دولة ما يتراوح ما بين 150 – 300 دولاراً شهرياً.

وفي المتغير المتصل لا يمكن حساب احتمال أن يأخذ المتغير قيمة معينة حيث أن هذا الاحتمال دائماً يساوي صفر. لأن هناك عدد لا نهائي من القيم فكيف يمكن حساب احتمال أن يأخذ المتغير إحدى هذه القيم؟ بمعنى أن دخل الفرد الشهري – مثلاً – يأخذ عدداً لا محدوداً من القيم ما بين الحد الأعلى والحد الأدنى لهذا المتغير. فما يمكن حسابه في المتغير المتصل هو احتمال أن تتراوح قيمته بين قيمتين، أو أن يكون أكبر من (أو أقل من) قيمة معينة. فمثلاً لا نستطيع حساب احتمال أن يكون دخل دولة ما هو ألف مليون دولار سنوياً، ولكننا يمكن أن نقول أن احتمال أن يكون هذا الدخل أكبر من، أو أقل من ذلك المبلغ، أو أنه يتراوح – مثلاً – ما بين ثمانمائة إلى ألف مليون دولار، لأن الدخل – وكما سبق وأن ذكرنا هو متغير متصل.

وتحسب الاحتمالات للمتغير المتصل باستخدام المساحات تحت منحنى الدالة أو المتغير (مثل المنحنى الطبيعي) وهو ما سوف نبينه لاحقاً.

التوزيع الاحتمالي (أو الإحصائي) Statistical Distribution

التوزيع الإحصائي: هو الشكل الذي تأخذه مجموعة البيانات، وشكل البيانات مهم جداً في تحليلها ووصفها كخطوة تسبق قرار استخدام أي أسلوب إحصائي.

ويرتبط التوزيع الإحصائي عادة بنوعين من البيانات المتصلة والمنفصلة، أي التوزيعات المنفصلة تناسب مقاييس البيانات الاسمية والرتبية، هناك بعض المقاييس المنفصلة ثنائية يوجد لها فقط قيمتين (نجاح أو فشل) هذه لا تسمى توزيعات طبيعية وإنما تسمى توزيعات ثنائية (توزيع ذو الحدين).

أما التوزيعات الإحصائية المتصلة فهي ذات أهمية كبيرة في العلوم الإحصائية وذلك لأن أغلب الاختبارات الإحصائية تتعامل مع هذا النوع من البيانات.

التوزيع الاحتمالي قد يأخذ شكل جدول أو دالة رياضية أو معادلة يبين كيف تتوزع الاحتمالات (والتي مجموعها الواحد الصحيح) على قيم المتغير المختلفة. أي أن التوزيع الاحتمالي (أو الإحصائي) والذي يسمى أيضاً دالة الاحتمال (Probability function) هو التوزيع أو الدالة التي تعطي احتمالات أن يأخذ المتغير العشوائي القيم المختلفة له.

فإذا كان المتغير العشوائي محل الدراسة متغيراً متقطعاً مثل عدد مرات انتخاب عضو ما، أو حجم الأسرة، كان التوزيع الاحتمالي له متقطعاً، أما إذا كان المتغير العشوائي متصلاً مثل دخل الدولة مثلاً، كان التوزيع الاحتمالي له متصلاً.

وهناك مجموعتان من التوزيعات الاحتمالية :

الأولى: التوزيعات الاحتمالية المنفصلة (Discrete Distributions):

الثانية: التوزيعات الاحتمالية المتصلة (Continuous Distributions):

التوزيعات الاحتمالية للمتغيرات المتصلة

هنا سوف نقوم بدراسة التوزيعات الاحتمالية المتصلة لأن أغلب الاختبارات الإحصائية تعتمد على هذا النوع من البيانات والتوزيعات، وسوف تتم الدراسة وفقاً للنقاط التالية:

تعريف المتغير العشوائي المستمر

التوزيع الاحتمالي للمتغير العشوائي المستمر

خصائص دالة الكثافة الاحتمالية للمتغير العشوائي المستمر

التمثيل البياني لدالة الكثافة الاحتمالية للمتغير العشوائي المستمر

دالة التوزيع للمتغير العشوائي المستمر

تعريف المتغير العشوائي المتصل (Continuous Random Variable)

هو متغير يأخذ قيماً متصلة وله عدد لا نهائي من القيم الممكنة داخل مجال محدود، فإذا كان X متغير عشوائي مستمر ويقع في المجال (a, b) أي أن $\{X = x : a < x < b\}$ فإن للمتغير X عدد لا نهائي من القيم تقع في المجال (a, b) . وتكون وحدات قياس هذا المتغير مستمرة مثل الزمن، الوزن، المسافة، الحجم، ... أمثلة:

- أطوال الطلاب بالجامعة تتراوح بين 150 – 200 سنتيمتر $\{X = x : 150 < x < 200\}$
- دخل الفرد السنوي في دولة ما بالدولار $\{X = x : 2000 < x < 3500\}$
- كمية الألبان التي تنتجها البقرة في اليوم باللتر $\{X = x : 10 < x < 40\}$
- فترة صلاحية حفظ الدجاج المبرد بالأيام $\{X = x : 1 < x < 5\}$
- وزن الجسم بالكيلو غرام للأعمار من 20-30 سنة $\{X = x : 55 < x < 80\}$

وفي المتغير المتصل لا يمكن حساب احتمال أن يأخذ المتغير قيمة معينة حيث أن هذا الاحتمال دائماً يساوي صفر لأن هناك عدد لا نهائي من القيم فكيف يمكن حساب احتمال أن يأخذ المتغير إحدى هذه القيم ؟ بمعنى أن احتمال أن يكون طول الطالب 162.35 سم هو 0 لأن متغير الطول يأخذ قيماً لانهائية ما بين الحد الأدنى والاعلى في المجال (150, 200).

والذي يمكن حسابه في المتغير المتصل هو احتمال أن تتراوح قيمته بين قيمتين، أو أن يكون أكبر من (أو أقل من) قيمة معينة. فمثلاً يمكن أن نقول أن احتمال أن يكون طول الطالب أقل من 180 سم كذا أو يتراوح بين 170 و 190 كذا.

وتحسب الاحتمالات للمتغير المتصل باستخدام المساحات تحت منحنى الدالة أو المتغير (مثل المنحنى الطبيعي) وهو ما سوف نبينه لاحقاً.

التوزيع الاحتمالي للمتغير العشوائي المستمر

وفي المقابل فإن التوزيعات الاحتمالية المتصلة لا يمكن أن توضع في جدول، وذلك لأن المتغير يأخذ عدداً لا نهائياً من القيم وبالتالي لا يمكن حصره، وإنما يعبر عنها دائماً بدوال رياضية أو برسم منحنيات هذه الدوال أو التوزيعات.

ويعبر عن الاحتمالات بدلالة المساحات تحت هذه المنحنيات. وقد تم حساب الاحتمالات المختلفة لأهم التوزيعات المتصلة وتم وضعها في جداول إحصائية ثابتة يمكن استخدامها في أي وقت وفي أي مكان حيث توفر من الجهد والوقت في حساب هذه الاحتمالات، حيث يمكن الكشف في هذه الجداول في ثوان معدودة والحصول على الاحتمال المطلوب.

خصائص دالة الكثافة الاحتمالية للمتغير العشوائي المستمر

$$1) f(x) \geq 0$$

دالة الكثافة الاحتمالية للمتغير المستمر تكتب كما يلي :

$$2) \int_{-\infty}^{+\infty} f(x)dx = 1$$

من البديهي إذا أن منحنى دالة الكثافة لا يمكن أن ينزل أسفل محور الم ع، كما أن المساحة الإجمالية بين المنحنى والمحور الأفقي تساوي الواحد.

هذه الخصائص تقيدنا في حساب احتمالات بعض المجالات من خلال احتمالات مجالات أخرى.

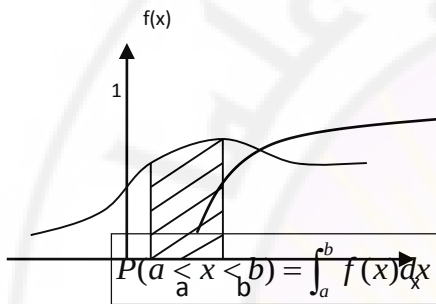
التمثيل البياني لدالة الكثافة للمتغيرة العشوائية المستمرة

لاحظ أنه بما أن X تأخذ عددا لا متناهيا من القيم داخل أي مجال، فإن احتمال قيمة بعينها هو احتمال يؤول

إلى الصفر. $P(X=x) \rightarrow 0$. لذلك فإن دالة الكثافة تستعمل لحساب

احتمال مجال. ويكون ذلك بحساب المساحة تحت منحنى $f(x)$ بين

حدود المجال.



التمثيل البياني لدالة الكثافة للمتغيرة العشوائية المستمرة

لاحظ أن إشارة التكامل هنا تقابل إشارة المجموع في حالة المتغيرات المتقطعة.

دالة التوزيع $F(x)$ للمتغير العشوائي المستمر

تعرف دالة التوزيع للمتغير المستمر كما يلي:

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(u) du$$

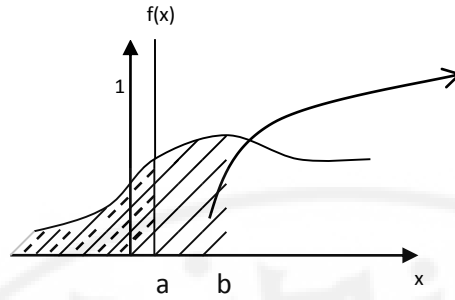
لدالة التوزيع أهمية أكبر بالنسبة للمتغير المستمر. السبب في ذلك أننا نهتم، في حالة المتغير المستمر، باحتمال

مجال وليس باحتمال نقطة، ولحساب احتمال مجال من الأيسر التعويض في دالة التوزيع بدلا من حساب

التكامل في كل مرة. يتضح ذلك من القاعدة التالية: بفرض a, b نقطتان من مجال تعريف X ، بحيث $b > a$.

لحساب احتمال أن تكون X تنتمي إلى المجال $[a, b]$:

$$P(a < x \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a)$$



حساب الاحتمال من خلال دالة التوزيع

أهم التوزيعات الاحتمالية المتصلة التي تستخدم فيها الجداول الإحصائية الجاهزة لحساب الاحتمالات المختلفة هي :

التوزيع الطبيعي	The normal distribution
التوزيع الطبيعي (القياسي) المعياري (أو توزيع Z)	
توزيع " ت " أو t	t - dist
توزيع " كاي 2 " أو X^2 أو مربع كاي	X^2 - dist
توزيع " ف " أو F	F - dist

وفيما يلي نتناول التوزيع الطبيعي والطبيعي المعياري بشيء من التفصيل. أما توزيعي t و X^2 و F

فسوف يأتي الحديث عنهما فيما بعد.

أولاً: التوزيع الطبيعي للمتغير العشوائي المستمر (Normal Distribution)

يعتبر التوزيع الطبيعي للمتغيرات العشوائية المستمرة من أهم التوزيعات الاحتمالية في علم الإحصاء لأنه يصف كثيراً من الظواهر التي تقابلنا في الحياة العملية مثل الأطوال و الأوزان و الأعمار و درجات الحرارة والدخول الشهرية ... و غيرها من الظواهر المتصلة .

وهذا التوزيع يعتبر أحد الأعمدة الأساسية لعلم الإحصاء وذلك لعدة أسباب منها:

1- لما له من تطبيقات كثيرة ومتنوعة كتوزيع احتمالي لكثير من الظواهر الطبيعية مثل الذكاء، وأطوال

الأشخاص وأوزانهم والظواهر الطبيعية (أو غير الطبيعية) الأخرى.

2- يعتبر التوزيع الطبيعي حجر الزاوية في الاستدلال الإحصائي ولاعتماد الكثير من الاختبارات الإحصائية والاستدلال الإحصائي بصفة عامة عليه.

3- وأيضاً لاشتقاق عدد من التوزيعات المهمة - مثل توزيعات t ، X^2 ، F من التوزيع الطبيعي.

معالم التوزيع الطبيعي:

يعتمد التوزيع الطبيعي على معلمتين هما:

$$1- \text{المتوسط } \mu = \text{Mean}$$

وهو يحدد موضع التوزيع (موضع تركيز قيم التوزيع على خط الأعداد الحقيقية).

$$2- \text{التباين } (\sigma^2) = \text{Variance}$$

وهو يقيس مدى تباعد أو تقارب (تذبذب) قيم التوزيع حول المتوسط. فكلما كانت قيمة التباين كبيرة، كلما كانت قيم التوزيع متباعدة عن المتوسط، والعكس بالعكس

دالة كثافة الاحتمال Probability Density Function

إذا كان X متغيراً عشوائياً متصلاً يتبع التوزيع الطبيعي بمتوسط μ وانحراف معياري σ فإن دالة كثافة الاحتمال للمتغير العشوائي X تعطى بالعلاقة:

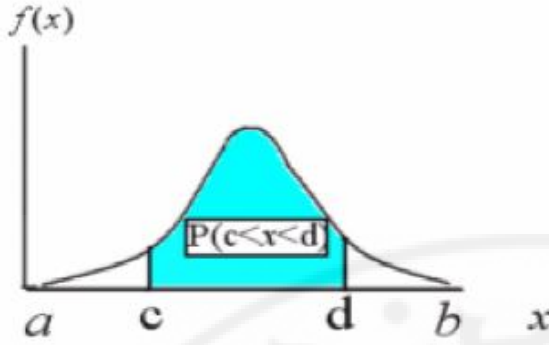
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} ; -\infty \leq x \leq \infty$$

فإنه يقال إن x تتبع توزيعاً طبيعياً متوسطه μ و تباينه σ^2 . و تكتب اختصاراً

$$X \sim N(\mu, \sigma^2)$$

ومن أجل حساب احتمال أن المتغير العشوائي X يقع في المجال (c, d) أي حساب $P(c \leq X \leq d)$

هذا يتطلب حساب المساحة أسفل المنحني من $x=c$ حتى $x=d$ كما هي مبينة بالشكل.



$$P(c \leq X \leq d) = \int_{x=c}^{x=d} f(x) dx$$

ودائماً في المتغيرات المستمرة يكون

$$P(X = \text{value}) = 0 \quad \text{الاحتمال}$$

مساوياً للصفر دائماً.

خصائص التوزيع الطبيعي:

يتميز التوزيع الطبيعي بالخصائص الآتية:

1- مقاييس النزعة المركزية الثلاثة (الوسط الحسابي والوسيط والمنوال) متساوية و تلتقى في نقطة واحدة

(في منتصف التوزيع μ)

2- المنحنى له نقطتا انقلاب عند $X = \mu \pm \sigma$

3- منحنى التوزيع الطبيعي منحنى ناقوسي الشكل و متماثل حول متوسطه، أي حول العمود الذي يمر

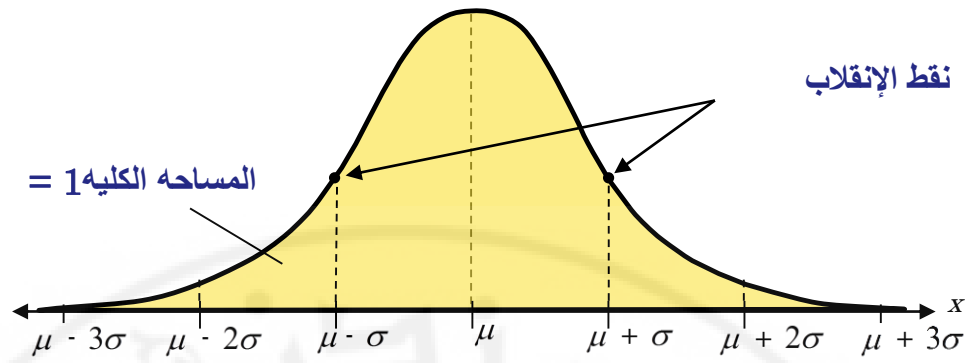
بقمته أي عند $X = \mu$ لذلك فإن هذا العمود يجزئ المنحنى الطبيعي القياسي إلى قسمين متماثلين في

الشكل والمساحة.

4- طرفا التوزيع تمتد من $-\infty$ إلى ∞

5- المساحة الكلية تحت منحنى التوزيع الطبيعي تساوى الواحد الصحيح وهذه المساحة تعبر عن مجموع

الاحتمالات.



المساحات تحت المنحنى الطبيعي:

بالنسبة للتوزيعات المتصلة فإنه تستخدم المساحات تحت منحنى التوزيع أو الدالة للحصول على

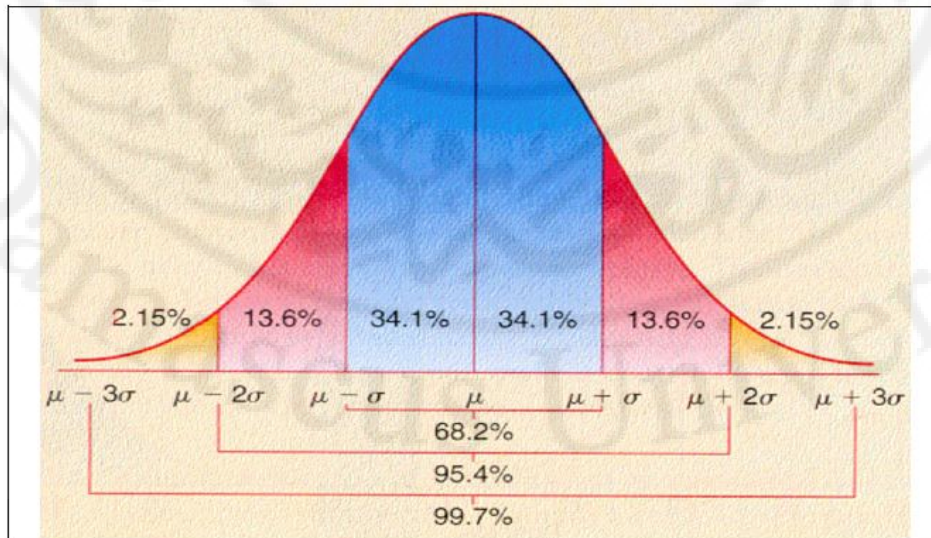
الاحتمالات المختلفة. وبالنسبة للتوزيع الطبيعي فإن المساحات تحت المنحنى تكون كما يلي:

68.26 % من إجمالي المساحة تحت المنحنى تكون محصورة بين $\mu - \sigma$, $\mu + \sigma$

95.44 % من إجمالي المساحة تحت المنحنى تكون محصورة بين $\mu - 2\sigma$, $\mu + 2\sigma$

99.72 % من إجمالي المساحة تحت المنحنى تكون محصورة بين $\mu - 3\sigma$, $\mu + 3\sigma$

المساحة الكلية تحت منحنى التوزيع = 1



هذا ويمكن التعبير عما سبق رمزيا كالآتي :

$$p(\mu - 1\sigma < x < \mu + 1\sigma) = 0.6826$$

$$p(\mu - 2\sigma < x < \mu + 2\sigma) = 0.9544$$

$$p(\mu - 3\sigma < x < \mu + 3\sigma) = 0.9974$$

وحيث أن استخدام هذه الدالة السابقة لحساب الاحتمالات المختلفة تكنتفه صعوبات رياضية كثيرة تعتمد على

معرفة تامة بعلم التكامل، فإنه لحسن الحظ يمكن تحويل أي توزيع طبيعي إلى توزيع طبيعي معياري له جداول

خاصة تعرف باسم جداول المساحات تحت منحنى التوزيع الطبيعي المعياري متى علم متوسطه وانحرافه المعياري

التوزيع الطبيعي المعياري (Standard Normal Distribution)

- التوزيع الطبيعي المعياري هو التوزيع الطبيعي متوسطه $\mu = 0$ وتباينه $\sigma^2 = 1$.
- يرمز للتوزيع الطبيعي المعياري بالرمز $N(0,1)$.
- يرمز للمتغير العشوائي الذي يتوزع طبيعياً معيارياً بالرمز Z ، ونكتب $Z \sim N(0,1)$.
- أي قيمه من قيم المتغير X يمكن تحويلها إلى المتغير Z الذي يتبع التوزيع الطبيعي القياسي باستخدام

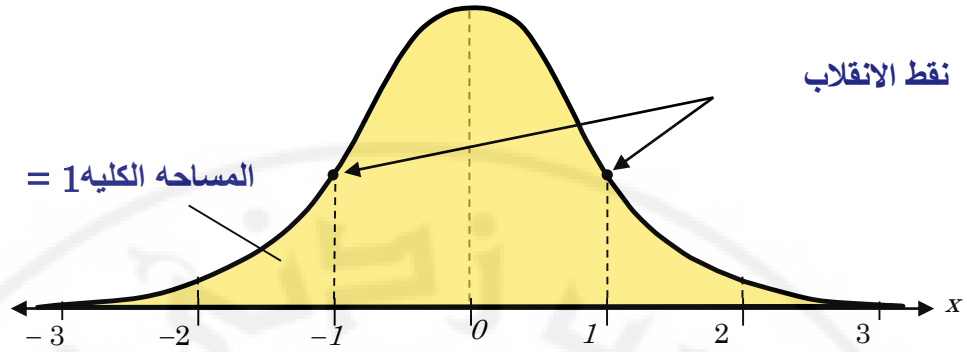
$$Z = \frac{X - \mu}{\sigma} \text{ :التحويلة}$$

ويمكن رسم المنحنى الاحتمالي للتوزيع الطبيعي القياسي باستخدام المعادلة:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \quad -\infty \leq z \leq \infty.$$

المنحنى الاحتمالي للتوزيع الطبيعي القياسي

الشكل الاتي يوضح المنحنى الاحتمالي للتوزيع الطبيعي القياسي.



وبالنسبة للتوزيع الطبيعي المعياري فإن المساحات تحت المنحنى تكون كما يلي:

68.26 % من إجمالي المساحة تحت المنحنى تكون محصورة بين -1, +1

95.44 % من إجمالي المساحة تحت المنحنى تكون محصورة بين -2, +2

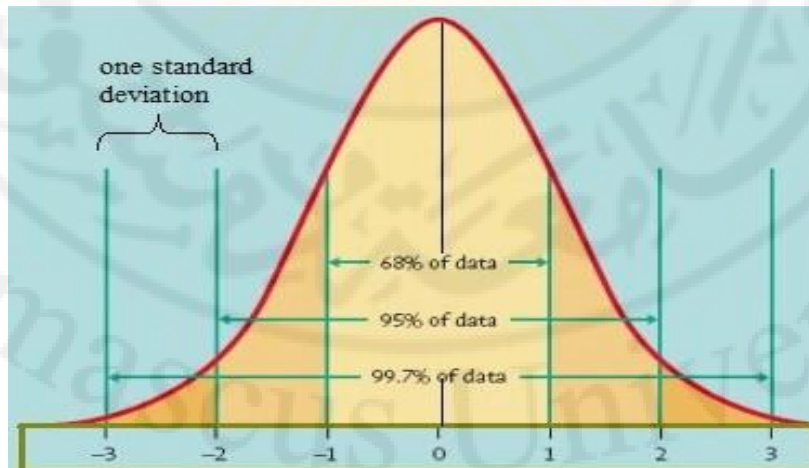
99.72 % من إجمالي المساحة تحت المنحنى تكون محصورة بين -3, +3

هذا ويمكن التعبير عما سبق رمزيا كالاتي :

$$P (-1 < Z < 1) = 0.6826$$

$$P (-2 < Z < 2) = 0.9544$$

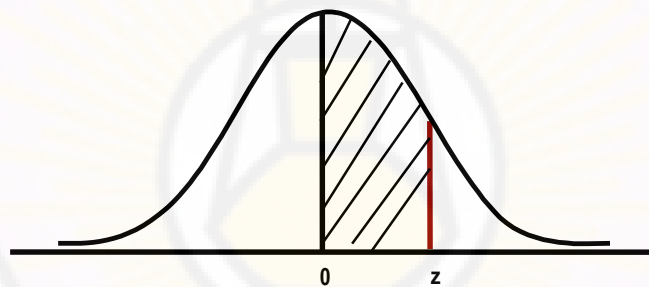
$$P ((-3 < Z < 3) = 0.9974$$



ويمكن حساب احتمال وقوع Z في أي فترة تريدها وذلك بإيجاد قيمه تكامل داله الكثافة الاحتمالية في هذه الفترة أي بإيجاد المساحة المحصورة تحت منحنى هذه الدالة داخل هذه الفترة.

ولأن متوسط التوزيع الطبيعي القياسي دائما يساوى صفر وتباينه يساوى واحد فقد تمكن الإحصائيون من حساب المساحات (أو الاحتمالات) تحت المنحنى الطبيعي المعياري والتي تقابل Z ووضعها في جدول يعطى احتمالات وقوع المتغير Z في فترة معينة. ونلاحظ أن الجدول يعطي قيم Z الموجبة حيث هي نفسها إذا كانت Z بالسالب وذلك لتماثل المنحنى

فمثلا يمكن بواسطة الجدول حساب الاحتمال $P(1 \leq Z \leq 2)$. والجدول يعطى احتمال وقوع Z بين الصفر وأي قيمه موجبه Z أي يعطى $P(0 \leq Z \leq z)$ وهي المساحة المظلة في الشكل الآتي:



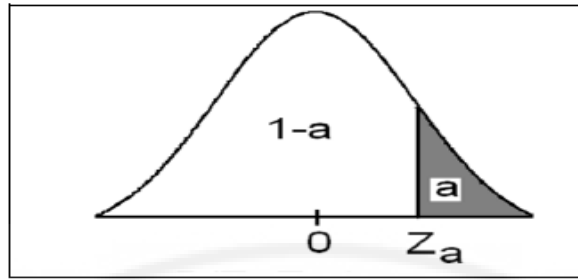
ملاحظات:

$$1. P(-\infty \leq Z \leq \infty) = 1$$

$$2. P(Z \leq 0) = P(Z \geq 0) = 0.5$$

القيم الجدولية للتوزيع الطبيعي المعياري (Z - Tabulated Value)

نرمز لقيمة المتغير العشوائي (Z) الذي يتوزع وفق الطبيعي المعياري والتي يليها (على يمينها) مساحة مقدارها a ويسبقها (على يسارها) مساحة مقدارها $(1-a)$ بالرمز Z_a : والشكل التالي يوضح هذا الترميز.



بسبب تماثل منحنى التوزيع الطبيعي المعياري حول الصفر فإن: $Z_a = Z_{1-a}$

كيفية استخدام جدول توزيع الاحتمالات المتجمعة للمتغير العشوائي Z

مثال (1): أوجد احتمال أن Z أقل من ($<$) 1.64

الحل: الجدول التالي جزء من جدول الاحتمالات المتجمعة للتوزيع المعتدل المعياري

Z	.00	.01	.02	.03	.04	.05
1.6					0.9495	

قواعد وامثلة لحساب قيم (Z) الجدولية في حالات مختلفة:

فالإحتمال المتجمع المناظر للقيمة 1.64 يوجد أمام الصف 1.6 وتحت العمود 0.04 (لاحظ أن $0.04 + 1.6$

$= 1.64$) وهي قيمة Z المطلوب إيجاد الاحتمال المتجمع عندها، وهذا الإحتمال هو 0.9495 ، أي أن

$P(Z < 1.64) = 0.9495$ وهذا هو الإحتمال المتجمع للمتغير Z من ($-$) إلى 1.64

يوجد جداول احصائية للتوزيع الطبيعي المعياري لقيم (Z) الموجبة وجداول أخرى لحساب قيم (Z) السالبة.

الجدول الاتي يعرض اهم حالات استخراج قيم (Z) الجدولية في حالات مختلفة مع امثلة:

$$P(Z > -0.8) = 1 - \Phi(-0.8) = 1 - 0.2119 = 0.7881$$

القاعدة	مثال
---------	------

1-	$P(Z < 0) = \Phi(0)=0.5$	$P(Z > 0) = 1 - \Phi(0)=1-0.5=0.5$
2-	$P(Z < a) = \Phi(a)$	$P(Z < 1.5) = \Phi(1.5)=0.9332$
3-	$P(Z < -a) = \Phi(-a)$	$P(Z < -1.2) = \Phi(-1.2)=0.1151$
4-	$P(a < Z < b) = \Phi(b)-\Phi(a)$	$P(1 < Z < 2.4) = \Phi(2.4)-\Phi(1)=0.9918-0.8413$
5-	$P(Z > a) = 1 - \Phi(a)$	$P(Z > 0.8) = 1 - \Phi(0.8)=1-0.7881=0.2119$
6-	$P(Z > -a) = 1 - \Phi(-a)$	$P(Z > -0.8) = 1 - \Phi(-0.8)=1-0.2119=0.7881$

مثال (2): إذا كان Z متغير عشوائي متصلًا يتبع التوزيع الطبيعي القياسي فأوجد

1. $P(Z \leq 1.54)$

2. $P(-1.8 \leq Z \leq 0)$

3. $P(1 \leq Z \leq 2)$

الحل:

1. $P(Z \leq 1.54) = 0.9382$

2. $P(-1.8 \leq Z \leq 0)$

$P(-1.8 \leq Z \leq 0) = \Phi(0) - \Phi(-1.8) = 0.5 - 0.359 = 0.4641$

3. $P(1 \leq Z \leq 2)$

$P(1 \leq Z \leq 2) = \Phi(2) - \Phi(1) = 0.9772 - 0.8413 = 0.1359$

حساب الاحتمالات في حالة التوزيع الطبيعي العادي

إذا كانت X تتبع توزيعاً طبيعياً متوسطه μ و تباينه σ^2 و أردنا حساب أي احتمال حول المتغير X فإننا نحوله أولاً إلى متغير يتبع التوزيع الطبيعي القياسي

إذا كانت $X \sim N(\mu, \sigma^2)$ فإن $Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$

و على ذلك يمكن حساب الاحتمال $P(a < X < b)$ كما يلي :

$$P(a \leq X \leq b) = P\left(\frac{a - \mu}{\sigma} \leq \frac{X - \mu}{\sigma} \leq \frac{b - \mu}{\sigma}\right)$$

$$= P\left(\frac{a - \mu}{\sigma} \leq Z \leq \frac{b - \mu}{\sigma}\right)$$

و هذا الاحتمال الأخير يمكن الحصول عليه من جدول التوزيع الطبيعي القياسي.

مثال (3): إذا كانت أطوال مجموعه من النباتات تتبع توزيعاً طبيعياً متوسطه 168 سم و انحرافه المعياري 6

سم. أخذنا عشوائياً أحد النباتات. ما هو احتمال أن يكون طوله:

1. أقل من 159 سم؟
2. أكبر من 180 سم؟
3. واقعاً في الفترة (165 , 174) ؟

الحل:

1- أقل من 159 سم

إذا جعلنا X ترمز لأطوال النباتات, فإن X تكون متغيراً عشوائياً يتبع توزيعاً طبيعياً متوسطه 168 سم وانحرافه المعياري 6 سم.

$$\begin{aligned} P(X \leq 159) &= P\left(\frac{x - \mu}{\sigma} \leq \frac{159 - 168}{6}\right) \\ &= P(Z \leq -1.5) = \Phi(1.5) = 0.0668 \end{aligned}$$

$$\begin{aligned} P(X \geq 180) &= P\left(\frac{x - \mu}{\sigma} \geq \frac{180 - 168}{6}\right) \\ &= P(Z \geq 2) = 1 - \Phi(2) = 1 - 0.9772 = 0.0228 \end{aligned}$$

2- واقعاً في الفترة (165 , 174) ؟

$$\begin{aligned} P(165 \leq X \leq 174) &= P\left(\frac{165 - 168}{6} \leq \frac{x - \mu}{\sigma} \leq \frac{174 - 168}{6}\right) \\ &= P(-0.5 \leq Z \leq 1) = \Phi(1) - \Phi(-0.5) = 0.8413 - 0.3085 = 0.5328 \end{aligned}$$

مثال (4): وجد أستاذ مادة الإحصاء أن متوسط الوقت الكافي الذي يحتاجه الطلاب لإكمال امتحانهم النهائي =

150 دقيقة بانحراف معياري قدره 30 دقيقة. المطلوب:

1. ما احتمال أن يكمل الطلاب امتحانهم بين 125 و 150 دقيقة.

2. ما احتمال أن يكمل الطلاب اختبارهم في 185 دقيقة أو أقل

3. ما احتمال أن يكمل الطلاب اختبارهم في وقت يزيد على 195 دقيقة

4. إذا كان عدد الطلاب 1000 طالب . أوجد عدد الطلاب الذين اكملوا امتحانهم في وقت يزيد على

185 دقيقة ؟

الحل :

1. احتمال أن يكمل الطلاب امتحانهم بين 125 و 150 دقيقة

$$\begin{aligned} P(125 \leq X \leq 150) \\ P(125 \leq X \leq 150) &= P\left(\frac{125-150}{30} \leq \frac{x-\mu}{\sigma} \leq \frac{150-150}{30}\right) \\ &= P(-0.83 \leq Z \leq 0) = \Phi(0) - \Phi(-0.83) = 0.5000 - 0.2033 = 0.2967 \end{aligned}$$

وعليه نقرر أن احتمال إكمال الطلاب لامتحانهم في فترة زمنية تتراوح بين 125 و 150 دقيقة = 29.7 % .

2- احتمال أن يكمل الطلاب اختبارهم في 185 دقيقة أو أقل

$$\begin{aligned} P(X \leq 185) &= P\left(\frac{x-\mu}{\sigma} \leq \frac{185-150}{30}\right) \\ &= P(Z \leq 1.17) = \Phi(1.17) = 0.8790 \end{aligned}$$

وعليه نقرر أن احتمال إكمال الطلاب لامتحانهم في فترة زمنية 185 دقيقة أو أقل = 87.9 %

3- ما احتمال أن يكمل الطلاب اختبارهم في وقت يزيد على 195 دقيقة.

أولاً: تحويل المسافة بين المتوسط 150 والدرجة الخام 195 إلى درجه معياريه باستخدام طريقة تحويل (Z).

$$\begin{aligned} P(X \geq 185) &= P\left(\frac{x-\mu}{\sigma} \geq \frac{195-150}{30}\right) \\ &= P(Z \geq 1.5) = 1 - \Phi(1.5) = 1 - 0.9332 = 0.0668 \end{aligned}$$

أي أن احتمال أن يكمل الطلاب اختيارهم في وقت يزيد على 195 دقيقة = 6.7 % تقريباً.

4- عدد الطلاب الذين اكملوا امتحانهم في وقت يزيد على 185 دقيقة :

$$\begin{array}{l} \text{عدد الطلاب الكلي} \\ \times P(X \geq 185) \\ 1000 \times 0.0668 = 66.8 \approx 66 \\ \text{طالب} \end{array}$$

الفصل الثالث

اختبار (T-Test)

(Student T-Distribution)

اختبار (T-Test)

يوجد العديد من التوزيعات الاحتمالية (الاختبارات الاحصائية) المستمرة أهمها:

1- اختبار التوزيع الطبيعي المعياري (Z)

2- اختبار توزيع ستودنت (t)

3- اختبار تحليل التباين (F)

4- اختبار الكاي مربع (X^2)

لقد تناولنا في الفصول السابقة التوزيع الطبيعي والطبيعي المعياري، أما في هذا الفصل سوف نقوم بدراسة توزيع

(t)

يعد اختبار (t) أحد أهم الاختبارات الاحصائية وأكثرها استخداماً في الأبحاث العملية والدراسات التربوية والنفسية.

اختبار توزيع ستودنت (t) إن هذا الإختبار يشبه اختبار التوزيع الطبيعي من حيث كونه متماثلاً حول

متوسطه الحسابي الذي يساوي الصفر و أحادي المنوال ، ولكن شكل التوزيع يختلف باختلاف حجم، لذلك

توجد عائلة من توزيعات (t) بحسب حجم العينة، وهذه التوزيعات تكون أكثر انبساطاً من التوزيع الطبيعي، مما يتطلب أن تكون قيم إختبار (t) الحرجة عند مستوى دلالة معين أكبر من القيم الحرجة للدرجات المعيارية المناظرة لها و بخاصة في حالة العينات الصغيرة.

تعريف إختبار (t): هو من أهم الاختبارات الاحصائية المعلمية والذي يستخدم في مقارنة متوسط عينة بقيمة مفترضة لمتوسط مجتمع، أو يستخدم في مقارنة متوسطي عينتين مستقلتين أو غير مستقلتين (مرتبطتين). إن الفضل في اكتشاف هذا الاختبار إلى العالم الأيرلندي (وليم كوست W.S Gosset) عام 1908 حيث نشر بحثاً باسم مستعار هو Student وعُرف هذا التوزيع بالاسم (Student T-Distribution) ويختصر باسم توزيع (t)، والذي توصل من خلاله إلى اشتقاق معادلة التوزيع الاحتمالي (t) بالعلاقة التالية:

$$t = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}}$$

4-1- استخدامات إختبار (T):

إن لإختبار (t) استخدامات كثيرة نذكر منها:

- 1) إختبار الفروق المعنوية بين متوسط عينة بمتوسط مجتمع.
- 2) إختبار الفروق المعنوية بين متوسطي عينتين من مجتمعين طبيعيين مجهولي التباين.
- 3) إختبار معنوية معامل الارتباط البسيط.
- 4) إختبار معنوية معاملات الانحدار البسيط والمتعدد.
- 5) تقدير فترات الثقة لمتوسط مجتمع طبيعي مجهول التباين

4-2- أنواع إختبار (t):

يوجد ثلاثة أنواع لإختبار (T) هي:

- 1) إختبار (T) لعينة واحدة (One sample T-Test).

2) اختبار (T) لعينتين مستقلتين (Independent sample T-Test)

3) اختبار (T) لعينتين مرتبطتين (Paired sample T-Test)

4-3- شروط استخدام اختبار (t).

1- نوع البيانات (مستوى قياس البيانات): من أجل استخدام هذا الاختبار يشترط ان تكون بيانات العينات

كمية (فئوية او نسبية)، حيث لا يمكن تطبيق اختبار (t) للبيانات الأسمية أو الرتبية.

2- الانحراف المعياري للمجتمع (σ) (أو للمجتمعين) غير معروف (مجهول).

3- مدى اعتدالية التوزيع التكراري لكل من عيني البحث: أي أن بيانات كل عينة تتبع التوزيع الطبيعي،

هذا يتطلب أن تكون البيانات خالية من القيم الشاذة أو المتطرفة.

4- عشوائية العينات: يجب أن تكون العينتين قد تم اختيارهما بطريقة عشوائية.

5- حجم كل عينة: يجب أن يزيد حجم كل من العينتين عن (5): مع العلم أن اختبار (t) مخصص للعينات

الصغيرة (التي يقل حجمها عن 30) ولكن هذا لا يمنع استخدامه للعينات الكبيرة، أما إذا قل حجم أي

من العينتين عن (5) فلا يجوز استخدام اختبار (t)، حيث أنه يمكن في حالات العينات الصغيرة جداً

يستعاض عنه بالاختبارات اللامعلمية (اللا بارامترية) التي تصلح للتوزيعات الحرة غير المقيدة باعتدالية

التوزيع مثل (اختبار مان- ويتني) للعينتين المستقلتين واختبار (ويلكوكسون) للعينتين المرتبطتين.

6- أن يكون حجم عيني الدراسة متقارباً (شرط التقارب): أي الفرق بين حجم عيني البحث قليل، فلا يكون

مثلاً حجم أحد العينتين 600 فرد والأخرى 70 فرد، لأن لحجم العينة أثر على مستوى الدلالة.

7- مدى تجانس العينتين أو تجانس التباين في المجتمعين ($\sigma_1^2 = \sigma_2^2$):

يتوجب هذا الافتراض أن يكون تباين المشاهدات في العينة الأولى (S_1^2) لا يختلف عن تباين المشاهدات

في العينة الثانية (S_2^2) وفي هذه الحالة يقال أن العينتين متجانستان ($S_1^2 = S_2^2$) أي لا يوجد فرق

معنوي بين تباين العينتين المراد حساب الفرق بينهما، وهذا الشرط (تجانس التباين للعينتين) يمكن

التغاضي عنه إذا كانتا متساويتان بالحجم أي $(n_1=n_2)$.

ويقاس مدى التجانس بين العينتين بالنسبة الفائية أي بقسمة التباين الأكبر على التباين الأصغر

$$F = \frac{S_1^2 \text{ التباين الأكبر}}{S_2^2 \text{ التباين الأصغر}}$$

ويتحقق الفرض الصفري للتجانس بين العينتين عندما تكون قيمة $(F=1)$ أي $(S_1^2 = S_2^2)$ ويقاس مدى تباعد

قيمة (F) عن الفرض الصفري بالكشف عن دلالة (F) وذلك بحساب درجات الحرية للعينتين للحصول على قيمة (F) .

ولفحص التجانس نستخدم قانون هارتلي ف $(\text{Hartley's } F \text{ max})$:

$$f_{\max} = \frac{\text{larger variance}}{\text{smaller variance}}$$

حيث:

larger variance التباين الأكبر
smaller variance التباين الأصغر

ويتم مقارنة قيمة $(f_{\max})$ المحسوبة مع قيمة $(f_{\max})$ الجدولية لاتخاذ قرار حول تجانس التباين، وبالتالي استخدام اختبار "ت" لعينتين مستقلتين.

مثال (1): إذا كان لدينا عينتان:

العينة الأولى: حجم العينة الأولى $(n_1=76)$ درجات الحرية $(df_1=75)$ ، وتباينها $(S_1^2 = 16.25)$

العينة الثانية: حجم العينة الأولى $(n_2=81)$ درجات الحرية $(df_2=80)$ ، وتباينها $(S_2^2 = 12.18)$

إن النسبة الفائية:

$$F = \frac{S_1^2 \text{التباين الأكبر}}{S_2^2 \text{التباين الأصغر}} = \frac{16.25}{12.18} = 1.23$$

ثم نقوم باستخراج القيمة الجدولية لتوزيع (F) بمستوى معنوية ($\alpha=0.05$) بدرجات حرية ($v_1=75$) للبسط و

$$F_{\alpha}(v_1, v_2) = F_{0.05}(75, 80) = 1.45 \text{ نجد أن قيمة } (v_2=80)$$

وبما أن قيمة (F) المحسوبة في المثال تساوي 1.23 وهي أصغر من القيمة الجدولية إذاً نقبل الفرض العدمي

القائل بأن ($S_1^2 = S_2^2$) أي لا يوجد فرق معنوي بين تباين العينتين (متجانستان)، وبالتالي يمكن تطبيق اختبار

(t) لقياس دلالة الفروق بين متوسط العينتين.

4-4- دالة الكثافة الاحتمالية لتوزيع (t):

دالة الكثافة الاحتمالية لتوزيع (t) بدرجات حرية (ν) تحسب بالمعادلة التالية:

$$f(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi} \Gamma(\frac{\nu}{2})} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

حيث: Γ هي دالة غاما والتي تساوي $\Gamma(n) = (n-1)!$

من أجل حساب احتمال أن يقع (t) في مجال معين (a, b) ذلك يتوجب حساب تكامل دالة الكثافة الاحتمالية

على هذا المجال وهذا طبعاً صعب جداً رياضياً لذلك عمد الاحصائيون الى وضع جداول احصائية تعرض القيم

المختلفة لتوزيع (t) عند درجات حرية مختلفة ووفقاً لقيم معينة من مستوى المعنوية (α)

4-5- المتوسط الحسابي والتباين لتوزيع (t):

المتوسط الحسابي: أو التوقع للمتغير العشوائي (t) يساوي صفر لكل درجات الحرية أي $E(t) = 0$

التباين: للمتغير العشوائي (t) بدرجات حرية أكثر من (2) يساوي $Var(t) = \sigma^2 = \frac{\nu}{\nu-2}$

الانحراف المعياري للمتغير العشوائي (t) يقترب من الواحد الصحيح كلما زادت درجات الحرية بحيث اذا كانت حجم العينة (n=30) فإن الانحراف المعياري لتوزيع t يساوي 1.035 أو أقل اي

$$\sigma = \sqrt{\frac{\nu}{\nu-2}} = \sqrt{\frac{29}{29-2}} = 1.035$$

مثال (2): اوجد المتوسط الحسابي والانحراف المعياري للمتغير العشوائي (X) الذي يتبع توزيع (t) بـ درجات حرية (v=8) وتكتب كما يلي: $X \sim t(8)$

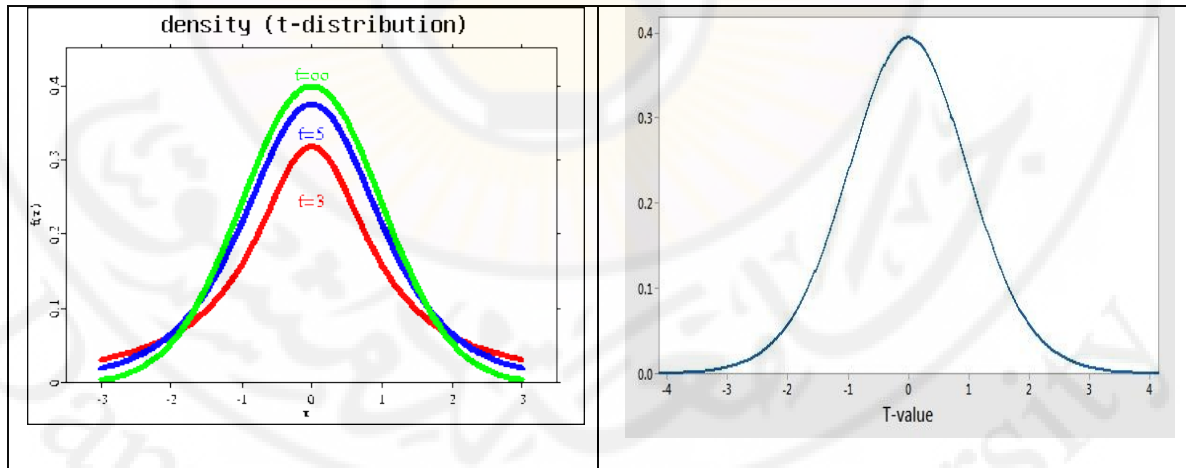
الحل: حسب خواص توزيع (t) ان المتوسط الحسابي لتوزيع t يساوي صفر

$$\sigma = \sqrt{\frac{\nu}{\nu-2}} = \sqrt{\frac{8}{8-2}} = 1.155$$

الانحراف المعياري

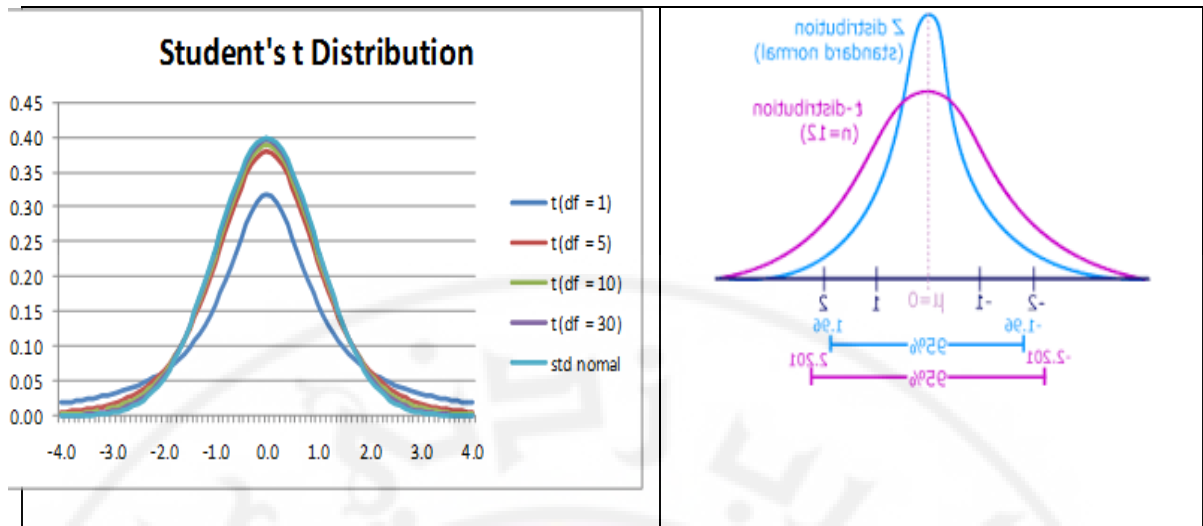
4-6- الشكل البياني لتوزيع (t):

نلاحظ من الشكل رقم (1) ان توزيع t متماثل حول العمود المقام من متوسطه (0) يقل تدريجياً كلما اتجهنا ناحيتي الذيلين الأيمن والأيسر، حيث تتحدد درجة حرية المتغير العشوائي t بـ (df=n-1).



الشكل (1): شكل توزيع (t)

ونلاحظ من الشكل السابق ان توزيع (t) يشبه توزيع (Z) فيما عدا أنه أكثر تشتتاً ويزداد تشتت توزيع (t) كلما قل حجم العينة (n)، أما إذا كانت (n) كبيرة فإن توزيع (t) يكون أقل تشتتاً وأكثر قرباً من شكل توزيع (Z)، وبتزايد درجات الحرية يقترب توزيع (t) من التوزيع الطبيعي المعياري (Z)، وهذا ما يوضحه الشكل:



4-7- خصائص توزيع (t):

- إن توزيع (t) هو توزيع احتمالي متصل أي يمكن حساب الاحتمالات بايجاد المساحات تحت منحنى التوزيع وكذلك $(-\infty < t < +\infty)$.
- يوجد عدد من توزيعات (t) أي لكل قيمة من درجات الحرية يوجد توزيع (t) مختلف.
- الشكل البياني لتوزيع (t) جرسى الشكل أي أن المنحنى متماثل حول العمود المقام من وسطه الحسابي (صفر) وأحادي المنوال (له قمة واحدة) وهو يشبه التوزيع الطبيعي المعياري.
- تحدد درجات الحرية شكل منحنى توزيع (t) أي كلما زادت درجات الحرية لتوزيع (t) يزداد ارتفاع المنحنى ويصبح أكثر تدبياً (أو أقل تفلطحاً) أي أقل تشتتاً ويقترب الانحراف المعياري لبيانات العينة (لتوزيع t) من من الواحد الصحيح، أي أن توزيع (t) اقترب من التوزيع الطبيعي المعياري بشكل كبير جداً.

4-8- الفرق بين توزيع t والتوزيع الطبيعي المعياري:

يمكن التفريق بين توزيع (t) والتوزيع الطبيعي المعياري (Z) بالنقاط التالية:

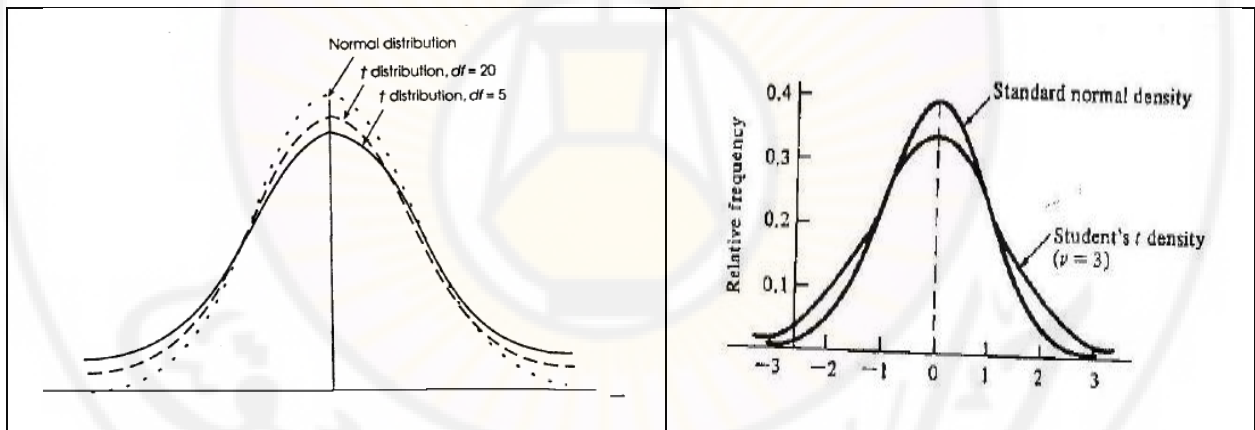
- 1) يوجد توزيع طبيعي معياري واحد لكل قيم المجتمع معالمه $(\mu=0, \sigma=1)$ ، بينما توزيع (t) هو توزيع لقيم العينة وكلما تغير حجم العينة وجد توزيع (t) جديد حسب حجم العينة.

(2) الانحراف المعياري للتوزيع الطبيعي المعياري هو الانحراف المعياري لقيم المجتمع بينما الانحراف المعياري لتوزيع (t) هو انحراف قيم العينة.

(3) إن توزيع (t) أكثر انبساطاً (تفلطحاً) من التوزيع الطبيعي المعياري وهذا الانبساط يقل ويقترب من التوزيع الطبيعي المعياري كلما زاد حجم العينة ويكادان يتطابقان التوزيعان عندما يكون حجم العينة $(n \geq 30)$.

(4) شكل التوزيع (t) يختلف باختلاف حجم العينة (n) لذلك توجد عائلة من توزيعات الاختبار (t) بحسب حجم العينة أو درجات الحرية $(df=n-1)$ ،

(5) أيضاً يختلف المتغير العشوائي (t) عن المتغير العشوائي الاعتدالي (X) (اي يتوزع وفقاً للتوزيع الطبيعي) بحيث يتحدد المتغير العشوائي الاعتدالي بمعلمتين هما الانحراف المعياري والمتوسط، بينما يتحدد المتغير العشوائي (t) بمعلم واحد فقط هو درجة الحرية (df).



اشتقاق المتغير العشوائي (T) من المتغير العشوائي الاعتدالي (X)

ومن أجل اشتقاق المتغير العشوائي (T) من المتغير العشوائي الاعتدالي (X)، فإن ذلك يتطلب معرفة قيمة المتوسط μ (للمجتمع الذي اخذت منه العينة) للمتغير العشوائي الاعتدالي، بينما لا نحتاج إلى معرفة انحرافه المعياري.

وتكون معادلة اشتقاق (التحويل) التوزيع الاحتمالي (T) كما يلي: $t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$

t: قيمة اختبار توزيع ستودينت

μ : قيمة متوسط المجتمع

$\bar{x}$: قيمة متوسط العينة

$\frac{s}{\sqrt{n}}$: الخطأ المعياري لمتوسط العينة

S: الانحراف المعياري لقيم العينة عن المتوسط الحسابي للعينة

طريقة استخراج قيمة توزيع (t) من جداول التوزيع الاحتمالي:

1- يتم حساب درجات الحرية (df) للتوزيع (T) حسب نوع الاختبار (عينة أم عينتين).

- اذا كانت عينة واحدة حجمها (n) تكون درجات الحرية (df=n-1)

- اذا كانت عينتين حجمهما (n1, n2) تكون درجات الحرية (df=n1+n2-2)

2- ثم يتم استخراج قيمة (t) الجدولية من الجداول الاحصائية بالمزاوجة بين درجات الحرية (df) ومستوى

المعنوية (α) كما يلي:

▪ إذا كان الاختبار من طرفين (أيسر و ايمن) تحسب (t) الجدولية عند مستوى المعنوية ($\alpha/2$) ودرجات

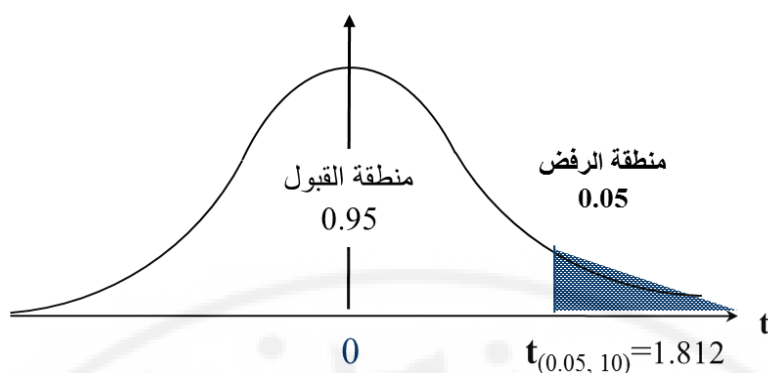
حرية (df=n-1) كما يلي: $t_{(\alpha/2, df)}$

▪ اذا كان نوع الإختبار من طرف واحد (أيسر او ايمن) تحسب (t) الجدولية عند مستوى المعنوية

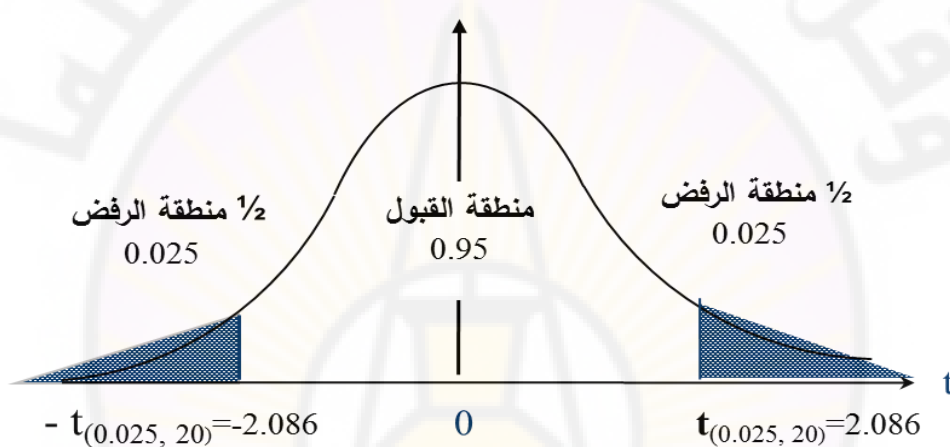
(α) ودرجات حرية (df=n-1) كما يلي: $t_{(\alpha, df)}$

مثال (3): اوجد قيمة كل من $t_{(0.05, 10)}$ و $t_{(0.025, 20)}$

- بالبحث في جدول (t) عند تقاطع السطر (10) مع العمود (0.05) نجد القيمة 1.812



- ولإيجاد $t_{(0.025, 20)}$ بالبحث في جدول (t) عند تقاطع السطر (20) مع العمود (0.025) نجد القيمة 2.086



4-9- حالات استخدام اختبار (T) لمقارنة المتوسطات:

يستخدم اختبار (T) في الحالات التالية:

الحالة الأولى: مقارنة متوسط عينة بمتوسط مجتمع:

يسمى اختبار (T) للعينة الواحدة (One sample T-Test).

الحالة الثانية: المقارنة بين متوسطي عينتين مستقلتين:

يسمى اختبار (T) لعينتين مستقلتين (Independent sample T-Test)

الحالة الثالثة: المقارنة بين متوسطي عينتين مرتبطتين:

يسمى اختبار (T) لعينتين مرتبطتين (Paired sample T-Test)

4-9-1 الحالة الأولى : اختبار (t) لمقارنة متوسط عينة بمتوسط مجتمع (One sample T-Test)

عند مقارنة متوسط عينة عشوائية بمتوسط المجتمع الأصلي الذي سحبت منه هذه العينة، يجب التمييز بين الاختلافات التالية:

1- إذا كان تباين المجتمع (σ^2) معلوماً عندها يستخدم اختبار (Z-Test) سواء كان حجم العينة كبير ($n > 30$) أم صغير ($n \leq 30$).

2- إذا كان تباين المجتمع (σ^2) مجهولاً وحجم العينة كبير ($n > 30$) سوف يتم استخدام تباين العينة (S^2) بدلاً عن تباين المجتمع (σ^2) وكذلك يتم استخدام اختبار (Z-Test) لاختبار معنوية الفرق بين متوسط العينة و متوسط المجتمع.

3- إذا كان تباين المجتمع (σ^2) مجهولاً وحجم العينة صغير ($n \leq 30$) يتم استخدام اختبار (t-Test) لاختبار معنوية الفرق بين متوسط العينة و متوسط المجتمع. ويمكن تلخيص الحالات السابقة بالجدول التالي:

جدول (4-1): طرائق حساب إحصاء الاختبار لمتوسط مجتمع.

حجم العينة	إحصائية الاختبار (القيمة المحسوبة)	تباين المجتمع
لا يشترط حجم العينة (غير مهم)	$Z = \frac{(\bar{x} - \mu_0)}{\frac{\sigma}{\sqrt{n}}}$	معلوم
$n > 30$	$Z = \frac{(\bar{x} - \mu_0)}{\frac{S}{\sqrt{n}}}$	غير معلوم

	$t = \frac{(\bar{X} - \mu_0)}{\frac{S}{\sqrt{n}}}$	$n \leq 30$
--	--	-------------

فإذا كان تباين المجتمع مجهول وحجم العينة أقل من (30) فيجب استخدام اختبار (t) للمقارنة بين متوسط عينة ومتوسط مجتمع وفق الخطوات التالية:

1- صياغة الفروض:

يمكن صياغة الفرضيات حول متوسط مجتمع على الشكل التالي:

● الفرض العدمي: $(H_0: \mu_1 = \mu_0)$ المتوسط المزعوم للمجتمع صحيح .

● الفرض العدمي: $(H_0: \mu_1 \neq \mu_0)$ المتوسط المزعوم للمجتمع غير صحيح .

2- حساب قيمة إحصائية الاختبار:

ولكن قبل البدء في استخدام اختبار (t-Test) على بيانات العينة يجب التأكد من استيفاء بيانات العينة

لافتراضات الاعتدالية، وبعد توفرها يمكن للباحث ان يستكمل إجراء اختبار (t-Test)

بما أن تباين المجتمع مجهول وحجم العينة أقل من (30) عندئذ يمكن تطبيق اختبار (t) حيث يتم حساب قيمة

اختبار (t-Test) والتي تسمى احصاءة الاختبار وفقا للقانون التالي:

$$t = \frac{\bar{x} - \mu_0}{\frac{S}{\sqrt{n}}} \sim t(n - 1)$$

أي قيمة احصاءة الاختبار تتوزع وفق التوزيع t بدرجة حرية (df=n-1)

$\bar{X}$: متوسط العينة.

μ_0 : متوسط المجتمع المفروض

S : الانحراف المعياري للعينة

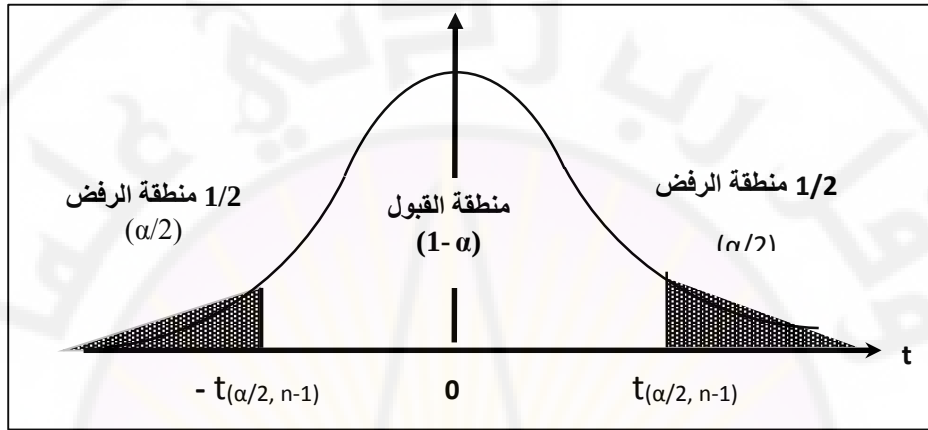
n : عدد افراد العينة.

$\frac{S}{\sqrt{n}}$: الخطأ المعياري لمتوسط العينة.

3- تحديد منطقتي القبول والرفض:

يتم تحديد منطقتي القبول والرفض بواسطة القيم الجدولية (النقاط الحرجة) التي تُستخرج من الجداول الإحصائية لتوزيع (t) بناءً على مستوى المعنوية (α) بدرجات حرية ($n-1$) كما هو مبين بالشكل التالي:

وتكون منطقة القبول من القيمة $-t(\alpha/2, n-1)$ حتى القيمة $+t(\alpha/2, n-1)$



4- المقارنة و اتخاذ القرار:

تتم مقارنة قيمة احصائية الاختبار المحسوبة سابقاً مع منطقتي القبول والرفض، فإذا وقعت قيمة إحصائية الاختبار (القيمة المحسوبة) في منطقة القبول نقبل الفرض العدم بأنه لا يوجد فروق معنوية بين متوسط العينة ومتوسط المجتمع، أما إذا وقعت في منطقة الرفض نرفض الفرض العدم ونقبل الفرض البديل بوجود فروق معنوية بين متوسط العينة ومتوسط المجتمع.

مثال (4):

ادعى مدرس مقرر الاحصاء ان متوسط درجات طلابه هو (68)، من اجل اختبار هذا الفرض تم سحب عينة عشوائية حجمه (20) من درجات الطلاب مبينة بالجدول (2).

المطلوب: هل ادعاء المدرس صحيح أم خاطئ عند مستوى معنوية (0.05)؟

الحل: $\mu = 68$ $n=20$

درجات الطلاب		العينة
62	70	

1- صياغة الفروض:

60	66	الملاحظات (درجات الطلاب)
64	62	
60	64	
58	74	
62	70	
68	65	
68	65	
60	65	
75	62	
20		حجم العينة (n_i)
65		متوسط العينة ($\bar{x}_i$)
21.68		تباين العينة (S_i^2)
4.66		الانحراف المعياري للعينة (S_i)

الفرض العدمي: $H_0 : \mu = 68$

الفرض البديل: $H_1 : \mu \neq 68$

2- حساب إحصاءة الاختبار

$$t = \frac{\bar{x} - \mu_0}{\frac{S}{\sqrt{n}}} = \frac{65 - 68}{\frac{4.66}{\sqrt{20}}} = -2.88$$

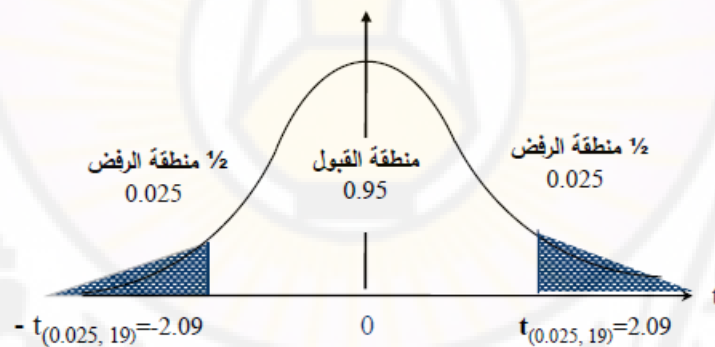
3- تحديد منطقتي القبول والرفض:

تحدد منطقتي القبول والرفض اعتماداً على قيمة

الجدولية ($t=2.09$) و على نوع الاختبار (ذو

طرفين) وعلى مستوى المعنوية ($\alpha=0.05$) وبالتالي

تكون منطقة القبول اكبر من $-t_{(\frac{\alpha}{2}, n-1)} = -2.09$ إلى $+t_{(\frac{\alpha}{2}, n-1)} = +2.09$



4- المقارنة واتخاذ القرار:

بمقارنة قيمة t المحسوبة ($t=-2.88$) بقيمة t الجدولية ($t=2.09$) نجد أن قيمة t المحسوبة تقع في منطقة

الرفض، فإن القرار هو رفض الفرض الصفري، أي ادعاء المدرس أن متوسط علامات طلابه (68) ادعاء

خاطئ. 4-9-2- الحالة الثانية: اختبار (t) للمقارنة بين متوسطي عينتين مستقلتين

في أغلب الدراسات نحتاج إلى مقارنة متوسطي عينتين مستقلتين كمقارنة متوسط أطوال الطلاب مع متوسط أطوال الطالبات. هنا بيانات كل من العينتين الطلاب والطالبات تحسب باستقلالية عن الأخرى، أي أنه لا يوجد أي علاقة ارتباط بين مشاهدات العينة الأولى ومشاهدات العينة الثانية. وهنا يمكن تمييز حالتين هما:

← **العينتين مستقلتين ومتجانستين (تباين المجتمعين متساويين):**

← **العينتين مستقلتين و غير متجانستين (تباين المجتمعين غير متساويين):**

أ- اختبار (T) للمقارنة بين عينتين مستقلتين ومتجانستين (تباين المجتمعين متساوي):

يستخدم هذا النوع من اختبار (T) للحكم على معنوية الفرق بين متوسطي عينتين مستقلتين. على افتراض أن تباين المجتمع الأول (σ_1^2) و تباين المجتمع الثاني (σ_2^2) غير معلومين ولكنهما متجانسين وأن حجم العينتين أو أحدهما صغير ($n_1, n_2 < 30$).

خطوات اختبار الـ (T) للمقارنة بين عينتين مستقلتين ومتجانستين:

1- صياغة الفروض:

الفرض العدمي: $H_0 : \mu_1 = \mu_2$

الفرض البديل: هو أحد الحالات الآتية: ($H_1 : \mu_1 \neq \mu_2$) أو ($H_1 : \mu_1 < \mu_2$) أو ($H_1 : \mu_1 > \mu_2$)

2- حساب احصاء الاختبار:

من أجل حساب احصاء الاختبار في هذه الحالة، يتم مزج تباين العينتين أو خلطهما معاً من خلال دمج (إضافة) مجموع المربعات في البيانات من كل من العينتين والقسمة على درجات الحرية الإجمالية لكل من العينتين أي ليتم حساب احصاء الاختبار (t المحسوبة) من العلاقة التالية:

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{S_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

أي أن احصائية الاختبار تتبع توزيع (t) بدرجات حرية تساوي $(n_1 + n_2 - 2)$

حيث (S_p^2) يمثل التباين المشترك (المرجح) للعينتين ويُحسب بإحدى المعادلتين التاليتين:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

$$S_p^2 = \frac{\sum(x_1 - \bar{x}_1)^2 + \sum(x_2 - \bar{x}_2)^2}{n_1 + n_2 - 2}$$

3- إيجاد قيمة (t) الجدولية:

يتم استخراج قيمة (t) الجدولية من الجداول الاحصائية كما يلي:

■ إذا كان الاختبار من طرفين (أيسر و أيمن) تحسب (t) الجدولية عند مستوى المعنوية $(\alpha/2)$ ودرجات

حرية $(df=n_1+n_2-2)$ كما يلي: $t_{(\alpha/2, n_1+n_2-2)}$

■ إذا كان نوع الإختبار من طرف واحد (أيسر أو أيمن) تحسب (t) الجدولية عند مستوى المعنوية

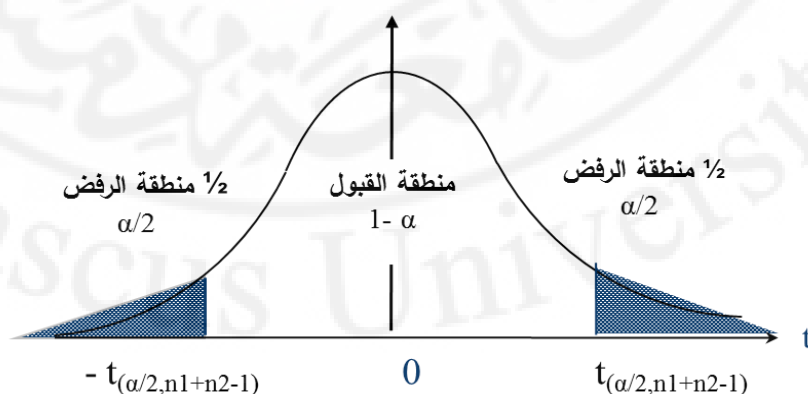
(α) ودرجات حرية $(df= n_1+n_2-2)$ كما يلي: $t_{(\alpha, n_1+n_2-2)}$

4- تحديد منطقتي القبول والرفض:

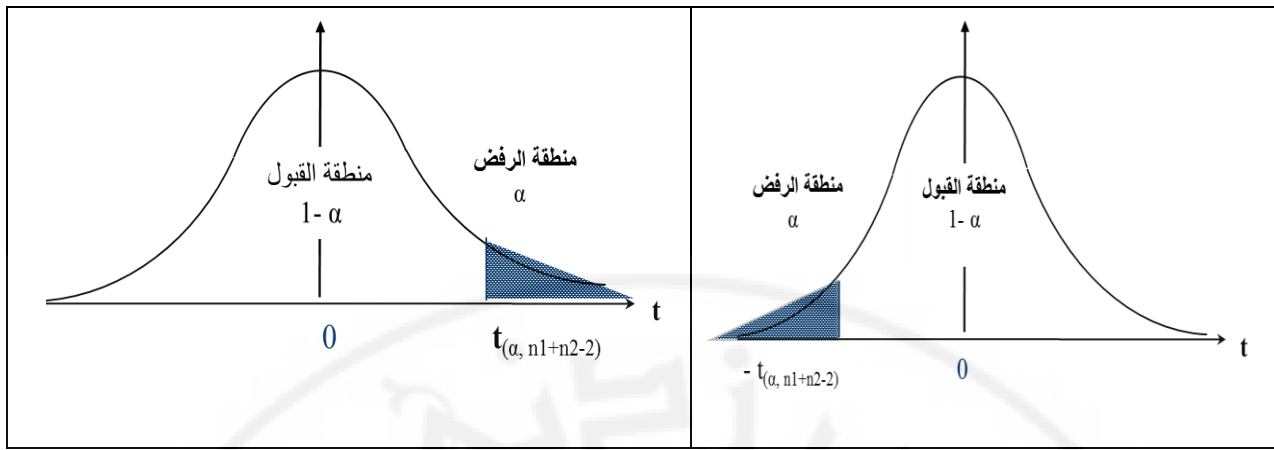
تحدد منطقتي القبول والرفض اعتماداً على قيمة (t) الجدولية و على نوع الاختبار (ذو طرف واحد One

Tail Test أو ذو طرفين Two Tail Test) وعلى مستوى المعنوية (α) .

- إذا كان الاختبار من طرفين تكون منطقتي القبول والرفض موضحة بالشكل التالي:



أما إذا كان الاختبار من طرف واحد تكون منطقة منطقتي القبول والرفض موضحة بالشكل التالي:



5- المقارنة واتخاذ القرار

يتم مقارنة قيمة الإحصائية (t) المحسوبة) بحدود منطقتي القبول والرفض فإذا وقعت قيمة الإحصائية داخل منطقة القبول فإن القرار هو: قبول الفرض الصفري أن الفرق بين متوسطي العينتين غير دال إحصائياً. أما إذا وقعت قيمة الإحصائية في منطقة الرفض فإن القرار هو رفض الفرض الصفري، وفي هذه الحالة نقبل الفرض البديل أن الفرق بين متوسطي العينتين دال إحصائياً.

مثال (5): نريد اختبار فيما إذا كان فروق بين علامات مقرر الاحصاء التطبيقي بين طلاب جامعة الشام الخاصة وعلامات طلاب جامعة دمشق، تم أخذ عينة من جامعة الشام حجمها (9) طلاب و عينة من جامعة دمشق حجمها (13) طالب ودونت العلامات بالجدول التالي:

العلامة	طلاب جامعة الشام	طلاب جامعة دمشق
الملاحظات (العلامات)	8	22
	17	15
	15	17
	6	16
	19	16
	11	14
	10	11
	13	10
	9	21
		10

8		
10		
12		
$n_2=13$	$n_1=9$	حجم العينة (n_i)
$(\bar{x}_2 = 14)$	$(\bar{x}_1 = 12)$	متوسط العينة ($\bar{x}_i$)
$(S_2^2 = 19)$	$(S_1^2 = 18.75)$	تباين العينة (S_i^2)

المطلوب: هل يوجد فرق معنوي بين علامات طلاب جامعة الشام وعلامات طلاب جامعة دمشق عند مستوى

معنوية ($\alpha=0.05$) مع العلم أن تباين المجتمعين متساويين.

الحل:

1- صياغة الفروض:

الفرض العدمي: $H_0 : \mu_1 = \mu_2$ متوسط علامات جامعة الشام يساوي متوسط علامات جامعة دمشق

الفرض البديل: ($H_1 : \mu_1 \neq \mu_2$) متوسط علامات جامعة الشام لا يساوي متوسط علامات جامعة دمشق

2- حساب احصاء الاختبار:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = \frac{8 \times 18.75 + 12 \times 19}{9 + 13 - 2} = \frac{378}{20} = 18.9$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{S_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{(12 - 14)}{\sqrt{18.9 \left(\frac{1}{9} + \frac{1}{13} \right)}} = \frac{-2}{1.88} = -1.1$$

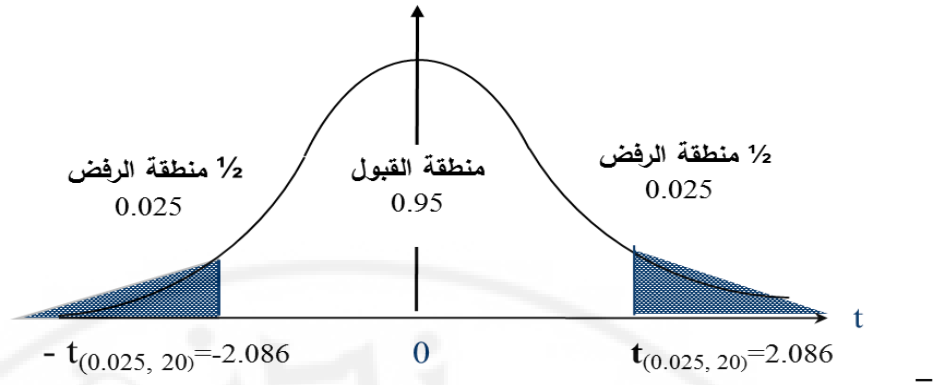
إيجاد قيمة (t) الجدولية:

بما ان الاختبار من طرفين تحسب (t) الجدولية كما يلي: $t_{(\frac{\alpha}{2}, n_1+n_2-2)} = t_{(0.025, 20)} = 2.086$

4- تحديد منطقتي القبول والرفض:

تحدد منطقتي القبول والرفض وفقاً لنوع الاختبار (ذو طرفين) وعلى مستوى المعنوية (α). تكون منطقة القبول

أكبر من (-2.086) وأصغر من $(+2.086)$



3- المقارنة واتخاذ القرار: بمقارنة قيمة t المحسوبة ($t = -1.1$) بقيمة t الجدولية ($t = 2.086$) نجد أن t

المحسوبة تقع في منطقة القبول، لذلك نقبل الفرض الصفري بأنه لا يوجد فرق معنوي بين متوسط علامات

طلاب جامعة الشام وطلاب جامعة دمشق بمستوى دلالة ($\alpha = 0.05$)

ب- اختبار (T) للمقارنة بين متوسطي عينتين مستقلتين وغير متجانستين:

في هذه الحالة تكون نتيجة اختبار (F) لتجانس التباين دالة احصائياً أي أن العينتين مستقلتان وغير متجانستين.

و بفرض ان البيانات في المجتمع الاول والثاني تخضع للتوزيع الطبيعي و أن تباين المجتمعين (σ_1^2)

(σ_2^2) مجهولين (غير معلومين) ولكنهما غير متجانسين أي ($\sigma_1^2 \neq \sigma_2^2$)، وحجم العينتين صغير فإن احصائية

الاختبار تتبع توزيع (t) وتحسب بالعلاقة التالية:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

وبما اننا نفترض أن ($H_0: \mu_1 = \mu_2$) فتصبح معادلة حساب احصاءة الاختبار كما يلي:

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

حيث إن احصائية الاختبار تتبع توزيع (t) بدرجات حرية تساوي:

$$df = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2 - 1}}$$

■ إذا كانت قيمة درجات الحرية (df) لمحسوبة بالمعادلة السابقة تمثل عدد حقيقي، فإنه من الضروري التقريب لأقرب عدد صحيح.

ثم يتم تطبيق باقي الخطوات من اختبارات الفروض بتحديد منطقتي القبول والرفض بعد استخراج قيمة (t) الجدولية وفقاً لدرجات الحرية، وأخيراً تتم مقارنة احصائية الاختبار بمنطقتي القبول والرفض ومن ثم اتخاذ القرار الاحصائي.

مثال (6): يرغب باحث زراعي في اختبار وجود فروق معنوية بين انتاجية صنفين من التفاح، لذلك قام باختيار (12) شجرة من الصنف الاول و (10) شجرات من الصنف الثاني عشوائياً وتم حساب انتاجية كل شجرة وتدوينها بالجدول التالي:

المطلوب: هل يوجد فروق معنوية بين متوسطي انتاج الصنفين عند مستوى معنوية ($\alpha=0.05$) مع العلم أن تباين مجتمعي الصنفين غير متجانسين ($\sigma_1^2 \neq \sigma_2^2$).

الحل :

الصنف الأول	الصنف الثاني	العينة
25	30	المشاهدات (كميات التفاح)
4	28	
42	22	
24	35	
32	48	
12	36	
58	32	
12	48	
42	35	
70	26	
35		
4		
12	10	حجم العينة (n_i)
30	34	متوسط العينة ($\bar{x}_i$)

1- صياغة الفروض:

الفرض العدمي: $H_0 : \mu_1 = \mu_2$

الفرض البديل: $(H_1 : \mu_1 \neq \mu_2)$

2- حساب احصاء الاختبار:

معادلة حساب احصاء الاختبار كما يلي:

432.9	73.55	تباين العينة (S_i^2)
20.8	8.58	الانحراف المعياري للعينة (S_i)

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

$$= \frac{30 - 34}{\sqrt{\frac{432.9}{12} + \frac{73.55}{10}}} = \frac{-4}{6.59} = -0.60$$

حيث إن احصائية الاختبار تتبع توزيع (t) بدرجات حرية تساوي:

$$df = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1-1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2-1}} = \frac{\left(\frac{432.9}{12} + \frac{73.55}{10}\right)^2}{\frac{\left(\frac{432.9}{12}\right)^2}{11} + \frac{\left(\frac{73.55}{10}\right)^2}{9}} = \frac{1885.7}{124.28} = 15.17$$

هذا يعني درجات الحرية (df=15).

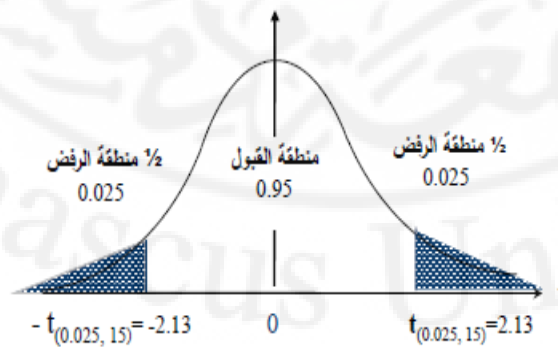
إيجاد قيمة (t) الجدولية:

بما ان الاختبار من طرفين تحسب (t) الجدولية كما يلي: $t_{(\frac{\alpha}{2}, n)} = t_{(0.025, 15)} = 2.13$

3- تحديد منطقتي القبول والرفض:

تحدد منطقتي القبول والرفض اعتماداً على قيمة (t) الجدولية (2.13) و تكون منطقة القبول اكبر من

(-2.13) واصغر من (+2.13)



4- المقارنة واتخاذ القرار

القرار الإحصائي: بمقارنة قيمة t المحسوبة ($t=-0.61$) بقيمة t الجدولية ($t=2.13$) نجد أن قيمة t المحسوبة تقع في منطقة القبول، لذلك نقبل الفرض الصفري القائل بأنه لا يوجد فرق معنوي بين متوسطي صنفَي التفاح بمستوى دلالة ($\alpha=0.05$)

4-9-3- الحالة الثالثة: اختبار الفرضيات للفرق بين متوسطي مجتمعين غير مستقلين (مرتبطتين).

هذا الاختبار يسمى اختبار للفرق بين متوسطي عينتين غير مستقلتين (مرتبطتين) أو اختبار الفروض للفرق بين الأزواج المتقابلة (Testing Hypotheses for the Difference between Matched Pairs) حيث يستخدم هذا النوع من الاختبارات للحكم على دلالة الفروق بين متوسطي عينتين مرتبطتين (Correlated Data)، مثل اختبار دلالة الفروق بين متوسط أداء الموظفين قبل التدريب وبعد التدريب. ولتوضيح هذا النوع من الاختبارات نورد المثال التالي:

يرغب باحثاً في معرفة فيما إذا كان إعطاء دورات تدريبية للموظفين الذين يعملون في قسم المبيعات حول أساليب التعامل مع الزبائن له دور فعال في زيادة عدد الوحدات المباعة من سلعة معينة. في هذه الحالة نأخذ بيانات حول عدد الوحدات المباعة من تلك السلعة قبل اتباع الدورة التدريبية وبعد اتباع الدورة، ثم نقوم بمقارنة متوسطي عدد الوحدات المباعة قبل وبعد اتباع الدورة التدريبية.

أيضاً على سبيل المثال يرغب باحث في معرفة ما إذا كان نظام غذائي معين في تخفيف الوزن لمجموعة من الأشخاص الذين يعانون من السمنة، في هذه الحالة نأخذ أوزانهم قبل وبعد اتباع النظام الغذائي. كذلك من التطبيقات العملية لهذا الاختبار الحالة الصحية لمجموعة مرضى قبل وبعد تناول دواء معين. يطبق هذا الاختبار للمقارنة بين مجموعتين من البيانات لعينة واحدة، ويطلق عليه اختبار العينات المرتبطة أو اختبار (قبلي - بعدي) بمعنى قبل وبعد البرنامج المطبق على العينة.

القيم قبل التجربة تسمى الدرجات القبليّة (X)، وتسمى القيم بعد تطبيق التجربة بالدرجات البعديّة (Y)، ويمكن وضع النتائج في جدول خاص على شكل أزواج متقابلة، فنحصل على عيّنتين مرتبطتين من القيم (لأن كل زوج يعود لشخص واحد) كما هي مبينة بالجدول (4-2).

جدول (4-2): البيانات القبليّة والبعديّة لاختبار الفرق بين متوسطي عيّنتين مرتبطتين.

رقم الشخص	1	2	3	4		i		n	المتوسط
الدرجات القبليّة (X)	x_1	x_2	x_3	x_4		x_i		x_n	$\bar{x}$
الدرجات البعديّة (Y)	y_1	y_2	y_3	y_4		y_i		y_n	$\bar{y}$
عينة الفروقات d_i	d_1	d_2	d_3	d_4		d_i		d_n	$\bar{d}$
									S_d

ثم يتم حساب الفروقات بين قيمتي كل زوج من العلاقة: $d_i = x_i - y_i$

ويحسب متوسط وتباين والانحراف المعياري لهذه الفروقات من العلاقتين:

$$\bar{d} = \frac{1}{n} \sum d_i$$

$$S_d^2 = \frac{\sum (d_i - \bar{d})^2}{n-1}$$

ولإجراء الاختبار حول الفرق بين العيّنتين يتم وضع الفرضيتين كما يلي:

1- صياغة الفروض

الفرض الصفري: $H_0: \mu_1 = \mu_2$ أو $H_0: \bar{D}_0 = 0$ حيث $\bar{D}_0 = \mu_1 - \mu_2$

الفرض البديل: $H_1: \mu_1 \neq \mu_2$ أو $\bar{D}_0 \neq 0$

يمكن للفرض البديل أن يأخذ إحدى الحالتين التاليتين أيضاً:

حيث:

d : هي متوسط الفرق بين أزواج المشاهدات.

S_d : هو الانحراف المعياري للفرق بين أزواج المشاهدات.

n : تمثل حجم العينة المسحوبة من المجتمع (عدد أزواج المشاهدات).

$\bar{D}_0$: هو الفرق بين متوسطي المجتمعين الاول والثاني

2- حساب قيمة إحصائية الاختبار

إن احصاءة الاختبار تتوزع وفقاً لتوزيع (t) بدرجات حرية (df=n-1) وتحسب بالعلاقة التالية:

$$t = \frac{\bar{d} - \bar{D}_0}{S_d/\sqrt{n}}$$

حيث $\bar{D}_0$ هي قيمة متوسط الفروقات المفترضة في المجتمع، وتؤخذ قيمتها الصفرية من فرضية العدم H_0 وهي

$\bar{D}_0 = 0$ فتصبح معادلة حساب احصاءة الاختبار على الشكل الآتي:

$$t = \frac{\bar{d}}{S_d/\sqrt{n}}$$

3- تحديد منطقتي القبول والرفض

تحدد منطقتي القبول والرفض باستخراج قيمة (t) الجدولية من جدول t عند درجات حرية تساوي (n -

(1)

عند مستوى معنوية يساوي (α) ،

4- المقارنة و اتخاذ القرار

تتم مقارنة قيمة احصاءة الاختبار المحسوبة بحدود منطقتي القبول والرفض، فإذا وقعت في منطقة

القبول يتم قبول الفرض العدمي (H_0) ونقول بأنه لا يوجد فرق معنوي بين الدرجات القبلية والبعدية، أما إذا كانت

قيمة احصائية الاختبار أكبر من (t) الجدولية أي تقع في منطقة الرفض وبالتالي نرفض الفرض العدمي ونقبل

الفرض البديل القائل أن هناك فرق معنوي بين القيم القبلية والبعدية.

مثال (7): البيانات التالية تمثل نتائج تجربة أجريت على عشرة أشخاص لاختبار مدى فعالية نظام خاص من

الغذاء لتخفيف الوزن، حيث تم قياس وزن الأشخاص قبل البدء في تطبيق النظام الغذائي وبعد تطبيقه بثلاثة

شهور فكانت النتائج موضحة بالجدول (2-6).

جدول (2-6): أوزان الأشخاص قبل وبعد تطبيق النظام الغذائي.

رقم الشخص	1	2	3	4	5	6	7	8	9	10
الأوزان قبل (X)	95	109	89	93	106	92	88	119	102	92
الأوزان بعد (Y)	90	96	85	87	104	85	76	103	95	84

المطلوب: اختبار ما اذا كان نظام الغذاء كان فعالاً في تخفيف الوزن مستخدماً مستوى دلالة ($\alpha=0.05$)

الحل:

1- صياغة الفروض

بفرض أن $d_i = x_i - y_i$ تمثل عينة الفرق بين أزواج البيانات للعينتين و $\bar{D}_0$ هو الفرق بين متوسطي المجتمعين الاول والثاني، يمكن صياغة الفروض كما يلي:

● الفرض الصفري: $H_0: \mu_x = \mu_y$ متوسط أوزان الأشخاص القبلية = متوسط أوزان الأشخاص البعدية
أي لا يوجد فرق معنوي بين القيم القبلية والبعدية.

● الفرض البديل: $H_1: \mu_x \neq \mu_y$ أو $H_1: \mu_x > \mu_y$ متوسط أوزان الأشخاص القبلية أكبر من متوسط أوزان الأشخاص البعدية، أي يوجد فرق معنوي بين الدرجات القبلية والبعدية.

يمكن صياغة الفرض البديل على الشكل $\bar{D}_0 > 0$ أي أن متوسط عينة الفروق أكبر من الصفر.

2- حساب قيمة إحصائية الاختبار

يتم حساب احصاءة الاختبار من العلاقة:

$$t = \frac{\bar{d} - \bar{D}_0}{\frac{S_d}{\sqrt{n}}}$$

وهذا يتطلب حساب متوسط الفروق ($\bar{d}$) والانحراف المعياري (S_d) لأزواج المشاهدات كما هي مبينة بالجدول (2-7).

جدول (2-7): حساب الانحراف المعياري لأوزان الأشخاص قبل وبعد تطبيق النظام الغذائي.

رقم الشخص	1	2	3	4	5	6	7	8	9	10
-----------	---	---	---	---	---	---	---	---	---	----

	92	102	119	88	92	106	93	89	109	95	الأوزان قبل (X)
	84	95	103	76	85	104	87	85	96	90	الأوزان بعد (Y)
80	8	7	16	12	7	2	6	4	13	5	الفروق (d=x-y)
0	0	-1	8	4	-1	-6	-2	-4	5	-3	(d-d̄)
172	0	1	64	16	1	36	4	16	25	9	(d-d̄)²

$$\bar{d} = \frac{80}{10} = 8$$

$$S_d = \sqrt{\frac{\sum (d - \bar{d})^2}{n-1}} = \sqrt{\frac{172}{9}} = 4.37$$

$$t = \frac{\bar{d}}{\frac{S_d}{\sqrt{n}}} = \frac{8}{4.37/\sqrt{10}} = 5.79$$

3- تحديد منطقتي القبول والرفض

تحدد منطقتي القبول والرفض باستخراج قيمة (t) الجدولية من جدول t عند درجات حرية تساوي (n - 1 = 9)

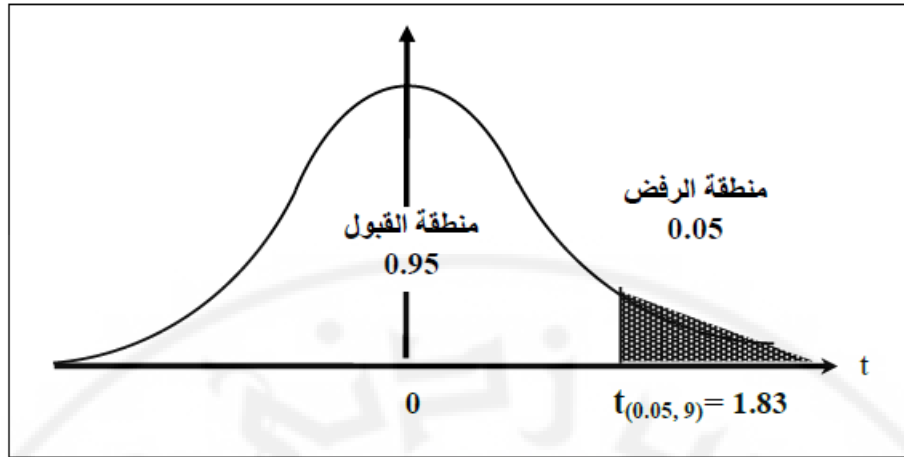
عند مستوى معنوية يساوي (α = 0.05)، أي أن $t_{\alpha,9} = t_{0.05,9} = 1.83$

تحدد منطقتي القبول والرفض باستخراج قيمة (t) الجدولية من جدول t عند درجات حرية تساوي (n - 1 = 9)

عند مستوى معنوية يساوي (α = 0.05) وبما أن الفرض البديل أكبر من هذا يعني أنه من طرف واحد نحو

اليمين أي القيمة الجدولية تكون أي $t_{(\alpha, n-1)} = t_{(0.05, 9)} = 1.83$ وبالتالي تتحدد منطقتي القبول

والرفض بالشكل (2-61).



شكل (2-16): التمثيل البياني لمنطقتي القبول والرفض للمثال (2-10).

4- المقارنة و اتخاذ القرار

بما أن قيمة الإحصائية تساوي 5.79 فإنها تقع في منطقة الرفض وبالتالي فإن القرار هو رفض الفرض العدمي وقبول الفرض البديل بأنه يوجد فرق معنوي بين أوزان الأشخاص قبل وبعد النظام الغذائي، وهذا يقودنا إلى النتيجة بأن النظام الغذائي المتبع فعال في انقاص الوزن.

الفصل الرابع

اختبارات الفروض الاحصائية

Testing of Statistical Hypotheses



اختبارات الفروض الاحصائية

Testing Statistical Hypotheses

2-1- مقدمة:

يُعد موضوع اختبارات الفروض الإحصائية من أهم المواضيع في مجال اتخاذ القرارات، وهو من أهم التطبيقات التي قدمها علم الإحصاء كحل للمشاكل العلمية المختلفة بشتى فروع العلوم، فمثلاً نحتاج كثيراً في الحياة العملية إلى التحقق من صحة فرض أو ادعاء ما، وقد يكون هذا التحقق سهلاً ومؤكداً مثل ادعاء شخص ما أن السماء تمطر بالخارج وللتحقق من صحة هذا الادعاء يكفي أن نذهب لخارج المبنى

ونتأكد هل السماء تمطر أم لا ؟ وأحياناً يكون التحقق من صحة الادعاء صعباً مثل ادعاء شخص أنه ذهب إلى أحد الأسواق التجارية في مدينة دمشق. وأحياناً أخرى يكون التحقق من الادعاء مستحيلاً مثل ادعاء شخص أنه قد أضرر شيئاً معيناً في نفسه.

وفي علم الاحصاء نسمي الادعاء بـ (فرض العدم) ونسمي التحقق منه بـ (اختبار فرض العدم) ونحتاج إلى عملية تطبيق اختبارات الفروض في كثير من الحالات مثل اختبار الفرض بأن متوسط دخل الفرد في دمشق هو 45000 ليرة سورية، أو اختبار ادعاء مدير مصنع الأقلام بأن نسبة القطع التالفة في مصنعه لا تتجاوز (3%)، أو ادعاء أحد الأطباء أن نسبة الشفاء في عيادته لا تقل عن (90%) وهكذا.

وبالتالي اختبارات الفروض الإحصائية هي وضع قواعد تمكن من التوصل إلى قرار بقبول أو رفض فرض عن معالم مجتمع واحد أو أكثر ومن خلالها يمكن لأي شخص أن يتخذ قرار برفض أو قبول فرض معين أو مجموعة من الفروض المتعلقة بمشكلة معينة موجودة في الحياة العامة.

مثال (2-1): بفرض أن إحدى شركات إنتاج المعلبات تكتب على المنتج أن وزنه الصافي (120) جرام وللتأكد من صحة هذا الادعاء تم أخذ عينة حجمها (100) علبة لاختبارها، ووجد أن متوسط أوزان هذه العينة هو (118.9) جرام، فهل يمكن الحكم على هذه الشركة بأنها تقوم بعملية غش أو كذب على المستهلك بناءً على هذه العينة؟ وأن جميع منتجات هذه الشركة من هذه المعلبات أقل من 120 جرام؟ الإجابة بالطبع لا؛ لا يمكن اتخاذ القرار إلا بعد إجراء إحصائي للتحقق من صحة ما كتب على المنتج وهذا الإجراء هو ما يسمى باختبار الفروض الاحصائية (Test of hypothesis) حول متوسط المجتمع.

مثال (2-2): نفرض أن باحثاً اجتماعياً ادعى أن متوسط أعمار طلاب الجامعة لا يختلف عن متوسط أعمار الطالبات. للتأكد من هذا الادعاء تم اختيار عينة عشوائية من بين الطلاب وعينة عشوائية من بين الطالبات وحسب متوسط العمر في كل عينة منهما، فإذا كان متوسط عمر الطالب هو 24 وكان متوسط عمر الطالبة هو 22 فهل يعني ذلك أن متوسط عمر الطالب أكبر من متوسط عمر الطالبة ؟ هل الفرق راجع لمجرد الصدفة ؟ متى يكون الفرق نتيجة للصدفة؟ متى يكون الفرق دالاً على وجود اختلاف حقيقي أو جوهري بين متوسطي المجتمعين الأصليين.

وقبل تناول كيفية إجراء الاختبارات الإحصائية نستعرض أولاً بعض المفاهيم والتعريفات الأساسية اللازمة لهذا الموضوع حتى تكون الصورة أكثر وضوحاً.

2-2- تعريف بعض المفاهيم المهمة:

هناك بعض المفاهيم المتعلقة باختبارات الفروض لابد من معرفتها:

- المقياس الإحصائي: هو قيمة وحيدة محسوبة من بيانات العينة أو المجتمع تحت شروط معينة.
 - الفرضية (Hypothesis): هي ادعاء حول صحة شيء ما قد يكون صحيحاً أو خطأ.
 - الفرضية الإحصائية (Statistical Hypothesis): هي عبارة عن ادعاء أو تخمين قد يكون صحيحاً أو خاطئاً حول معلمة أو أكثر لمجتمع أو لمجموعة من المجتمعات، ويكون المطلوب اختبار صحة هذا الادعاء أو التخمين.
 - اختبارات الفروض الإحصائية (اختبار المعنوية): هي طريقة معيارية لاختبار ادعاء (فرض) حول معلمة لمجتمع أو أكثر معتمدة في ذلك على بيانات عينة مسحوبة من ذلك المجتمع.
- إن الغرض من اختبارات الفروض هو وصول الباحث إلى قرار بخصوص فرض معين حول معلمة المجتمع، مع احتمال الوقوع في خطأ يحدده الباحث وذلك بتحديد مناطق رفض وقبول فرض العدم أسفل منحني التوزيع. تعتمد اختبارات الفروض على بيانات العينة وفرض قيمة معينة لمعلمة من معالم المجتمع حيث يكون الاختبار: هل هناك فرق معنوي بين قيمة معلمة المجتمع المفروضة والقيمة المقدرة لها من خلال بيانات العينة؟ فإذا كان هناك فرق فهل يرجع هذا الفرق إلى خطأ المعاينة؟ أم هو فرق حقيقي معنوي (Significant). فإذا كان الفرق معنوياً فيكون القرار هو عدم قبول الفرض العدمي وعليه فإننا نقبل بالفرض البديل. أما إذا كان الفرق غير معنوي فإننا نقبل الفرض العدمي.

2-3- مفهوم اختبارات الفروض الإحصائية المعلمية واللامعلمية:

تتقسم اختبارات الفروض الإحصائية إلى قسمين:

- أولاً: اختبارات الفروض الإحصائية المعلمية (Parametric Tests):** وفي هذا القسم يكون معلوم لدينا التوزيع الذي تتبعه البيانات وما إذا كان توزيعاً متصلًا أم منفصلاً (متقطعاً) ويكون المطلوب هو اختبار فروض حول معالم المجتمع مثل الفروض المتعلقة بالوسط الحسابي، النسبة، التباين، معامل الارتباط، وهذا النوع من الفروض قد لا تتعلق بمعالم المجتمع ولكن تتعلق بأشياء أخرى قد تكون وصفية مثل العلاقة بين التعليم والتدخين.
- ثانياً: اختبارات الفروض الإحصائية اللامعلمية (Non Parametric Test):** في كثير من التجارب والأبحاث يكون لدينا بيانات واقعية يصعب من خلالها التعرف على التوزيع الذي تتبعه هذه البيانات، ومن هنا نشأت الحاجة إلى ما يعرف باختبارات الفروض اللامعلمية حيث لا تحتاج مثل الاختبارات معرفة شكل التوزيع الذي تتبعه البيانات محل الدراسة، كما يفضل استخدامها عندما يكون حجم العينة المسحوبة من المجتمع صغيراً نسبياً. و سوف نهتم في هذا الفصل باختبارات الفروض الإحصائية المعلمية فقط.

البيانات المعلمية:

تكون البيانات معلميه إذا كانت تحقق الشرطين التاليين:

1-إذا كانت مقاييس البيانات كمية(نسبي أو فئوي).

2- تتبع التوزيع الطبيعي.

2-4- أنواع الفروض الإحصائية:

تنقسم الفروض الإحصائية إلى قسمين هما:

الفرض العدمي (الصفري) (H_0) (Null Hypothesis): هو عبارة عن جملة لفظية أو رياضية تعبر عن معلمة المجتمع (أو معلمتي مجتمعين أو أكثر) ويرمز للفرض العدمي (H_0)، ويتضمن الفرض العدمي علامة من العلامات التالية: ($=, \leq, \geq$)

وصياغة الفرض العدم تختلف باختلاف حالة الدراسة، فمثلاً إذا كان الغرض من الدراسة هي اختبار ادعاء حول متوسط مجتمع ما، يكون الفرض العدمي هو افتراض ان الادعاء صحيح أي أن متوسط المجتمع الحقيقي هو فعلاً نفس المتوسط المزعوم ($H_0: \mu = \mu_0$)

أما إذا كان الغرض هو اختبار تساوي معلمتي مجتمعين في مثال أعمار الطلاب وطالبات الجامعة فإن فرض العدم هو: ($H_0: \mu_1 = \mu_2$) أي يفترض عدم وجود اختلاف بين متوسطي اعمار الطلاب والطالبات.

الفرض البديل (H_1) (Alternative Hypothesis):

هو عبارة عن جملة لفظية أو رياضية تعبر عن معلمة المجتمع (أو معلمتي مجتمعين أو أكثر) و هو الفرض الذي يضعه الباحث كبديل عن فرض العدم ويتم قبوله في حال رفض الفرض الصفري ويرمز للفرض البديل (H_1)، ويتضمن الفرض البديل علامة من العلامات التالية ($\neq, >, <$).

فمثلاً في مثال أعمار الطلاب وطالبات الجامعة فإن الفرض البديل هو ($H_1: \mu_1 \neq \mu_2$) أي أنه يوجد اختلاف حقيقي وليس ظاهري بين متوسط أعمار الطلاب و متوسط أعمار الطالبات.

يهدف اختبار الفرضيات الإحصائية إلى اتخاذ قرار حول ما إذا كان الفرض الصفري -Null Hypothesis- (H_0) مقبول أم مرفوض ويتم ذلك باستخدام دالة اختبار إحصائية مناسبة.

إن قبول الفرض الصفري لا يعني بالضرورة أنه صحيح وإنما لا يوجد أدلة كافية من بيانات العينة لرفضه، كما أن رفضه لا يعني بالضرورة أنه خاطئ بل يعني أن الإحصائية المحسوبة من العينة كانت بعيدة عن المعلمة المناظرة لها في المجتمع مما يدفعنا إلى رفض الفرض الصفري.

اختيار مستوى المعنوية α ودرجة الثقة $(1 - \alpha)$:

إن القرار الاحصائي الذي سوف يتم اتخاذه بناء على الاختبار الإحصائي لا يمكن اعتباره صحيح 100 % فهناك مقدار من الخطأ لأن المعلومات التي يتم اتخاذ القرار بناء عليها مأخوذة من عينة وليس من المجتمع الأصلي.

في اختبار فرض معين، فإن مقدار ثقتنا في القرار الاحصائي المتخذ بالرفض أو القبول يسمى بدرجة الثقة ويرمز له بالرمز $(1 - \alpha)$ كما وأن مقدار عدم الثقة يسمى بمستوى المعنوية (أو احتمال الخطأ)، (Significance Level) ويرمز له بالرمز (α) . أي أنه إذا كانت درجة الثقة 95% فإن مستوى المعنوية $(\alpha=0.05)$ وهي تمثل مساحة منطقة الرفض تحت منحنى التوزيع وتكون أما على صورة ذيل واحد جهة اليمين أو اليسار أو ذيلين متساويين في المساحة واحد جهة اليمين والثاني جهة اليسار.

مستوى المعنوية (احتمال الخطأ): اعتاد الباحثون على مستوى معنوية $(\alpha=0.05)$ أو $(\alpha=0.01)$ ، حسب نوع التجربة وهذه النسبة مرتبطة ومكملة بمعامل الثقة.

درجة (معامل) الثقة: هي درجة الثقة في النتائج، اعتاد الباحثون على مستوى ثقة 95% وأغلب درجات الثقة المستخدمة هي (90% أو 95% أو 99%).

درجة الحرية: عدد مفردات العينة -1 أي $(n-1)$.

وعادة يحدد الباحث مستوى المعنوية أو درجة الثقة قبل البدء في عملية الاختبار. ومفهوم مستوى المعنوية مرتبط بمفهوم الخطأ من النوع الأول والخطأ من النوع الثاني كما يلي:

2-4-1- أنواع الأخطاء:

الخطأ من النوع الأول (α) والخطأ من النوع الثاني (β) :

إن أي قرار احصائي يمكن ان ينتج عنه نوعان من الخطأ وذلك وفقاً لقيمة مستوى المعنوية (the level of significance) (α) المحددة مسبقاً.

1- الخطأ من النوع الأول (a type I error) :

يحدث هذا النوع من الخطأ عند رفض الفرض العدمي وهو في الحقيقة صحيح باحتمال مقداره (α) . والخطأ من النوع الأول يمكن التحكم فيه، والذي يحدده الباحث قبل الاختبار والذي يمثل مستوى المعنوية.

2- الخطأ من النوع الثاني (a type II error) :

يحدث هذا النوع من الخطأ عند قبول الفرض العدمي وهو في الحقيقة خاطئ وذلك باحتمال مقداره (β) . الجدول (1-2) يوضح أنواع الأخطاء.

جدول (1-2): الخطأ من النوع الأول (α) والخطأ من النوع الثاني (β) .

في الواقع (الحقيقة) Population		القرار	
الفرض الصفري خاطئ (H_0 خطأ)	الفرض الصفري صحيح (H_0 صحيح)	Sample العينة	رفض H_0 قبول H_0
القرار صائب (power of test probability = $1 - \beta$)	خطأ من النوع الأول = α		
خطأ من النوع الثاني = β	القرار صائب (power of test probability = $1 - \alpha$)		

هنا عندما نقول أن الفرض الصفري بالحقيقة صحيح (H_0 صحيح) نقصد بها انه بناءً على بيانات المجتمع الحقيقية للظاهرة كان صحيحاً، بينما القرار الاحصائي بقبول أو رفض الفرض العدمي يتم بناءً على المقاييس التي يتم حسابها من بيانات العينة المسحوبة من ذلك المجتمع، وبالتالي يمكن لهذا القرار ان يكون صائب أو خاطئ بنسب مختلفة.

2-4-2- القرارات الاحصائية:

إن القرار الاحصائي بقبول أو رفض الفرض العدمي يتم بناءً على المقاييس التي يتم حسابها من بيانات العينة المسحوبة من ذلك المجتمع، وبالتالي يمكن لهذا القرار ان يكون صائب أو خاطئ بنسب مختلفة. عند اختبار فرض عدم (H_0) ضد الفرض البديل (H_1) عندها يمكن تقسيم القرارات الاحصائية إلى أربع فئات كما يلي:

1- أن يكون القرار الاحصائي هو قبول فرض عدم وهو في الحقيقة صحيحاً....وهذا قرار صائب. ويكون

$$P(\text{accept } H_0 | H_0 \text{ is correct}) = (1 - \alpha)$$

2- أن يكون القرار الاحصائي هو رفض فرض عدم وهو في الحقيقة صحيحاً....وهذا قرار خاطئ. (هذا

الخطأ من النوع الأول : فيكون احتمال رفض الفرض عدم وهو في الحقيقة صحيح (احتمال وقوع خطأ

$$P(\text{reject } H_0 | H_0 \text{ is correct}) = \alpha$$

3- أن يكون القرار الاحصائي رفض فرض عدم وهو في الحقيقة خطأ....وهذا قرار صائب. فيكون احتمال

$$P(\text{reject } H_0 | H_0 \text{ is uncorrected}) = (1 - \beta)$$

4- أن يكون القرار الاحصائي هو قبول فرض عدم وهو في الحقيقة خطأ... وهذا قرار خاطئ (الخطأ من

النوع الثاني : قبول (H_0) عندما يكون (H_0) خاطئ ويرمز لحجم هذا الخطأ بالرمز β).

$$P(\text{accept } H_0 | H_0 \text{ is uncorrected}) = \beta$$

و الخطأ من النوع الأول ألفا (α) يمكن أن يقلل بتقليل مستوى المعنوية (α). أما الخطأ من النوع الثاني بيتا (β) يمكن أن يقلل بزيادة حجم العينة، والخطأ من النوع الأول أكثر خطورة من الخطأ من النوع الثاني، حيث ان الخطأ من النوع الثاني لا يحدده الباحث بل يتم حسابه ويعتمد على خمسة عوامل وهي:

(1) القيمة الحقيقية لمعلمة المجتمع.

(2) قيمة (α) المختارة

(3) نوع الاختبار (ذات ذيلين أو ذيل واحد).

(4) الانحراف المعياري للعينة.

(5) حجم العينة.

قوة الاختبار Power Of The Test:

☛ قوة الاختبار هي عبارة عن قدرة الاختبار على رفض الفرض الصفري عندما يكون في الحقيقة خاطئ

وتساوي: $\text{power of test} = (1 - \beta)$ (وهذه هي القوة في الاختبار)

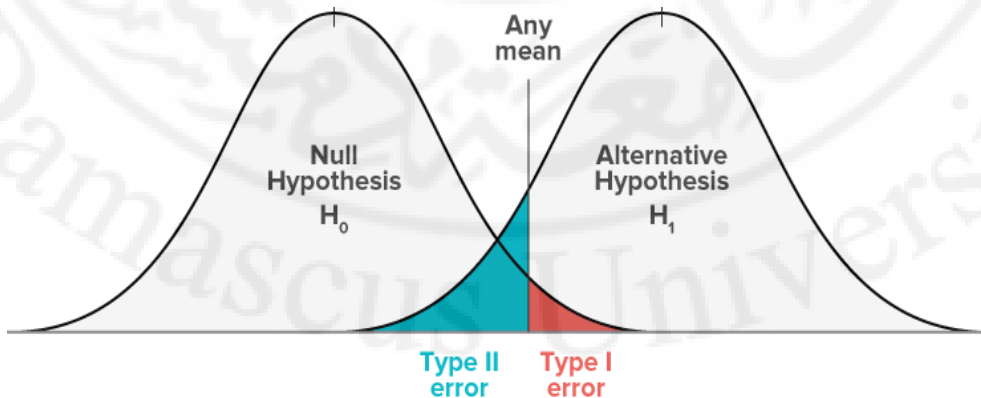
☛ وهناك علاقة ارتباط عكسية بين قوة الاختبار و الخطأ من النوع الثاني (β) أي اذا زاد الاول ينقص الثاني

والعكس صحيح، أي أن $\beta = 1 - \text{power of test}$.

العلاقة بين الخطأ من النوع الاول و الخطأ من النوع الثاني

العلاقة بين الخطأ من النوع الاول و الخطأ من النوع الثاني علاقة عكسية، أي كلما زاد الخطأ من النوع

الأول ينقص الخطأ من النوع الثاني (β) والعكس صحيح كما بالشكل (1-2):



شكل (1-2): العلاقة بين الخطأ من النوع الاول و الخطأ من النوع الثاني.

2-5- خطوات إجراء اختبار الفروض الإحصائية:

الاختبار الإحصائي قد يكون متعلقاً بعينة واحدة أو عينتين أو أكثر ولذلك يجب أن يمر الاختبار بعدة خطوات يمكن إيجازها في التالي:

- 1- صياغة الفروض الإحصائية حول معالم مجتمع و تحديد مستوى المعنوية.
- 2- حساب قيمة احصاء الاختبار بالاعتماد على عينة مسحوبة من مجتمع الدراسة.
- 3- تحديد منطقتي القبول والرفض اعتماداً على القيمة الجدولية لتوزيع بيانات المجتمع (Z أو t أو F).
- 4- المقارنة واتخاذ القرار بقبول أو رفض الفرض العدمي بناءً على موقع قيمة احصاء الاختبار بالنسبة لمناطق الرفض.

أولاً: صياغة الفرضيات الإحصائية:

الفكرة الأساسية في اختبار الفروض هي تقسيم المساحة تحت المنحنى إلى منطقتين: أحدهما تسمى منطقة القبول أي منطقة قبول الفرض العدمي. والأخرى تسمى منطقة الرفض، أي منطقة رفض الفرض العدمي والتي تسمى أحياناً بالمنطقة الحرجة (Critical region)، والنقطة الجديرة بالملاحظة هنا هي أن منطقة القبول تمثل مساحة قدرها يساوي درجة الثقة $(1-\alpha)$ ، بينما تمثل منطقة الرفض مساحة تساوي مستوى المعنوية (α) . بفرض أن (θ) معلمة مجتمع المجهولة وكانت (θ_0) تمثل قيمة معلمة المجتمع المفترضة (المزعومة) يمكن صياغة الفرضيات على إحدى الحالات الثلاثة الموضحة بالجدول (2-2):

جدول (2-2): حالات صياغة الفروض.

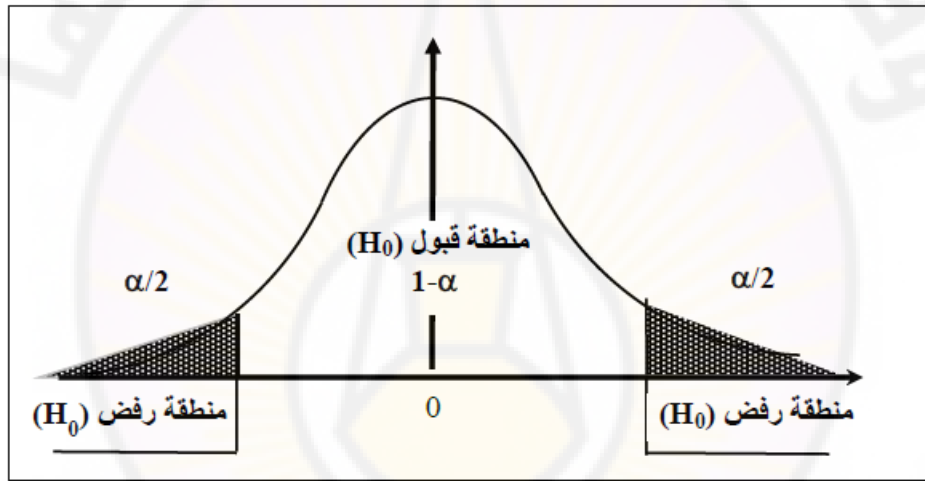
الحالة الأولى	الحالة الثانية	الحالة الثالثة
$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0 \end{cases}$	$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta > \theta_0 \end{cases}$	$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta < \theta_0 \end{cases}$
اختبار الفرضيات ذو الطرفين	اختبار الفرضيات ذو طرف واحد نحو اليمين	اختبار الفرضيات ذو طرف واحد نحو اليسار

إذا كان الفرض البديل يأخذ شكل "لا يساوي" فإن منطقة الرفض تكون موزعة على طرفي المنحنى بالتساوي، ويسمى الاختبار في هذه الحالة "اختبار الطرفين"

الحالة الأولى: اختبار الفرضيات ذو الطرفين (ذيلين):

تكون صياغة الفرضيات في هذه الحالة على الشكل: $\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta \neq \theta_0 \end{cases}$

في هذه الحالة يأخذ الفرض البديل شكل "لا يساوي" أي أن معلمة المجتمع لا تساوي القيمة المدعاة. وهنا تكون منطقة الرفض موزعة على طرفي المنحنى بالتساوي ومساحة كل طرف $(\alpha/2)$ ، أي أن مساحة منطقة الرفض (α) ، بينما تكون منطقة القبول في المنتصف ومساحتها $(1-\alpha)$ ، ويسمى الاختبار في هذه الحالة "اختبار الطرفين" كما هو مبين بالشكل (2-2):



شكل (2-2): التمثيل البياني لاختبار الفرضيات ذو طرفين.

مثلاً إذا زعمت شركة انتاج المصابيح الكهربائية أن متوسط عمر المصباح هو (1000) ساعة عمل، ومن أجل التأكد من هذا الإعاء تم أخذ عينة من هذه المصابيح وتم قياس عمرها الافتراضي ومن ثم يمكن صياغة الفرضيات حول علاقة متوسط العينة بمتوسط المجتمع (μ) كما يلي:

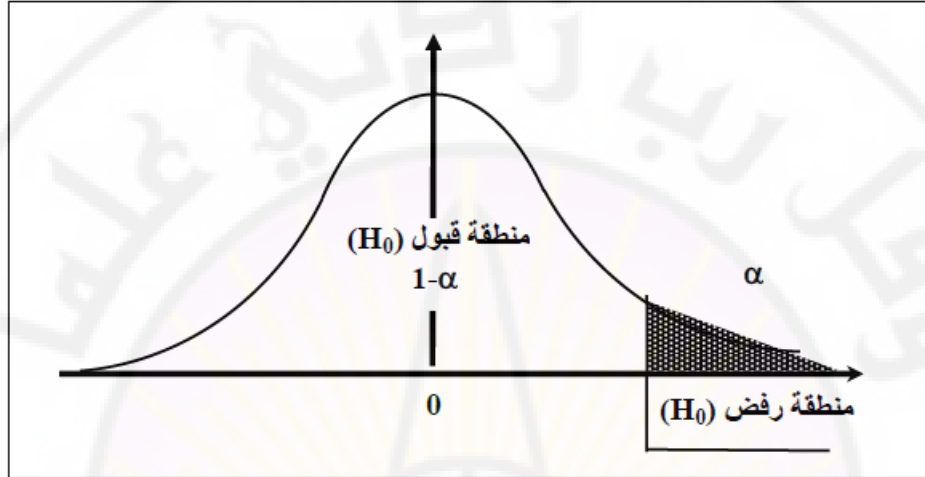
☉ الفرض العدمي: $(H_0: \mu = 1000)$ يعني أن متوسط عمر المصباح فعلاً يساوي 1000 ساعة عمل.

☉ الفرض البديل: $(H_1: \mu \neq 1000)$ يعني أن متوسط عمر المصباح لا يساوي 1000 ساعة عمل.

الحالة الثانية: اختبار الفرضيات ذو طرف واحد نحو اليمين:

تكون صياغة الفرضيات في هذه الحالة على الشكل: $\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta > \theta_0 \end{cases}$

في هذه الحالة الفرض البديل يأخذ شكل "أكبر من" أي أن معلمة المجتمع أكبر من القيمة المدعاة، هنا تكون منطقة الرفض مركزة بالكامل في الطرف الأيمن للمنحنى ومساحتها تساوي (α) ، ويسمى الاختبار في هذه الحالة اختبار ذو طرف واحد نحو اليمين، والذي يمثلته الشكل (3-2):



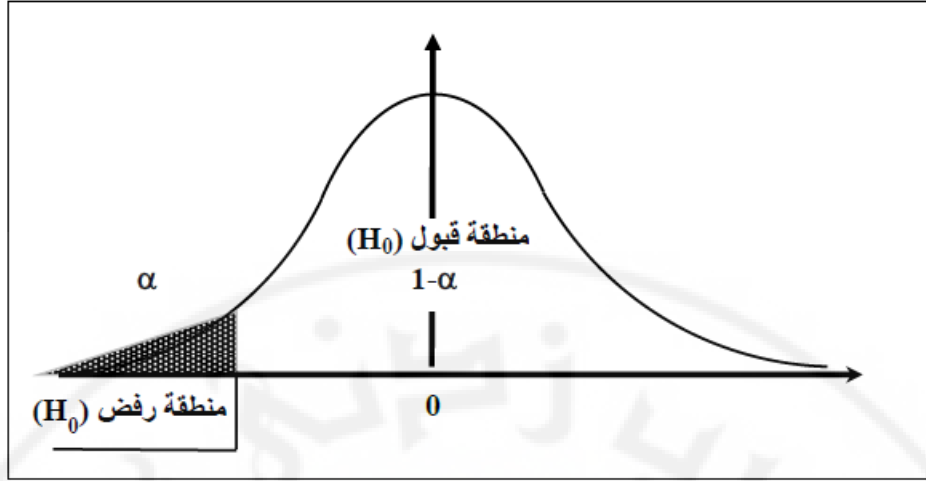
شكل (3-2): التمثيل البياني لاختبار الفرضيات ذو طرف واحد نحو اليمين.

وبالطريقة نفسها لمثال المصابيح السابق يتم صياغة الفروض، أي يكون الفرض العدمي نفسه، بينما الفرض البديل هو $(H_1: \mu > 1000)$ بمعنى أن عمر المصباح أكبر من (1000) ساعة. وبالتالي فإن مستوى المعنوية والذي يساوي مثلاً 5% مركز في الطرف الأيمن من المنحنى.

الحالة الثالثة: اختبار الفرضيات ذو طرف واحد نحو اليسار:

تكون صياغة الفرضيات في هذه الحالة على الشكل: $\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta < \theta_0 \end{cases}$

في هذه الحالة الفرض البديل يأخذ شكل "أصغر من" أي أن معلمة المجتمع أصغر القيمة المدعاة، هنا تكون منطقة الرفض مركزة بالكامل في الطرف الأيسر للمنحنى ومساحتها تساوي (α) . ويسمى الاختبار في هذه الحالة اختبار ذو طرف واحد نحو اليسار، والذي يمثلته الشكل (4-2):



شكل (2-4): التمثيل البياني لاختبار الفرضيات ذو طرف واحد نحو اليسار.

كذلك بالطريقة نفسها لمثال المصاييح السابق يتم صياغة الفروض، أي يكون الفرض العدمي نفسه، بينما الفرض البديل هو $(H_1: \mu < 1000)$ بمعنى أن عمر المصباح أقل من (1000) ساعة. وبالتالي فإن مستوى المعنوية والذي يساوي مثلاً 5% مركز في الطرف الأيسر المنحني.

ثانياً: حساب قيمة إحصائية الاختبار

إحصاء الاختبار (Test Statistic): هي القيمة المحسوبة من بيانات العينة باستخدام ما يعرف بدالة الاختبار.

ويتوقف شكل إحصائية الاختبار على العوامل التالية:

- ← توزيع المجتمع، وهل هو طبيعي أم لا، وهل تباينه معروف أم لا.
- ← حجم العينة، وهل هو كبير أم صغير.
- ← الفرض العدمي المراد اختباره وهل هو عن الوسط أو النسبة أو التباين أو الارتباط... الخ.

ثالثاً: تحديد منطقتي القبول والرفض:

تحدد منطقتي القبول والرفض اعتماداً على قيم جدولية من الجداول الإحصائية والتي تعتمد على:

أ- توزيع المعاينة (هل هو طبيعي أو t أو F ...)

ب- نوع الاختبار (ذو طرف واحد One Tail Test أو ذو طرفين Two Tail Test)

ج- ومستوى المعنوية (وهل هو 1% أو 5% أو غير ذلك).

حيث أن مساحة منطقة القبول تساوي قيمة درجة الثقة $(1-\alpha)$ ، بينما مساحة منطقة الرفض تساوي قيمة مستوى المعنوية (α) .

القيمة الحرجة: هي القيمة التي تفصل بين منطقتي القبول والرفض، وهي تمثل القيمة الجدولية من الجداول الاحصائية التي تمثل توزيع العينة.

رابعاً: المقارنة و اتخاذ القرار:

عملية المقارنة تتم بمقارنة قيمة إحصائية الاختبار المحسوبة بحدود منطقتي القبول والرفض، فإذا وقعت قيمة الإحصائية داخل منطقة القبول فإن القرار الاحصائي هو قبول الفرض الصفري، أما إذا وقعت قيمة الإحصائية في منطقة الرفض فإن القرار هو رفض الفرض الصفري وقبول الفرض البديل. مع ملاحظة أن القرار مرتبط بمستوى المعنوية المحدد، بمعنى أن القرار الاحصائي قد يتغير إذا تغير مستوى المعنوية المستخدم (وفي بعض الحالات قد لا يتغير القرار بتغير مستوى المعنوية).

وعادةً توجد طريقتين لاتخاذ القرار في الاختبارات الاحصائية وهي:

- 1- حساب احصاء الاختبار ومقارنتها بمنطقتي القبول والرفض بناء على نوع الاختبار ذو طرف واحد أو ذو طرفين ومن ثم اتخاذ القرار الاحصائي بقبول أو رفض الفرض العدمي.
- 2- حساب ما يسمى بالقيمة الاحتمالية (P-value) ويرمز لها في بعض البرامج الإحصائية بالرمز (Sig.) إذا كان الاختبار ذو طرف واحد تقارن قيمة (Sig. one-tail) بالقيمة (α)، أما إذا كان الاختبار ذو طرفين تقارن قيمة (Sig. 2-tailed) بالقيمة ($\alpha/2$). وفي كلتا الحالتين اذا كانت قيمة (Sig.) أصغر من مستوى المعنوية نرفض الفرض العدمي، أما اذا كانت أكبر نقبل الفرض العدمي.

2-6- تصنيف اختبارات الفروض الاحصائية:

يمكن تقسيم اختبارات الفروض وفقاً للمقاييس أو المعالم الاحصائية إلى ثلاث مجموعات من الاختبارات هي:

☉ اختبارات الفروض المتعلقة بالأوساط الحسابية

☉ اختبارات الفروض المتعلقة بالنسب

☉ اختبارات الفروض المتعلقة بالتباينات

ويمكن تصنيفها بشكل أكثر تفصيلاً كما يأتي:

اختبارات الفروض المتعلقة بالأوساط الحسابية

أولاً: اختبارات الفروض لمتوسط مجتمع μ (عينة واحدة)

■ تباين المجتمع معلوم

▪ تباين المجتمع مجهول وحجم العينة كبير

▪ تباين المجتمع مجهول وحجم العينة صغير

ثانياً: اختبارات الفروض للفرق بين متوسطي مجتمعين

▪ اختبار الفرضيات للفرق بين متوسطي مجتمعين مستقلين

لـ تباين المجتمعين معلومين (العينات صغيرة أو كبيرة)

لـ تباين المجتمعين مجهولين وحجمي العينتين كبيرين

لـ تباين المجتمعين مجهولين ولكنهما متساويين وحجم إحدى العينتين على الأقل صغيراً.

لـ تباين المجتمعين مجهولين ولكنهما غير متساويين وحجم إحدى العينتين على الأقل صغيراً.

▪ اختبار الفرضيات للفرق بين متوسطي مجتمعين غير مستقلين (مرتبطتين)

اختبارات الفروض المتعلقة بالنسب

أولاً: اختبارات الفروض لنسبة مجتمع

ثانياً: اختبارات الفروض للفرق بين نسبتي مجتمعين:

2-7- اختبارات الفروض المتعلقة بالأوساط الحسابية:

في هذا القسم سوف يتم دراسة اختبارات الفروض المتعلقة بالأوساط الحسابية للمجتمعات الاحصائية طبقاً لمعلومية تباين هذه المجتمعات وأحجام العينات المسحوبة منها. وفيما يلي نورد أهم الحالات التي يمكن تطبيق اختبارات الفروض فيها وهنا يمكن التمييز بين حالتين هما:

أولاً: اختبارات الفروض لمتوسط مجتمع μ (عينة واحدة)

ثانياً: اختبارات الفروض للفرق بين متوسطي مجتمعين

2-7-1 اختبارات الفروض لمتوسط مجتمع μ (عينة واحدة):

3- صياغة الفروض:

يمكن صياغة الفرضيات حول متوسط مجتمع على إحدى الحالات الثلاثة الموضحة بالجدول (2-3).

جدول (2-3): حالات صياغة الفروض لمتوسط مجتمع.

الحالة الاولى	الحالة الثانية	الحالة الثالثة
---------------	----------------	----------------

$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu < \mu_0 \end{cases}$	$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu > \mu_0 \end{cases}$	$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$
اختبار الفرضيات ذو طرف واحد نحو اليسار	اختبار الفرضيات ذو طرف واحد نحو اليمين	اختبار الفرضيات ذو الطرفين

2- حساب قيمة إحصائية الاختبار

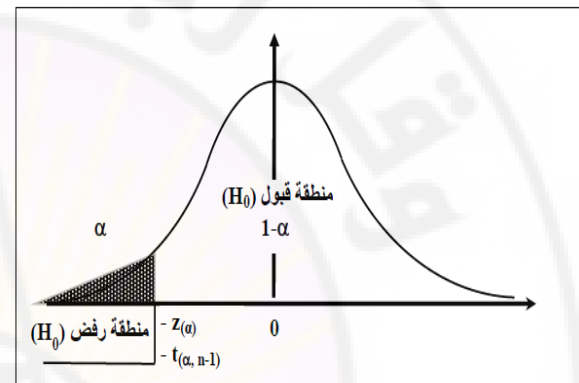
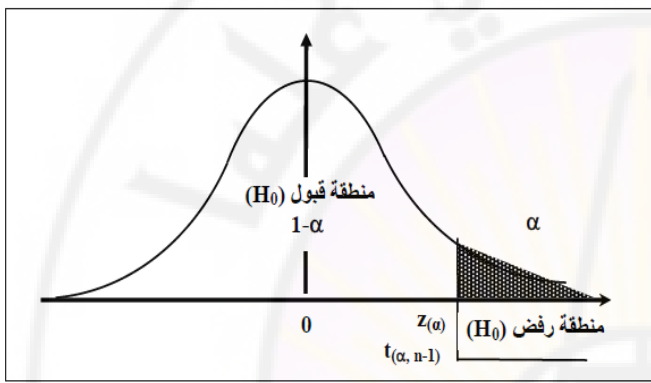
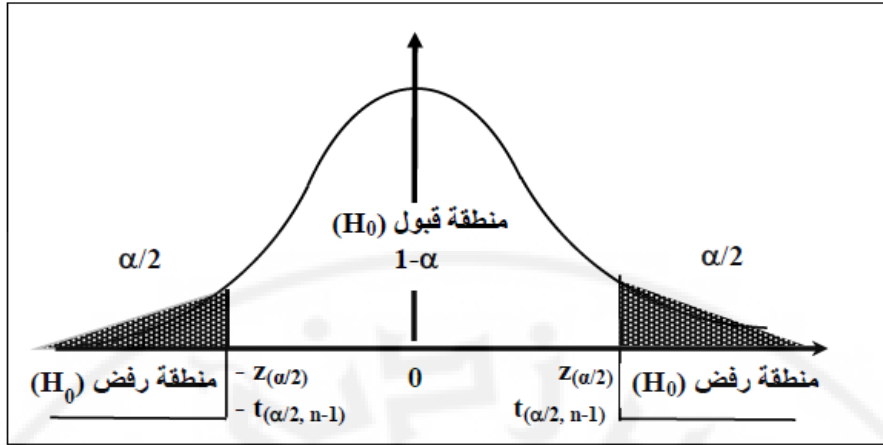
يتم حساب احصائية الاختبار التي تتحدد وفقاً لحجم العينة و لمعلومية تباين المجتمع أو عدم معلوميته باستخدام المتوسط الحسابي للعينة ($\bar{x}$) ومتوسط المجتمع عند صحة الفرض العدم (μ_0) كما هو مبين بالجدول (4-2):

جدول (4-2): طرائق حساب احصاءة الاختبار لمتوسط مجتمع.

تباين المجتمع	احصائية الاختبار (القيمة المحسوبة)	حجم العينة
معلوم	$Z = \frac{(\bar{x} - \mu_0)}{\frac{\sigma}{\sqrt{n}}}$	لا يشترط حجم العينة (غير مهم)
غير معلوم	$Z = \frac{(\bar{x} - \mu_0)}{\frac{S}{\sqrt{n}}}$	$n > 30$
	$t = \frac{(\bar{x} - \mu_0)}{\frac{S}{\sqrt{n}}}$	$n \leq 30$

3- تحديد منطقتي القبول والرفض:

يتم تحديد منطقتي القبول والرفض بواسطة القيم الجدولية التي يتم استخراجها من الجداول الاحصائية بناءً على مستوى المعنوية (α) وتوزيع المعاينة إما التوزيع الطبيعي المعياري (Z) أو توزيع t بدرجات حرية ($n-1$) كما هو مبين بالشكل (5-2):



شكل (2-5): التمثيل البياني لحالات اختبارات الفروض لمتوسط مجتمع.

4- المقارنة و اتخاذ القرار:

إذا وقعت قيمة احصائية الاختبار (القيمة المحسوبة) في منطقة القبول نقبل الفرض العدم، أما إذا وقعت في منطقة الرفض نرفض الفرض العدم ونقبل الفرض البديل.

مثال (2-3): تدعي هولندا أن متوسط الدخل الأسبوعي لمواطنيها يساوي (72 دولار) أي ($\mu_0=72$) بانحراف معياري مقداره ($\sigma=14$) دولار.

من أجل اختبار هذا الفرض (الادعاء) اختيرت عينة عشوائية حجمها ($n=49$) شخصاً من هولندا، فتبين أن الوسط الحسابي لدخول أفراد العينة الأسبوعية هو ($\bar{x}=75$) دولاراً.

المطلوب: طبق خطوات اختبارات الفروض للتحقق من صحة الفرض العدم بأن متوسط الدخل الأسبوعي لمواطني هولندا يساوي 72 دولاراً مقابل الفرض البديل أنه لا يساوي 72 وذلك بمستوى معنوية ($\alpha=0.05$).

الحل:

1- صياغة الفروض:

الفرض العدمي: ($H_0: \mu = 72$) يعني أن متوسط دخل الفرد في هولندا هو فعلاً (72) دولار اسبوعياً.

أي أنه لا يوجد فرق معنوي بين متوسط الدخل الحقيقي (μ) ومتوسط الدخل المزعوم (المدعاة) (μ_0) للمجتمع.

• الفرض البديل: ($H_1: \mu \neq 72$) يعني أن متوسط دخل الفرد في هولندا لا يساوي (72) دولار اسبوعياً.

2- **حساب احصائية الاختبار:** بما أن العينة كبيرة وتباين المجتمع معلوم فإن الإحصائية في حالة اختبار

$$Z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

$$\text{حيث: } (\sigma=14) \quad (\bar{x}=75) \quad (\mu_0=72) \quad (n=49)$$

وبالتعويض نحصل على :

$$Z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{75 - 72}{\frac{14}{\sqrt{49}}} = \frac{3}{2} = 1.5$$

أي أن قيمة احصاء الاختبار هي تساوي 1.5

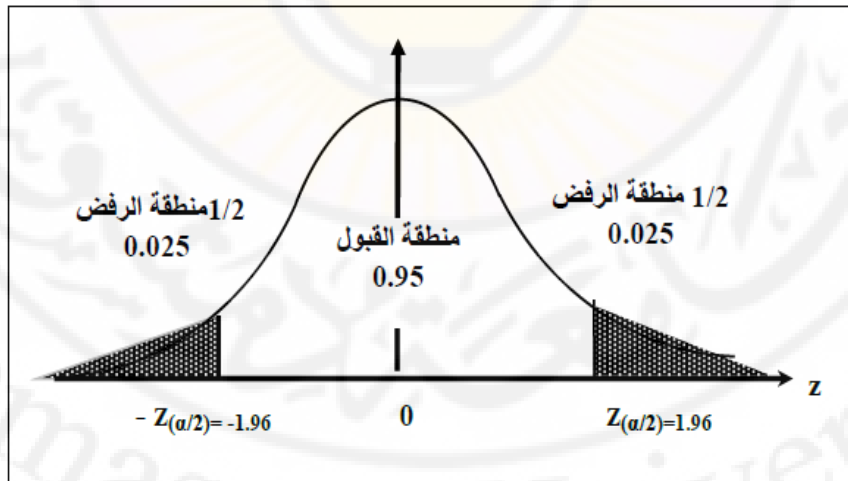
3- **تحديد منطقتي القبول والرفض:** يتم تحديد منطقتي القبول والرفض من القيم الجدولية التي يتم الحصول

عليها من التوزيع الطبيعي المعياري عند مستوى معنوية ($\alpha=0.05$) وبما أن الفرض البديل هو " لا

يساوي" أي هو اختبار ذو الطرفين إذا قيم Z الجدولية تكون كالتالي: $Z_{\alpha/2} = Z_{0.025} = 1.96$

وتكون منطقة القبول ضمن المجال $(-1.96, +1.96)$ كما في الشكل (2-2)

(6).



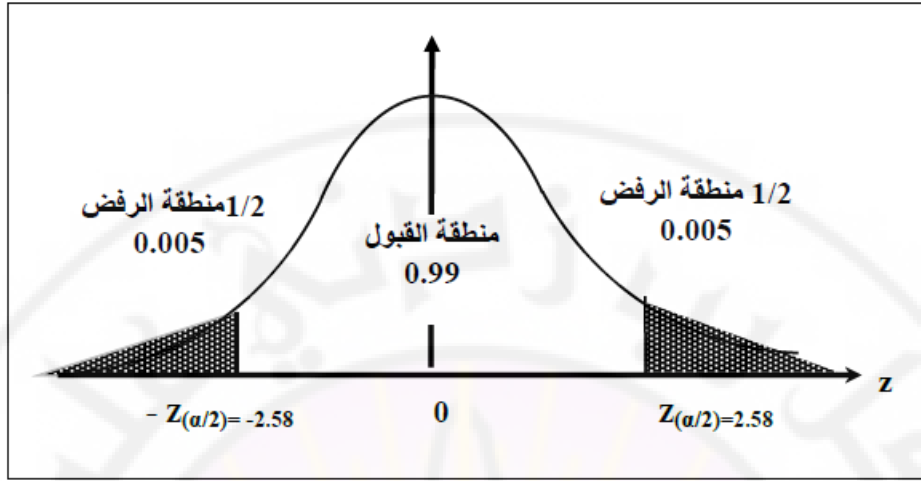
شكل (2-6): التمثيل البياني لمنطقتي القبول والرفض للمثال (2-3).

4- **المقارنة والقرار:** بمقارنة قيمة الإحصائية المحسوبة والتي تساوي 1.5 بحدود منطقتي القبول والرفض نجد

أنها تقع في منطقة القبول لذلك فإن القرار هو قبول الفرض العدمي بأن متوسط دخول الأفراد الأسبوعية في هذه

الدولة يساوي 72 دولاراً وذلك بمستوى معنوية 5 %.

ملاحظة: لو استخدمنا مستوى معنوية 1% بدلاً من 5% فإن حدود منطقتي القبول والرفض تصبح كما هي موضحة بالشكل (7-2).



شكل (7-2): التمثيل البياني لمنطقتي القبول والرفض للمثال (3-2) بمعنوية (0.01).

وبمقارنة قيمة الإحصائية 1.5 بحدود منطقتي القبول والرفض نجد أنها تقع في منطقة القبول أي أن القرار هو نفسه قبول الفرض العدمي ولن يتغير بل يتأكد باستخدام مستوى معنوية 1%. هل النتيجة ستكون نفسها لو كانت $(\bar{x}=76)$ ؟

مثال (4-2): إذا كان متوسط استهلاك الفرد السوري في عام 2000 من الدجاج حسب تقارير وزارة الصحة هو (12) كيلوجرام بانحراف معياري $(\sigma=6)$ كيلوجرامات. أجرى أحد الباحثين دراسة في عام 2020 على عينة حجمها $(n=49)$ فرداً فوجد أن متوسط الاستهلاك للفرد هو $x=14$ كيلوجرام. **المطلوب:**

هل تشير الدراسة الحالية أن متوسط الاستهلاك ارتفع عما كان عليه في عام 2000. أو بمعنى آخر هل متوسط الاستهلاك الحالي (2020) هو نفس الاستهلاك في عام 2000 أم أكبر.

الحل: من أجل حل هذا التمرين يجب أن نجيب على السؤال التالي:

هل يوجد فروق معنوية بين متوسط الاستهلاك القديم (12) و الاستهلاك الجديد (14) ؟

$$(n=49) \quad (\mu_0=12) \quad (\bar{x}=14) \quad (\sigma=6) \quad (\alpha=0.05)$$

1- صياغة الفروض:

⊖ **فرض العدم:** $(H_0: \mu=12)$ متوسط الاستهلاك الحالي نفس الاستهلاك القديم.

⊙ **الفرض البديل:** ($H_1: \mu > 12$) متوسط الاستهلاك الحالي أكبر من الاستهلاك القديم.

وبالتالي هنا اختبار الفرضيات ذو طرف واحد نحو اليمين:

2- حساب قيمة إحصائية الاختبار

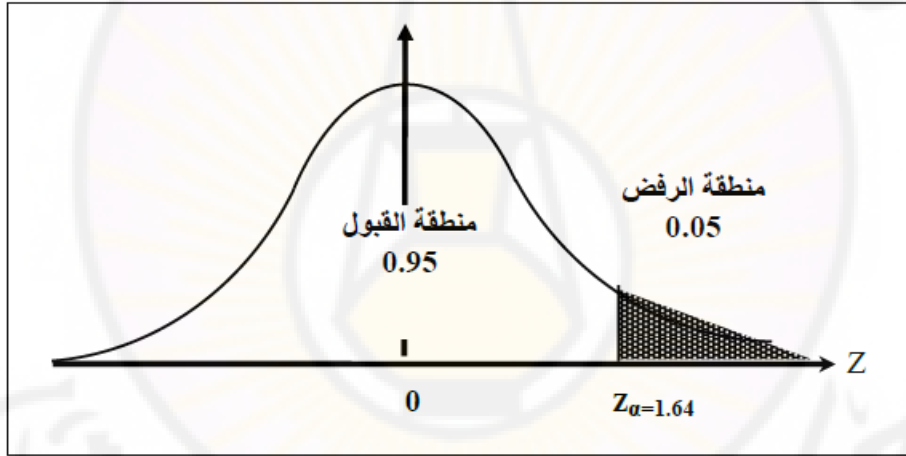
بما أنه تباين المجتمع معلوم و حجم العينة كبير تكون احصائية الاختبار كما يلي:

$$Z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{14 - 12}{\frac{6}{\sqrt{49}}} = 2.33$$

3- تحديد منطقتي القبول والرفض

يتم تحديد منطقتي القبول والرفض من التوزيع الطبيعي المعياري عند مستوى معنوية ($\alpha=0.05$) وبما أن

الفرض البديل هو " اكبر من " أي هو اختبار الفروض ذو الطرف الايمن اذاً قيم Z_α الجدولية التي تقع علي اليمين وتساوي ($Z_\alpha=1.645$) Z_α كما يبين الشكل (2-8).



شكل (2-8): التمثيل البياني لمنطقتي القبول والرفض للمثال (2-4).

4- المقارنة و اتخاذ القرار

بما أن قيمة احصاءة الاختبار المحسوبة (2.33) أكبر من القيمة الحرجة (1.645) كما بالشكل، فإنها تقع في منطقة الرفض. وبذلك نرفض فرض العدم ونقبل الفرض البديل أي أن متوسط استهلاك الفرد من لحوم الدواجن في الوقت الحالي قد ارتفع قياساً باستهلاك الفرد عام 2000 وذلك عند مستوى دلالة ($\alpha=0.05$).

مثال (2-5): اذا كان متوسط ضغط الدم للأشخاص الطبيعيون (75)، أدعى أحد الأطباء أنه من الآثار الجانبية لاستخدام عقار معين هو انخفاض ضغط الدم إلى أقل من الطبيعي (75).

ومن أجل التحقق من هذا الادعاء تم سحب عينة عشوائية حجمها (25) مريض ممن تناولوا العقار وتم قياس ضغطهم فوجد أن متوسط ضغط الدم للعينة ($\bar{x}=65$) وانحراف معياري قدره ($S=5$) عند مستوى معنوية ($\alpha=0.05$).

المطلوب: قم بتطبيق اختبارات الفروض الاحصائية من أجل تأكيد أو نفي ادعاء الطبيب بأن العقار يؤدي إلى خفض الضغط.

الحل: من أجل حل هذا التمرين يجب أن نجيب على السؤال التالي:

هل يوجد فروق معنوية بين متوسط الضغط قبل العقار (75) و متوسط الضغط بعد العقار (65) ؟

$$(\alpha=0.05) \quad (S=5) \quad (\mu_0=75) \quad (n=25) \quad (\bar{x}=65)$$

1- صياغة الفروض:

⊖ الفرض العدمي: ($H_0: \mu = 75$) استخدام العقار لا يؤدي إلى انخفاض ضغط الدم.

⊕ الفرض البديل: ($H_1: \mu < 75$) استخدام العقار يؤدي إلى انخفاض ضغط الدم.

في هذه الحالة اختبار الفرضيات ذو طرف واحد نحو اليسار:

2- حساب قيمة إحصائية الاختبار

بما أن تباين المجتمع (σ) مجهول وحجم العينة صغير ($n < 30$)، فعلياً أن نستخدم توزيع t (بدرجة

حرية $n - 1 = 24$) تكون إحصائية الاختبار كما يلي:

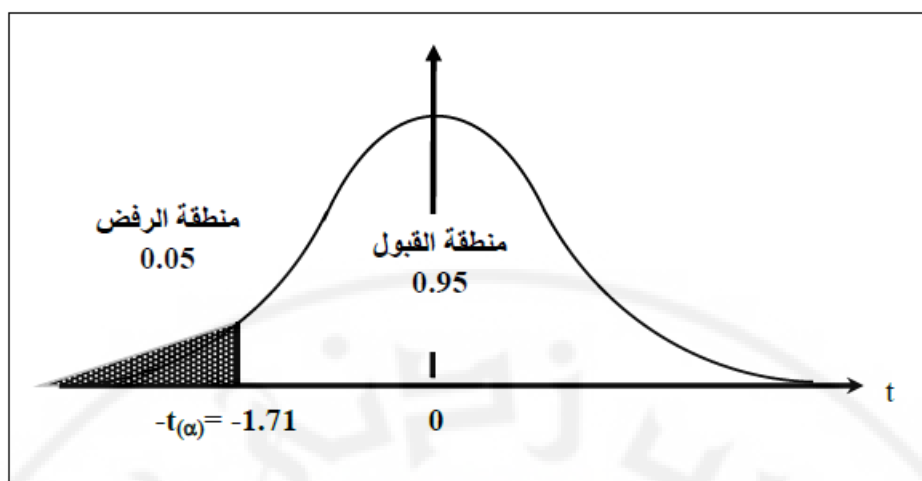
$$t = \frac{\bar{x} - \mu_0}{\frac{S}{\sqrt{n}}} = \frac{65 - 75}{\frac{5}{\sqrt{25}}} = -10$$

3- تحديد منطقتي القبول والرفض

يتم تحديد منطقتي القبول والرفض من توزيع (t) عند مستوى معنوية ($\alpha=0.05$) ودرجات حرية ($n-$

$1=24$) وبما أن الفرض البديل هو "أصغر من" أي هو اختبار الفروض ذو الطرف الأيسر إذاً قيمة (t) الجدولية

تساوي $t_{(\alpha, n-1)} = t_{(0.05, 24)} = 1.71$ كما بالشكل (2-9).



شكل (2-9): التمثيل البياني لمنطقتي القبول والرفض للمثال (2-5).

4- المقارنة و اتخاذ القرار

بما أن القيمة المحسوبة ($t=-10$) أصغر من القيمة الجدولية ($t=-1.71$) وهي تقع في منطقة الرفض، إذاً نرفض الفرض العدم القائل إن استخدام العقار لا يؤدي إلى انخفاض ضغط الدم، ونقبل الفرض البديل بأن استخدام العقار يؤدي إلى انخفاض ضغط الدم، وهذا يؤكد أدعاء الطبيب.

2-7-2- اختبارات الفروض للفرق بين متوسطي مجتمعين:

قد يرغب الباحث في إجراء اختبار فيما إذا كان متوسط الدخل في سوريا يساوي متوسط الدخل في لبنان، أو إجراء اختبار عما إذا كان متوسط عمر الفرد في سوريا يساوي متوسط عمر الفرد في مصر وهكذا... بمعنى آخر قد يرغب الباحث في إجراء اختبار عما إذا كان متوسط المجتمع الأول يساوي متوسط المجتمع الثاني، ومن أجل اختبار الفروض الاحصائية بين متوسطي مجتمعين يلجأ الباحث إلى سحب عينتين من المجتمعين وعندها يكون هدف الباحث هو تطبيق اختبارات الفروض للفرق بين متوسطي عينتين تمثلان مجتمعين وهنا يمكن التمييز بين نوعين من الاختبارات وفقاً لاستقلال أو ارتباط العينات:

- أولاً: اختبار الفرضيات للفرق بين متوسطي مجتمعين مستقلين.
- ثانياً: اختبار الفرضيات للفرق بين متوسطي مجتمعين غير مستقلين (مرتبطين).

أولاً: اختبار الفرضيات للفرق بين متوسطي مجتمعين مستقلين.

توجد أربع حالات مختلفة في حالة اختبار الفرضيات المتعلقة بالفرق بين متوسطي مجتمعين مستقلين

هي:

الحالة الأولى: تباين المجتمعين معلومين (العينات صغيرة أو كبيرة)

الحالة الثانية: تباين المجتمعين مجهولين وحجمي العينتين كبيرين

الحالة الثالثة: تباين المجتمعين مجهولين ولكنهما متساويين ($\sigma_1 = \sigma_2$) وحجم إحدى العينتين على الأقل صغيراً.

الحالة الرابعة: تباين المجتمعين مجهولين ولكنهما غير متساويين ($\sigma_1 \neq \sigma_2$) وحجم إحدى العينتين على الأقل صغيراً.

الحالة الأولى: تباين المجتمعين معلومين (العينات صغيرة أو كبيرة)

بافتراض أن كل مجتمع من المجتمعين له توزيع طبيعي وتباينه معلوم، وخصائص كل منهما كما يلي:

الخصائص المجتمع	المجتمع الأول	المجتمع الثاني
المتوسط	μ_1	μ_2
التباين	σ_1^2	σ_2^2
التوزيع	$X_1 \sim N(\mu_1, \sigma_1^2)$	$X_2 \sim N(\mu_2, \sigma_2^2)$

1- صياغة الفروض

في هذه الحالة تكون صياغة الفروض على الشكل التالي:

● الفرض العدمي : ($H_0: \mu_1 = \mu_2$) متوسط المجتمع الأول يساوي متوسط المجتمع الثاني.

● الفرض البديل : ($H_1: \mu_1 \neq \mu_2$) أن المتوسطين غير متساويين:

ويمكن للباحث استخدام أكبر من أو أقل من بدلاً من لا يساوي إذا كان لديه معلومات تشير إلى ضرورة ذلك.

$$\text{or: } H_1: \mu_1 < \mu_2$$

$$\text{or: } H_1: \mu_1 > \mu_2$$

2- حساب قيمة إحصائية الاختبار

إذا كان متوسط العينة الأولى ($\bar{x}_1$) مسحوبة من المجتمع الأول الذي متوسطه (μ_1) وتباينه (σ_1^2).

ومتوسط العينة الثانية ($\bar{x}_2$) مسحوبة من المجتمع الثاني الذي متوسطه (μ_2) وتباينه (σ_2^2) على الترتيب، تحسب

إحصائية الاختبار كما يلي:

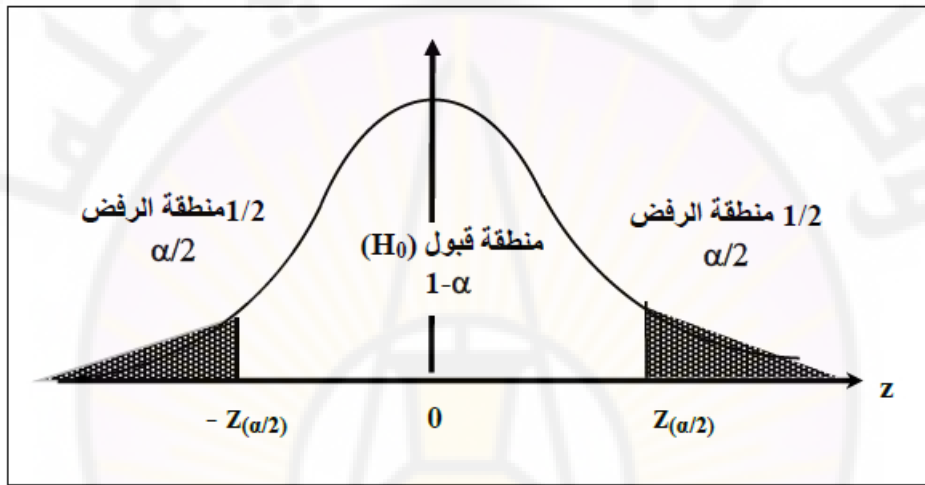
$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

وبما أننا نفترض أن ($H_0: \mu_1 = \mu_2$) في الفرض العدم فتصبح معادلة حساب إحصاء الاختبار كما يلي:

$$Z = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

3- تحديد منطقتي القبول والرفض

بما ان احصاءة الاختبار تتوزع وفقاً للتوزيع الطبيعي المعياري (Z) فإنه يتم تحديد منطقتي القبول والرفض بواسطة القيم الجدولية لتوزيع (Z) عند مستوى معنوية (α) وكذلك حسب صياغة الفرض البديل هل هو اختبار من طرفين أم من طرف واحد نحو اليمين أم من طرف واحد نحو اليسار، مثلاً إذا كان الاختبار من طرفين تكون منطقتي القبول والرفض موضحة بالشكل (10-2).



شكل (10-2): التمثيل البياني لاختبار الفروض للفرق بين متوسطي مجتمعين تباينهما معلوم.

4- المقارنة و اتخاذ القرار

نقارن قيمة الإحصائية بحدود منطقتي القبول والرفض، فإذا وقعت في منطقة القبول نقبل الفرض العدمي، وإذا وقعت في منطقة الرفض نرفض الفرض العدمي، ونقبل الفرض البديل.

مثال (2-6): في دراسة لإحدى المنظمات ادعت أن متوسط دخل الفرد في سوريا يساوي متوسط دخل الفرد بالأردن مع العلم أن تباين الدخل في سوريا يساوي ($\sigma_1^2=60$) دولار وتباين الدخل في الاردن يساوي ($\sigma_2^2=32$). ومن أجل التأكد من صحة هذا الادعاء تم سحب عينة من سوريا حجمها ($n_1=100$) شخص فكان متوسط الدخل في هذه العينة هو ($\bar{x}_1=35$) دولار، كذلك تم سحب عينة من الاردن حجمها ($n_2=80$) فكان متوسط الدخل فيها ($\bar{x}_2=29$) دولار.

المطلوب: طبق اختبارات الفروض لاختبار هل متوسط الدخل في سوريا يساوي متوسط الدخل بالأردن بمستوى معنوية 5% مقابل الفرض البديل أنهما غير متساويين.

الحل:

$$(\bar{x}_1=35) \quad (\bar{x}_2=29) \quad (n_1=100) \quad (n_2=80) \quad (\sigma_1^2=60) \quad (\sigma_2^2=32)$$

1- صياغة الفروض

- ☉ الفرض العدمي : $(H_0: \mu_1 = \mu_2)$ متوسط الدخل في سوريا يساوي متوسط الدخل بالأردن.
- ☉ الفرض البديل : $(H_1: \mu_1 \neq \mu_2)$ متوسط الدخل في سوريا لا يساوي متوسط الدخل بالأردن.

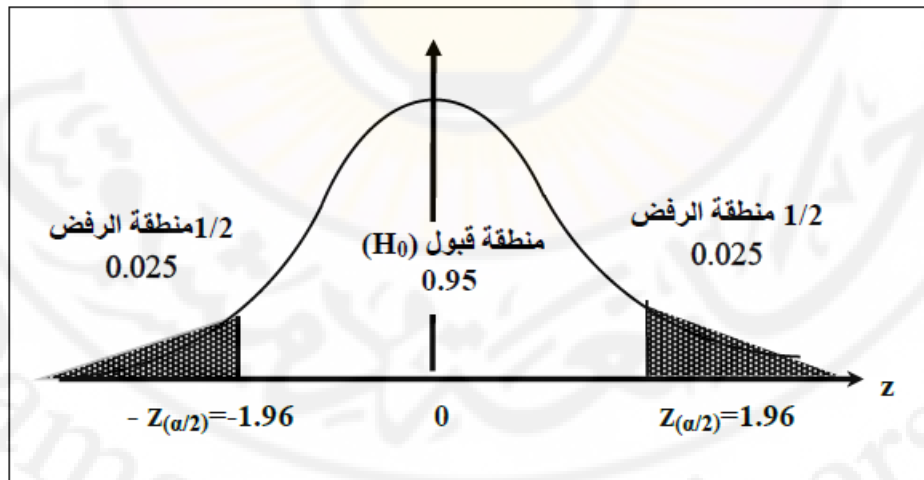
2- حساب قيمة إحصائية الاختبار

بما أن تباين المجتمعين معلومين، يتم حساب إحصائية الاختبار بالمعادلة الآتية:

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{35 - 29}{\sqrt{\frac{60}{100} + \frac{32}{80}}} = 6$$

3- تحديد منطقتي القبول والرفض

يتم تحديد منطقتي القبول والرفض بواسطة القيم الجدولية التي نحصل عليها من جدول التوزيع الطبيعي Z لأن العينات كبيرة، وبما أن الفرض البديل هو " لا يساوي " أي هو اختبار ذو الطرفين إذاً قيم Z الجدولية تكون كالتالي: $Z_{\alpha/2} = Z_{0.025} = 1.96$ عند مستوى معنوية $(\alpha=0.05)$.
وتكون منطقة القبول ضمن المجال $(-1.96, +1.96)$ كما في الشكل (2-11).



شكل (2-11): التمثيل البياني لمنطقتي القبول والرفض للمثال (2-6).

أي أن منطقة القبول تبدأ من -1.96 إلى +1.96 ومنطقة الرفض هي القيم التي أصغر من -1.96 والتي أكبر من +1.96.

4- المقارنة و اتخاذ القرار

بما أن قيمة الإحصائية (والتي تساوي 6) تقع في منطقة الرفض فإن القرار هو رفض الفرض العدمي وقبول الفرض البديل بمستوى معنوية 5% أي أننا نرفض الفرض القائل بأن متوسط دخل الفرد في سوريا يساوي متوسط دخل الفرد في الاردن وذلك بمستوى معنوية $(\alpha=0.05)$.

الحالة الثانية: باين المجتمعين مجهولين وحجمي العينتين كبيرين

إذا كان تباين المجتمعين (σ_1^2, σ_2^2) مجهولين وحجمي العينتين كبيرين فإن احصاء الاختبار يتم حسابها بنفس الطريقة في الحالة الأولى بعد التعويض عن تباين المجتمعين (σ_1^2, σ_2^2) بتباين العينتين (S_1^2, S_2^2) وتصبح الاحصاءة كما يلي:

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

وبما أننا نفترض أن $(H_0: \mu_1 = \mu_2)$ فتصبح معادلة حساب احصاء الاختبار كما يلي:

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

مثال (7-2): ادعى باحث اجتماعي أن متوسط ذكاء طلاب الجامعة الخاصة أقل من متوسط ذكاء طلاب جامعة دمشق. من أجل اختبار صحة هذا الفرض (الادعاء)، اختيرت عينة عشوائية من $(n_1=60)$ طالب من الجامعة الخاصة، فوجد أن متوسط ذكائهم $(\bar{x}_1=69)$ درجة وتباين قدره $(S_1^2=230)$ درجة، كذلك تم اختيار عينة عشوائية أخرى من $(n_2=85)$ طالب من جامعة دمشق فوجد أن متوسط ذكائهم $(\bar{x}_2=74)$ درجة وتباين قدره $(S_2^2=215)$ درجة.

المطلوب: طبق طريقة اختبارات الفروض لاختبار الفرض القائل أن متوسط ذكاء طلاب الجامعة الخاصة أقل من متوسط ذكاء طلاب جامعة دمشق وذلك بمستوى معنوية $(\alpha = 0.05)$.

الحل:

المجتمع الاول: الجامعة الخاصة: (μ_1) $(S_1^2=230)$ $(n_1=60)$ $(\bar{x}_1=69)$

المجتمع الثاني: جامعة دمشق: (μ_2) $(S_2^2=215)$ $(n_2=85)$ $(\bar{x}_2=74)$

1- صياغة الفروض

الفرض العدمي: $(H_0: \mu_1 = \mu_2)$ متوسط ذكاء طلاب الجامعة الخاصة يساوي متوسط ذكاء جامعة دمشق

© الفرض البديل: ($H_1: \mu_1 < \mu_2$) متوسط ذكاء طلاب الجامعة الخاصة أقل من متوسط ذكاء جامعة دمشق

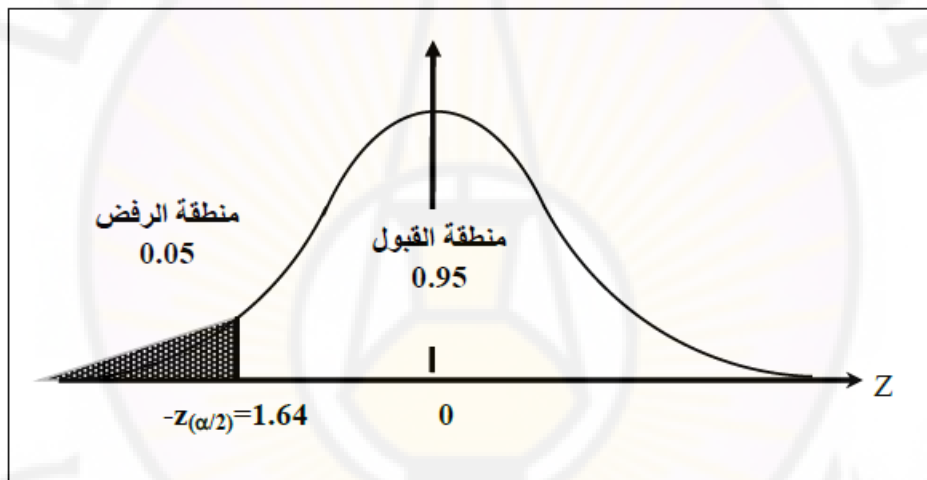
2- حساب قيمة إحصائية الاختبار

نحسب احصاء الاختبار من العلاقة:

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{69-74}{\sqrt{\frac{230}{60} + \frac{215}{85}}} = \frac{-5}{\sqrt{6.36}} = -1.98$$

3- تحديد منطقتي القبول والرفض

هنا اختبار الفرضيات ذو طرف واحد نحو اليسار، لذلك نحدد القيمة الحرجة التي تحدد منطقة القبول والرفض وذلك بإيجاد قيمة ($-Z_\alpha$) من جداول التوزيع الطبيعي عند مستوى معنوية معنوية ($\alpha = 0.05$)، أي إيجاد قيمة ($-Z_\alpha = -1.65$) وبالتالي الشكل (12-2) يوضح حدود منطقتي القبول والرفض.



شكل (12-2): التمثيل البياني لمنطقتي القبول والرفض للمثال (7-2).

4- المقارنة و اتخاذ القرار

بما أن قيمة Z المحسوبة ($Z = -1.98$) أصغر من قيمة Z الجدولية ($-Z_\alpha = -1.65$) أي أن قيمة احصاء الاختبار تقع في منطقة الرفض. وبذلك نرفض فرض العدم ونقبل الفرض البديل القائل أن متوسط ذكاء طلاب الجامعة الخاصة أقل من متوسط ذكاء جامعة دمشق.

الحالة الثالثة: تباين المجتمعين مجهولين ولكنهما متساويين ($\sigma_1^2 = \sigma_2^2$) وحجم احدى العينتين على الاقل صغيراً.

بفرض ان البيانات في المجتمع الاول والثاني تخضع للتوزيع الطبيعي و أن تباين المجتمعين (σ_1^2, σ_2^2) غير معلومين ولكنهما متجانسين أي $(\sigma_1^2 = \sigma_2^2)$ ، وحجم العينتين صغير فإن احصائية الاختبار تتبع توزيع (t) بدرجات حرية تساوي $(n_1 + n_2 - 2)$.

ويمكن تطبيق اختبار الفروض من أجل الحكم على معنوية الفرق بين متوسطي عينتين صغيرتين وفقاً للشروط السابقة باتباع الخطوات الآتية:

1- صياغة الفروض

⊖ الفرض العدمي : $(H_0: \mu_1 = \mu_2)$ متوسط المجتمع الأول يساوي متوسط المجتمع الثاني.

⊖ الفرض البديل : $(H_1: \mu_1 \neq \mu_2)$ متوسطي المجتمعين غير متساويين.

2- حساب قيمة إحصائية الاختبار

احصائية الاختبار تتبع توزيع (t) بدرجات حرية تساوي $(n_1 + n_2 - 2)$ وبفرض أن تباين المجتمع الاول يساوي تباين المجتمع الثاني تحسب الاحصاءة من المعادلة الآتية:

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{S_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

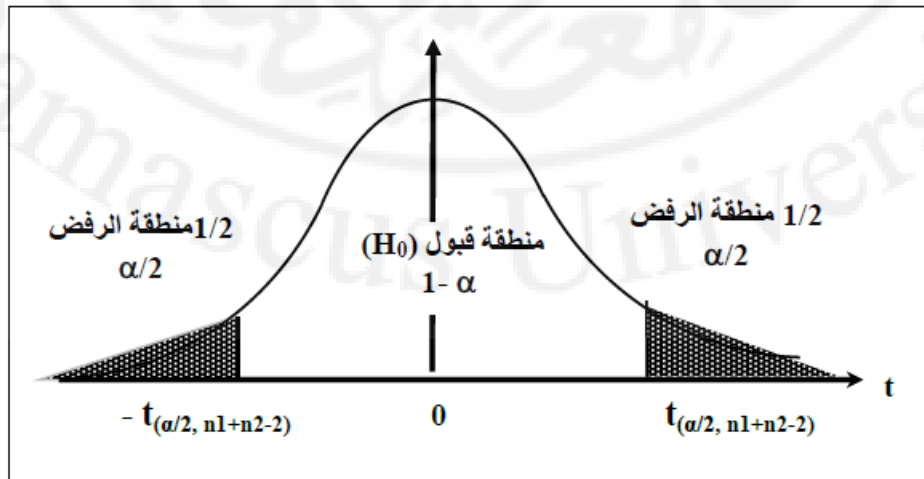
حيث (S_p^2) يمثل التباين المشترك للعينتين ويحسب بالمعادلة الآتية:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

3- تحديد منطقتي القبول والرفض

يتم تحديد منطقتي القبول والرفض بالقيم الحرجة التي نستخرجها من جداول التوزيع (t) عند درجات

حرية تساوي $(n_1 + n_2 - 2)$ وعند مستوى معنوية يساوي $(\alpha/2)$ في هذه الحالة لان الاختبار من طرفين كما في الشكل (2-13).



شكل (2-13): شكل اختبار الفروض للفرق بين متوسطي مجتمعين تباينهما مجهولان ومتساويان والعينتان صغيرتان

4- المقارنة و اتخاذ القرار

نقارن قيمة الإحصائية المحسوبة بحدود منطقتي القبول والرفض، فإذا وقعت في منطقة القبول نقبل الفرض العدمي، وإذا وقعت في منطقة الرفض نرفض الفرض العدمي، ونقبل الفرض البديل.

مثال (2-8): ادعى باحث زراعي أن متوسط إنتاج شجرة الزيتون في محافظة حماة يساوي متوسط إنتاج شجرة الزيتون في محافظة حمص.

من أجل اختبار صحة هذا الفرض (الادعاء)، اختيرت عينة من $(n_1=10)$ اشجار زيتون من حماة، فوجد أن متوسط انتاجهم $(\bar{x}_1=28)$ كغ وتباين قدره $(S_1^2=50)$ ، كذلك تم اختيار عينة عشوائية أخرى من $(n_2=10)$ اشجار زيتون من حمص فوجد أن متوسط انتاجهم $(\bar{x}_2=26)$ كغ وتباين قدره $(S_2^2=30)$.

المطلوب: طبق طريقة اختبارات الفروض لاختبار الفرض القائل إن متوسط إنتاج شجرة الزيتون في محافظة حماة يساوي متوسط إنتاج شجرة الزيتون في محافظة حمص بمستوى معنوية $(\alpha=0.05)$ مع العلم أن تبايني المجتمعين متجانسين $(\sigma_1^2=\sigma_2^2)$.

الحل:

المجتمع الاول: محافظة حماة:	(μ_1)	$(S_1^2=50)$	$(n_1=10)$	$(\bar{x}_1=28)$
المجتمع الثاني: محافظة حمص:	(μ_2)	$(S_2^2=30)$	$(n_2=10)$	$(\bar{x}_2=26)$

1- صياغة الفروض

الفرض العدمي : $(H_0: \mu_1 = \mu_2)$ متوسط إنتاج شجرة الزيتون في حماة يساوي متوسط إنتاج شجرة الزيتون في حمص.

الفرض البديل : $(H_1: \mu_1 \neq \mu_2)$ متوسط إنتاج شجرة الزيتون في حماة لا يساوي متوسط إنتاج شجرة الزيتون في حمص.

2- حساب قيمة إحصائية الاختبار

بما أن العينتين صغيرتين وأن تباين المجتمعين متجانسين، والمجتمعين طبيعيين، في هذه الحالة إن

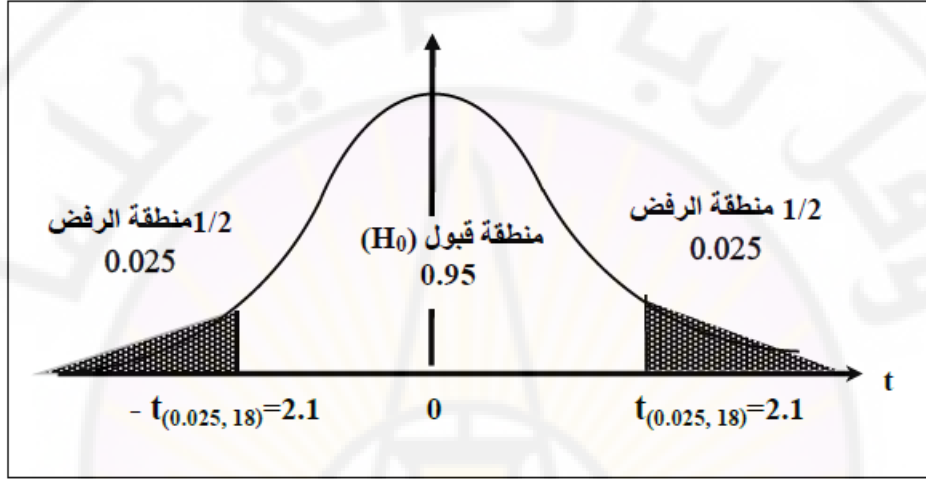
إحصائية الاختبار تتبع توزيع (t) بدرجات حرية تساوي $(n_1 + n_2 - 2)$ وتحسب كما يلي:

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} = \frac{9 \times 50 + 9 \times 30}{10+10-2} = \frac{740}{18} = 40$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{S_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{(28-26)}{\sqrt{40 \left(\frac{1}{10} + \frac{1}{10} \right)}} = \frac{2}{\sqrt{8}} = 0.7$$

3- تحديد منطقتي القبول والرفض

تتحدد منطقتي القبول والرفض باستخراج قيمة (t) الجدولية من جدول t عند درجات حرية تساوي $(n_1 + n_2 - 2)$ والتي تساوي 18 أي $t_{(0.025, 18)} = 2.101$ وتتمثل منطقتي القبول والرفض بالشكل (14-2).



شكل (14-2): التمثيل البياني لمنطقتي القبول والرفض للمثال (8-2).

أي أن منطقة القبول تبدأ من -2.101 وحتى + 2.101

4- المقارنة و اتخاذ القرار

بما أن قيمة الإحصائية تساوي 0.7 فإنها تقع في منطقة القبول وبالتالي فإن القرار هو قبول الفرض العدمي بأن متوسط إنتاج شجرة الزيتون في حماة يساوي متوسط إنتاج شجرة الزيتون في حمص وذلك بمستوى معنوي 5%.

الحالة الرابعة: تباين المجتمعين مجهولين ولكنهما غير متساويين $(\sigma_1^2 \neq \sigma_2^2)$ وحجم إحدى العينتين على الأقل صغيراً.

بفرض أن البيانات في المجتمع الأول والثاني تخضع للتوزيع الطبيعي و أن تباين المجتمعين (σ_1^2, σ_2^2) مجهولين (غير معلومين) ولكنهما غير متجانسين أي $(\sigma_1^2 \neq \sigma_2^2)$ ، وحجم العينتين صغير فإن إحصائية الاختبار تتبع توزيع (t) وتحسب كما يلي:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

وبما أننا نفترض أن $(H_0: \mu_1 = \mu_2)$ فتصبح معادلة حساب احصاءة الاختبار كما يلي:

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

حيث إن احصائية الاختبار تتبع توزيع (t) بدرجات حرية تساوي

$$df = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2 - 1}}$$

■ إذا كانت قيمة درجات الحرية (df) المحسوبة بالمعادلة السابقة تمثل عدد حقيقي، فإنه من الضروري التقريب لأقرب عدد صحيح.

■ من المعلوم أن درجات الحرية تكون دائماً أكبر من حجم العينة الأصغر وبالتالي يمكن أن نستخدم حجم أصغر عينة كدرجات حرية بدلاً من استخدام المعادلة السابقة لصعوبة حسابها أي أن:

درجات الحرية = العدد الأصغر لـ $(n_1 - 1)$ أو $(n_2 - 2)$.

وأخيراً يمكن تطبيق خطوات اختبار الفروض على هذه الحالة بالطريقة نفسها التي طبقت على الحالات السابقة.

مثال (9-2): إذا كنت مديراً لمصنع ملابس، وادعى مدير الإنتاج أن الوردية الليلية تنتج أكثر من الوردية النهارية، ومن أجل اختبار صحة هذا الادعاء قمت بأخذ عينات عشوائية مستقلة من (20) يوماً إنتاجياً مفترضاً أن المعاينة أخذت من مجتمعين طبيعيين مستقلين وتباينهما مجهولين ولكنهما غير متساويين $(\sigma_1^2 \neq \sigma_2^2)$ ، فكانت النتائج عند مستوى معنوية $(\alpha = 0.01)$ كما يلي:

	وردية النهار (قطعة)	وردية الليل (قطعة)	متوسط المجتمع
μ_1	$\bar{x}_2 = 245$	$\bar{x}_1 = 250$	
μ_2	$S_2 = 10$	$S_1 = 15$	
	الانحراف المعياري		

المطلوب: طبق طرق اختبارات الفروض الاحصائية لتثبت أو تنفي صحة ادعاء مدير الإنتاج بأن إنتاج الوردية الليلية أكثر من النهارية.

1- صياغة الفروض

(μ_1): يمثل متوسط انتاج الوردية الليلية.

(μ_2): يمثل متوسط انتاج الوردية النهارية.

⊖ الفرض العدمي : ($H_0: \mu_1 = \mu_2$) متوسط انتاج الوردية النهارية يساوي متوسط انتاج الوردية الليلية.

⊖ الفرض البديل : ($H_1: \mu_1 > \mu_2$) متوسط انتاج الوردية النهارية أقل من متوسط انتاج الوردية الليلية.

2- حساب قيمة إحصائية الاختبار

بما أن العينات صغيرة، وأن تباين المجتمعين غير متجانسين ($\sigma_1^2 \neq \sigma_2^2$)، وأن المجتمعين طبيعيان، فإن إحصائية الاختبار تتبع توزيع (t) وتحسب كما يلي:

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{(250 - 245)}{\sqrt{\frac{5^2}{20} + \frac{10^2}{20}}} = 2$$

3- تحديد منطقتي القبول والرفض

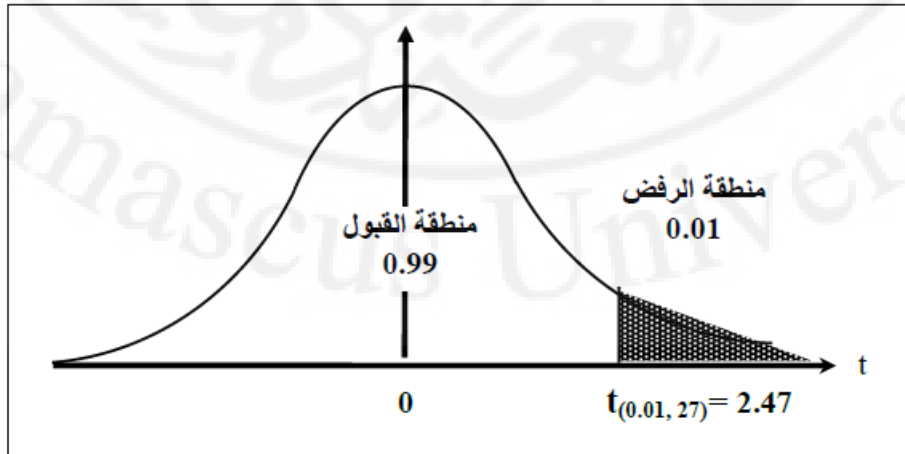
لتحديد منطقتي القبول والرفض باستخراج قيمة (t) الجدولية من جدول t عند درجات حرية تحسب من العلاقة:

$$df = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2 - 1}} = \frac{\left(\frac{5^2}{20} + \frac{10^2}{20}\right)^2}{\frac{\left(\frac{5^2}{20}\right)^2}{19} + \frac{\left(\frac{10^2}{20}\right)^2}{19}} \approx 27.94$$

بما أن درجات الحرية = 27.94 لذلك نقوم بتقريب لأقل عدد صحيح، أي يصبح (df=27) وبما أن الفرض

البديل أكبر من هذا يعني انه من طرف واحد نحو اليمين أي القيمة الجدولية تكون

$t_{(\alpha, 27)} = t_{(0.01, 27)} = 2.47$ وبالتالي تتحدد منطقتي القبول والرفض بالشكل (2-15).



شكل (2-15): التمثيل البياني لمنطقتي القبول والرفض للمثال (2-9).

4- المقارنة و اتخاذ القرار

بما أن قيمة الإحصائية تساوي 2 فإنها تقع في منطقة القبول وبالتالي فإن القرار هو قبول الفرض العدمي بأن أن متوسط انتاج الوردية النهارية يساوي متوسط انتاج الوردية الليلية وهذا ينفي ادعاء مدير الانتاج ان انتاج الوردية الليلية أكثر من الوردية النهارية

ثانياً: اختبار الفرضيات للفرق بين متوسطي مجتمعين غير مستقلين (مرتبطتين).

هذا الاختبار يسمى اختبار الفرضيات للفرق بين متوسطي عينتين غير مستقلتين (مرتبطتين) أو اختبار الفروض للفرق بين الأزواج المتقابلة (Testing Hypotheses for the Difference between Matched Pairs)

حيث يستخدم هذا النوع من الاختبارات للحكم على دلالة الفروق بين متوسطي عينتين مرتبطتين (Correlated Data)، مثل اختبار دلالة الفروق بين متوسط أداء الموظفين قبل التدريب وبعد التدريب. ولتوضيح هذا النوع من الاختبارات نورد المثال التالي:

يرغب باحثاً في معرفة فيما إذا كان إعطاء دورات تدريبية للموظفين الذين يعملون في قسم المبيعات حول أساليب التعامل مع الزبائن له دور فعال في زيادة عدد الوحدات المباعة من سلعة معينة. في هذه الحالة نأخذ بيانات حول عدد الوحدات المباعة من تلك السلعة قبل اتباع الدورة التدريبية وبعد اتباع الدورة، ثم نقوم بمقارنة متوسطي عدد الوحدات المباعة قبل وبعد اتباع الدورة التدريبية.

أيضاً على سبيل المثال يرغب باحث في معرفة ما إذا كان نظام غذائي معين في تخفيف الوزن لمجموعة من الأشخاص الذين يعانون من السمنة، في هذه الحالة نأخذ أوزانهم قبل وبعد اتباع النظام الغذائي. كذلك من التطبيقات العملية لهذا الاختبار الحالة الصحية لمجموعة مرضى قبل وبعد تناول دواء معين.

يطبق هذا الاختبار لمقارنة نتائج إجابات أو علامات عينة من الأشخاص قبل التجربة وبعدها، لذلك تسمى القيم الأولى بالقيم أو الدرجات القبلية (X)، وتسمى القيم الثانية بالقيم البعدية (Y)، ويمكن وضع النتائج في جدول خاص على شكل أزواج متقابلة، فنحصل على عينتين مرتبطتين من القيم (لأن كل زوج يعود لشخص واحد) كما هي مبينة بالجدول (2-5).

جدول (2-5): البيانات القبلية والبعدية لاختبار الفرضيات للفرق بين متوسطي مجتمعين غير مستقلين.

رقم الشخص	1	2	3	4		i		n	المتوسط
-----------	---	---	---	---	-------	---	-------	---	---------

	$\bar{X}$	x_n		x_i		x_4	x_3	x_2	x_1	الدرجات القبلية (X)
	$\bar{y}$	y_n		y_i		y_4	y_3	y_2	y_1	الدرجات البعدية (Y)
S_d	$\bar{d}$	d_n		d_i		d_4	d_3	d_2	d_1	عينة الفروقات d_i

ثم يتم حساب الفروقات بين قيمتي كل زوج من العلاقة: $d_i = x_i - y_i$

ويحسب متوسط وتباين والانحراف المعياري لهذه الفروقات من العلاقتين: $\bar{d} = \frac{1}{n} \sum d_i$

$$S_d^2 = \frac{\sum (d_i - \bar{d})^2}{n-1}$$

ولإجراء الاختبار حول الفرق بين العينتين يتم وضع الفرضيتين كما يلي:

3- صياغة الفروض

الفرض الصفري: $H_0: \mu_1 = \mu_2$ أو $H_0: \bar{D}_0 = 0$ حيث $\bar{D}_0 = \mu_1 - \mu_2$

الفرض البديل: $H_1: \mu_1 \neq \mu_2$ أو $\bar{D}_0 \neq 0$

يمكن للفرض البديل أن يأخذ إحدى الحالتين التاليتين أيضاً:

$$H_1: \bar{D}_0 < 0 \quad \text{أو} \quad H_1: \mu_1 < \mu_2$$

$$H_1: \bar{D}_0 > 0 \quad \text{أو} \quad H_1: \mu_1 > \mu_2$$

حيث:

d : هي متوسط الفرق بين أزواج المشاهدات.

S_d : هو الانحراف المعياري للفرق بين أزواج المشاهدات.

n : تمثل حجم العينة المسحوبة من المجتمع (عدد أزواج المشاهدات).

$\bar{D}_0$: هو الفرق بين متوسطي المجتمعين الأول والثاني

4- حساب قيمة إحصائية الاختبار

إن احصاءة الاختبار تتوزع وفقاً لتوزيع (t) بدرجات حرية (df=n-1) وتحسب بالعلاقة التالية:

$$t = \frac{\bar{d} - \bar{D}_0}{S_d / \sqrt{n}}$$

حيث $\bar{D}_0$ هي قيمة متوسط الفروقات المفترضة في المجتمع، وتؤخذ قيمتها الصفريّة من فرضية العدم H_0 وهي

$\bar{D}_0 = 0$ فتصبح معادلة حساب احصاءة الاختبار على الشكل الآتي:

$$t = \frac{\bar{d}}{S_d / \sqrt{n}}$$

3- تحديد منطقتي القبول والرفض

تتحدد منطقتي القبول والرفض باستخراج قيمة (t) الجدولية من جدول t عند درجات حرية تساوي (n - 1)

(1)

عند مستوى معنوية يساوي (α) ،

5- المقارنة و اتخاذ القرار

تتم مقارنة قيمة احصاء الاختبار المحسوبة بحدود منطقتي القبول والرفض، فإذا وقعت في منطقة القبول يتم قبول الفرض العدمي (H_0) ونقول بأنه لا يوجد فرق معنوي بين الدرجات القبلية والبعديّة، أما إذا كانت قيمة احصائية الاختبار أكبر من (t) الجدولية أي تقع في منطقة الرفض وبالتالي نرفض الفرض العدمي ونقبل الفرض البديل القائل أن هناك فرق معنوي بين القيم القبلية والبعديّة.

مثال (2-10): البيانات التالية تمثل نتائج تجربة أجريت على عشرة أشخاص لاختبار مدى فعالية نظام خاص من الغذاء لتخفيف الوزن، حيث تم قياس وزن الأشخاص قبل البدء في تطبيق النظام الغذائي وبعد تطبيقه بثلاثة شهور فكانت النتائج موضحة بالجدول (2-6).

جدول (2-6): أوزان الأشخاص قبل وبعد تطبيق النظام الغذائي.

رقم الشخص	1	2	3	4	5	6	7	8	9	10
الأوزان قبل (X)	95	110	90	94	107	93	89	120	103	92
الأوزان بعد (Y)	90	96	85	87	104	85	76	103	95	84

المطلوب: اختبار ما إذا كان نظام الغذاء كان فعالاً في تخفيف الوزن مستخدماً مستوى دلالة ($\alpha=0.05$)

الحل:

2- صياغة الفروض

بفرض أن $d_i = x_i - y_i$ تمثل عينة الفرق بين أزواج البيانات للعينتين و $\bar{D}_0$ هو الفرق بين متوسطي المجتمعين الأول والثاني، يمكن صياغة الفروض كما يلي:

الفرض الصفري: $H_0: \mu_x = \mu_y$ أو $H_0: \bar{D}_0 = 0$ لا يوجد فرق معنوي بين القيم القبلية والبعديّة.

الفرض البديل: $H_1: \bar{D}_0 > 0$ أي أنه يوجد فرق معنوي بين الدرجات القبلية والبعديّة.

مع ملاحظة أننا افترضنا أن $\bar{D}_0 > 0$ في الفرضية البديلة لاختبار ما إذا كان متوسط الوزن قبل اتباع النظام الغذائي يزيد عن متوسط الوزن بعد اتباع النظام الغذائي.

2- حساب قيمة إحصائية الاختبار

يتم حساب احصاء الاختبار من العلاقة $t = \frac{\bar{d} - D_0}{\frac{S_d}{\sqrt{n}}}$ وهذا يتطلب حساب متوسط الفروق ($\bar{d}$) والانحراف

المعياري (S_d) لأزواج المشاهدات كما هي مبينة بالجدول (7-2).

جدول (7-2): حساب الانحراف المعياري لأوزان الأشخاص قبل وبعد تطبيق النظام الغذائي.

رقم الشخص	1	2	3	4	5	6	7	8	9	10	
الأوزان قبل (X)	95	109	89	93	106	92	88	119	102	92	
الأوزان بعد (Y)	90	96	85	87	104	85	76	103	95	84	
الفروق ($d=x-y$)	5	13	4	6	2	7	12	16	7	8	80
($d-\bar{d}$)	-3	5	-4	-2	-6	-1	4	8	-1	0	0
($d-\bar{d}$) ²	9	25	16	4	36	1	16	64	1	0	172

$$\bar{d} = \frac{80}{10} = 8 \quad S_d = \sqrt{\frac{\sum (d-\bar{d})^2}{n-1}} = \sqrt{\frac{172}{9}} = 4.37$$

$$t = \frac{\bar{d}}{\frac{S_d}{\sqrt{n}}} = \frac{8}{4.37/\sqrt{10}} = 5.79$$

3- تحديد منطقتي القبول والرفض

تحدد منطقتي القبول والرفض باستخراج قيمة (t) الجدولية من جدول t عند درجات حرية تساوي ($n - 1$)

$$(1=9)$$

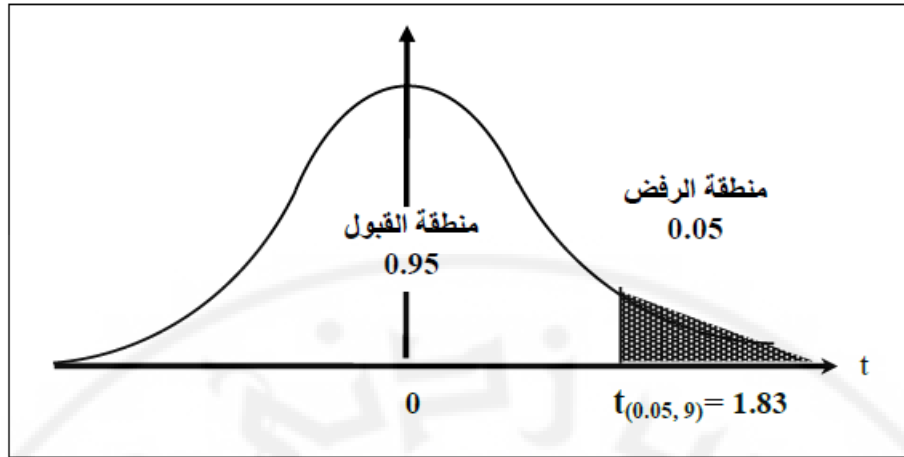
عند مستوى معنوية يساوي ($\alpha = 0.05$)، أي أن $t_{\alpha,9} = t_{0.05,9} = 1.83$

تحدد منطقتي القبول والرفض باستخراج قيمة (t) الجدولية من جدول t عند درجات حرية تساوي ($n - 1 = 9$)

عند مستوى معنوية يساوي ($\alpha = 0.05$) وبما أن الفرض البديل أكبر من هذا يعني أنه من طرف واحد نحو

اليمين أي القيمة الجدولية تكون أي $t_{(\alpha, n-1)} = t_{(0.05, 9)} = 1.83$ وبالتالي تتحدد منطقتي القبول

والرفض بالشكل (2-61).



شكل (2-16): التمثيل البياني لمنطقتي القبول والرفض للمثال (2-10).

5- المقارنة و اتخاذ القرار

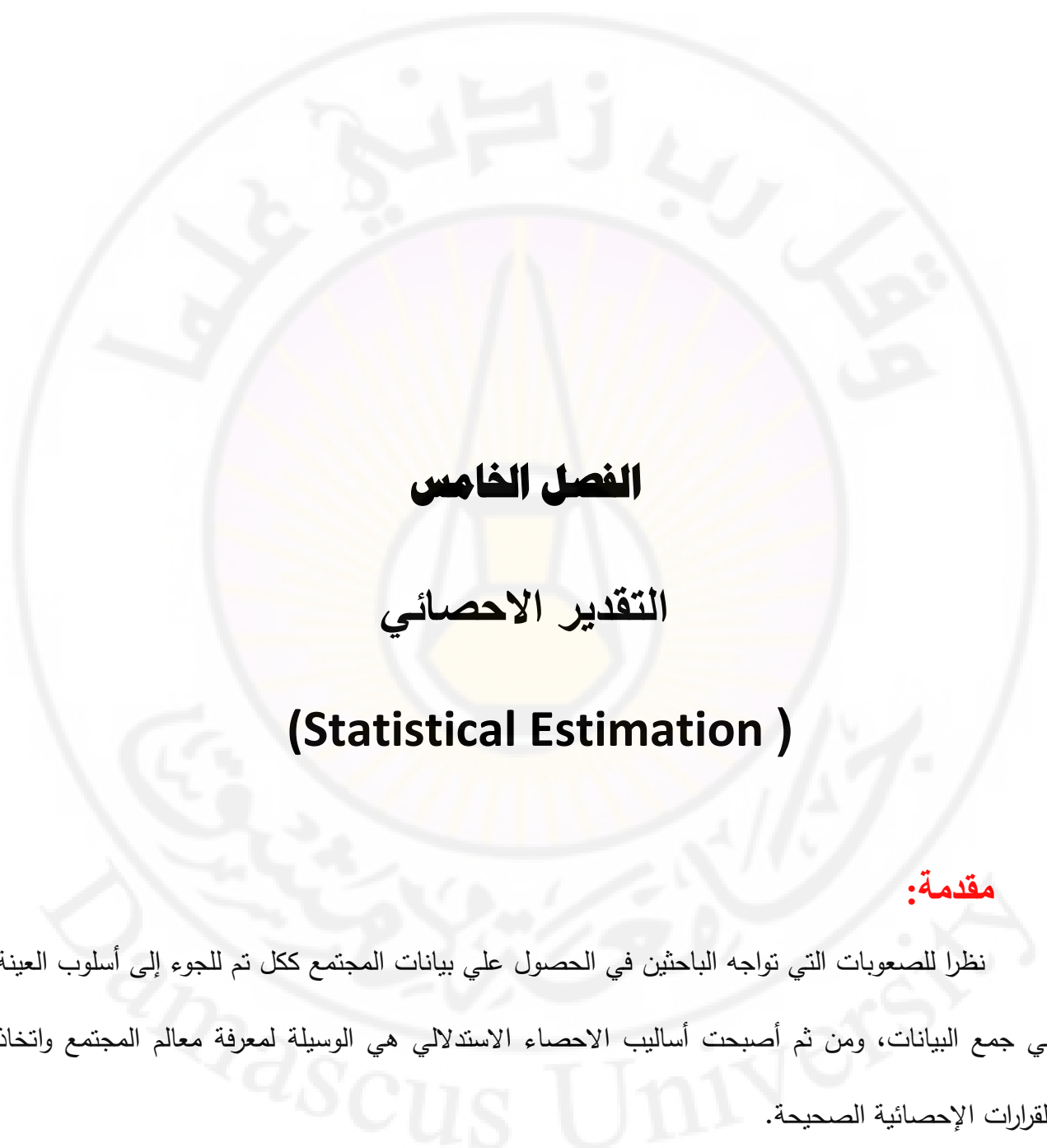
بما أن قيمة الإحصائية تساوي 5.79 فإنها تقع في منطقة الرفض وبالتالي فإن القرار هو رفض الفرض العدمي وقبول الفرض البديل بأنه يوجد فرق معنوي بين أوزان الاشخاص قبل وبعد النظام الغذائي، وهذا يقودنا إلى النتيجة بأن النظام الغذائي المتبع فعال في انقاص الوزن.

The logo of Damascus University is a large, faint watermark in the background. It is a circular emblem with a purple outer ring containing the university's name in Arabic and English. The center features a yellow sunburst with a stylized building or monument in the middle.

الفصل الخامس

التقدير الاحصائي

(Statistical Estimation)



الفصل الخامس

التقدير الاحصائي

(Statistical Estimation)

مقدمة:

نظرا للصعوبات التي تواجه الباحثين في الحصول علي بيانات المجتمع ككل تم اللجوء إلى أسلوب العينة في جمع البيانات، ومن ثم أصبحت أساليب الاحصاء الاستدلالي هي الوسيلة لمعرفة معالم المجتمع واتخاذ القرارات الإحصائية الصحيحة.

الإحصاء الاستدلالي (Statistical Inferences): هو الاستدلال واستنتاج المقاييس الإحصائية للمجتمع (والتي تعتبر مجهولة للباحث) من البيانات والمقاييس الإحصائية الخاصة بالعينة العشوائية والتي تعتبر معروفة أو متاحة للباحث.

فالإحصاء الاستدلالي يتم باستخدام عينة عشوائية مسحوبة من المجتمع وذلك لاستحالة التعامل مع المجتمع ككل. حيث ينقسم الإحصاء الاستدلالي إلى فرعين هما:

1- التقدير الإحصائي (Statistical Estimation): هو عملية تقدير أحد معالم المجتمع المجهولة

(مثل الوسط الحسابي أو الانحراف المعياري) بناءً على بيانات عينة مسحوبة من المجتمع ويتم التقدير ب (نقطة أو فترة ثقة). يستخدم أسلوب التقدير لتقدير معالم المجتمع إذا كان الهدف هو تحديد قيمة معلمة مجهولة (Unknown Parametric).

2- اختبارات الفروض (Hypotheses Testing): اختبارات الفروض تستخدم بهدف الوصول إلى قرار

بشأن رفض أو عدم رفض فرض إحصائي عن معلمة مزعومة (Hypothesized parametric).

التقدير الإحصائي (Statistical Estimation):

هو عملية تقدير أحد معالم المجتمع المجهولة (مثل الوسط الحسابي μ أو الانحراف المعياري σ أو النسبة P) بناءً على بيانات عينة مسحوبة من ذلك المجتمع. هذا يعني من أجل تقدير معالم المجتمع المجهولة مثل (μ, σ, P) نسحب عينة من هذا المجتمع ثم نحسب مقاييس العينة $(\bar{x}, \hat{P}, S)$ والتي تعتبر تقريباً لمعالم المجتمع الحقيقي.

مثل تقدير متوسط دخل الدولة، أو متوسط عمر الفرد، أو نسبة البطالة.

وهناك نوعان (أو أسلوبان) للتقدير الإحصائي هما:

■ الأول: تقدير النقطة (أو القيمة الواحدة) (Point Estimation)

في هذا النوع يتم استخدام عينة عشوائية مسحوبة من مجتمع الدراسة، ثم يتم حساب قيمة واحدة فقط (مقياس) تستخدم لتقدير المعلمة المجهولة.

■ الثاني: تقدير الفترة (أو فترة التقدير أو الثقة) (Interval Estimation)

يتم استخدام عينة عشوائية مسحوبة من مجتمع الدراسة، ثم يتم تحديد فترة معينة تقع فيها معلمة المجتمع المجهولة.

أولاً: التقدير بنقطة (Point Estimation):

هي عملية حساب قيمة واحدة من بيانات العينة، وتستخدم هذه القيمة الواحدة كتقدير لمعلمة المجتمع المجهولة. مثلاً لو أخذنا الوسط الحسابي للدخل في العينة كتقدير لمتوسط الدولة نكون قد حصلنا على تقدير نقطة لمتوسط دخل الدولة. وكمثال آخر لو أخذنا نسبة المدخنين في العينة كتقدير لهذه النسبة في المجتمع نكون حصلنا على تقدير نقطة للنسبة في المجتمع.

كذلك لو أخذنا الوسط الحسابي لأطوال الطلاب في عينة من طلاب جامعة دمشق والبالغ (172 سم)، يمكن اعتبار هذه القيمة كتقدير لمتوسط أطوال الطلاب في مجتمع جامعة دمشق نكون قد حصلنا على تقدير نقطة لمتوسط أطوال الطلاب.

الخطأ المعياري للتقدير بنقطة:

أثناء التقدير بنقطة قد تكون قيمة الإحصاء (المقياس المحسوب من العينة) المستخلصة من بيانات العينة لا تساوي قيمة المعلمة (μ) الحقيقية للمجتمع والمطلوب تقديرها.

الخطأ المعياري (Standard Error): هو الفرق المطلق بين قيمة الإحصاء المحسوبة من بيانات العينة وقيمة المعلمة الحقيقية المراد تقديرها للمجتمع.

والجدول الآتي يبين الأخطاء المعيارية لأهم التقديرات لمعالم المجتمع.

جدول لأهم تقدير بنقطة لبعض معلمات المجتمع والخطأ المعياري للتقدير.

المعلمة المراد تقديرها	الرمز	أفضل تقدير بنقطة	الرمز	الخطأ المعياري للتقدير
الوسط الحسابي للمجتمع	μ	الوسط الحسابي للعيينة	$\bar{x}$	$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$
النسبة في المجتمع	P	النسبة في العينة	$\hat{p} = \frac{r}{n}$	$\sigma_{\hat{p}} = \sqrt{\frac{\hat{p}\hat{q}}{n}}$
الفرق بين متوسطي مجتمعين مستقلين	$\mu_1 - \mu_2$	تباين المجتمعين معلومين	$(\bar{x}_1 - \bar{x}_2)$	$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
		تباين المجتمعين مجهولين والعينات كبيرة	$(\bar{x}_1 - \bar{x}_2)$	$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$

$\sigma_{\bar{x}_1 - \bar{x}_2} = S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$	$(\bar{x}_1 - \bar{x}_2)$	تباين المجتمعين مجهولين ومتجانسين والعينات صغيرة		
$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$	$(\bar{x}_1 - \bar{x}_2)$	تباين المجتمعين مجهولين وغير متجانسين والعينات صغيرة		
$\sigma_{\bar{x}_1 - \bar{x}_2} = \frac{S_d}{\sqrt{n}}$	$(\bar{x}_1 - \bar{x}_2)$			الفرق بين متوسطي مجتمعين مرتبطتين
$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$	$\hat{p}_1 - \hat{p}_2$	الفرق بين نسبتي عينتين	$P_1 - P_2$	الفرق بين نسبتي مجتمعين
$\sigma_{S^2}^2 = S^2 \sqrt{\frac{2}{n-1}}$	S^2	تباين العينة (S^2)	σ^2	تباين المجتمع

أهم أنواع التقديرات بنقطة:

أهم أنواع التقديرات بنقطة هي:

- 1- تقدير متوسط مجتمع (μ) غير معلوم بواسطة متوسط عينة ($\bar{x}$).
- 2- تقدير نسبة مجتمع (P) غير المعلومة بواسطة نسبة عينة ($\hat{P}$).
- 3- تقدير تباين مجتمع (σ^2) غير معلوم بواسطة تباين عينة (S^2).

• تقدير متوسط مجتمع (μ) بواسطة متوسط عينة ($\bar{x}$).

يتم استخدام الوسط الحسابي للعينة ($\bar{x}$) كتقدير نقطة لمتوسط المجتمع (μ).

ومن المتوقع أن لا تكون قيمة متوسط العينة مساوية لقيمة متوسط المجتمع، ولكن قريب منها. ويسمى الفرق

المطلق بين قيمة الإحصاء ($\bar{x}$) وقيمة المعلمة (μ) الفعلية المراد تقديرها باسم الخطأ المعياري.

الخطأ المعياري لتقدير متوسط مجتمع (Standard Error-SE):

أثناء تقدير متوسط مجتمع من بيانات عينة مسحوبة منه، يتم استخدام الوسط الحسابي للعينة ($\bar{x}$) كتقدير

نقطة لمتوسط المجتمع (μ)، ومن المتوقع أن لا تكون قيمة متوسط العينة مساوية لقيمة متوسط المجتمع، ولكن

قريب منها. ويسمى الفرق المطلق بين متوسط العينة ($\bar{x}$) و قيمة متوسط المجتمع الفعلية (μ) المراد تقديرها باسم الخطأ المعياري.

تعريف الخطأ المعياري لمتوسط العينة:

الخطأ المعياري لمتوسط العينة هو الانحراف المعياري لتوزيع مجتمع متوسطات العينات ويرمز له بالرمز $(\sigma_{\bar{x}})$.

تعريف آخر: الخطأ المعياري لمتوسط العينة هو انحراف متوسطات العينات عن متوسط مجتمعها وبحسب بالعلاقة:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

وغالباً ما تكون قيمة الانحراف المعياري للمجتمع (σ) غير معلومة، لذا نقدرها بقيمة الانحراف المعياري للعينة (S) عندما تكون كبيرة الحجم. فيصبح لدينا تقدير للخطأ المعياري للمتوسط باستخدام الانحراف المعياري للعينة بالصيغة التالية:

$$\sigma_{\bar{x}} = \frac{S}{\sqrt{n}}$$

مثال (1): أوجد الخطأ المعياري للوسط الحسابي لعينة حجمها ($n=49$) تم سحبها من مجتمع له انحراف معياري يساوي ($\sigma=14$).

$$\text{الحل: الخطأ المعياري لمتوسط العينة يساوي } \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{14}{\sqrt{49}} = 2$$

مثال (2): تم قياس مستوى الأداء على مهارة معينة لأفراد عينة حجمها ($n=36$)، فكان الوسط الحسابي والانحراف المعياري للعينة $\bar{x} = 30$ و $S=9$ على الترتيب. أوجد الخطأ المعياري لمتوسط العينة.

الحل: لأن قيمة الانحراف المعياري للمجتمع (σ) مجهولة، لذا نقدرها بقيمة الانحراف المعياري للعينة (S).

فيصبح لدينا تقدير للخطأ المعياري للوسط الحسابي للعينة يساوي:

$$\sigma_{\bar{x}} = \frac{S}{\sqrt{n}} = \frac{9}{\sqrt{36}} = 1.5$$

⊖ تقدير نسبة مجتمع (P) بواسطة نسبة عينة ($\hat{p}$)

من أجل تقدير نسبة مجتمع (P) من بيانات عينة مسحوبة منه، يتم استخدام نسبة العينة ($\hat{p}$) كتقدير نقطة لنسبة المجتمع (P). ومن المتوقع أن لا تكون قيمة نسبة العينة ($\hat{p}$) مساوية لنسبة المجتمع (P) ولكن قريب منها. ويسمى الفرق المطلق بين نسبة العينة ($\hat{p}$) وقيمة نسبة المجتمع الفعلية (P) المراد تقديرها باسم الخطأ المعياري لتقدير نسبة المجتمع.

⊖ الخطأ المعياري لتقدير نسبة مجتمع (Standard Error-SE):

⊖ تعريف الخطأ المعياري لنسبة عينة:

⊖ الخطأ المعياري لتقدير نسبة مجتمع (p): يساوي الفرق المطلق بين نسبة العينة ($\hat{p}$) المحسوبة من بيانات العينة وقيمة نسبة المجتمع الفعلية (P) المراد تقديرها ويرمز له بالرمز ($\sigma_{\hat{p}}$)، ويعطى

$$\sigma_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}}$$

: نسبة المجتمع (p)

($\hat{p}$): نسبة العينة

$$\hat{q} = 1 - \hat{p}$$

مثال (3): عينة عشوائية حجمها (n=144) ناخباً سحبت من إحدى المدن فوجد أن عدد المؤيدين لمرشح معين في العينة هو 60 ناخباً، أوجد الخطأ المعياري لنسبة هذه.

$$\hat{p} = \frac{60}{144} = 0.42$$

$$\sigma_{\hat{p}} = \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} = \sqrt{\frac{0.42 * 0.58}{144}} = 0.0041$$

⊖ تقدير تباين المجتمع (σ^2) غير المعلوم بواسطة تباين العينة (S^2)

يتم استخدام تباين العينة (S^2) كمقدر لتباين المجتمع المجهول (σ^2).

الخطأ المعياري لتقدير تباين المجتمع: الخطأ المعياري لتباين المجتمع يساوي $\sigma_{S^2}^2 = S^2 \sqrt{\frac{2}{n-1}}$.

ثانياً: التقدير بفترة (Interval Estimation):

وجدنا فيما سبق أن تقدير النقطة نادراً ما يساوي المعلمة التي نرغب في تقديرها، لذلك نلجأ إلى استخدام النوع الثاني وهو التقدير بالفترة.

التقدير بفترة: هي عملية استخدام بيانات العينة لتقدير فترة (مدى) تقع داخلها معلمة المجتمع المجهولة، وهذه الفترة تسمى فترة الثقة (Confidence intervals) ولها حدين أدنى وأعلى **يسميان بحدي الثقة** و تتضمن قيم عددها غير محدود أو لا نهائياً في أغلب الحالات.

يمكن حساب احتمال وقوع معلمة المجتمع في الفترة المقدرة، هذا الاحتمال يسمى درجة الثقة التي يرمز لها بـ $(1 - \alpha)$ وهي غالباً تكون 95% أو 99%.

درجة الثقة $(1 - \alpha)$ = احتمال وقوع معلمة المجتمع في الفترة المقدرة

ومكمل هذه القيمة يسمى مستوى المعنوية ويرمز له بالرمز (α) .

إذا كانت درجة الثقة 95% مثلاً حول متوسط مجتمع (μ) فهذا يعني أنه لدينا ثقة بنسبة 95% أن المتوسط (μ) للمجتمع يقع في هذه الفترة.

مثال (4): إذا قدرنا فترة الثقة لأطوال الطلاب كما يلي (160 185) سم بدرجة ثقة 95% فإن هذه الفترة فترة

الثقة لأطوال الطلاب بدرجة الثقة (95%) وإن احتمال أن يقع متوسط أطوال طلاب الحقيقة (متوسط المجتمع)

ان تقع ضمن فترة الثقة باحتمال (0.95)، مع ملاحظة أن هذه الفترة (160 185) تحتوي على عدد لا نهائي من الاطوال، بمعنى أن العدد لا يقتصر فقط على الأعداد الصحيحة والتي تشمل السنتيمترات، ولكنها تشمل أيضاً المليمترات وكسورها.. الخ.

أهم التقديرات بفترة:

- تقدير فترة الثقة لمتوسط مجتمع (μ):
- تقدير فترة الثقة للفرق بين متوسطي مجتمعين ($\mu_1 - \mu_2$):
- تقدير فترة الثقة لنسبة مجتمع (P).
- تقدير فترة الثقة للفرق بين نسبتي مجتمعين ($P_1 - P_2$).

تقدير فترة الثقة لمتوسط المجتمع (μ):

تقدير فترة الثقة لمتوسط المجتمع (μ) غير المعروف بواسطة الوسط الحسابي للعينة ($\bar{x}$). أي إيجاد الحدين الأدنى والاعلى اللذان يقع بينهما متوسط المجتمع (μ) باحتمال قدره $(1 - \alpha)$.

و هنا يمكن التميز بين ثلاث حالات:

- ← تقدير فترة الثقة لمتوسط مجتمع طبيعي تباينه (σ^2) معلوم.
- ← تقدير فترة الثقة لمتوسط مجتمع طبيعي تباينه (σ^2) مجهول وحجم العينة كبير.
- ← تقدير فترة الثقة لمتوسط مجتمع طبيعي تباينه (σ^2) مجهول وحجم العينة صغير.

أولاً: تقدير فترة الثقة لمتوسط مجتمع طبيعي تباينه (σ^2) معلوم.

إذا كان المجتمع المراد تقدير متوسطه يتوزع وفق التوزيع الطبيعي وتباينه (σ^2) معلوماً، فيتم سحب عينة عشوائية من ذلك المجتمع ويكون متوسطها يساوي ($\bar{x}$). في هذه الحالة إن فترة الثقة للوسط الحسابي للمجتمع تأخذ الشكل التالي:

فترة تقدير متوسط المجتمع = الوسط الحسابي للعينة ($\pm$) الخطأ المعياري للوسط $\times$ معامل الثقة $Z_{(\alpha/2)}$

$$\mu = \bar{x} \pm Z_{(\frac{\alpha}{2})} \sigma_{\bar{x}}$$

حيث إن القيمة $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ يمثل الخطأ المعياري لمتوسط العينة، فإن فترة تقدير الوسط تأخذ الشكل التالي:

$$\mu = \bar{x} \pm Z_{(\frac{\alpha}{2})} \frac{\sigma}{\sqrt{n}}$$

أو هذا الشكل:

$$\bar{x} - Z_{(\frac{\alpha}{2})} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + Z_{(\frac{\alpha}{2})} \frac{\sigma}{\sqrt{n}}$$

أي أنه احتمال أن يقع متوسط المجتمع الحقيقي (μ) في فترة التقدير يساوي درجة الثقة ($1 - \alpha$).

$$P \left[\bar{x} - Z_{(\frac{\alpha}{2})} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + Z_{(\frac{\alpha}{2})} \frac{\sigma}{\sqrt{n}} \right] = 1 - \alpha$$

كيفية تحديد معامل الثقة $Z_{(\alpha/2)}$

تحدد قيمة معامل الثقة حسب درجة الثقة ($1 - \alpha$) أو مستوى المعنوية (α) كما يلي:

معامل الثقة $Z_{(\alpha/2)}$	درجة الثقة ($1 - \alpha$)	مستوى المعنوية (α)
1.65	90%	10%
1.96	95 %	5%
2.58	99%	1%

هناك علاقة وثيقة بين درجة الثقة ($1 - \alpha$) و معامل الثقة، أي أنه بناءً على درجة الثقة المطلوبة يتحدد معامل

الثقة $Z_{(\alpha/2)}$ الذي يساوي الدرجة أو القيمة المعيارية للتوزيع الطبيعي المعياري والذي يمكن استخراجه من جدول

التوزيع الطبيعي المعياري، حيث تكون القيمتان (± 1.96) هما معاملات الثقة من أجل درجة ثقة (95%) بينما

القيمتان (± 2.58) تمثلان معاملات الثقة من أجل درجة ثقة (99%)، الجدول التالي يمثل أهم معاملات الثقة حسب درجات الثقة:

مثال (5): بفرض انه لدينا مجتمع أطوال طلاب جامعة دمشق يتوزع طبيعياً ومعروف الانحراف المعياري لأطوال الطلاب وهو (25) سم. تم سحب عينة عشوائية من الطلاب حجمها (81) طالب فكان متوسط طول الطالب (175) سم.

المطلوب: أوجد تقدير فترة الثقة 95% و فترة الثقة 99% لأطوال طلاب جامعة دمشق.

أو بصيغة أخرى: تقدير الوسط الحسابي للمجتمع (اطوال الطلاب) اذا كانت:

درجة الثقة 95% ب- درجة الثقة 99%.

الحل: من أجل الحالة الأولى بدرجة الثقة (95%)

نطبق العلاقة:

$$\bar{x} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$175 - 1.96 \frac{25}{9} \leq \mu \leq 175 + 1.96 \frac{25}{9}$$

$$169.56 \leq \mu \leq 180.44$$

هذا يعني أن متوسط المجتمع (أطوال الطلاب) يقع بالمجال (169.56, 180.44) سم، أي أننا واثقون باحتمال قدره (95%) بأن متوسط المجتمع يقع ضمن فترة الثقة أعلاه.

$$P(169.56 \leq \mu \leq 180.44) = 0.95$$

من أجل الحالة الثانية بدرجة الثقة (99%)

$$\bar{x} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$175 - 2.58 \frac{25}{9} \leq \mu \leq 175 + 2.58 \frac{25}{9}$$

$$167.83 \leq \mu \leq 182.17$$

$$P(167.83 \leq \mu \leq 182.17) = 0.99$$

هذا يعني أن متوسط المجتمع (أطوال الطلاب) يقع بالمجال (167.83, 182.17) سم، أي أننا واثقون باحتمال قدره (99%) بأن متوسط المجتمع يقع ضمن فترة الثقة أعلاه.

$$P(167.83 \leq \mu \leq 180.17) = 0.99$$

ثانياً: تقدير فترة الثقة لمتوسط مجتمع طبيعي تباينه (σ^2) مجهول وحجم العينة كبير.

إذا كان التباين أو الانحراف المعياري للمجتمع (σ) غير معروف وهو غالباً ما يحدث في الواقع، و إذا كان حجم العينة كبير أي ($n \geq 30$) فيمكن استخدام الانحراف المعياري للعينة (S) بدلاً من الانحراف المعياري للمجتمع (σ) وتصبح فترة تقدير الوسط الحسابي للمجتمع كما يلي:

$$\mu = \bar{x} \mp Z_{\left(\frac{\alpha}{2}\right)} \frac{S}{\sqrt{n}}$$

$$\bar{x} - Z_{\left(\frac{\alpha}{2}\right)} \frac{S}{\sqrt{n}} \leq \mu \leq \bar{x} + Z_{\left(\frac{\alpha}{2}\right)} \frac{S}{\sqrt{n}}$$

مثال (6): من أجل تقدير فترة الثقة لمتوسط الاستهلاك الشهري للأسرة في مدينة دمشق، تم اختيار عينة عشوائية حجمها (100) أسرة فكان متوسط الاستهلاك (250) دولار بانحراف معياري (25) دولار.

المطلوب: قدر فترة الثقة لمتوسط **المتوسط** الاستهلاك الشهري للأسرة في دمشق بدرجة ثقة 95%

الحل : بما أن فترة تقدير الوسط الحسابي للمجتمع هي:

$$\mu = \bar{x} \mp Z_{\left(\frac{\alpha}{2}\right)} \frac{S}{\sqrt{n}}$$

حيث: حجم العينة ($n=100$) الوسط الحسابي للعينة ($\bar{x} = 250$) و الانحراف المعياري للعينة ($S=25$).

وبما أن درجة الثقة هي (95%) فإن معامل الثقة $Z_{(\frac{\alpha}{2})} = 1.96$ وبالتالي فإن فترة الثقة لتقدير الوسط الحسابي للاستهلاك الشهري.

$$\bar{x} - Z_{(\frac{\alpha}{2})} \frac{S}{\sqrt{n}} \leq \mu \leq \bar{x} + Z_{(\frac{\alpha}{2})} \frac{S}{\sqrt{n}}$$

$$250 - 1.96 \frac{25}{10} \leq \mu \leq 250 + 1.96 \frac{25}{10}$$

$$245.1 \leq \mu \leq 254.9$$

أي أن الوسط الحسابي للاستهلاك الشهري للأسرة في مدينة دمشق يتراوح بين 245.1 دولاراً كحد أدنى، 254.9 كحد أعلى، وذلك بدرجة ثقة 95%.

مثال (7): مصنع لإنتاج المصابيح الكهربائية، اختير من إنتاجه عينة حجمها (100) مصباح لتقييم جودة الإنتاج، حيث حدد مدير المصنع معيار الجودة أن يتراوح عمر المصباح بين (1000) و(1100). إذا كان الوسط الحسابي لعمر المصباح في العينة المختارة (1200) ساعة وانحرافه المعياري (250) ساعة. قدر بدرجة ثقة 95% متوسط عمر المصباح من إنتاج المصنع كله.

الحل: حيث أن الانحراف المعياري للمجتمع σ مجهول نقدره باستخدام الانحراف المعياري للعينة.

$$n = 100 \quad \bar{x} = 1200 \quad S = 250 \quad 1 - \alpha = 0.95 \quad Z_{\frac{\alpha}{2}} = 1.96$$

$$\bar{x} - Z_{(\frac{\alpha}{2})} \frac{S}{\sqrt{n}} \leq \mu \leq \bar{x} + Z_{(\frac{\alpha}{2})} \frac{S}{\sqrt{n}}$$

$$1200 - 1.96 \frac{250}{10} \leq \mu \leq 1200 + 1.96 \frac{250}{10}$$

$$1151 \leq \mu \leq 1249$$

$$P(1151 \leq \mu \leq 1249) = 0.95$$

أي أن متوسط عمر المصباح من إنتاج المصنع كله يتراوح بين 1151 و1249 ساعة بدرجة ثقة 95%. وتدل هذا التقدير أن جودة المصابيح عالية من هذا المصنع.

ثالثاً: تقدير فترة الثقة لمتوسط مجتمع طبيعي تباينه (σ^2) مجهول وحجم العينة صغير.

إذا كان المجتمع المراد تقدير متوسطه يتوزع وفق التوزيع الطبيعي وتباينه (σ^2) مجهول، و تم سحب عينة عشوائية صغيرة حجمها $(n < 30)$ من ذلك المجتمع و كان متوسطها يساوي $(\bar{x})$ فإن فترة الثقة للوسط (μ) تعطى بالعلاقة الآتية:

$$\mu = \bar{x} \pm t_{\left(\frac{\alpha}{2}, n-1\right)} \frac{S}{\sqrt{n}}$$

$$\bar{x} - t_{\left(\frac{\alpha}{2}, n-1\right)} \times \frac{S}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\left(\frac{\alpha}{2}, n-1\right)} \times \frac{S}{\sqrt{n}}$$

أو بشكل آخر.

$$\left(\bar{x} - t_{\left(\frac{\alpha}{2}, n-1\right)} \times \frac{S}{\sqrt{n}}, \quad \bar{x} + t_{\left(\frac{\alpha}{2}, n-1\right)} \times \frac{S}{\sqrt{n}} \right)$$

مثال (8): إذا كان انتاج مصنع الأقلام يخضع للتوزيع الطبيعي، أخذت عينة عشوائية حجمها (25) قلم من

خط الانتاج فكان متوسط طول الخط الذي يكتبه القلم $(\bar{X} = 17)$ متر بانحراف معياري $(S=2)$ متر.

والمطلوب: أوجد فترة 95% ثقة للوسط الحسابي (μ) ؟

الحل : $\alpha/2 = 2.5\%$ $\alpha = 5\%$

$$\bar{x} - t_{\left(\frac{\alpha}{2}, n-1\right)} \times \frac{S}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\left(\frac{\alpha}{2}, n-1\right)} \times \frac{S}{\sqrt{n}}$$

$$17 - 2.15 \times \frac{2}{\sqrt{25}} \leq \mu \leq 17 + 2.15 \times \frac{2}{\sqrt{25}}$$

$$17 - 2.15 \times \frac{2}{\sqrt{25}} \leq \mu \leq 17 + 2.15 \times \frac{2}{\sqrt{25}}$$

$$16.17 \leq \mu \leq 17.86$$

$$P(16.17 \leq \mu \leq 17.86) = 0.95$$

تقدير فترة الثقة للفرق بين متوسطي مجتمعين $(\mu_1 - \mu_2)$:

هنا يمكن التمييز بين أربع حالات وهي:

الحالة الأولى: تقدير فترة الثقة للفرق بين متوسطي مجتمعين مع معلومية تباينهما (σ_1^2, σ_2^2) .

الحالة الثانية: تقدير فترة الثقة للفرق بين متوسطي مجتمعين مجهولي التباينين ولكنهما متجانسين $(\sigma_1^2 = \sigma_2^2)$ وحجم العينتين صغير، مع تبعية المجتمعين للتوزيع الطبيعي.

الحالة الثالثة: تقدير فترة الثقة للفرق بين متوسطي مجتمعين مجهولي التباينين وغير متجانسين $(\sigma_1^2 \neq \sigma_2^2)$ وحجم العينتين صغير، مع تبعية المجتمعين للتوزيع الطبيعي.

الحالة الرابعة: تقدير فترة الثقة للفرق بين متوسطي مجتمعين غير مستقلين (مرتبطتين)، مع تبعية المجتمعين للتوزيع الطبيعي.

الحالة الأولى: فترة الثقة للفرق بين متوسطي مجتمعين مع معلومية تباينهما (σ_1^2, σ_2^2) .

إذا كانت $(x_1, x_1, \dots, x_{n1})$ عينة عشوائية مسحوبة من مجتمع يتبع التوزيع الطبيعي $N \sim (\mu_1, \sigma_1^2)$ وكانت $(x_1, x_1, \dots, x_{n2})$ عينة عشوائية مسحوبة من مجتمع يتبع التوزيع الطبيعي $N \sim (\mu_2, \sigma_2^2)$ مستقل عن التوزيع الأول فإنه يمكننا تقدير فترة الثقة للفرق بين متوسطي المجتمعين كما يلي:

$$(\mu_1 - \mu_2) = (\bar{x}_1 - \bar{x}_2) \mp Z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

هذا الحد يمثل الخطأ المعياري للفرق بين متوسط العينتين. $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

أو يمكن كتابة العلاقة بشكل آخر كما يلي:

$$(\bar{x}_1 - \bar{x}_2) - Z_{\frac{\alpha}{2}} \times \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq (\mu_1 - \mu_2) \leq (\bar{x}_1 - \bar{x}_2) + Z_{\frac{\alpha}{2}} \times \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

فتكون فترة الثقة للفرق بين متوسطي المجتمعين

$$\left[(\bar{x}_1 - \bar{x}_2) - Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, (\bar{x}_1 - \bar{x}_2) + Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right]$$

في حالة عدم معلومية تباينات المجتمعات الأصلية (σ_1^2, σ_2^2) يمكن استخدام تباينات العينات (S_1^2, S_2^2)

دون التأثير على جودة فترات الثقة بشرط ان تكون العينات كبيرة الحجم، وتحسب فترة الثقة بالعلاقة التالية:

$$(\bar{x}_1 - \bar{x}_2) - Z_{\frac{\alpha}{2}} \times \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \leq (\mu_1 - \mu_2) \leq (\bar{x}_1 - \bar{x}_2) + Z_{\frac{\alpha}{2}} \times \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

مثال (9): اذا تم اختيار (75) طالبة و اختيار (50) طالب، فكان متوسط درجات الطالبات في مقرر

الاحصاء $(\bar{x}_1 = 80)$ بانحراف معياري $(S_1 = 7)$ درجة وكان متوسط درجات الطلاب $(\bar{x}_2 = 70)$

بانحراف معياري $(S_2 = 6)$

المطلوب: اوجد فترة الثقة للفرق بين متوسطي درجات الطلاب والطالبات بدرجة ثقة (0.95).

$$(\bar{x}_1 - \bar{x}_2) - Z_{\frac{\alpha}{2}} \times \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \leq (\mu_1 - \mu_2) \leq (\bar{x}_1 - \bar{x}_2) + Z_{\frac{\alpha}{2}} \times \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

$$(80 - 70) - 1.96 \times \sqrt{\frac{49}{75} + \frac{36}{50}} \leq (\mu_1 - \mu_2) \leq (80 - 70) + 1.96 \times \sqrt{\frac{49}{75} + \frac{36}{50}}$$

$$(80 - 70) - 1.96 \times (1.172) \leq (\mu_1 - \mu_2) \leq (80 - 70) + 1.96 \times (1.172)$$

$$10 - 2.297 \leq (\mu_1 - \mu_2) \leq 10 + 2.297$$

$$7.703 \leq (\mu_1 - \mu_2) \leq 12.297$$

الحالة الثانية: فترة الثقة للفرق بين متوسطي مجتمعين مجهولي التباين ولكنهما متجانسين ($\sigma_1^2 = \sigma_2^2$) وحجم العينتين صغير، مع تبعية المجتمعين للتوزيع الطبيعي.

هنا يتم استخدام يمثل التباين المشترك (S_p^2) للعينتين ويُحسب بإحدى المعادلتين التاليتين:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

$$S_p^2 = \frac{\sum(x_1 - \bar{x}_1)^2 + \sum(x_2 - \bar{x}_2)^2}{n_1 + n_2 - 2}$$

وفترة الثقة تعطى بالعلاقة:

$$(\mu_1 - \mu_2) = (\bar{x}_1 - \bar{x}_2) \mp t_{(\frac{\alpha}{2}, n_1 + n_2 - 2)} \times S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

أو بالعلاقة

$$(\bar{x}_1 - \bar{x}_2) - t_{(\frac{\alpha}{2}, n_1 + n_2 - 2)} \times S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \leq (\mu_1 - \mu_2)$$

$$\leq (\bar{x}_1 - \bar{x}_2) + t_{(\frac{\alpha}{2}, n_1 + n_2 - 2)} \times S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

مثال (10): من أجل معرفة الفرق الحقيقي بين متوسط انتاج شجرة الزيتون في محافظة حماة ومتوسط إنتاج شجرة الزيتون في محافظة حمص.

تم اختيار عينة عشوائية من $(n_1=16)$ شجرة زيتون من محافظة حماة، فُوجِدَ أنَّ مُتَوَسِّطَ إنتاجها $(\bar{x}_1 = 28)$ كغ بانحراف معياري $(S_1=3)$ كغ، كذلك تمَّ اختيار عينة عشوائية من $(n_2=14)$ شجرة زيتون من محافظة حمص فُوجِدَ أنَّ مُتَوَسِّطَ إنتاجها $(\bar{x}_2 = 26)$ كغ وتباين يساوي $(S_2^2=8)$.

المطلوب: أوجد قيمة الفرق الحقيقي باحتمال (0.95) بين متوسط انتاج شجرة الزيتون في حماة ومتوسط انتاج شجرة الزيتون في حمص. أو يمكن كتابة المطلوب بصيغة أخرى كما يلي: أوجد فترة الثقة للفرق بين متوسطي انتاج شجرة الزيتون في حماة وفي حمص بدرجة ثقة (0.95) . مع العلم أنَّ تبايني المجتمعين متجانسين $(\sigma_1^2=\sigma_2^2)$.

الحل:

المجتمع الأول: محافظة حماة: (μ_1) $(S_1^2=9)$ $(n_1=16)$ $(\bar{x}_1 = 28)$

المجتمع الثاني: محافظة حمص (μ_2) $(S_2^2=8)$ $(n_2=14)$ $(\bar{x}_2 = 26)$

فترة الثقة تعطى بالعلاقة:

$$(\mu_1 - \mu_2) = (\bar{x}_1 - \bar{x}_2) \mp t_{(\frac{\alpha}{2}, n_1+n_2-2)} \times S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$(\mu_1 - \mu_2) = (28 - 26) \mp 2.048 \times 2.92 \sqrt{\frac{1}{16} + \frac{1}{14}}$$

$$(\mu_1 - \mu_2) = 2 \mp 2.19$$

فتكون فترة الثقة (4.19) (0.188)

الحالة الثالثة: فترة الثقة للفرق بين متوسطي مجتمعين مجهولي التباينين وغير متجانسين ($\sigma_1^2 \neq \sigma_2^2$)

وحجم العينتين صغير، مع تبعية المجتمعين للتوزيع الطبيعي.

عندما يكون المجتمعين غير متجانسين ($\sigma_1^2 \neq \sigma_2^2$) فإن فترة الثقة للفرق بين متوسطي المجتمعين تعطى بالعلاقة:

$$(\mu_1 - \mu_2) = (\bar{x}_1 - \bar{x}_2) \mp t_{(\frac{\alpha}{2}, df)} \times \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

أو بالعلاقة

$$(\bar{x}_1 - \bar{x}_2) - t_{(\frac{\alpha}{2}, df)} \times \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} < (\mu_1 - \mu_2) < (\bar{x}_1 - \bar{x}_2) + t_{(\frac{\alpha}{2}, df)} \times \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

ويمكن حساب قيمة درجات الحرية (df) بالعلاقة التالية:

$$df = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \right)^2}{\frac{\left(\frac{S_1^2}{n_1} \right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_2} \right)^2}{n_2 - 1}}$$

مثال (11):

في دراسة احصائية لمكتب الارصاد الجوية في سوريا وجد ما يلي:

- معدل الهطولات المطرية خلال شهر شباط لـ (15) سنة سابقة في مدينة حلب يساوي (150) ملم بانحراف معياري قدره (50) ملم.

- معدل الهطولات المطرية خلال شهر شباط لـ (10) سنوات سابقة في مدينة دمشق يساوي (130) ملم بانحراف معياري قدره (30) ملم.

المطلوب: اوجد فترة الثقة للفرق بين متوسطي الهطولات المطرية في حلب ودمشق بدرجة ثقة (0.95).

الحل:

فترة الثقة للفرق بين متوسطي الهطولات المطرية في حلب ودمشق بدرجة ثقة (0.95).

$$(\bar{x}_1 - \bar{x}_2) - t_{(\frac{\alpha}{2}, df)} \times \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \leq (\mu_1 - \mu_2) \leq (\bar{x}_1 - \bar{x}_2) + t_{(\frac{\alpha}{2}, df)} \times \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

$$(150 - 130) - 2.06 \times \sqrt{\frac{2500}{15} + \frac{900}{10}} \leq (\mu_1 - \mu_2) \leq (150 - 130) + 2.06 \times \sqrt{\frac{2500}{15} + \frac{900}{10}}$$

$$(20) - 33 \leq (\mu_1 - \mu_2) \leq (20) + 33$$

$$-13 \leq (\mu_1 - \mu_2) \leq 53$$

أي أن فترة الثقة (53 -13)

الحالة الرابعة: فترة الثقة للفرق بين متوسطي مجتمعين غير مستقلين (مرتبطتين)، مع تبعية المجتمعين للتوزيع الطبيعي.

في حالة العينات المرتبطة يمكن حساب فترة الثقة بالعلاقة التالية:

$$(\mu_1 - \mu_2) = (\bar{x}_1 - \bar{x}_2) \mp t_{(\frac{\alpha}{2}, n-1)} \times \frac{S_d}{\sqrt{n}}$$

وبما أن

$$\bar{d} = |\bar{x}_1 - \bar{x}_2| \quad \text{تمثل متوسط عينة الفروق}$$

فتصبح العلاقة السابقة على الشكل التالي

$$(\mu_1 - \mu_2) = \bar{d} \mp t_{(\frac{\alpha}{2}, n-1)} \times \frac{S_d}{\sqrt{n}}$$

أو

$$\bar{d} - t_{(\frac{\alpha}{2}, n-1)} \times \frac{S_d}{\sqrt{n}} < (\mu_1 - \mu_2) < \bar{d} + t_{(\frac{\alpha}{2}, n-1)} \times \frac{S_d}{\sqrt{n}}$$

مثال (12): من أجل اختبار تأثير زيادة الرواتب على انتاجية العمال، تم قياس الانتاجية الشهرية لعشرة عمال،

قبل وبعد زيادة رواتبهم فكانت النتائج موضحة بالجدول (2).

جدول (2): انتاجية العمال قبل وبعد زيادة الرواتب.

رقم العامل	1	2	3	4	5	6	7	8	9	10
الانتاجية قبل الزيادة - بالقطعة (X)	90	96	85	87	104	85	76	103	95	84
الانتاجية بعد الزيادة - بالقطعة (Y)	95	109	89	93	106	92	88	119	102	92

المطلوب: اوجد فترة الثقة للفرق بين متوسطي الانتاجية قبل وبعد الزيادة لعينة العمال بدرجة ثقة (0.95)

الحل:

فترة الثقة للفرق بين متوسطي الانتاجية قبل وبعد الزيادة لعينة العمال مستخدماً مستوى دلالة ($\alpha=0.05$)

$$(\bar{x}_1 - \bar{x}_2) - t_{(\frac{\alpha}{2}, n-1)} \times \frac{S_d}{\sqrt{n}} \leq (\mu_1 - \mu_2) \leq (\bar{x}_1 - \bar{x}_2) + t_{(\frac{\alpha}{2}, n-1)} \times \frac{S_d}{\sqrt{n}}$$

متوسط الانتاجية للعمال قبل الزيادة : ($\bar{x}_1 = 90.5$)

متوسط الانتاجية للعمال بعد الزيادة : ($\bar{x}_2 = 98.5$)

$$\bar{d} = |\bar{x}_1 - \bar{x}_2| = |90.5 - 98.5| = 8$$

$$(\bar{x}_1 - \bar{x}_2) - t_{(\frac{\alpha}{2}, n-1)} \times \frac{S_d}{\sqrt{n}} \leq (\mu_1 - \mu_2) \leq (\bar{x}_1 - \bar{x}_2) + t_{(\frac{\alpha}{2}, n-1)} \times \frac{S_d}{\sqrt{n}}$$

$$(\mu_1 - \mu_2) = \bar{d} \mp t_{(\frac{\alpha}{2}, n-1)} \times \frac{S_d}{\sqrt{n}}$$

$$\bar{d} - t_{(\frac{\alpha}{2}, n-1)} \times \frac{S_d}{\sqrt{n}} \leq (\mu_1 - \mu_2) \leq \bar{d} + t_{(\frac{\alpha}{2}, n-1)} \times \frac{S_d}{\sqrt{n}}$$

$$8 - 2.26 \times \frac{4.37}{\sqrt{10}} \leq (\mu_1 - \mu_2) \leq 8 + 2.26 \times \frac{4.37}{\sqrt{10}}$$

$$8 - 3.123 \leq (\mu_1 - \mu_2) \leq 8 + 3.123$$

$$4.876 \leq (\mu_1 - \mu_2) \leq 11.123$$

3- تقدير فترة الثقة لنسبة المجتمع (P)

إن تقدير النسبة في المجتمع (p) تعتبر من الحالات المهمة لقياس الظواهر المختلفة مثل نسبة البطالة، ونسبة الأمية ونسبة المدخنين... وغيرها. ونظراً لأنه من الصعوبة بمكان في كثير من الأحيان حساب هذه النسبة مباشرة من المجتمع، فإننا غالباً ما نلجأ لتقدير هذه النسبة من عينة عشوائية مسحوبة من هذا المجتمع.

نسبة العينة (Sample Proportion) : $\hat{P} = \frac{r}{n}$ حيث (n) حجم العينة و (r) عدد المفردات من العينة المدروسة التي تحمل صفة معينة (عدد الطلبة المدخنين).

نسبة المجتمع (Population Proportion) : $P = \frac{R}{N}$ حيث (N) حجم المجتمع و (R) عدد المفردات من المجتمع التي تحمل صفة معينة (عدد الطلبة المدخنين).

وفي هذه الحالة تحسب فترة الثقة لتقدير نسبة المجتمع بالعادلة التالية:

فترة تقدير نسبة المجتمع = نسبة العينة $\pm$ الخطأ المعياري لنسبة العينة مضروباً في معامل الثقة $Z_{\alpha/2}$

ويمكن حسابها بالخطوات التالية:

أ- احسب النسبة في العينة ($\hat{P}$).

ب- احسب الخطأ المعياري للنسبة والتي تساوي في هذه الحالة : $\sigma_{\hat{p}} = \sqrt{\frac{\hat{p}\hat{q}}{n}}$ مع العلم $\hat{q} = 1 - \hat{p}$

ج- اضرب الخطأ المعياري للنسبة في معامل الثقة ($Z_{\alpha/2}$) المستخرج من جدول التوزيع الطبيعي المعياري.

د. نحدد فترة الثقة لنسبة المجتمع بالمعادلة التالية:

$$P = \hat{p} \pm Z_{(\frac{\alpha}{2})} \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}}$$

$$\hat{p} - Z_{(\frac{\alpha}{2})} \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} \leq P \leq \hat{p} + Z_{(\frac{\alpha}{2})} \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}}$$

$$P \left[\hat{p} - Z_{(\frac{\alpha}{2})} \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} \leq P \leq \hat{p} + Z_{(\frac{\alpha}{2})} \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} \right] = 1 - \alpha$$

حيث $q=1-p$.

مثال (13): لإيجاد فترة الثقة باحتمال 95% لنسبة الطلاب الذين لديهم ضعف في البصر، أخذت عينة عشوائية

حجمها (100) طالب فتيين أن من لديهم ضعف في البصر كان (15) طالب. أوجد فترة الثقة المطلوبة:

الحل

:

$$\hat{p} = \frac{15}{100} = 0.15$$

ويمكن حساب فترة الثقة للنسبة من العلاقة التالية:

$$\hat{p} - Z_{(\frac{\alpha}{2})} \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} \leq P \leq \hat{p} + Z_{(\frac{\alpha}{2})} \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}}$$

$$0.15 - 1.96 \times \sqrt{\frac{0.15(0.85)}{100}} \leq P \leq 0.15 + 1.96 \times \sqrt{\frac{0.15(0.85)}{100}}$$

$$0.08 \leq P \leq 0.22$$

أي أن فترة الثقة هي:

$$(0.08, 0.22)$$

مثال (14): عينة عشوائية حجمها (144) ناخباً سحبت من إحدى المدن فوجد أن عدد المؤيدين في العينة لمرشح معين هو (60) ناخباً، أوجد فترة تقدير لنسبة المؤيدين لهذا المرشح في المدينة كلها بدرجة ثقة (0.95).

الحل: نحسب أولاً نسبة المؤيدين للمرشح في العينة ($\hat{P}$) التي نحصل عليها بقسمة عدد المؤيدين له على العدد الكلي للعينة (حجم العينة) أي أن:

$$\hat{p} = \frac{60}{144} = 0.42$$

وحيث أن درجة الثقة المطلوبة هي (0.95) فإن معامل الثقة هو ($Z_{\alpha/2}=1.96$) وفترة تقدير نسبة المؤيدين لهذا المرشح في المدينة تأخذ الشكل التالي:

$$\hat{P} = 0.42 \quad \hat{q} = 1 - \hat{P} = 1 - 0.42 = 0.58 \quad n=144$$

نعوض بالقانون:

$$\hat{P} - Z_{(\frac{\alpha}{2})} \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} \leq P \leq \hat{P} + Z_{(\frac{\alpha}{2})} \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}}$$

$$0.42 - 1.96 \sqrt{\frac{0.42 * 0.58}{144}} \leq P \leq 0.42 + 1.96 \sqrt{\frac{0.42 * 0.58}{144}}$$

$$0.34 \leq P \leq 0.50$$

أي أن نسبة المؤيدين للمرشح في المدينة تتراوح بين (0.34 0.5) وذلك بدرجة ثقة (95%) بمعنى آخر أن نسبة مؤيدي هذا المرشح في هذه المدينة لا تتجاوز (50%) كحد أعلى، وبالتالي ففرصته في الفوز كمرشح قد لا تكون كبيرة وذلك بدرجة ثقة (95%) وهذا التقدير أو الحكم لا تتجاوز نسبة الخطأ فيه 5 %.

4- تقدير فترة الثقة للفرق بين نسبتي مجتمعين $(P_1 - P_2)$

نسبة المجتمع الاول (P_1) : نسبة الذين يملكون خاصية معينة في المجتمع الأول (مثلاً: الاشخاص الذين يلبسون نظارات طبية في المجتمع الاول).

نسبة المجتمع الثاني (P_2) : نسبة الذين يملكون خاصية معينة في المجتمع الثاني (مثلاً: الاشخاص الذين يلبسون نظارات طبية في المجتمع الثاني).

الفرق بين نسبتي المجتمعين $(P_1 - P_2)$

نسبة العينة الاولى $(\hat{P}_1)$: نسبة الذين يملكون خاصية معينة في العينة الأولى، $\hat{P}_1 = \frac{r_1}{n_1}$

نسبة العينة الثانية $(\hat{P}_2)$: نسبة الذين يملكون خاصية معينة في العينة الثانية، $\hat{P}_2 = \frac{r_2}{n_2}$

الفرق بين نسبتي العينتين $(\hat{P}_1 - \hat{P}_2)$

لتقدير فترة الثقة للفرق بين نسبتي مجتمعين نطبق العلاقة:

$$(p_1 - p_2) = (\hat{P}_1 - \hat{P}_2) \mp Z_{\left(\frac{\alpha}{2}\right)} \sqrt{\frac{\hat{P}_1 \cdot \hat{Q}_1}{n_1} + \frac{\hat{P}_2 \cdot \hat{Q}_2}{n_2}}$$

$$(\hat{P}_1 - \hat{P}_2) - Z_{\left(\frac{\alpha}{2}\right)} \sqrt{\frac{\hat{P}_1 \cdot \hat{Q}_1}{n_1} + \frac{\hat{P}_2 \cdot \hat{Q}_2}{n_2}} \leq (P_1 - P_2) \leq (\hat{P}_1 - \hat{P}_2) + Z_{\left(\frac{\alpha}{2}\right)} \sqrt{\frac{\hat{P}_1 \cdot \hat{Q}_1}{n_1} + \frac{\hat{P}_2 \cdot \hat{Q}_2}{n_2}}$$

مثال (15): تم سحب عينة عشوائية حجمها (120) طالب من جامعة دمشق فتبين أن عدد المدخنين (24)

طالب، وسحبت عينة حجمها (90) طالب من جامعة حلب فتبين أن عدد المدخنين (17) طالب.

المطلوب : اوجد فترة ثقة للفرق بين نسبة الطلاب المدخنين في جامعة دمشق وحلب (المجتمعين) بدرجة ثقة

95%.

الحل:

$$(r_1 = 24) \quad (n_1=120) \quad \hat{P}_1 = \frac{r_1}{n_1} = \frac{24}{120} = 0.2 \quad \text{المجتمع الأول: جامعة دمشق:}$$

$$(r_2 = 17) \quad (n_2=90) \quad \hat{P}_2 = \frac{r_2}{n_2} = \frac{17}{90} = 0.188 \quad \text{المجتمع الثاني: جامعة حلب:}$$

0.188 الحل:

$$\hat{P}_1 = \frac{r_1}{n_1} = \frac{24}{120} = 0.2$$

$$\hat{P}_2 = \frac{r_2}{n_2} = \frac{17}{90} = 0.188$$

نسبة العينة الاولى ($\hat{P}_1$): نسبة الطلاب المدخنين في جامعة دمشق.

نسبة العينة الثانية ($\hat{P}_2$): نسبة الطلاب المدخنين في جامعة حلب.

فترة ثقة (95%) للفرق بين نسبة المدخنين بين طلاب جامعة دمشق وحلب (المجتمعين)

$$(\hat{P}_1 - \hat{P}_2) - Z_{\left(\frac{\alpha}{2}\right)} \sqrt{\frac{\hat{P}_1 \cdot \hat{q}_1}{n_1} + \frac{\hat{P}_2 \cdot \hat{q}_2}{n_2}} \leq (P_1 - P_2) \leq (\hat{P}_1 - \hat{P}_2) + Z_{\left(\frac{\alpha}{2}\right)} \sqrt{\frac{\hat{P}_1 \cdot \hat{q}_1}{n_1} + \frac{\hat{P}_2 \cdot \hat{q}_2}{n_2}}$$

$$(0.2 - 0.188) - 1.96 \sqrt{\frac{0.2 \cdot 0.8}{120} + \frac{0.188 \cdot 0.812}{90}} \leq (P_1 - P_2) \leq (0.2 - 0.188) +$$

$$1.96 \sqrt{\frac{0.2 \cdot 0.8}{120} + \frac{0.188 \cdot 0.812}{90}}$$

$$(0.2 - 0.188) - 0.108 \leq (P_1 - P_2) \leq (0.2 - 0.188) + 0.108$$

$$-0.096 \leq (P_1 - P_2) \leq 0.12$$

أي أن فترة الثقة هي (0.096, 0.12)

جدول لأفضل تقدير بنقطة لبعض معالم المجتمع والخطأ المعياري للتقدير.

التقدير بنقطة وبفترة ثقة لبعض معالم المجتمع						
التقدير بفترة الثقة		تقدير بنقطة لبعض معالم المجتمع والخطأ المعياري للتقدير				
		الخطأ المعياري للتقدير بنقطة	الرمز	أفضل تقدير بنقطة	الرمز	المعلمة المراد تقديرها
$\bar{x} - Z_{\alpha} \frac{\sigma}{\sqrt{n}}$	σ معلوم	$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$	$\bar{x}$	الوسط الحسابي للعيينة	μ	الوسط الحسابي للمجتمع

$\bar{x} - Z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}$	σ مجهول (n) كبيرة					
$\bar{x} - t_{(\frac{\alpha}{2}, n-1)} \frac{S}{\sqrt{n}}$	σ مجهول (n) صغيرة					
$\hat{P} \pm Z_{\alpha/2} \sqrt{\frac{\hat{P}\hat{Q}}{n}}$	$\sigma_{\bar{x}} = \sqrt{\frac{Pq}{n}}$	$\hat{P} = \frac{r}{n}$	النسبة في العينة	P	النسبة في المجتمع	
$(\bar{x}_1 - \bar{x}_2) \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	تباين المجتمعين معلومين	الفرق بين متوسطي العينتين	$\mu_1 - \mu_2$	الفرق بين متوسطي مجتمعين	
$(\bar{x}_1 - \bar{x}_2) \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$	$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$	تباين المجتمعين مجهولين والعينات كبيرة				
$(\bar{x}_1 - \bar{x}_2) \pm t_{(\frac{\alpha}{2}, n_1+n_2-2)} \times S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$	$\sigma_{\bar{x}_1 - \bar{x}_2} = S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$	تباين المجتمعين مجهولين ومتجانسين والعينات صغيرة				
		تباين المجتمعين مجهولين وغير متجانسين والعينات صغيرة				
$(\hat{P}_1 - \hat{P}_2) \pm Z_{\alpha/2} \sqrt{\frac{\hat{P}_1\hat{Q}_1}{n_1} + \frac{\hat{P}_2\hat{Q}_2}{n_2}}$	$\sigma_{\hat{P}_1 - \hat{P}_2} = \sqrt{\frac{P_1q_1}{n_1} + \frac{P_2q_2}{n_2}}$	$\hat{P}_1 - \hat{P}_2$	الفرق بين نسبتي العينتين	$P_1 - P_2$	الفرق بين نسبتي مجتمعين	

مثال (9): أخذت عينة عشوائية حجمها 9 من مجتمع طبيعي $N(\mu_1, 25)$ ثم أخذت عينة عشوائية حجمها

10 من مجتمع طبيعي $N(\mu_2, 40)$ مستقل عن الأول، فإذا أعطيت العينة الأولى وسطاً حسابياً = 32،

بينما أعطيت العينة الثانية وسطاً حسابياً = 47 أوجد :

أ- فترة ثقة 95% للفرق بين الوسطين $(\mu_1 - \mu_2)$ ؟

ب- فترة ثقة 90% للفرق بين الوسطين $(\mu_2 - \mu_1)$ ؟

الحل: المعطيات المجتمع الأول المجتمع الثاني

$$\sigma_2^2 = 40$$

$$\sigma_1^2 = 25$$

$$n_2 = 10$$

$$n_1 = 9$$

$$\bar{Y} = 47$$

$$\bar{X} = 32$$

المطلوب : أ- فترة 95% ثقة للفرق $\mu_1 - \mu_2$ ؟

وبتطبيق العلاقة نجد أن:

$$[(32 - 47) - 1.96 \times \sqrt{\frac{25}{9} + \frac{40}{10}}, (32 - 47) + 1.96 \sqrt{\frac{25}{9} + \frac{40}{10}}]$$

$$(-15 - 1.96 \times \sqrt{\frac{25}{9} + 4}, -15 - 1.96 \times \sqrt{\frac{25}{9} + 4}) = (-20.1, -9.9)$$

ب- فترة ثقة 90% للفرق بين $\mu_2 - \mu_1$ ؟

بتطبيق العلاقة نجد أن:

$$[(47 - 32) - 1.64 \times \sqrt{\frac{25}{9} + \frac{40}{10}}, (47 - 32) + 1.64 \sqrt{\frac{25}{9} + \frac{40}{10}}]$$

$$(15 - 1.64 \times \sqrt{\frac{25}{9} + 4}, 15 + 1.64 \times \sqrt{\frac{25}{9} + 4}) = (10.73, 19.27)$$

الحل:

المجتمع الأول: محافظة حماه: (μ_1) $(S_1^2=9)$ $(n_1=16)$ $(\bar{x}_1 = 28)$

المجتمع الثاني: محافظة حمص (μ_2) $(S_2^2=8)$ $(n_2=14)$ $(\bar{x}_2 = 26)$

3- صياغة الفروض:

الفرض العدمي: $(H_0: \mu_1 = \mu_2)$ متوسط إنتاج شجرة الزيتون في حماة يساوي متوسط إنتاج شجرة الزيتون في حمص.

الفرض البديل: $(H_1: \mu_1 \neq \mu_2)$ متوسط إنتاج شجرة الزيتون في حماة لا يساوي متوسط إنتاج شجرة الزيتون في حمص.

4- حساب قيمة إحصائية الاختبار:

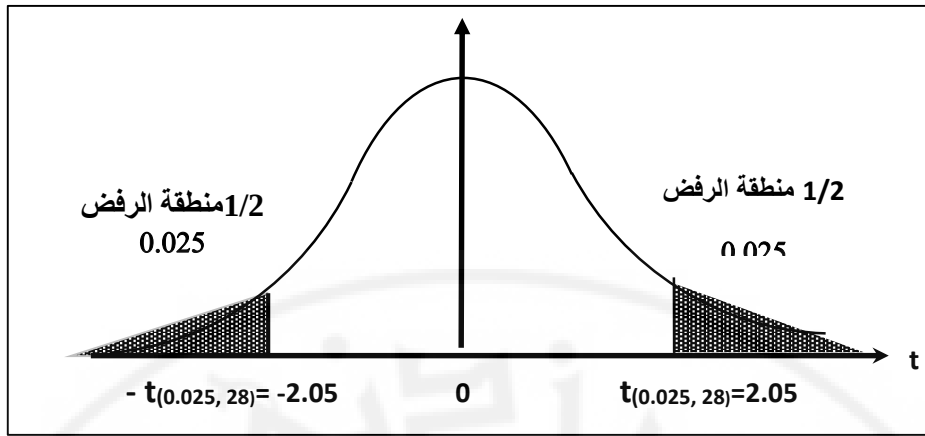
بما أن العينتين صغيرتين وتباين المجتمعين متجانسين، والمجتمعين طبيعيين، في هذه الحالة تتبع إحصائية الاختبار توزيع (t) بدرجات حرية تساوي $(n_1 + n_2 - 2)$ وتُحسب كما يلي:

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} = \frac{15 \times 9 + 13 \times 8}{16+14-2} = \frac{740}{28} = 8.55$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{S_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{(28-26)}{\sqrt{8.55 \left(\frac{1}{16} + \frac{1}{14} \right)}} = \frac{2}{\sqrt{8}} = 1.869$$

3- تحديد منطقتي القبول والرفض :

تحدد منطقتي القبول والرفض باستخراج قيمة (t) الجدولية من جدول (t) عند درجات حرية تساوي (n_1+n_2-2) والتي تساوي 18 أي $t_{(0.025, 28)} = 2.048$ وتتمثل منطقتي القبول والرفض بالشكل (5-15).



شكل (5-15): التمثيل البياني لمنطقتي القبول والرفض.

4- المقارنة و اتخاذ القرار:

بما أن قيمة الإحصائية تساوي (1.869) فإنها تقع في منطقة القبول وبالتالي فإن القرار هو قبول الفرض العدمي بأن متوسط إنتاج شجرة الزيتون في حماة يساوي متوسط إنتاج شجرة الزيتون في حمص وذلك بمستوى معنوية ($\alpha = 0.05$).

مثال (5-8): ادعى باحث زراعي أن متوسط إنتاج شجرة الزيتون في محافظة حماة يساوي متوسط إنتاج شجرة الزيتون في محافظة حمص.

اختيرت عينة من ($n_1=16$) شجرة زيتون من حماة، فوجد أن متوسط إنتاجها ($\bar{x}_1 = 28$) كغ وتباينه ($S_1^2=9$)، كذلك تم اختيار عينة عشوائية أخرى من ($n_2=14$) شجرة زيتون من حمص فوجد أن متوسط إنتاجها ($\bar{x}_2 = 26$) كغ وتباينه ($S_2^2=8$).

المطلوب: اوجد فترة الثقة للفرق بين متوسطي إنتاج شجرة الزيتون في حماة وفي حمص بدرجة ثقة (0.95).

مع العلم أن تبايني المجتمعين متجانسين ($\sigma_1^2=\sigma_2^2$).

الحل:

المجتمع الأول: محافظة حماه: (μ_1) $(S_1^2=9)$ $(n_1=16)$ $(\bar{x}_1 = 28)$

المجتمع الثاني: محافظة حمص (μ_2) $(S_2^2=8)$ $(n_2=14)$ $(\bar{x}_2 = 26)$

5- صياغة الفروض:

الفرض العدمي: $(H_0: \mu_1 = \mu_2)$ متوسط إنتاج شجرة الزيتون في حماة يساوي متوسط إنتاج شجرة الزيتون في حمص.

الفرض البديل: $(H_1: \mu_1 \neq \mu_2)$ متوسط إنتاج شجرة الزيتون في حماة لا يساوي متوسط إنتاج شجرة الزيتون في حمص.

6- حساب قيمة إحصائية الاختبار:

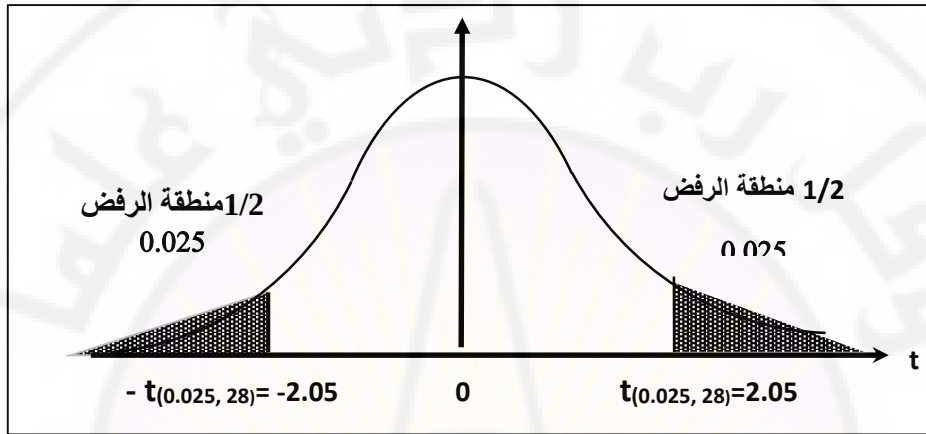
بما أن العينتين صغيرتين وتباين المجتمعين متجانسين، والمجتمعين طبيعيتين، في هذه الحالة تتبع إحصائية الاختبار توزيع (t) بدرجات حرية تساوي $(n_1 + n_2 - 2)$ وتحسب كما يلي:

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} = \frac{15 \times 9 + 13 \times 8}{16+14-2} = \frac{740}{28} = 8.55$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{S_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{(28-26)}{\sqrt{8.55 \left(\frac{1}{16} + \frac{1}{14} \right)}} = \frac{2}{\sqrt{8}} = 1.869$$

3- تحديد منطقتي القبول والرفض:

تحدّد منطقتي القبول والرفض باستخراج قيمة t الجدوليّة من جدول t عند درجات حريّة تساوي (n_1+n_2-2) والنّبي تساوي 18 أي $t_{(0.025, 28)} = t_{(\frac{\alpha}{2}, n_1+n_2-2)} = 2.048$ وتتمثّل منطقتي القبول والرفض بالشكل (5-15).



شكل (5-15): التمثيل البياني لمنطقتي القبول والرفض.

4- المقارنة و اتخاذ القرار:

بما أنّ قيمة الإحصائيّة تساوي (1.869) فإنّها تقع في منطقة القبول وبالتالي فإنّ القرار هو قبول الفرض العدمي بأنّ متوسط إنتاج شجرة الزيتون في حماة يساوي متوسط إنتاج شجرة الزيتون في حمص وذلك بمستوى معنويّة $(\alpha = 0.05)$.

فترة الثقة تعطى بالعلاقة:

$$(\mu_1 - \mu_2) = (\bar{x}_1 - \bar{x}_2) \mp t_{(\frac{\alpha}{2}, n_1+n_2-2)} \times S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$(\mu_1 - \mu_2) = (28 - 26) \mp 2.048 \times 2.92 \sqrt{\frac{1}{16} + \frac{1}{14}}$$

$$(\mu_1 - \mu_2) = 2 \mp 2.19$$

فتكون فترة الثقة

$$(0.188 \quad 4.19)$$

الحالة الثالثة: فترة الثقة للفرق بين متوسطي مجتمعين مجهولي التباينين وغير متجانسين ($\sigma_1^2 \neq \sigma_2^2$) وحجم العينتين صغير، مع تبعية المجتمعين للتوزيع الطبيعي.

فترة الثقة تعطى بالعلاقة:

$$(\bar{x}_1 - \bar{x}_2) - t_{(\frac{\alpha}{2}, df)} \times \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} < (\mu_1 - \mu_2) < (\bar{x}_1 - \bar{x}_2) + t_{(\frac{\alpha}{2}, df)} \times \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

ويمكن حساب قيمة درجات الحرية (df) بالعلاقة التالية:

$$df = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \right)^2}{\frac{\left(\frac{S_1^2}{n_1} \right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_2} \right)^2}{n_2 - 1}}$$

مثال:

في دراسة احصائية لمكتب الارصاد الجوية في سوريا وجد ما يلي:

- معدل الهطولات المطرية خلال شهر شباط لـ (15) سنة سابقة في مدينة حلب يساوي (150) ملم بانحراف معياري قدره (50) ملم.

- معدل الهطولات المطرية خلال شهر شباط لـ (10) سنوات سابقة في مدينة دمشق يساوي (130) ملم بانحراف معياري قدره (30) ملم.

المطلوب:

1- هل يوجد فروق معنوية بين متوسط الهطولات المطرية في حلب ودمشق، مع العلم أن العينتين سحبتا من مجتمعين طبيعيين غير متجانسين ($\sigma_1^2 \neq \sigma_2^2$)

- اوجد فترة الثقة للفرق بين متوسطي الهطولات المطرية في حلب ودمشق بدرجة ثقة (0.95).

5- صياغة الفروض:

الفرض العدمي: $H_0 : \mu_1 = \mu_2$

الفرض البديل: $(H_1 : \mu_1 \neq \mu_2)$

6- حساب احصاء الاختبار:

معادلة حساب احصاء الاختبار كما يلي:

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{150 - 130}{\sqrt{\frac{2500}{15} + \frac{900}{10}}} = \frac{20}{16.02} = 1.25$$

حيث إن احصائية الاختبار تتبع توزيع (t) بدرجات حرية تساوي:

$$df = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1-1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2-1}} = \frac{\left(\frac{2500}{15} + \frac{900}{10}\right)^2}{\frac{\left(\frac{2500}{15}\right)^2}{14} + \frac{\left(\frac{900}{10}\right)^2}{9}} = \frac{65876.06}{2884} = 22.84 \approx 23$$

هذا يعني درجات الحرية (df=23).

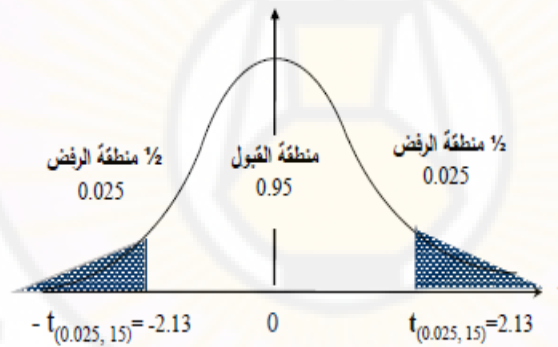
الجدولية: t (إيجاد قيمة

بما ان الاختبار من طرفين تحسب (t) الجدولية كما يلي: $t_{(\frac{\alpha}{2}, n)} = t_{(0.025, 23)} = 2.06$

7- تحديد منطقتي القبول والرفض:

تحدد منطقتي القبول والرفض اعتماداً على قيمة (t) الجدولية (2.06) و تكون منطقة القبول اكبر من

(-2.06) واصغر من (+2.06)



8- المقارنة واتخاذ القرار

القرار الإحصائي: بمقارنة قيمة t المحسوبة (t=1.25) بقيمة t الجدولية (t=2.06) نجد أن قيمة t المحسوبة

تقع في منطقة القبول، لذلك نقبل الفرض الصفري القائل بأنه لا يوجد فرق معنوي بين متوسطي صنف

التفاح بمستوى دلالة ($\alpha=0.05$)

- اوجد فترة الثقة للفرق بين متوسطي الهطولات المطرية في حلب ودمشق بدرجة ثقة (0.95).

$$(\bar{x}_1 - \bar{x}_2) - t_{(\frac{\alpha}{2}, df)} \times \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} < (\mu_1 - \mu_2) < (\bar{x}_1 - \bar{x}_2) + t_{(\frac{\alpha}{2}, df)} \times \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

$$(150 - 130) - 2.06 \times \sqrt{\frac{2500}{15} + \frac{900}{10}} < (\mu_1 - \mu_2) < (150 - 130) + 2.06 \times \sqrt{\frac{2500}{15} + \frac{900}{10}}$$

$$(20) - 33 < (\mu_1 - \mu_2) < (20) + 33$$

$$-13 < (\mu_1 - \mu_2) < 53$$

أي أن فترة الثقة (53 -13)

مثال (7): من أجل اختبار تأثير زيادة الرواتب على انتاجية العمال، تم قياس الانتاجية الشهرية لعشرة عمال، قبل وبعد زيادة رواتبهم فكانت النتائج موضحة بالجدول (2).

جدول (2): انتاجية العمال قبل وبعد زيادة الرواتب.

رقم العامل	1	2	3	4	5	6	7	8	9	10
الانتاجية قبل الزيادة - بالقطعة (X)	90	96	85	87	104	85	76	103	95	84
الانتاجية بعد الزيادة - بالقطعة (Y)	95	109	89	93	106	92	88	119	102	92

المطلوب: اختبار فيما اذا كانت الزيادة بالرواتب لها تأثير معنوي على زيادة الانتاجية،

اوجد فترة الثقة للفرق بين متوسطي الانتاجية قبل وبعد الزيادة لعينة العمال مستخدماً مستوى دلالة ($\alpha=0.05$)

الحل:

3- صياغة الفروض

بفرض أن $d_i = x_i - y_i$ تمثل عينة الفرق بين أزواج البيانات للعينتين و $\bar{D}_0$ هو الفرق بين متوسطي المجتمعين الاول والثاني، يمكن صياغة الفروض كما يلي:

⊖ الفرض الصفري: $H_0: \mu_x = \mu_y$ متوسط انتاجية العمال قبل الزيادة = متوسط انتاجية العمال بعد

الزيادة أي لا يوجد فرق معنوي بين القيم القبلية والبعدية.

⊖ الفرض البديل:

⊖ $H_1: \mu_x < \mu_y$ أو $H_1: \mu_x \neq \mu_y$

متوسط انتاجية العمال قبل الزيادة أقل من متوسط انتاجية العمال بعد الزيادة، أي يوجد فرق معنوي بين الدرجات القبلية والبعدية.

يمكن صياغة الفرض البديل على الشكل $\bar{D}_0 > 0$ أي أن متوسط عينة الفروق أكبر من الصفر.

2- حساب قيمة إحصائية الاختبار

يتم حساب احصاءة الاختبار من العلاقة:

$$t = \frac{\bar{d} - \bar{D}_0}{\frac{S_d}{\sqrt{n}}}$$

وهذا يتطلب حساب متوسط الفروق ($\bar{d}$) والانحراف المعياري (S_d) لأزواج المشاهدات كما هي مبينة بالجدول (3).

جدول (3): حساب الانحراف المعياري لانتاجية العمال قبل وبعد الزيادة على الرواتب.

رقم العامل	1	2	3	4	5	6	7	8	9	10	
انتاجية قبل الزيادة - بالقطعة	90	96	85	87	104	85	76	103	95	84	
(X)											

	92	102	119	88	92	106	93	89	109	95	الانتاجية بعد الزيادة - بالقطعة (Y)
-80	-8	-7	-16	-12	-7	-2	-6	-4	-13	-5	الفروق (d=x-y)
0	0	1	-8	-4	1	6	2	4	-5	3	(d-d̄)
172	0	1	64	16	1	36	4	16	25	9	(d-d̄) ²

$$\bar{d} = \frac{-80}{10} = -8$$

$$S_d = \sqrt{\frac{\sum(d-\bar{d})^2}{n-1}} = \sqrt{\frac{172}{9}} = 4.37$$

$$t = \frac{\bar{d}}{\frac{S_d}{\sqrt{n}}} = \frac{-8}{4.37/\sqrt{10}} = -5.79$$

3- تحديد منطقتي القبول والرفض

تحدد منطقتي القبول والرفض باستخراج قيمة (t) الجدولية من جدول t عند درجات حرية تساوي (n - 1=9)

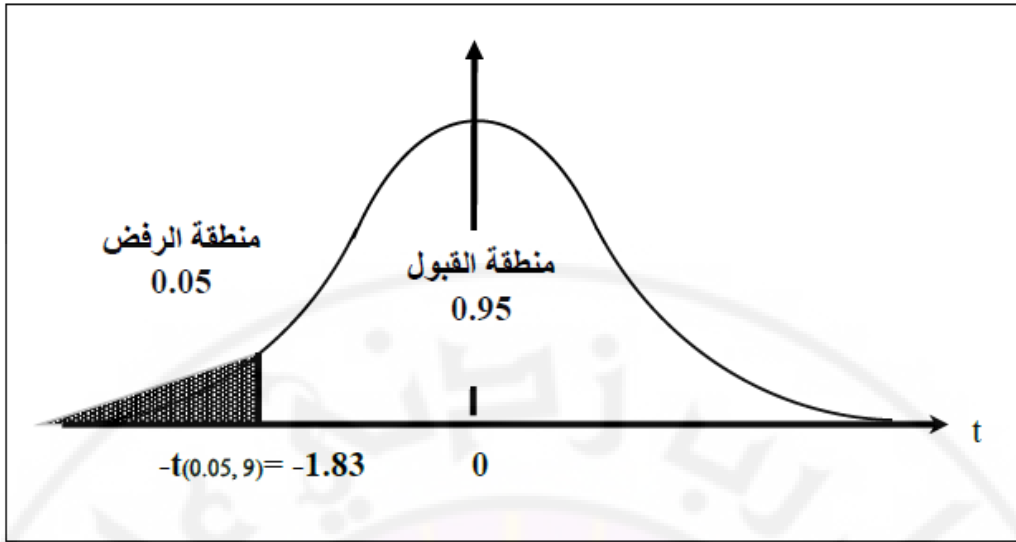
$$t_{\alpha,9} = t_{0.05,9} = -1.83 \text{ أي أن } (\alpha = 0.05) \text{ عند مستوى معنوية يساوي}$$

تحدد منطقتي القبول والرفض باستخراج قيمة (t) الجدولية من جدول t عند درجات حرية تساوي (n - 1=9)

عند مستوى معنوية يساوي (α = 0.05) وبما أن الفرض البديل أكبر من هذا يعني أنه من طرف واحد نحو

اليمين أي القيمة الجدولية تكون أي $t_{(\alpha, n-1)} = t_{(0.05, 9)} = -1.83$ وبالتالي تتحدد منطقتي القبول

والرفض بالشكل (3).



شكل (2-16): التمثيل البياني لمنطقتي القبول والرفض للمثال (2-10).

6- المقارنة و اتخاذ القرار

بما أن قيمة الإحصائية تساوي 5.79- فإنها تقع في منطقة الرفض وبالتالي فإن القرار هو رفض الفرض العدمي وقبول الفرض البديل بأنه يوجد فرق معنوي بين متوسط انتاجية العمال قبل الزيادة و متوسط انتاجية العمال بعد الزيادة وهذا يقودنا إلى النتيجة بأن الزيادة فعالة في زيادة الانتاجية.

اوجد فترة الثقة للفرق بين متوسطي الانتاجية قبل وبعد الزيادة لعينة العمال مستخدماً مستوى دلالة ($\alpha=0.05$)

$$(\bar{x}_1 - \bar{x}_2) - t_{(\frac{\alpha}{2}, n-1)} \times \frac{S_d}{\sqrt{n}} < (\mu_1 - \mu_2) < (\bar{x}_1 - \bar{x}_2) + t_{(\frac{\alpha}{2}, n-1)} \times \frac{S_d}{\sqrt{n}}$$

متوسط الانتاجية للعمال قبل الزيادة : ($\bar{x}_1 = 90.5$)

متوسط الانتاجية للعمال بعد الزيادة : ($\bar{x}_2 = 98.5$)

$$\bar{d} = |\bar{x}_1 - \bar{x}_2| = |90.5 - 98.5| = 8$$

$$(\bar{x}_1 - \bar{x}_2) - t_{(\frac{\alpha}{2}, n-1)} \times \frac{S_d}{\sqrt{n}} < (\mu_1 - \mu_2) < (\bar{x}_1 - \bar{x}_2) + t_{(\frac{\alpha}{2}, n-1)} \times \frac{S_d}{\sqrt{n}}$$

$$\bar{d} - t_{(\frac{\alpha}{2}, n-1)} \times \frac{S_d}{\sqrt{n}} < (\mu_1 - \mu_2) < \bar{d} + t_{(\frac{\alpha}{2}, n-1)} \times \frac{S_d}{\sqrt{n}}$$

$$8 - 1.83 \times \frac{4.37}{\sqrt{10}} < (\mu_1 - \mu_2) < 8 + 1.83 \times \frac{4.37}{\sqrt{10}}$$

$$8 - 2.53 < (\mu_1 - \mu_2) < 8 + 2.53$$

$$5.47 < (\mu_1 - \mu_2) < 10.53$$

