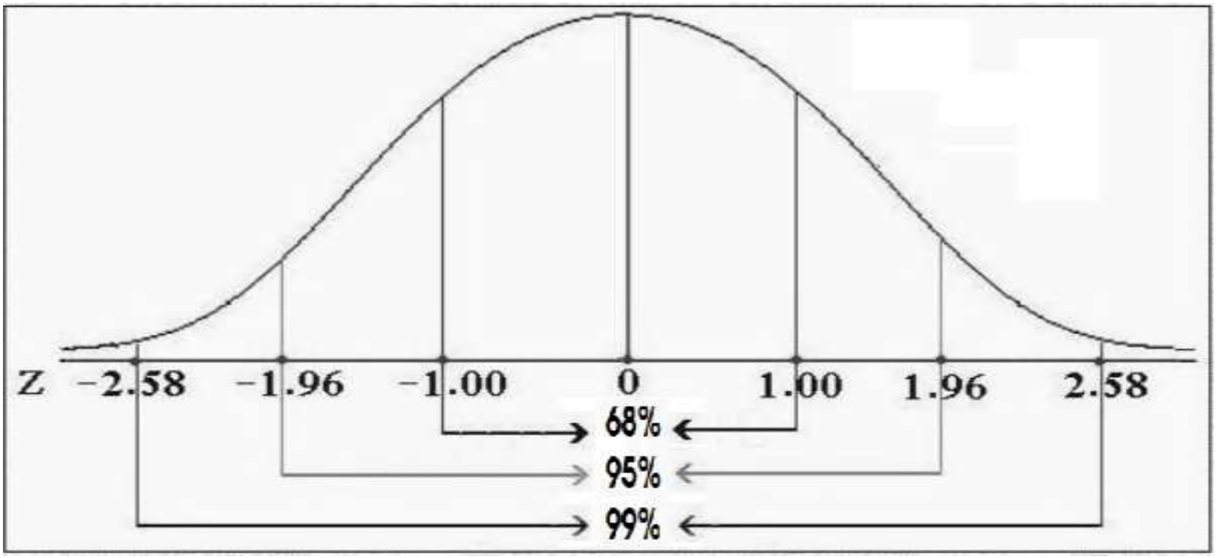
جامعة البليدة 2 – لونيسي علي

كلية العلوم الإنسانية والاجتماعية

الإحصاء الاستدلالي

لطلبة السنة الأولى ميدان العلوم الاجتماعية (S2)



الأستاذ: زيدان محمد

السنة الجامعية : 2023 / 2024

فهرس المحتويات

مقدمة

عند عامة الناس الإحصاء يعني العد والحساب وأحسانا جمع المعطيات. لكن المفهوم الحقيقي للإحصاء هو عملية وصف البيانات وإعادة تشكيلها بطريقة تسهل قراءتها ثم تهيئ لتساعد في اتخاذ القرارات.

الإحصاء إذن هو علم يهتم بالمعلومات والبيانات وهدفه الأساسي هو جمع المعلومات وتبويبها وتنظيمها وعرضها على شكل جداول وعلى شكل رسومات بيانية كما يهتم علم الإحصاء بتلخيص هذه المعلومات في قيم تسمى مؤشرات ومقاييس وكذا تحليلها واستخلاص النتائج منها واستخدامها في اتخاذ القرارات.

ونستعمل الأساليب الإحصائية في الكثير من الميادين مثل العلوم الإنسانية والاجتماعية والعلوم الاقتصادية وعلم الطب والصيدلة وغيرها من مجالات الزراعة والصناعة والإدارة والأعمال، ولكل ميدان أساليبه وحاجياته.

ويمكن تقسيم الإحصاء إلى قسمين هما، قسم الإحصاء الوصفي (Descriptive Statistic) و قسم الإحصاء الاستدلالي أو التطبيقي (*Inferential Statistic*)، يهتم القسم الأول بجمع المعطيات وتنظيمها وتبويبها وتلخيصها وعرضها في جداول وبيانيا بينما يهتم القسم الثاني استنباط النتائج واتخاذ القرارات وتعميم النتائج على المجتمع.

في هذه المطبوعة سوف نتطرق لبعض المفاهيم ذات العلاقة بالإحصاء الاستدلالي ثم نتناول طرق اختيار عينة ممثلة للمجتمع. بعد ذلك نتعلم كيف نستخرج مجال الثقة للمتوسط ومجال الثقة للنسبة في المجتمع، وفي الأخير نتطرق لاختبارات الفروض الأكثر استعمالا مثل اختبار z، اختبار t واختبار كاي تربيع وغيرها.

لم ننسى أن نذكر الطالب في هذه المطبوعة بكيفية حساب بعض المؤشرات الأساسية مثل النسبة والمتوسط الحسابي والانحراف المعياري التي لا يمكننا الاستغناء عنها في الإحصاء الاستدلالي.

نقدم في ملاحق المطبوعة مجموعة من الجداول الإحصائية الأكثر استعمالا مثل جدول الأرقام العشوائية وجدول التوزيع الطبيعي المعياري وجدول ستودنت وجدول كاي تربيع.

1 – مدخل للإحصاء:

1 – 1 – مدخل إلى الإحصاء والمنهج الإحصائي:

الإحصاء هو علم يبحث في كيفية جمع المعطيات (البيانات، الملاحظات) المتعلقة بمختلف الظواهر الاقتصادية والاجتماعية والديموغرافية والطبية وغيرها، ثم تفريغها وتصنيفها في جداول منظمة تعطي قراءة ملخصة لهذه المعطيات، ثم عرض هذه الجداول في أشكال بيانية لتسهيل قراءتها، ثم وصفها وتحليلها باستخدام مقاييس مختلفة، وأخيرا استخدام هذه المعطيات في اتخاذ قرارات مناسبة أو التحقق من فرضية معينة أو رقابة مدى تطابق دراسة نظرية معينة بالنتائج الفعلية.

إحصاء وصفي وإحصاء استدلالي

ينقسم الإحصاء إلى إحصاء وصفي وإحصاء استدلالي. الإحصاء الوصفي يختزل مجموعة البيانات إلى معلومة أو اثنين تميزان كل البيانات. كما يتعلق الإحصاء الوصفي بتقديم مجموعة البيانات على شكل جداول أو رسوم بيانية. أما الإحصاء الاستدلالي (ويشمل التقدير واختبار الفروض) يتعلق باستخلاص تعميمات عن خواص المجتمع من واقع خواص عينة مأخوذة من هذا المجتمع. ومن ثم فإن الإحصاء الاستدلالي يتضمن تعليلا استقرائيا (وذلك على نقيض التعليل الاستنباطي الذي يستنبط خواص الجزء مبتدئا بالكل).

مراحل البحث الإحصائي:

مرحلة التحضير:

-1- فهم الهدف وتحديد المشكل ثم طرح الفرضية (**لماذا نحصي**)،

-2- تحديد المجتمع الإحصائي (المجتمع المعني بهذه الدراسة والذي نلتزم بحدوده في جمع البيانات) (**أين نحصي**)،

-3- تحديد الوحدة (المفردة) الإحصائية التي تمثل عنصرا من المجتمع الإحصائي (**من نحصي**)،

-4- تحديد المتغيرات الإحصائية وذلك بجردها وتحديد نوعها ومختلف قيمها (**ماذا نحصي**)،

مرحلة جمع البيانات:

5- نحضر استمارة أسئلة لكل وحدة إحصائية،

6- نجمع البيانات، نملأ استمارة لكل وحدة إحصائية،

مرحلة التفريغ والعرض:

7- نقوم بتفريغ البيانات (الاستمارات) في جداول جامعة،

8- من أجل كل فرضية نقوم بتكوين جداول تكرارية تجيب عليها،

9- إذا تلزم الأمر، نعرض الجداول في أشكال بيانية موضحة،

مرحلة التحليل:

10- نقوم بالوصف و التحليل باستخدام المقاييس المناسبة،

مرحلة الاستقراء (تعميم النتائج و اتخاذ القرارات):

11- نستعمل طرق الإحصاء الاستقرائي لاتخاذ القرارات المناسبة في مدى قبول أو رفض فرضية ما.

نشير إلى أن بعض المختصين يقسمون علم الإحصاء إلى قسمين: "**الإحصاء الوصفي**" و يتناول جمع وعرض و تحليل البيانات و "**الإحصاء الاستقرائي**" الذي يتناول تعميم النتائج و اتخاذ القرارات.

1 – 2 – مثال لتوضيح ضرورة الإحصاء

كيف يمكن لمدير شركة تنتج مصابيح كهربائية أن يلخص ويصف لاجتماع مجلس الإدارة نتائج اختبار عمر عينة من 100 مصباح من إنتاج الشركة؟

إن عرض البيانات (الخام) من عمر كل مصباح في العينة أمر غير ملائم ويستغرق وقتا طويلا من أعضاء المجلس لتقويمها. ويمكن بدلا من ذلك أن يقوم من ذلك أن يقوم المدير باختزال البيانات بتوضيح أن متوسط عمر المصابيح التي تم اختبارها في العينة هو 360 ساعة وأن 95% من المصابيح التي فحصت قد عاشت بين 320 و 400 ساعة. ويكون المدير بذلك قد قدم معلومتين (متوسط العمر

وانتشار المفردات حول القيمة الوسطى) تميزان عمر المصابيح المائة التي تم فحصها. وقد يرى المدير أيضا أن يصف البيانات باستخدام جدول أو رسم بياني يوضح عدد أو نسبة المصابيح موزعة على فئات طول كل منها عشرة ساعات طبقا للعمر الذي قضاه كل مصباح. ويكون مثل هذا العرض الجدولي أو البياني مفيدا في الإلمام العام السريع بالبيانات. وبتلخيص وتوصيف البيانات على النحو الموضح يكون المدير قد استخدم الإحصاء الوصفي. وجدير بالذكر أن الإحصاء الوصفي يمكن استخدامه في تلخيص وتوصيف أي مجموعة من البيانات سواء كانت عينة (كالمثال السابق) أو مجتمعا (عندما تكون مفردات المجتمع معروفة ويمكن قياس خواصها).

لماذا قد يرغب هذا المدير أن يتطرق إلي الاستدلال الإحصائي؟ وماذا يتضمن هذا الاستدلال وماذا يتطلب.؟

1. تتطلب مراقبة جودة الإنتاج أن يكون لدى المدير فكرة جيدة تماما عن متوسط عمر المصابيح الكهربائية التي تنتجها الشركة والانتشار حول هذا المتوسط. غير أن فحص جميع المصابيح الكهربائية يؤدي إلى تدمير إنتاج الشركة كله. وحتى عندما لا يؤدي الفحص إلى تدمير المنتج، فان فحص الإنتاج كله يكون عادة باهظ التكلفة ويستغرق وقتا طويلا. وعليه، فان الإجراء المتبع هو أخذ عينة من الإنتاج والاستدلال على خواص وصفات الإنتاج كله (المجتمع) من الصفات المناظرة العينة المسحوبة من المجتمع.

2. يتطلب الاستدلال الإحصائي أولا أن تكون العينة ممثلة للمجتمع الذي تؤخذ منه. فإذا كانت الشركة تنتج المصابيح الكهربائية في مصانع مختلفة، باستخدام أكثر من وردية واحدة، وباستخدام مواد خام مشتراة من أكثر من مورد فهذه جميعا يجب أن تمثل في العينة بنسبة مساهمتها في الإنتاج الكلي للشركة. فباستخدام متوسط عمر المصابيح في العينة والانتشار حول هذا المتوسط يمكن لمدير الشركة أن يقدر، باحتمال 95% أن يكون تقديره صحيحا واحتمال 5% ان يكون تقديره خاطئا، ان متوسط العمر لكل المصابيح التي تنتجها الشركة يقع بين 320 و 400 ساعة (مثلا). وكبديل يمكن للمدير أن يستخدم معلومات العينة لكي يختبر، باحتمال 95% أن يكون على صواب، واحتمال 5% أن يكون على خطأ، إن متوسط العمر في مجتمع جميع المصابيح التي تنتجها الشركة اكبر من 320 ساعة. وسواء في التقدير أو في اختبار متوسط المجتمع باستخدام بيانات العينة يكون المدير مستخدما للاستدلال الإحصائي.

1 – 3 – تعريف الاحصائي الاستدلالي

الاستدلال الإحصائي هو استدلال عن مجتمع من خلال أخذ عينة عشوائية منه. و يمكن تعريف الاستدلال الإحصائي بأنه الاجراء (أو العملية) التي يتم من خلالها الحصول على معلومات من عينة (إحصاءات العينة) والوصول (إطلاق أحكام) عن المجتمع (معالم المجتمع). وهو يشمل:

- تقدير معالم المجتمع (التقدير بنقطة والتقدير بفترة أو بمجال).
- اختبار الفرضية (أو اختبار المعنوية الاحصائية)
- التنبؤ

<u>أمثلة على استخدام الإحصاء الاستدلالي:</u>

- اختبار فعالية دواء جديد: يتم اختبار الدواء على عينة من المرضى، ثم يتم استخدام الإحصاء الاستدلالي لتحديد ما إذا كان الدواء فعالًا بشكل ملحوظ.
- تقدير متوسط دخل الفرد في بلد ما: يتم جمع بيانات عن دخل عينة من السكان، ثم يتم استخدام الإحصاء الاستدلالي لتقدير متوسط دخل الفرد في البلد ككل.
- التنبؤ بنتائج الانتخابات: يتم جمع بيانات عن استطلاعات الرأي، ثم يتم استخدام الإحصاء الاستدلالي للتنبؤ بنتائج الانتخابات.

1 – 4 - مصطلحات ومفاهيم ضرورية:

1 - 4 – 1 – المجتمع الإحصائي: الأفراد أو الأشياء محل الدراسة. فإذ كنا نريد إجراء دراسة حول طول الطلبة الجامعيين الجزائريين، فيكون المجتمع الإحصائي في هذه الحالة الطلبة الجامعيين الجزائريين.

1 - 4 – 2 – العينة: هي جزء من المجتمع الإحصائي نلجأ إليها في حالة عدم إمكانية دراسة كل أفراد المجتمع ومن أجل تقليص تكاليف وجهد ووقت الدراسة.

ونعرف عينة البحث بأنها المجموعة المصغرة التي يختارها الباحث لإجراء دراسته عليها، وتعميم النتائج التي يحصل عليها من هذه الدراسة على المجتمع الأصلي، ولكي تكون نتائج دراسته متوافقةً مع المجتمع الأصلي عليه أن يكون حذراً باختيار العينة.

1 - 4 – 3 – الوحدة الإحصائية: هي فرد من المجتمع الإحصائي. ففي مثالنا السابق الوحدة الإحصائية هي طالب جامعي جزائري.

1 - 4 – 4 – المتغير الإحصائي (الخاصية الإحصائية): هي تلك الميزة أو الصفة التي نريد دراستها على الوحدة الإحصائية. في مثالنا السابق، المتغير الإحصائي هو الطول.

والمتغير الإحصائي له عدة أنواع:

المتغير الإحصائي

- **كيفي** (غير قابل للقياس)
  - **اسمي** (الترتيب اختياري)
    مثل:
    - اللون
    - الجنسية
    - الحالة المدنية
  - **ترتيبي** (الترتيب إجباري)
    مثل:
    - مستوى الدراسي
    - تقدير الطالب
    - رضى المستهلك
- **كمي** (قابل للقياس)
  - **منقطع** (منفصل) (غير قابل للتجزئة)
    مثل:
    - عدد الأبناء
    - عدد الفرق
  - **مستمر** (متصل) (قابل للتجزئة)
    مثل:
    - الطول
    - الوزن

1 - 4 - 5 - الإحصاء المعلمي والاحصاء اللامعلمي:

الإحصاء المعلمي (Parametric Tests) هو مجموعة من الاختبارات الإحصائية التي تشترط تحقق افتراضات معينة حول المجتمع الذي تأخذ منه عينة الدراسة. و يشترط في الإحصاء المعلمي أن تكون بيانات المجتمع محل الدراسة موزعة توزيعا طبيعيا. بعض أنواع الإحصاءات المعلمية: معامل ارتباط بيرسون ، اختبار التباين (الأحادي والمتعدد) ، اختبار z ، اختبار t

الإحصاء اللامعلمي (Non-parametric Tests) هي مجموعة من الاختيارات الإحصائية (البديلة) بغرض اختبار صحة أو عدم صحة فرض معين دون الأخذ في الاعتبار شكل توزيع بيانات المجتمع محل الدراسة. ويفضل استخدام هذا النوع من الاختبارات (الاختبارات اللامعلمية) في الحالات التالية:

- عدم معرفة شكل توزيع بيانات المجتمع محل الدراسة،
- حجم العينة صغيرا جدا،
- إذا كانت البيانات وصفية ولا يمكن التعبير عنها بصورة رقمية،
- بيانات ذات طبيعة ترتيبية.

بعض أنواع الاختبارات اللامعلمية: اختبار فريدمان، كا تربيع، كولموجروف سميرنوف، مان ويتني، ويلكوكسن، كروسكال وليس واختيار فاي وغيرها.

مزايا الاحصاء اللامعلمي:

1- قلة الافتراضات و لا تحتاج الى شروط كثيرة.
2- سهولة العمليات الحسابية المستخدمة.
3- امكانية تطبيقها على البيانات الوصفية.
4- امكانية تطبيقها على العينات الصغيرة.
5- تستخدم في حالة صعوبة الحصول على بيانات دقيقة والسرعة في جمع البيانات.

2 – طرق اختيار عينة عشوائية

الخطوة الأولى التي يقوم بها الباحث هي اختيار عينة البحث. ولا بد على العينة أن تكومن عشوائية، مما يعني منح جميع أفراد المجتمع فرصاً متساوية ليكونوا في العينة من تكوين عينة البحث العلمي. من أجل ذلك على الباحث اتباع طرق محددة نذكرها في مل يلي:

2 – 1 – العينة العشوائية البسيطة:

لتبسيط مفهوم هذه العينة: لديك من مجموعة من الأشخاص وأردت أن تأخذ منهم عينة متكونة من أربعة أشخاص عشوائيا، فإنك تسجل أسماء كل الأشخاص في قصاصات وتضعها في صندوق وتقوم بتحريك الأوراق ثم تخرج منهم أربعة. هذه العملية التي تسمى بالقرعة هي العينة العشوائية البسيطة إلا أننا عوض كتابة الأسماء في قصاصات نستعمل جدول الأرقام العشوائية لاختيار الأشخاص لكن المبدأ هو نفسه.

نستعمل هذا النوع من طرق المعاينة لما يكون أفراد المجتمع متجانسين ولدينا قائمة لهم.

نعطي أرقاما متسلسلة لعناصر (وحدات) المجتمع المراد دراسته،

نحدد عدد الأعمدة التي سنستخدمها من الجدول العشوائي للحصول على الأرقام المطلوبة، ويتوقف هذا على حجم المجتمع.

نحدد نقطة البداية في الجداول العشوائية، ثم نبدأ باختيار أول رقم من الجدول من نقطه البداية التي حددناها شرط أن يكون من ضمن الأعمدة التي قمنا باختيارها.

نحدد عناصر المجتمع ذات الأرقام التي تم اختيارها لتمثل وحدات العينة البسيطة العشوائية.

مثال:

طلبة السنة الأولى عددهم 950 طالب، قم باختيار عينة من 10 طلبة.

أولا نقوم بالحصول على قائمة كل الطلبة

ثم نقوم بترقيم الطلبة من 1 إلى 950

نستعمل جدول الأرقام العشوائية (أنظر الملحق رقم 1) ونختار الأرقام الثلاثة الأولى من العمود الأول (يمكننا اختيار عمود آخر كما يمكننا اختيار الثلاث أرقام الوسطى أو الأخيرة)

نقرأ الأرقام عموديا ونأخذ كل رقم ينتمي للمجال (1 – 950)

نتحصل على العينة التالية : الطلبة الذين أرقامهم : 056 – 581 – 271 – 187 – 804 – 283 – 676 – 280 – 168 – 507 (لاحظ: الرقم 985 لم نأخذه لأنه ليس لدينا طالب بهذا الرقم).

على الباحث الآن الاتصال بهؤلاء الطلبة لكي يسلم لهم استمارة البحث أو يجري معهم مقابلة، وليس من حقه استبدال طالب وجده غائبا الآن بآخر حاضر بل عليه أن يتصل به أو ينتظره حتى يحضر.

2 – 2 – العينة العشوائية الطبقية

اختيار العيِنة العشوائية الطبقية أسلوبٌ لاختيار العيِّنات يُستخدَم عندما يكون من المُمكن تقسيم المجتمع الإحصائي بشكل طبيعي إلى مجموعات أو طبقات صغيرة مختلفة غير متداخِلة.

تُؤخَذ عيّنات عشوائية بسيطة من كلِّ طبقة منفردة وتُجمَّع معًا لتكوين العيِّنة الكلية. يعكس حجم العيِّنة العشوائية من كلّ طبقةٍ حجمَ هذه الطبقة داخل المجتمع الإحصائي. ومن ثَمَّ تكون الطبقات مُمثَّلة في العيِّنة النهائية بنفس النِّسَب التي تُمثَّل بها داخل المجتمع الإحصائي.

مثال:

طلبة السنة الأولى عددهم 1000 طالب منهم 700 طالبة، قم باختيار عينة من 10 طلبة.

نلاحظ أن المجتمع يتكون من طبقتين (طلبة عددهم 300 و طالبات عددهم 700) إذن لا بد من استعمال العينة العشوائية الطبقية.

بما أننا نريد اختيار عينة حجمها 10 من مجتمع حجمه 1000 يعني بنسبة 1 من 100

يعني نأخذ 1 من 100 من الذكور (3) و 1 من 100 من الإناث (7)

نستعمل العينة العشوائية البسيطة لاختيار 3 طلبة من بين 300

نستعمل العينة العشوائية البسيطة لاختيار 7 طالبات من بين 700

3 طلبة + 7 طالبات الذين حصلنا عليهم يمثلون العينة العشوائية الطبقية التي نبحث عنها.

2 – 3 – العينة المنتظمة

وهي العينة التي يتم اختيارها في حالة المجتمعات الواسعة وعندما تكون مفردات هذه المجتمعات مرتبة بتسلسل معين او موضوعة في قوائم فنعمل على اختيار العينة المنتظمة. ونختارها وفق الخطوات الآتية:

- ترقيم أفراد المجتمع من رقم 1 إلى آخر مفردة.
- تقسيم المجتمع إلى عدد من الأقسام ويكون عدد هذه الأقسام مساوي لحجم العينة.
- يتم تحديد عدد المفردات في كل قسم وذلك بقسمة حجم المجتمع على عدد الأقسام.
- يتم اختيار مفردة واحدة فقط بطريقة الاختيار العشوائي من القسم الأول.
- لاستخراج رقم المفردة الثانية في العينة يتم إضافة رقم المفردة الأولى إلى الرقم الذي يمثل عدد المفردات في كل قسم والذي تم استخراجه في الخطوة رقم 3.
- لاستخراج رقم المفردة الثالثة في العينة يتم إضافة رقم المفردة الثانية إلى الرقم الذي يمثل عدد المفردات في كل قسم. وهكذا يتم استخراج بقية مفردات العينة.

مثال:

نريد استخراج عينة عدد أفرادها 10 من تلاميذ نهائي علوم والبالغ عددهم 200، بطريقة العينة المنتظمة. حدّد أرقام أفراد العينة المختارة.

الحل:

حجم المجتمع = 200 ، حجم العينة = 10

نرقم أفراد المجتمع من 1 – 200.

نقسم المجتمع إلى 10 أقسام (لأن حجم العينة هو 10).

سيكون عدد الأفراد في كل قسم 200/10 = 20 فرد

نختار المفردة الأولى بطريقة عشوائية، ونفرض أن رقم المفردة المختارة هو 13.

رقم المفردة الأولى = 13 (بطريقة عشوائية)

رقم المفردة الثانية = 13 + 20 = 33

رقم المفردة الثالثة = 33 + 20 = 53

رقم المفردة الرابعة = 53 +20 = 73

وهكذا...فستكون الأرقام المختارة في العينة هي:

13, 33, 53, 73, 93, 113, 133, 153, 173, 193

نلجأ لنستعمل العينة العشوائية المنتظمة لما لا نملك قائمة لأفراد المجتمع فلا يمكننا استعمال العينة البسيطة ولا الطبقية التي تشترط وجود قائمة لمفردات المجتمع المدروس. فمثلا عند اختيار عينة من الطلبة الذين يدخلون إلى المطعم المركزي (لا نملك قائمة لهم نستعمل العينة العشوائية المنتظمة بالطريقة التالية:

إذا كنا نعلم أن المطعم يقدم 1000 وجبة (تقريبا) ونريد اختيار عينة من 100 شخص، نقسم 1000 على 100 نحصل على 10 يعني من كل 10 أشخاص يخرجون من المطعم نختار واحدا. الأول نختاره عشوائيا و البقية بانتظام.

2 – 4 – العينة العنقودية

العينة العنقودية هي نوع من أنواع العينات العشوائية، وتعتمد على تقسيم التجمعات السكانية الكبيرة إلى عدة مجموعات بهدف تسهيل جمع بيانات البحث وتحليلها، ومن ثم يأخذ الباحث عينات عشوائية من المجموعات الفرعية بالاستعانة بأنواع أخرى من العينات.

مثال:

نريد إجراء بحث على تلاميذ السنة الرابعة متوسط في ولاية البليدة. إذا لم يمكننا الوصول إلى كل هؤلاء التلاميذ لأنه ليست لدينا سلطة مدير التعليم بالولاية لأنه لو توفرت لدينا هذه الامكانية فإنه غير مسموح لنا باستعمال العينة. إذا كان لا بد من استعمال العينة فسوف نستعمل العينة العنقودية لأننا نعلم أن الولاية تحتوي على مجموعة من البلديات وكل بلدية تحتوي على مجموعة من المتوسطات.

طريقة أخذ العينة الطبقية:

إذا كنا نريد أخذ عينة حجمها مثلا 120 تلميذ.

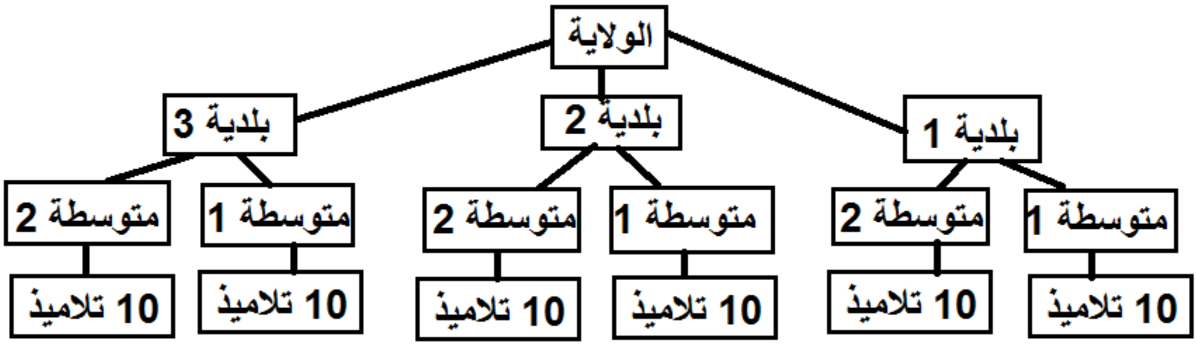
نختار ثلاث (3) بلديات عشوائيا من بين بلديات الولاية

من كل بلدية تم اختيارها، نختار متوسطتين (2) من بين مجموع متوسطاتها.

من كل متوسطة تم اختيارها نختار عشوائيا عشرة (10) تلاميذ من تلاميذ السنة الرابعة.

نكون في الأخير قد اخترنا:

120 تلميذ بالطريقة العنقودية:



في كل مرحلة (اختيار ثلاث بلديات أو اختيار متوسطتين أو اختيار عشرة تلاميذ) نستعمل العينة العشوائية البسيطة.

3 – تذكير بطريقة حساب النسبة، المتوسط الحسابي والانحراف المعياري والتباين وكذا الارتباط والانحدار

المطلوب من الطالب أن يكون ملما بالإحصاء الوصفي قبل الشروع في تعلم الإحصاء الاستدلالي. على الأقل ما زال متذكرا البديهيات والأساسيات التي تعلمها في الإحصاء الوصفي. في هذا الجزء من المطبوعة هذه نذكر الطالب بكيفية حساب النسبة والمتوسط الحسابي والانحراف المعياري لكن بطريقة مختصرة.

3 – 1 – تذكير بطريقة حساب النسبة:

أبسط أمر يتعلمه المرء هو حساب النسبة والنسبة المئوية، فنسبة الجزء من الكل هي قسمة عدد مفردات الجزء على عدد مفردات الكل و النسبة المئوية هي تلك المسبة مضروبة في 100 مع إضافة إشارة %.

مثال1: إذا علمت أن قسما به 50 طالب من بينهم 15 طالب و 35 طالبة. أوجد النسبة والنسبة المئوية للذكور في هذا القسم.

نسبة الذكور = عدد الذكور قسمة العدد الكلي = 15 ÷ 50 = 0.3

النسبة المئوية للذكور = النسبة x 100 = 30%

مثال2: إذا علمت أن نسبة الذكور من بين المواليد الجدد هي 0.525 (أو 52.5%) من مجموع المواليد الجدد، فما هو عدد المواليد الذكور إذا علمت أن عدد كل المواليد هو 10000.

عدد المواليد الذكور = العدد الكلي x النسبة = 10000 x 0.525 = 5250 ذكر

يعني عدد الإناث هو 10000 – 5250 = 4750 أنثى.

3 – 2 – تذكير بطريقة حساب المتوسط الحسابي:

المتوسط الحسابي، أو الوسط الحسابي، وأحياناً المعدّل (arithmetic mean).
في الرياضيات والإحصاء هو قيمة تتجمع حولها كل قيم المجموعة ويمكن من خلالها الحكم على بقية قيم المجموعة، فتكون هذه القيمة هي الوسط الحسابي.

يحسب الوسط الحسابي بجمع قيم عناصر المجموعة، ويقسم المجموع على عدد العناصر.

على سبيل المثال، لنفرض بأن لدينا العينة التالية $(x_1, \ldots, x_n)$

حيث n هو حجم العينة، فالوسط الحسابي $\bar{x}$ لهذه العينة يحسب بالعلاقة التالية:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}(x_1 + \cdots + x_n)$$

للملاحظة: المتوسط الحسابي لكل المجموعة يرمز له بالرمز الإغريقي μ

مثال على حساب المتوسط الحسابي: أحسب المتوسط الحسابي للبيانات التالية

لفئات	التكرار
5 – 2	2
8 – 5	2
11 – 8	1
14 – 11	3
17 – 14	2
20 – 17	1
المجموع	**11**

من أجل إيجاد المتوسط الحسابي نتبع الخطوات التالية :

أولاً: نجد مركز كل فئة

ثانياً: نضرب مركز كل فئة في تكراراها

ثالثاً: نجمع حواصل ضرب مركز كل فئة تكرارها

رابعاً: نقسم الناتج على التكرار الكلي

الفئات	مراكز الفئات x_i	التكرار n_i	مراكز الفئات التكرار n_ix_i
5 − 2	3.5	2	7
8 − 5	6.5	2	13
11 − 8	9.5	1	9.5
14 − 11	12.5	3	37.5
17 − 14	15.5	2	31
20 − 17	18.5	1	18.5
المجموع		**11** n	**116.5** $\sum n_ix_i$

$$\bar{X} = \frac{1}{n}\sum n_ix_i = \frac{116.5}{11} = 10.59$$

3 – 3 – تذكير بطريقة حساب الانحراف المعياري (التباين Variance):

التباين هو أحد مقاييس التشتت، وأكثرها استخداما في النواحي التطبيقية، ويعبر عن متوسط مربعات انحرافات القيم عن وسطها الحسابي. ويحسب التباين من بيانات العينة كتقدير لتباين المجتمع، فإذا كانت قراءات عينة عشوائية حجمها n هي ، $x_1, x_2, \dots, x_n$ ، فإن

تباين العينة ويرمز له بالرمز s^2 هو: .... $S^2 = \frac{\sum(x_i-\bar{x})^2}{n-1}$

أما تباين المجتمع الذي حجمه N فهو: .... $\sigma^2 = \frac{\sum(x_i-\mu)^2}{N}$

حيث أن $\bar{x}$ هو الوسط الحسابي لقراءات العينة، أي أن: .... $\bar{x} = \frac{1}{n}\sum x_i$

و أن μ هو الوسط الحسابي لقراءات العينة، أي أن: .... $\mu = \frac{1}{N}\sum x_i$

الانحراف المعياري: Standard Deviation

لجأ الإحصائيون إلى مقياس منطقي يأخذ في الاعتبار الجذر ألتربيعي للتباين، لكي يناسب وحدات قياس المتغير، وهذا المقياس هو الانحراف المعياري: $S = \sqrt{\frac{\sum(x_i-\bar{x})^2}{n-1}}$

معامل الاختلاف: Coefficient of Variation (C.V)

إن معامل الاختلاف (أفضل معاملات التشتت) يعتمد على أفضل مقاييس النزعة المركزية وأفضل مقاييس التشتت. إنه يوضح نسبة حصة كل وحدة من وحدات الوسط الحسابي والانحراف المعياري وعليه عند إجراء مقارنة بين قيم مجموعتين ثم مقارنة معامل الاختلاف للمجموعة الأولى مع معامل الاختلاف للمجموعة الثانية، عندئذ يقال عن المجموعة بأنها أكثر تجانسا إذا كان معامل الاختلاف اقل من الأخر.

يعرف معامل الاختلاف و الذي نرمز له بالرمز (C.V) كالآتي: $CV = {S}/{\bar{X}}$

خصائص التباين:

- لا يمكن حسابه من الجداول التكرارية المفتوحة، يتأثر بالقيم المتطرفة، قابل للعمليات الجبرية،
- إذا كان تباين المتغير x هو V(x)، فإن من أجل كل عددين حقيقيين a و b لدينا :

$V(b)= 0$ $\quad V(a.x) = a^2.V(x)$ $\quad V(a.x + b) = a^2.V(x)$

- إذا كان المتغيران $x:(x_1,x_2,\ldots,x_m)$ و $y:(y_1,y_2,\ldots,y_n)$ متغيران تباينهما على التوالي:

$S_x^2 = \frac{1}{m}\sum(x_i-\bar{x})^2$ و $S_y^2 = \frac{1}{n}\sum(y_i-\bar{y})^2$ فإن التباين المشترك للظاهرتين معا ونرمز له بالرمز

S_p^2 (Pooled variance): يعرف كما يلي: .... $S_p^2 = \frac{1}{m+n}\left(mS_x^2 + nS_y^2 + \frac{mn}{m+n}(\bar{x}-\bar{y})^2\right)$

وغالبا ما نستخدم الصيغة التالية لحساب التباين المشترك:.............. $S_p^2 = \frac{(m-1)S_x^2+(n-1)S_y^2}{m+n-2}$

- إذا كان المتغيران x و y متغيران مستقلان فإن: .... $V(x \pm y) = V(x) + V(y)$

الصيغة العملية للتباين:

$$S^2 = \frac{1}{n-1}\sum(x_i-\bar{x})^2 = \frac{1}{n-1}\sum[x_i^2 - 2\bar{x}x_i + \bar{x}^2] = \frac{1}{n-1}[\sum x_i^2 - 2\bar{x}\sum x_i + \sum \bar{x}^2]$$
$$= \frac{1}{n-1}[\sum x_i^2 - 2n\bar{x} + n\bar{x}^2] = \frac{1}{n-1}[\sum x_i^2 - n\bar{x}^2]$$

بنفس الطريقة يكون تباين المجتمع كالآتي: .... $\sigma^2 = \frac{1}{N}\sum x_i^2 - \mu^2$

3 – 4 – مسألة لمراجعة الإحصاء الوصفي:

لتكن المعطيات التالية:

الفئات	التكرارات
4 – 2	3
6 – 4	5
8 – 6	10
10 – 8	2

1 – أحسب المتوسط الحسابي

2 – أحسب المنوال

3 – أحسب الوسيط

4 – أحسب الربيع الأول والثالث

5 – أحسب الانحراف المتوسط

6 – أحسب الانحراف المعياري

7 – أحسب معامل الاختلاف

الحل:

من أجل الحسابات نقدر قيمة x_i بمنتصف الفئة،

من أجل حساب المتوسط μ نحسب العمود n_ix_i

من أجل حساب الانحراف المتوسط نضيف العمود $|x_i-\mu|$ ثم العمود $n_i|x_i-\mu|$

من أجل حساب الانحراف المعياري نضيف العمود $(x_i-\mu)^2$ ثم العمود $n_i(x_i-\mu)^2$

من أجل أشباه الوسيط نحسب التكرار المتجمع الصاعد

وبالتالي نتحصل على الجدول التالي:

الفئات	x_i	n_i	n_ix_i	$\lvert x_i-\mu \rvert$	$n_i\lvert x_i-\mu \rvert$	$(x_i-\mu)^2$	$n_i(x_i-\mu)^2$	F_i (↑)
4 – 2	3	3	9	3,1	9,3	9,61	28,83	3
6 – 4	5	5	25	1,1	5,5	1,21	6,05	8
8 – 6	7	10	70	0,9	9	0,81	8,1	18
10 – 8	9	2	18	2,9	5,8	8,41	16,82	20
المجموع		20	122		29,6		58,8	
المتوسط			**6,1**		**1,48**		**2,99**	

ملاحظة: xi هي منتصف الفئة، نقوم بجمع حدي الفئة ثم نقسم على 2

من الجدول يمكننا استنتاج الآتي:

1 – **المتوسط الحسابي:**

$$\overline{X} = \frac{\sum n_i x_i}{n} = \frac{122}{20} = 6.1$$

2 – **المنوال:** المنوال هو القيمة الأكثر تكرارا، يمكننا ملاحظة أن المنوال ينتمي للفئة المنوالية [6 – 8] ومنه المنوال يساوي:

$$M_o = L + C\frac{\Delta_1}{\Delta_1+\Delta_2} = 6 + 2\frac{5}{5+8} = 6.77$$

L: الحد الأدنى للفئة المنوالية،

C: طول الفئة المنوالية،

Δ_1 : الفرق بين تكرار الفئة المنوالية والفئة السابقة، و Δ_2 الفرق بين تكرار الفئة المنوالية والفئة الموالية لها.

3 – **الوسيط:** هو القيمة التي تقسم المجتمع إلى نصفين (أي القيمة n/2). يمكننا من خلال التكرار المتجمع الصاعد ملاحظة أن هذه القيمة تنتمي للفئة الوسيطية [8 – 6] ومنه:

$$M_e = L + C\frac{{}^{n}/_{2} - F_{e-1}}{n_e} = 6 + 2\frac{10-8}{10} = 6.4$$

L: الحد الأدنى للفئة الوسيطية،

C: طول الفئة الوسيطية،

F_{e-1} : التكرار المتجمع الصاعد للفئة السابقة، و

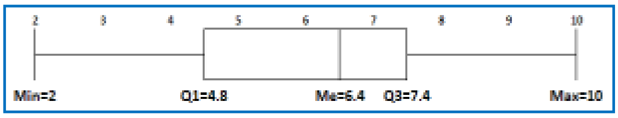
n_e : تكرار الفئة الوسيطية.

4 – **الربيع الأول والثالث**، يحسبان بنفس طريقة حساب الوسيط، كالآتي:

$$Q_1 = L + C\frac{{}^{n}/_{4} - F_{Q1-1}}{n_{Q1}} = 4 + 2\frac{5-3}{5} = 4.8$$

$$Q_3 = L + C\frac{{}^{3n}/_{4} - F_{Q3-1}}{n_{Q3}} = 6 + 2\frac{15-8}{10} = 7.4$$

يمكننا الآن تمثيل المعطيات في شكل علبة كما يلي:



5 – **الانحراف المتوسط:**

$$DM = \frac{1}{n}\sum n_i\,|x_i - \bar{x}| = \frac{29.6}{20} = 1.48$$

6 – **الانحراف المعياري:**

$$S^2 = \frac{1}{n}\sum n_i\,(x_i - \bar{x})^2 = \frac{58.8}{20} = 2.99$$

$$S = \sqrt{2.99} = 1.73$$

لحساب الانحراف المعياري، يمكن استعمال العلاقة

$$S^2 = \frac{1}{n}\sum n_i\,{x_i}^2 - \bar{x}^2$$

7 – **معامل الاختلاف:**

$$CV = \frac{S}{\bar{x}} = \frac{1.73}{6.1} = 0.28 = 28\%$$

3 – 5 – الارتباط والانحدار

ا- **تعاريف هامة:**

معامل الارتباط Auto Correlation:

في أحوال كثير يواجه الباحث حالات تتطلب دراسة متغيرين أو أكثر في آن واحد لبيان طبيعة ونوع العلاقة التي تربط هذه المتغيرات، لذا سوف ندرس مقاييس تحدد درجة و نوع وشكل العلاقة بين متغيرين أو أكثر.

الارتباط الخطي Linear Correlation:

الارتباط الخطي هو وجود متغيرين أو أكثر مرتبطين بعلاقات خطية معينة على سبيل المثال العلاقة بين طول الشخص و وزنه. فإذا كان التغير في إحدى المتغيرات يؤثر في تغير متغير أخر عندئذ يقال أن هذه المتغيرات مرتبطة فيما بينها.

الارتباط الخطى البسيط Simple Correlation:

إذا كان الغرض من التحليل هو تحديد نوع وقوة العلاقة بين متغيرين، يستخدم تحليل الارتباط، وأما إذا كان الغرض هو دراسة وتحليل أثر أحد المتغيرين على الآخر، يستخدم تحليل الانحدار، وفي هذا الفصل يتم عرض أسلوب تحليل الارتباط الخطي البسيط، أي في حالة افتراض أن العلاقة بين المتغيرين تأخذ الشكل الخطي.

الغرض من تحليل الارتباط الخطي البسيط:

الغرض من تحليل الارتباط الخطي البسيط هو تحديد نوع وقوة العلاقة بين متغيرين، ويرمز له في حالة المجتمع بالرمز ρ (رو)، وفي حالة العينة بالرمز r، وحيث أننا في كثير من النواحي التطبيقية نتعامل مع بيانات عينة مسحوبة من المجتمع، سوف نهتم بحساب معامل الارتباط في العينة r كتقدير لمعامل الارتباط في المجتمع، ومن التحديد السابق للغرض من معامل الارتباط، نجد أنه يركز على نقطتين هما:

- نوع العلاقة: وتأخذ ثلاث أنواع حسب إشارة معامل الارتباط كما يلي:

$(r < 0)$ علاقة عكسية، $(r > 0)$ علاقة طردية، $(r = 0)$ انعدام العلاقة بين المتغيرين.

- قوة العلاقة: نحكم على قوة العلاقة من حيث درجة قربها أو بعدها عن (±1)، حيث أن قيمة معامل الارتباط تقع في المدى$(-1 < r < 1)$، وقد صنف بعض الإحصائيين درجات لقوة العلاقة يمكن تمثيلها على الشكل التالي:

ارتباط طردي					ارتباط عكسي				
قوي جدا	فوي	متوسط	ضعيف	ضعيف جدا	ضعيف جدا	ضعيف	متوسط	قوي	قوي جدا
1	0.9	0.7	0.5	0.3	0.3−	0.5−	0.7−	0.9−	1−

ب- **<u>معامل الارتباط الخطي البسيط (بيرسون)</u>** Simple Linear Correlation Coefficient

معامل الارتباط الخطي البسيط هو القيمة العددية للعلاقة الخطية بين متغيرين وتحسب كالآتي:

$$r_{xy} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{(n\sum x_i^2 - (\sum x_i)^2)(n\sum y_i^2 - (\sum y_i)^2)}}$$

مثال1: لتكن المعطيات التالية ونريد حساب معامل الارتباط الخطي البسيط لبيرسون:

x	1	2	5	3	2	1	4
y	2	4	6	3	5	4	8

الحسابات:

	x	y	x.y	x^2	y^2
	1	2	2	1	4
	2	4	8	4	16
	5	6	30	25	36
	3	3	9	9	9
	2	5	10	4	25
	1	4	4	1	16
n=7	4	8	32	16	64
المجموع	$\sum x_i = 18$	$\sum y_i = 32$	$\sum x_i y_i = 95$	$\sum x_i^2 = 60$	$\sum y_i^2 = 170$
تربيع المجموع	$(\sum x_i)^2 = 324$	$(\sum y_i)^2 = 1024$			
n x المجموع			$n\sum x_i y_i = 665$	$n\sum x_i^2 = 420$	$n\sum y_i^2 = 1190$

$$r_{xy} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{(n\sum x_i^2 - (\sum x_i)^2)(n\sum y_i^2 - (\sum y_i)^2)}}$$

$$= \frac{665 - 18 \times 32}{\sqrt{(420 - 324)(1190 - 1024)}} = 0.7$$

اختبار معامل الارتباط:

إذا كان معامل الارتباط r محسوب من معطيات عينة يعني هو تقدير لمعامل الارتباط الحقيقي في المجتمع الكلي، في هذه الحالة نقوم باختبار مدى معنوية هذا المعامل (يعني هل معامل الارتباط الحقيقي يختلف فعلا عن الصفر) باستعمال اختبار ستودنت أو اختبار T كما يلي:

H0 : $\rho=0$ (عدم وجود ارتباط خطي بين x و y) ، H1 : $\rho\neq 0$

نقوم بحساب t0 كما يلي: $t_0 = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.7\sqrt{7-2}}{\sqrt{1-0.7^2}} = 2.19$

نستخرج t الجدولية عند مستوى دلالة $\alpha=0.05$ و درجة حرية $ddl=n-2=5$ $t_{0.05,5} =$ 2.05

القرار : إذا كانت القيمة المطلقة ل t0 المحسوبة أكبر من قيمة t الجدولية نرفض H0 ونقبل H1 والعكس صحيح. في حالتنا هذه (نرفض H0 ونقبل H1) أي فعلا هناك علاقة خطية بين المتغيرين x و y.

ج- **معامل الارتباط الجزئي** Partial Correlation Coefficient

يلاحظ في بعض الأحيان و جود ارتباط بين متغيرين يعزى جزئيا إلى ارتباط متغير ثالث مرتبط مع كلا المتغيرين. لنفرض إن المتغيرين X1,X2 مرتبطين و هناك متغير أخر X3 مرتبط مع كلا المتغيرين و إننا نرغب في قياس درجة الارتباط بين X1,X2 باستبعاد اثر المتغير الثالث X3 على كلا المتغيرين. إن المعامل الذي يقيس درجة ارتباط متغيرين باستبعاد اثر الثالث يدعى بمعامل الارتباط الجزئي. على سبيل المثال عدد السجائر المدخنة X1 يوميا و الإصابة بنوع معين من أمراض الرئة X2 يتأثر بارتباط عمر المدخن X3 وعدد سنوات التدخين X4 مع كل من X1,X2. ويمكن إيجاد قيمة معامل الارتباط الجزئي وفق ما يلي: افرض إن X1,X2,X3 تمثل ثلاثة متغيرات عشوائية بحيث

أن المتغيرين X1,X2 مرتبطين و أن X3 مرتبط مع كل من X1,X2 وافرض أن على أساس عينة من المفردات حجمها n تم الحصول على القياسات المتناظرة لهذه المتغيرات ((X1i , X2i , X3i), i=1,2,3………n) وأن معامل الارتباط البسيط بين متغيرين منهما هو r12 , r13 , r23 فان معامل الارتباط الجزئي بين X1,X2 باستبعاد اثر الثالث X3 هو:

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}}$$

وفي حالة وجود أربعة متغيرات ونرغب في حساب الارتباط الجزئي بين الأول و الثاني باستبعاد اثر الثالث والرابع فان هذا المعامل هو: $r_{12.34} = \frac{r_{12.3} - r_{14.3}r_{24.3}}{\sqrt{(1-r_{14.3}^2)(1-r_{24.3}^2)}}$

د- **<u>معامل الارتباط الرتب لسبيرمان</u>** Spearman’s rank correlation Coefficient

إذا كان X,Y متغيرين من النوع الوصفي و كانت البيانات المتحصل عليها من x,y على أساس عينة حجمها n هي بهيئة صفات غير قابلة للقياس الكمي. وإذا كانت (Xi , Yi , i=1,2,3,…n) ممكنة الترتيب تصاعديا أو تنازليا وفق معيار معين يمتاز به كل متغير، عندئذ استنادا لهذا الترتيب ولكل متغير يمكن تخصيص قيم سلسلة الأعداد الطبيعية (1,2,3, …,n) لصفات الترتيب بحيث إن كل صفة يخصص لها أحد إعداد هذه السلسة، في حالة عدم تكرار أيه صفة منها (وقد يكون التخصيص تنازلي). القانون المستخدم لإيجاد الارتباط لمثل هذه الحالات:

$$r_{xy} = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)} \quad , \quad d_i = x_i - y_i$$

مثال1: لتكن المعطيات التالية ونريد حساب معامل الارتباط لسبيرمان:

(x , y) = { (1 , 2), (2 , 4), (3 , 8), (4 , 7), (5 , 9) }

المعطيات تتحول إلى رتب:

D^2	D	رتب y	رتب x
0	0	1	1
0	0	2	2
1	−1	4	3
1	1	3	4
0	0	4	5

مجموع مربعات الفروق =2 ، مجموع الأزواج=N=5 ،

$$r_{xy} = 1 - \frac{6\sum d_i^2}{n(n^2-1)} = 1 - \frac{6 \times 2}{5(5^2-1)} = 1 - \frac{12}{120} = 0.9$$

مثال2: لتكن المعطيات التالية ونريد حساب معامل الارتباط لسبيرمان:

4	1	2	3	5	2	1	**x**
8	4	5	3	6	4	2	**y**

المعطيات تتحول إلى رتب:

	4	1	2	3	5	2	1	x
	8	4	5	3	6	4	2	y
	6	1	3	5	7	3	1	رتب X
	7	3	5	2	6	3	1	رتب Y
	1	−2	−2	3	1	0	0	D
19	1	4	4	9	1	0	0	D^2

$$r_{xy} = 1 - \frac{6\sum d_i^2}{n(n^2-1)} = 1 - \frac{6 \times 19}{7(7^2-1)} = 1 - \frac{114}{336} = 0.66$$

هـ - نموذج الانحدار الخطي البسيط:

يستخدم نموذج الانحدار البسيط العلاقة بين متغير تابع Y ومتغير مستقل أو مفسر X هذه العلاقة تسمح بشرح قيم Y بواسطة قيم مأخوذة من X. وتعرف العلاقة العامة للانحدار البسيط بـ:

Y = a + bX

نلاحظ أن $\hat{a}$ تمثل تقاطع خط الانحدار مع محور العينات Y (عندما X=0). وأن $\hat{b}$ تسمى معامل الانحدار أي مقياس يوضح مقدار تغير **Y** إذا ما تغيرت X بوحدة واحدة و يلاحظ إن $\hat{b}$ قد تكون موجبة أو سالبة.

تقدير المعالم a و b :

$$b = \frac{\text{Covar}(x,y)}{\text{Var}(x)}$$

$$b = \frac{\frac{\sum x_i y_i}{n} - \bar{x}\bar{y}}{\frac{\sum x_i^2}{n} - \bar{x}^2} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2}$$

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

معامل الارتباط الخطي:

يستعمل معامل الارتباط الخطي r لمعرفة درجة الارتباط بين تغيرات y و تغيرات x و هو محصور في المجال [−1 , +1].

* r =1 هناك ارتباط كلي بين x و y.

* r = −1 هناك ارتباط كلي سالب.

* r = 0 لا توجد علاقة بين تغيرات x وy

| r | قريبة من 1 العلاقة قوية بين x و y

| r | قريبة من 0 العلاقة ضعيفة بين x و y

أما الصيغة الجبرية لمعامل الارتباط فهي:

$$r = \frac{\text{Covar}(x,y)}{\sqrt{\text{Var}(x)}\sqrt{\text{Var}(y)}}$$

$$r = \frac{\frac{\sum x_i y_i}{n} - \bar{x}\bar{y}}{\sqrt{\frac{\sum x_i^2}{n} - \bar{x}^2}\sqrt{\frac{\sum y_i^2}{n} - \bar{y}^2}}$$

$$r = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sqrt{\sum x_i^2 - n\bar{x}^2}\sqrt{\sum x_i^2 - n\bar{x}^2}}$$

مثال: أوجد معادلة خط الانحدار y على x وقيمة معامل الارتباط الخطي من البيانات التالية:

x	1	2	5	3	2	1	4
y	2	4	6	3	5	4	8

الحل: نشكل الجدول التالي:

x	y	$x_i y_i$	x_i^2	y_i^2
1	3	3	1	9
2	4	8	4	16
5	6	30	25	36
3	3	9	9	9
2	5	10	4	25
1	4	4	1	16
4	8	32	15	64
18	**33**	**96**	**60**	**175**
2.57	**4.71**	**13.71**	**8.57**	**25**

السطر ما قبل الأخير فيه المجاميع، والأخير المعدلات (المجاميع قسمة عدد القيم 7).

– معادلة خط الانحدار:

$$b = \frac{\frac{\sum x_i y_i}{n} - \bar{x}\bar{y}}{\frac{\sum x_i^2}{n} - \bar{x}^2} = \frac{13.71 - 2.57 \times 4.71}{8.57 - 2.57^2} = 0.814$$

$$a = \bar{y} - b\bar{x} = 4.71 - 0.814 \times 2.57 = 2.62$$

ومنه معادلة خط الانحدار تكتب على الشكل:

$$Y = 2.62 + 0.814X$$

– معامل الارتباط الخطي:

يمكن حساب r كما سبق شرح ذلك:

$$r = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sqrt{\sum x_i^2 - n\bar{x}^2}\sqrt{\sum y_i^2 - n\bar{y}^2}} = \frac{12.7143}{18.034} = 0.7$$

يمكن حساب *r* في Excel باستعمال العلاقة:

=COEFFICIENT.CORRELATION(Plage A; Plage B)=0.7

ج) – اختبار t لدلالة ميل معادلة خط الانحدار :

إن اختبار الميل يتطلب حساب القيمة t_0 كما يلي:

$$t_0 = \frac{b}{\sqrt{\frac{\frac{S_y^2}{S_x^2} - b^2}{n-2}}}$$

ثم مقارنة هذه القيمة مع القيمة t الجدولية المستخرجة من جدول ستودنت عند درجة حرية n−2 ومستوى دلالة α.

4 – التقدير بنقطة والتقدير بفترة (بمجال):

4 – 1 – التقدير بنقطة:

- هو طريقة لتقدير قيمة معلمة مجتمعية (مثل المتوسط، أو الانحراف المعياري، أو النسبة) باستخدام **قيمة واحدة**.
- تُستخدم عادةً **إحصائية العينة** (مثل المتوسط الحسابي للعينة) كنقطة تقدير لقيمة المعلمة المجتمعية.

4 – 2 – التقدير بفترة (بمجال):

- هو طريقة لتقدير قيمة معلمة مجتمعية باستخدام **مجال من القيم**.
- تُستخدم هذه الطريقة عندما **لا نكون متأكدين تمامًا من القيمة الحقيقية للمعلمة**.
- يحدد مجال الثقة من خلال **مستوى الثقة**، وهو احتمال أن يحتوي المجال على القيمة الحقيقية.

4 – 3 – مثال: نفترض أننا نريد تقدير متوسط طول الطلاب في مدرسة معينة.

التقدير بنقطة:

- نقوم بقياس طول 100 طالب عشوائيًا من المدرسة.
- نحسب المتوسط الحسابي لأطوال الطلبة في العينة، لنفترض أنه 170 سم.
- نقدر أن متوسط طول جميع الطلاب في المدرسة هو 170 سم.

التقدير بفترة:

- بعد إيجاد مجال الثقة (نتعلم ذلك لاحقا) نعط النتيجة التالية (على سبيل المثال): نثق بنسبة 95% أن متوسط طول جميع الطلاب في المدرسة يقع بين 167.5 سم و 173.5 سم.

4 – 4 – رموز ومصطلحات في الإحصاء الاستدلالي:

- الإحصاء الاستدلالي: الاستدلال الإحصائي هو عملية استنتاج معالم المجتمع من إحصائيات العينة. وهو نوعان التقدير واختبار الفروض، التقدير هو عملية تقدير معلمة المجتمع من إحصائية العينة. وهناك التقدير بنقطة والتقدير بمجال. أما التقدير بنقطة فهو عبارة عن عدد واحد فقط فمثلا $\bar{X}$هي تقدير لـ μ و S هي تقدير لـ δ أما التقدير بمجال فيعطي مجالا يسمى بمجال الثقة ونعين له معاملا يسمى بمعامل الثقة.

2 - رموز ومصطلحات:

N.................حجم المجتمع

n.................حجم العينة

μ.................المتوسط الحسابي للمجتمع

x .... المتوسط الحسابي للعينة

P.................النسبة الموجودة في المجتمع

p.................النسبة الموجودة في العينة (و أحيانا نستعمل f)

σ.................الانحراف المعياري للمجتمع

S.................الانحراف المعياري للعينة

$\hat{\sigma}$.................تقدير الانحراف المعياري للمجتمع

$\sigma_{\bar{x}}$ ....الخطأ المعياري للمتوسط

z_{α}.................معامل الثقة (التوزيع الطبيعي)

t_{α}.................معامل الثقة (توزيع ستودنت)

e = الدقة ..... (خطأ التقدير) μ-x أو P- p

ddl : n-1 درجة الحرية

5 – مجال الثقة لمتوسط المجتمع:

إذا أخذنا من مجتمع عينة وتحصلنا على متوسطها وانحرافها المعياري، وأردنا تقدير بمجال وفي حدود ثقة معينة متوسط هذا المجتمع؟

فإن مجال الثقة لمتوسط المجتمع يعطى بالعلاقة:

$\mu = \bar{x} \pm$ (معامل الثقة)x(الخطأ المعياري لمتوسط)

1- الخطأ المعياري للمتوسط يعطى بالعلاقة: $\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{s}{\sqrt{n-1}}$

بالنسبة للمجتمعات المحدودة ذات الحجم N وعندما تكون $n>0.05N$ نضرب الخطأ المعياري في معامل التصحيح: $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}\sqrt{\dfrac{N-n}{N-1}}$

2- معامل الثقة: معامل الثقة يستخرج من جدول التوزيع الطبيعي z_α إذا كانت $n>30$، و يستخرج من جدول توزيع ستودنت t_α إذا كانت $n<30$ والمجتمع يتبع التوزيع الطبيعي. و نستعمل نظرية تشبتشيف في خلاف ذلك.

وفي ما يلي بعض القيم Z_α الأكثر استعمالا:

10%	5%	1%	0.5%	نسبة الخطأ α
90%	95%	99%	99.5%	معامل الثقة 1 – α
1.645	1.96	2.58	2.81	Z_α

مثال1: في حدود ثقة 95% أوجد مجال الثقة للمتوسط أوزان الأشخاص في المجتمع إذا علمت أن متوسط عينة متكونة من 50 شخص هو = 70 كغ بانحراف معياري قدره S=15.4 كغ.

$$\mu \in \bar{X} \pm 1.96 \frac{S}{\sqrt{n-1}} = 70 \pm 1.96 \frac{15.4}{\sqrt{49}} = 70 \pm 4.312 = [65.688 ; 74.312]$$

مثال2: : عينة من 8 عناصر ، متوسطها 20.11 وانحرافها المعياري 0.42 . أوجد مجال الثقة لمتوسط المجتمع في حدود ثقة 95% ؟

من جدول ستودنت و من أجل: $\alpha=0.05$ و $ddl=8-1=7$ نحصل على $t_\alpha = 2.365$.

و منه يكون مجال ثقة لمتوسط المجتمع :

$$\left[20.11 - 2.365 \frac{0.422}{2.828} \quad ; \quad 20.11 + 2.365 \frac{0.422}{2.828}\right] = [19.75 ; 20.46]$$

ملاحظة : بما أن الانحراف المعياري (S) يتناسب عكسا مع $\sqrt{n}$، فمجال الثقة يقل طوله كلما كبرت n

وفي ما يلي تفصيل للخطوات السابقة:

5 – 1 – حساب المتوسط الحسابي للعينة:

تأكد من أن البيانات المستخدمة لحساب المتوسط الحسابي تمثل المجتمع بشكل دقيق.
استخدم الآلة الحاسبة أو برنامجًا إحصائيًا لحساب المتوسط الحسابي بدقة.

5 – 2 – حساب الانحراف المعياري للعينة:

تأكد من استخدام صيغة الانحراف المعياري المناسبة (معلوم أو غير معلوم). استخدم الآلة الحاسبة أو برنامجًا إحصائيًا لحساب الانحراف المعياري بدقة.

5 – 3 – تحديد مستوى الثقة:

حدد مستوى الثقة المناسب لدراستك (عادةً 95% أو 99%). استخدم قيمة مستوى الثقة لتحديد قيمة Z من جدول Z (جدول التوزيع الطبيعي).

5 – 4 – حساب هامش الخطأ (marge d'erreur):

استخدم صيغة هامش الخطأ المناسبة (معلوم أو غير معلوم). تأكد من استخدام قيم الانحراف المعياري وحجم العينة بشكل صحيح.

5 – 5 – حساب مجال الثقة:

استخدم صيغة مجال الثقة: **متوسط العينة ± هامش الخطأ**

$$\mu = \bar{X} \pm marge\ d'erreur$$

حيث:

μ : متوسط المجتمع

$\bar{X}$: متوسط العينة

تأكد من استخدام نفس وحدات القياس في جميع الخطوات.

نصائح لتحسين دقة مجال الثقة:

زيادة حجم العينة: كلما زاد حجم العينة، زادت دقة مجال الثقة.

التحقق من توزيع البيانات: تأكد من أن البيانات تتبع توزيعًا طبيعيًا أو تقاربًا طبيعيًا.

استخدام اختبارات الفرضيات: اختبر ما إذا كانت فروضك حول متوسط المجتمع صحيحة.

5 – 6 – كيفية حساب هامش الخطأ (marge d'erreur) للمتوسط:

يعتمد حساب هامش الخطأ على نوع البيانات (معلوم أو غير معلوم) ومستوى الثقة.

5 – 6 – 1 – البيانات معلومة الانحراف المعياري:

إذا كان تباين المجتمع معلوما أو كان حجم العبنة كبير بدرجة كافية ($n \geq 30$) فإننا نستخدم العلاقة الموالية لحساب هامش الخطأ:

$$Marge\ d'erreur = z * \frac{\sigma}{\sqrt{n}}$$

و يكون متوسط المجتمع ينتمي للمجال:

$$\mu \in \left[\bar{X} \pm z * \frac{\sigma}{\sqrt{n}}\right]$$

حيث:

- z : قيمة Z من جدول Z (جدول التوزيع الطبيعي)، تعتمد على مستوى الثقة.
- σ : الانحراف المعياري: انحراف معياري للمجتمع (معلوم).
- n : حجم العينة: عدد العناصر في العينة.

5 – 6 – 2 – البيانات غير معلومة الانحراف المعياري:

تقدير متوسط المجتمع μ باستخدام التوزيع t :

تناولنا فيما سبق التقدير الإحصائي للوسط الحسابي للمجتمع في الحالات التي يكون فيها الانحراف المعياري للمجتمع δ معلوماً، و (أو) أن العينة كبيرة بدرجة كافية ($n \geq 30$). ولكن إذا كانت العينة صغيرة بمعنى أن حجمها أقل من (30) مفردة، والانحراف المعياري للمجتمع الطبيعي غير معلوم (مجهول)، فإن التوزيع الإحصائي المتبع في مثل هذه الحالات هو ما يطلق عليه توزيع ستيودنت أي " توزيع t " فعند تقدير متوسط عمر الناخب في مدينة ما عن طريق سحب عينة صغيرة (حجمها أقل من 30 ناخب) التوزيع الطبيعي يكون في مثل هذه الحالات غير مناسب لصغر حجم العينة أولاً، ثم عدم معرفة الانحراف المعياري لعمر الناخب ثانياً. لذا فإن الأسلوب الإحصائي المتبع في حالات كهذه هو استخدام " توزيع t " والذي يسميه البعض "توزيع العينات الصغيرة".

ولعل الاختلاف الأساسي بين توزيع t والتوزيع الطبيعي هو أن الانحراف المعياري للعينة هو المستخدم في الأول بدلا من الانحراف المعياري للمجتمع في الثاني، وفيما عدا ذلك فالتوزيعان متماثلان وكلما زادت قيمة n كلما اقترب توزيع t من توزيع z ويعتمد توزيع t على ما يعرف بدرجات الحرية.

درجات الحرية :

تعرف درجات الحرية بأنها عدد المشاهدات المستقلة في العينة والتي تساوي حجم العينة مطروحاً منه عدد القيود أو معالم المجتمع التي يتم تقديرها من بيانات العينة، وبصفة عامة إذا كان عدد القيود k فإن درجات الحرية تساوي $n - k$.

ويمكن تحديد الشروط الثلاثة لاستخدام توزيع t كما يلي:

- أن يكون المجتمع المسحوبة منه العينة له توزيع طبيعي.
- والانحراف المعياري للمجتمع δ غير معروف (أو مجهول).
- والعينة صغيرة (حجمها أقل من 30 مفردة).

خلاصة: إن مجال الثقة للوسط الحسابي للمجتمع في حالة δ مجهولة وحجم العينة أقل من 30 يأخذ الشكل التالي:

$$\left[\bar{x} \pm t_{\frac{\alpha}{2},n-1} \frac{S}{\sqrt{n-1}}\right]$$

مثال: لنفترض أننا نريد حساب هامش الخطأ لعينة من 100 شخص بمستوى ثقة 95%.

معلومة الانحراف المعياري: انحراف معياري للمجتمع σ = 5

Z = 1.96 (من جدول Z)

$$Marge\ d'erreur = 1.96 * \frac{5}{\sqrt{100}} = 0.98$$

غير معلومة الانحراف المعياري: انحراف معياري للعينة s = 6

t = 1.984 (من جدول t، درجات الحرية = 99)

$$Marge\ d'erreur = 1.984 * \frac{6}{\sqrt{99}} = 1.19$$

في هذا المثال، يكون هامش الخطأ أكبر عندما يكون الانحراف المعياري غير معلوم.

5 – 6 – 3 – استعمال تصحيح Cochran إذا كان حجم المجتمع معلوما:

إذا كان حجم المجتمع معلوما لا بد من تصحيح مجال الثقة وبالأخص إذا كان حجم المجتمع ليس كبير جدا مقارنة بحجم العينة (إذا كان حجم العينة أكبر من 5% من حجم المجتمع). وطريقة التصحيح هي مبينة في المثال التالي:

مثال : لنفترض أن لدينا البيانات التالية:

- مجتمع: 100 شخص
- عينة: 30 شخص
- متوسط العينة: 150 سم
- انحراف معياري للعينة: 5 سم
- مستوى الثقة: 95%

نريد حساب مجال الثقة لمتوسط المجتمع مع التصحيح:

1. حساب عامل التصحيح f:

$$f = \frac{N-n}{N-1} = \frac{100-30}{100-1} = \frac{70}{99} = 0.7$$

2. حساب القيمة الحرجة z: $z\_\alpha/2 = z\_0.025 = 1.96$ (من جدول z)

3. حساب مجال الثقة:

$$\mu = \bar{X} \pm z\frac{S}{\sqrt{n}}\sqrt{f} = 150 \pm 1.96\frac{5}{\sqrt{30}}\sqrt{0.7} = 150 \pm 1.5$$

مجال الثقة = متوسط العينة ± ($z\_\alpha/2$ x الخطأ المعياري x √f)

= 150 سم ± 1.5 سم

التفسير:

مع مستوى ثقة 95%، يمكننا أن نكون متأكدين بنسبة 95% من أن متوسط المجتمع يقع بين 148.5 سم و 151.5 سم.

أخذت طريقة Cochran بعين الاعتبار حجم المجتمع المحدود (100 شخص) عند حساب مجال الثقة.

5 – 6 – 4 – تقدير مجال الثقة للفرق ما بين متوسطي مجتمعين معلومي التباين:

إذا أخذنا من مجتمعين طبيعيين تباينهما $\delta_1{}^2$ و $\delta_2{}^2$ عينتين n_1 و n_2 وتحصلنا على المتوسطين الحسابيين X_1، و X_2 ، وأردنا أن نقدر بمجال وفي حدود ثقة معينة الفرق ما بين متوسطي المجتمعين μ_1 و μ_2 . فإن:

$$\mu_1 - \mu_1 = X_1 - X_2 \pm \text{(معامل الثقة)(خطأ معياري)}$$

معامل الثقة: من التوزيع الطبيعي (z)،

الخطأ المعياري: $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

مثال: أخذنا عينة حجمها 22 من القسم 1 وكان متوسط أطوالهم 119.38سم وأخذنا عينة أخرى حجمها 25 من القسم 2 فكان متوسط أطوالهم 126سم. إذا علمت أن الانحراف المعياري لأطوال القسمين 1 و 2 هو 4.5سم و4.13سم على التوالي.

المطلوب: أوجد في حدود 90% ثقة، مجال الثقة حول الفرق ما بين متوسطي أطوال المجتمعين.

الحل: معامل الثقة في حدود 90%=1.645، الخطأ المعياري:

$$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{4.5^2}{22} + \frac{5.13^2}{25}} = 1.405$$

مجال الثقة: $\mu_1 - \mu_2 = 119.38 - 126 \pm (1.645)(1.405) = -6.62 \pm 2.311$

أي $[-4.309 , -8.931]$

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \rightarrow N(0,1)$$

6 – مجال الثقة لنسبة

مجال الثقة لنسبة هو نطاق من القيم التي نثق بنسبة معينة (مثل 95%) أن النسبة الحقيقية للمجتمع تقع فيها.

إذا أخذنا من مجتمع عينة وتحصلنا على نسبة، وأردنا أن نقدر بمجال وفي حدود ثقة معينة النسبة الحقيقية الموجودة في المجتمع؟

إذا كان حجم العينة n حيث المتغير المدروس X يمكنه أن يأخذ فقط القيم 0 و 1 ، حيث نسبة ظهور x=1 هي p=k/n حيث k هو عدد المرات التي ظهر فيها x=1.

فإن مجال الثقة للنسبة في المجتمع هو: **P = p ± (معامل الثقة)x(الخطأ المعياري لنسبة)**

معامل الثقة نستخرجه بنفس الطريقة التي استعملناها في تقدير متوسط المجتمع.

الخطأ المعياري $\sigma_p = \sqrt{\frac{p(1-p)}{n}}$ ويضرب في معامل التصحيح $\sqrt{\frac{N-n}{N-1}}$ إذا كان $n > 0.05$ N

مثال: في مدرسة تضم 1000 تلميذ، أخذنا عينة من 35 تلميذ ووجدنا أن 7 منهم لديهم التهابات معدية. أوجد نسبة المصابين في كل المدرسة بنسبة خطأ 5% .

الحل: الصفة المدروسة هنا لا تأخذ سوى قيمتين (x=1 للتلاميذ المصابين وx=0 للتلاميذ الغير مصابين)

نسبة التلاميذ المصابين في العينة هي p=7/35=0.2 . و يصبح مجال الثقة بمعامل ثقة 95% كالآتي:

$$0.2-1.96\sqrt{\frac{0.2(1-0.2)}{35}} \leq P \leq 0.2+7.96\sqrt{\frac{0.2(1-0.2)}{35}}$$

يعني: $0.068 \leq P \leq 0.332$

وفي ما يلي خطوات حساب مجال الثقة لنسبة:

6 – 1 – حساب النسبة الموجودة في العينة:

نقوم بحساب نسبة العينة (النسبة المئوية) من خلال قسمة عدد النجاحات في العينة على حجم العينة.

6 – 2 – حساب الخطأ المعياري للنسبة:

$$\sqrt{\frac{p(1-p)}{n}}$$

حيث:

- p: نسبة العينة.
- n: حجم العينة.

6 – 3 – تحديد مستوى الثقة:

نحدد مستوى الثقة المناسب لدراستنا (عادةً 95% أو 99%).

نستخدم قيمة مستوى الثقة لتحديد قيمة Z من جدول Z.

6 – 4 – حساب هامش الخطأ:

نستخدم صيغة هامش الخطأ لنسبة:

$$marge\ d'erreur = z * \sqrt{\frac{p(1-p)}{n}}$$

6 – 5 – حساب مجال الثقة:

نستخدم صيغة مجال الثقة: **نسبة العينة ± هامش الخطأ**

مثال:

لنفترض أننا نريد حساب مجال الثقة لنسبة من عينة من 100 شخص بمستوى ثقة 95%.

نسبة العينة = 50% (50 نجاحًا من 100)

انحراف معياري للعينة = $0.05 = \sqrt{\frac{0.5(1-0.5)}{100}}$

Z = 1.96 (من جدول Z)

هامش الخطأ = 1.96 * (0.05) = 0.098

مجال الثقة = 50% ± 0.098% = (49.902%، 50.098%)

في هذا المثال، نثق بنسبة 95% أن النسبة الحقيقية للمجتمع تقع بين 49.902% و 50.098%.

6 – 6 – استعمال تصحيح Cochran إذا كان حجم المجتمع معلوما:

نستعمل نفس التصحيح وبنفس الطريقة التي استعملناها من أجل تصحيح مجال الثقة للمتوسط يعني نضرب الخطأ المعياري في: $\sqrt{{}^{N-n}/_{N-1}}$

6 – 7 – نصائح لتحسين دقة مجال الثقة لنسبة:

زيادة حجم العينة: كلما زاد حجم العينة، زادت دقة مجال الثقة.

التحقق من توزيع البيانات: تأكد من أن البيانات تتبع توزيعًا ثنائيًا.

استخدام اختبارات الفرضيات: اختبر ما إذا كانت فروضك حول نسبة المجتمع صحيحة.

7 – حساب حجم العينة الضروري

في الكثير من الأحيان يتساءل الباحث عن حجم العينة الضروري لإجراء بحثه، من أجل ذلك لا بد للباحث أن تكون لديه بعض المعلومات عن المجتمع (التشتت) وعن الدقة التي يريد الوصول إليها .

العوامل التي تحدد حجم العينة

حجم العينة من أهم الأسئلة التي ينبغي الجواب عنها. وهو مرتبط بعدة عوامل هي:

1 – حجم المجتمع الإحصائي الذي ستسحب منه العينة. كلما كان حجم المجتمع كبيراً كلما تطلب ذلك أن يكون حجم العينة كبيراً. فالعينة المتكونة من 40 طالبا تعد ممثلة تمثيلاً صادقا لفوج حجمه 50 طالبا ولكنها ليست كذلك لمدرسة بها 1000 طالبا.

2 – درجة الاختلاف بين مفردات المجتمع الإحصائي. فإذا كانت درجة الاختلاف كبيرة بين أفراد ذلك المجتمع استدعى الأمر زيادة حجم العينة والعكس صحيح.

3 – نسبة الخطأ المسموح به ودرجة الثقة التي يرغب الباحث في توافرها في النتائج التي يصل إليها من دراسته للعينة. اعتاد الباحثون أن يقبلوا حجم العينة الذي يستطيعون بنسبة ثقة 95% أن يعتمدوا على البيانات التي يوفرها لبحثهم وتساعدهم في استخلاص نتائج يمكن تعميمها على مجتمع الدراسة.

لتحديد حجم العينة الضروري لدراستك، عليك مراعاة العوامل التالية:

1. مستوى الثقة:

هو مستوى التأكد من أن النتائج التي ستحصل عليها من العينة تعكس بدقة نتائج المجتمع الأصلي.

مستوى الثقة 95%: هو المستوى الأكثر شيوعًا ويستخدم في معظم الدراسات. في الدراسات التي تتطلب دقة عالية نستعمل مستوى الثقة 99%.

2. هامش الخطأ = (précision) = e :

هو أقصى فرق مقبول بين نتائج العينة ونتائج المجتمع الأصلي. إذا كنا نبحث عن نسبة نستعمل احيانا هامش الخطأ 5% و احيانا 3%. واحيانا نستعمل ما يتطلبه البحث. وإذا كنا نبحث عن متوسط نحدد أقصى فرق مقبول بين متوسط المجتمع ومتوسط العينة حسب طبيعة البحث وطبيعة المعطيات وكذا حسب ما يطلبه صاحب البحث.
نسمي هذا المقدار بالدقة .

3. النسبة المتوقعة (في الدراسات التي نبحث فيها عن النسبة):

إذا كانت نسبة الذين يملكون الخاصية التي نهتم بدراستها غير معروفة: نفترض أنها 50%، وهي النسبة التي تعطي أكبر حجم للعينة.
إذا كانت النسبة معروفة (من دراسات سابقة أو استطلاع تجريبي): يمكن استخدامها في حساب حجم العينة بدقة أكبر.

4. التباين (في الدراسات التي نبحث فيها عن المتوسط):

هو مقياس لمدى انتشار البيانات حول المتوسط.

إذا كان التباين غير معروفا: يمكن تقديره من دراسات سابقة أو باستخدام عينات تجريبية.
وإذا كان معروفا: يمكن استخدامه في حساب حجم العينة بدقة أكبر.

7 – 1 – حجم العينة الضروري لتقدير متوسط المجتمع:

على الباحث أن تكون لديه فكرة عن الانحراف المعياري للمجتمع و عن الدقة المرغوبة

الدقة = µ – X = (معامل الثقة)(الخطأ المعياري)

$$\Pr\acute{e}cision = \mu - \overline{X} = z \times \frac{\sigma}{\sqrt{n}}$$

$$\Rightarrow n = \left(\frac{z \times \sigma}{\Pr ecision} \right)^2$$

مثال: يرغب مدير في تقدير عدد الدقائق التي يأخذها العمال لإنجاز عملية معينة في حدود ±3 دقائق و بدرجة ثقة 90% ويعلم هذا المدير من تجربته السابقة أن الانحراف المعياري 15 دق

ما هو حجم العينة الضروري؟

الحل: الحد الأدنى لحجم العينة المطلوب هو:

الدقة = ± 3 = ± (معامل الثقة)(الخطأ المعياري)

$$\Rightarrow n = \left(\frac{z \times \sigma}{Precision} \right)^2 = \left(\frac{1.64 \times 15}{3} \right)^2 \approx 68$$

الدقة = Précision هي أقصى فرق نسمح به بين متوسط العينة و متوسط المجتمع

<u>إذا كان حجم المجتمع معلوما:</u>

إذا كان حجم المجتمع N معلوما، نقوم بتصحيح حجم العينة المحصل عليه كالآتي:

$$n_{\text{مصححة}} = \frac{n}{1 + \frac{n}{N}}$$

مثال: نريد حساب حجم العينة الضروري من أجل تقدير متوسط أوزان 900 طالب:

لحساب حجم العينة الضروري لتقدير متوسط أوزان 900 طالب، نحتاج إلى معرفة بعض المعلومات الإضافية:

- مستوى الثقة: يُشير مستوى الثقة إلى احتمال صحة تقديرنا للمتوسط الحقيقي. فمثلاً، مستوى ثقة 95% يعني أنّنا واثقون بنسبة 95% من أنّ متوسطنا المُقدّر يقع ضمن نطاق معين حول المتوسط الحقيقي.

- هامش الخطأ: يُشير هامش الخطأ إلى أقصى فرق يمكن أن يوجد بين متوسطنا المُقدّر والمتوسط الحقيقي. فمثلاً، هامش خطأ 2 كيلوغرام يعني أنّ متوسطنا المُقدّر قد يختلف عن المتوسط الحقيقي بحد أقصى 2 كيلوغرام في أي اتجاه.

- الانحراف المعياري: يُشير الانحراف المعياري إلى مدى تباعد أوزان الطلاب عن المتوسط.

خطوات الحل:

تحديد مستوى الثقة المطلوب: 95% هو مستوى الثقة الأكثر شيوعًا.
قد نحتاج إلى مستوى ثقة أعلى (مثل 99%) إذا كانت الدراسة حساسة.

تحديد هامش الخطأ المقبول: يعتمد هامش الخطأ على دقة التقدير المطلوبة.
هامش خطأ 2 كيلوغرام قد يكون مقبولاً لدراسة تقريبية، بينما قد نحتاج إلى هامش خطأ أصغر (مثل 1 كيلوغرام) لدراسة دقيقة.

تقدير الانحراف المعياري: إذا لم يكن لدينا معلومات سابقة عن الانحراف المعياري، يمكن استخدام قيمة افتراضية 10 كيلوغرام. يمكن الحصول على قيمة أكثر دقة من خلال دراسة سابقة أو من خلال تحليل عينة صغيرة من الطلاب.

حساب حجم العينة:

يمكن استخدام الصيغة التالية لحساب حجم العينة: $n = \frac{z^2 \times S^2}{e^2}$

حيث:

n: حجم العينة

Z: قيمة Z-score لمستوى الثقة المطلوب (1.96 لمستوى ثقة 95%)

s: الانحراف المعياري (10 كيلوغرام افتراضياً)

e: هامش الخطأ (2 كلغ)

$$n = \frac{z^2 \times S^2}{e^2} = \frac{1.96^2 \times 10^2}{2^2} = 97\text{طالب}$$

تصحيح حجم العينة :

بما أن حجم المجتمع معلوم (N = 900 طالب)

فلا بد من تصحيح حجم العينة الضروري كالآتي:

$$n_{\text{مصححة}} = \frac{n}{1 + \frac{n}{N}} = \frac{97}{1 + \frac{97}{900}} = 88$$

حجم العينة المصحح هو 88 طالبًا.

7 – 2 - حجم العينة الضروري لتقدير نسبة في المجتمع:

نستعمل كل الخطوات السابقة، الخطأ المعياري الذي هنا يساوي $\sqrt{\frac{P(1-P)}{n}}$ ونحن طبعا نجهل قيمة p لذا نحسب قيمة هذا الخطأ المعياري بوضع p=0.5. (كما قلنا سابقا أحيانا نستعمل الحرف f للتعبير عن النسبة في العينة)

$$Précision = \mathrm{e} = \mathrm{P} - \mathrm{p} = \mathrm{z} \times \sqrt{\frac{p(1-p)}{n}}$$

$$e^2 = \mathrm{z}^2 \times \frac{p(1-p)}{n}$$

$$n = \mathrm{z}^2 \times \frac{p(1-p)}{e^2}$$

الدقة = e = Précision : هي أقصى فرق نسمح به بين النسبة الموجودة في العينة و النسبة الحقيقية الموجودة في المجتمع (إذا لم تكن لدينا فكرة نختار e=0.05)

إذا كان حجم المجتمع معلوما:

عند حساب حجم العينة من مجتمع إحصائي معلوم (نعرف عدد أفراده N)، نحسب حجم العينة على أساس أن حجم المجتمع غير معلوم بالعملية الحسابية السابقة ثم نقوم بعد ذلك بتصحيح حجم العينة باستخدام معادلة تصحيح العينة التالية:

$$n_{\text{مصححة}} = \frac{n}{1+\frac{n}{N}}$$

8 – اختبار الفروض الإحصائية:

8 – 1 - مقدمة:

نستعمل اختبارات الفروض الإحصائية من أجل تقدير معالم المجتمع وكذا من أجل وضع قواعد تمكن من التوصل إلي قرار بقبول أو رفض خاصية أو بالمعني الإحصائي فرض عن معالم مجتمع واحد أو أكثر.

الاختبارات الإحصائية قد تدور حول معالم المجتمع المجهولة مثل الفروض المتعلقة بالوسط الحسابي، النسبة، التباين، معامل الارتباط،... وفى هذه الحالة يطلق على هذه الاختبارات اسم الاختبارات المعلمية Parametric Tests . وقد تكون فروضا لا تتعلق بمعالم المجتمع ولكن تتعلق بأشياء اخرى قد تكون وصفية مثل العلاقة بين التعليم والتدخين، خضوع نتائج معينه لنظريه معينه، العلاقة بين لون العينين ولون الشعر،.... وفى هذه الحالة يسمى الاختبار باسم الاختبار اللامعلمي Non Parametric Tests.

تعتبر الفروض الإحصائية بمثابة اقتراح عن معالم المجتمع موضوع الدراسة، والتي ما زالت غير معلومة للباحث، فهي حلول ممكنة لمشكلة البحث.

سنركز اهتمامنا في البداية على الاختبارات المعلمية. وخلاصة عملية اختبار الفروض هي كالآتي:

بعد أن نسحب عينة عشوائية من المجتمع محل الدراسة، بحيث تكون هذه العينة ممثلة تمثيلا جيدا لهذا المجتمع، نحاول اتخاذ قرار ما بشأن خواص توزيع المجتمع (المتوسط – التباين – النسبة). ولكي نصل إلى قرار إحصائي لابد من وضع فروض عن هذه الخواص ثم نختبر مدى صحة الفرض من عدمه عن طريق العينة.

8 – 2 – نظرية اختبار الفروض:

وتنقسم اختبارات الفروض الإحصائية إلى قسمين:

أولا: اختبارات الفروض الإحصائية المعلمية : التوزيع الذي تتبعه البيانات يكون معلوما، ويكون المطلوب هو اختبار فروض حول معالم المجتمع. (تكون البيانات معلمية إذا كانت مقاييسها كمية وتتبع التوزيع الطبيعي).

ثانيا: اختبارات الفروض الإحصائية اللامعلمية : بيانات واقعية يصعب من خلالها التعرف على التوزيع الذي تتبعه، في هذه الحالات لا نحتاج إلى معرفة شكل التوزيع الذي تتبعه البيانات محل الدراسة، كما يفضل استخدامها عندما يكون حجم العينة المسحوبة من المجتمع صغيرًا نسبياً .

8 – 3 – الفرضية الصفرية والفرضية البديلة:

الفرضية الإحصائية تتعلق بعبارة عن إحدى معالم المجتمع مثل الوسط أو النسبة أو التباين أو غيرها. أو عدة معالم مثل المقارنة بين معلمتين أو أكثر.

هناك نوعان من الفرضيات، الفرضية الصفرية أو الابتدائية وهي الفرضية التي تبنى على أمل أن يتخذ قرار بعدم صحتها، نرمز لها بـ H_0. والبديلة H_1 .

مثال: يدعي احد المصانع أن معدل عمر المصباح 1000 ساعة. أردت اختبار هذا الادعاء, فكيف تكتب الفرضية الصفرية والفرضية البديلة.

الحل: إذا كان متوسط عمر المصابيح هو μ, فإن الفرضية الصفرية هي $H_0: \mu =$ 1000. أما الفرضية البديلة فقد تكون: $H_0: \mu \neq 1000$ أو $H_0: \mu < 1000$ أو $H_0: \mu > 1000$.

8 – 4 –أنــواع الأخطاء : إن أي قرار إحصائي يمكن أن ينتج عنه نوعان من الخطأ:

الخطأ من نوع الأول α والخطأ من النوع الثاني β:

خطأ من النوع الأول : يحدث هذا النوع من الأخطاء عندما نقوم برفض H_0 في حين أنها صحيحة، وذلك باحتمال مقدار α (وتسمى α بمستوى المعنوية وهي تأخذ قيما صغيرة مثلا ((0.05),(0,01)) .

خطأ من النوع الثاني : يقع مثل هذا الخطأ عندما نقبل H_0 في حين أنها خطأ وذلك باحتمال مقداره β.

ويمكن تلخيص القرارات الإحصائية بالجدول التالي :

الفرض / القرار	قبول H_0	رفض H_0
H_0 صحيحة	قرار صحيح	خطأ من النوع الأول α
H_0 خاطئة	خطأ من النوع الثاني β	قرار صحيح

فإذا كانت $\alpha = 0.05$ نقول أن مستوى الدلالة للاختبار هو 5 %. وعادة ما يكون مستوى الدلالة 1% أو 5% وهما الأكثر استعمالا في اختبار الفرضيات.

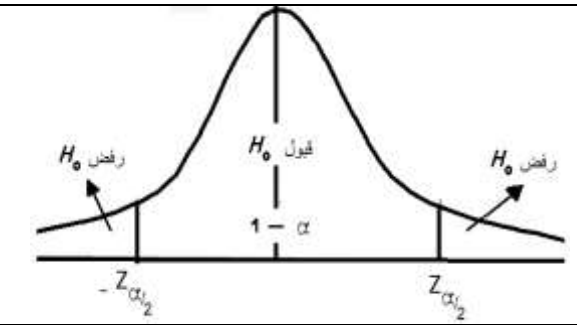
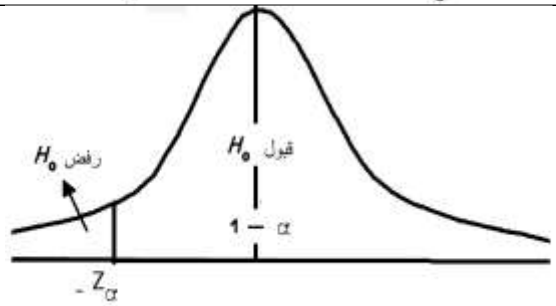
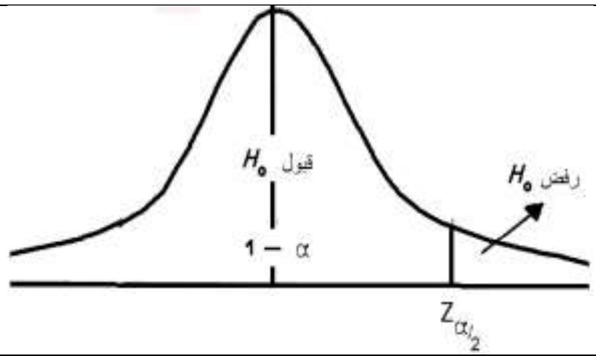
لاختبار صحة H_0 يجب أن يكون توزيع الإحصاءة معروفا ويتم تقسيم المجال المقابل لهذه الدالة إلى قسمين (منطقتين):

منطقة القبول: حيث يتم قبول H_0 ويكون احتمال حدوث قيم الإحصاءة $(1-\alpha)$ كبيرا.

المنطقة الرفض: حيث يتم رفض H_0 ويقبل الفرض البديل ويكون احتمال حدوث قيم الإحصاءة α صغيرا.

8 – 5 – أمثلة توضيحية لمناطق القبول والرفض:

الأشكال التالية يمكن من خلالها توضيح مناطق الرفض والقبول وذلك حسب نوع الفرض البديل, وسوف نوضح ذلك باستخدام المتوسط μ (متوسط المجتمع):

الاختبار من طرفين	$H_0: \mu = \mu_0$ $H_0: \mu \neq \mu_0$	
الاختبار من طرف واحد أدنى	$H_0: \mu = \mu_0$ $H_0: \mu < \mu_0$	
الاختبار من طرف واحد أعلى	$H_0: \mu = \mu_0$ $H_0: \mu > \mu_0$	

8 – 6 – درجات الحرية :

إذا كان لدينا مجتمع ما ونريد تقدير عدد من معالم هذا المجتمع كالمتوسط والانحراف المعياري إلى آخره وتم سحب عينة من البيانات المستقلة التي تمثل ذلك المجتمع حجمها n فإن درجات الحرية التي يرمز لها بالرمز v تساوي حجم العينة مطروحا منه عدد المعالم المراد تقديرها ويمكن التعبير عن ذلك من خلال العلاقة التالية : $v = n - k$

حيث k هي عدد المعالم المقدرة .

8 – 7 – أهم الفروض الإحصائية: في ما يلي أهم الإحصاءات المستخدمة في الاختبارات:

الاختبار	العينات	إحصاء الاختبار	توزيع الإحصاء
متوسط مجتمع غير محدود و تباينه معلوم	كبيره $(n \geq 30)$	$z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$	$N(0,1)$
متوسط مجتمع محدود حجمه N و تباينه معلوم	كبيره	$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}}$	$N(0,1)$
متوسط مجتمع طبيعي و تباينه مجهول	صغيره $n < 30$	$t = \frac{\bar{x}-\mu}{s/\sqrt{n}}$	$t_{(n-1)}$
اختبار النسبة في مجتمع	كبيره	$Z = \frac{\hat{p}-p}{\sqrt{p(1-p)/n}}$	$N(0,1)$
تباين مجتمع طبيعي و تباينه معلوم	صغيره	$X^2 = \frac{(n-1)S^2}{\sigma^2}$	$X^2_{(n-1)}$
النسبة بين تباين مجتمعين طبيعيين	صغيره	$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$	$F_{(n_1-1,}$
الفرق بين متوسطي مجتمعين طبيعيين معلومي التباين	صغيره	$Z = \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$	$N(0,1)$

الفرق بين متوسطي مجتمعين مجهولين التباين	كبيره	$Z = \frac{(\overline{x_1} - \overline{x_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$	N(0,1)
الفرق بين متوسطي مجتمعين طبيعيين مجهولين التباين وبفرض تساوي التباين $\sigma_1^2 = \sigma_2^2 = \sigma_p^2$	صغيره	$t = \frac{(\overline{x_1} - \overline{x_2}) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ $S_p^2 = \frac{(n_1-1)\,S_1^2 + (n_1-1)S_2^2}{n_1 + n_2 - 2}$	$t_{(n_1+n_2-2}$
الفرق بين متوسطين مجتمعين غير مستقلين	مزدوجة وكبيره	$Z = \frac{\bar{D} - D}{S_{\bar{p}}/\sqrt{n}}$	N(0,1)
الفرق بين متوسطي مجتمعين طبيعي غير مستقلين	مزدوجة وصغيره	$t = \frac{\bar{D} - D}{S_{\bar{p}}/\sqrt{n}}$	$t_{(n-1)}$
الفرق بين نسبتين في مجتمعين مختلفين	كبيره	$Z = \frac{(\widehat{p_1} - \widehat{p_2}) - (p_1 - p_2)}{\sqrt{p(1-p)\,(1/n_1 + 1/n_2)}}$ $P = \frac{n_1\,p_1 + n_2\,p_2}{n_1 + n_2}$	N(0,1)

8 – 8 - الرموز المستعملة في اختبارات الفروض:

$\bar{x}$: ... متوسط البيانات المشاهدة من العينة،

μ : ... متوسط البيانات في المجتمع،

μ_0 :... قيمة متوسط المجتمع التي نقوم باختبارها،

s : .... الانحراف المعياري للعينة،

σ :.... الانحراف المعياري للمجتمع،

n : .... حجم العينة المسحوبة من المجتمع،

N :.... حجم المجتمع،

$\hat{P}$:... نسبة الظاهرة المدروسة في العينة (مقدر بنقطة للنسبة في المجتمع)،

P :... نسبة الظاهرة المدروسة في المجتمع،

D :... الفرق بين مشاهدتين مزدوجتين $d_1 = x_1 - y_1$ ،

$\bar{D}$: .. متوسط الفرق بين مشاهدتين مزدوجتين،

α :.... (مستوى المعنوية)حيث يتم على أساسه تحديد القيم الحرجة $\alpha/_2$ و $-Z\,\alpha/_2$

أو z_α

8 – 9 – خطوات اختبار الفروض:

الخطوة الأولى: تحديد الفرضية الصفرية (فرضية العدم) H_0 والفرضية البديلة H_1

الخطوة الثانية: اختيار المستوى المعنوي المناسب للاختبار وهو احتمال رفض H_0 عندما يكون صحيحا. α (تحديد نسبة الثقة)

الخطوة الثالثة: اختيار مقياس إحصائي مناسب للاختبار سواء كان اختبار t أو التوزيع الطبيعي أو أي اختبار آخر.

الخطوة الرابعة: حساب احصاءة الاختبار من العينة. القيمة المحسوبة

الخطوة الخامسة: تحديد المنطقة الحرجة , من التوزيع الطبيعي او التوزيع t حسب مستوى المعنوية ودرجات الحرية.

الخطوة السادسة: تقدير رفض أو قبول H_0, فإذا كان المقياس الإحصائي المحسوب (القيمة المحسوبة) يقع في المنطقة الحرجة ترفض H_0, أما إذا وقع خارج تلك المنطقة فانه لا يمكن رفض الفرض H_0 بناء على بيانات العينة المدروسة.

8 – 10 –نظرية توزيع المعاينة للمتوسط:

إذا أخذت عينة عشوائية حجمها n من مجتمع طبيعي $N(\sigma, \mu^2)$ وكانت σ^2 معلومة فإن إحصاء الاختبار للفرضية $H_0: \mu = \mu_0$ هو $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$ وهو يخضع للتوزيع الطبيعي المعياري حيث $\bar{x}$ هو الوسط الحسابي للعينة .

9 – اختبار الفرضيات المتعلقة بالوسط:

9-1- اختبار الفرضيات المتعلقة بالوسط لمجتمع طبيعي تباينه معلوم:

سوف نستعمل التوزيع الطبيعي كتقريب لتوزيع وسط العينة $\bar{X}$.

إحصائية الاختبار $\boldsymbol{Z} = \frac{\bar{\boldsymbol{X}}-\boldsymbol{\mu_0}}{\boldsymbol{\sigma}/\sqrt{\boldsymbol{n}}}$ تتبع التوزيع الطبيعي المعياري *N(0,1)*

نقوم بالاختبار حسب نوع الاختبار (اختبار من طرف واحد أو من طرفين).

خطوات الاختبار: اختبر عند مستوى الدلالة α الفرضية الصفرية $H_0: \mu = \mu_0$ مقابل الفرضية البديلة:

$(i)H_1: \mu > \mu_0\ ,\quad (ii)H_1: \mu < \mu_0\ ,\quad (iii)H_1: \mu \neq \mu_0$

إحصائية الاختبار: بفرض أن H_0 صحيحة فإن $z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ يخضع لتوزيع طبيعي معياري

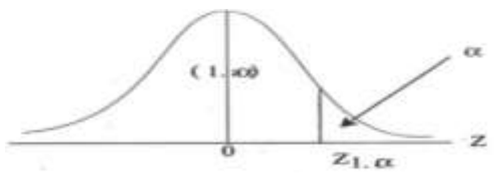
القيم الحرجة ومنطقة الرفض:

(*i*) نرفض H_0 إذا كان $Z > Z_{1-\alpha}$ بالنسبة للفرضية البديلة $H_1: \mu > \mu_0$

القيمة الحرجة $Z_{1-\alpha}$ للاختبار ذي الطرف الواحد

للفرضية البديلة $\mu > \mu_0$ بمستوى الدلالة α

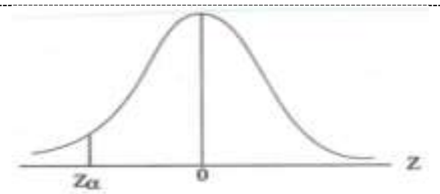
.



(*ii*) نرفض الفرضية H_0 إذا كان $Z > Z_\alpha$ بالنسبة للفرضية البديلة $H_1: \mu < \mu_0$

القيمة الحرجة z_α بالنسبة

للفرضية البديلة $H_1: \mu < \mu_0$

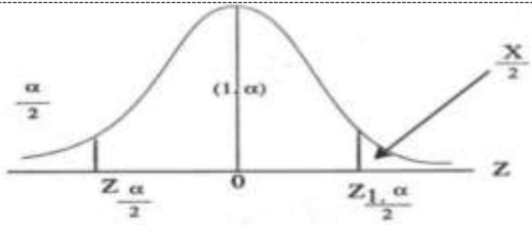


لاحظ أن القيمة الحرجة Z_α هي نفسها $-Z_{1-\alpha}$ بسبب التماثل .

(iii) نرفض الفرضية H_0 إذا كان $Z > Z_{1-\frac{\alpha}{2}}$ أو $Z < Z_{\frac{\alpha}{2}}$ للاختبار ذي الطرفين للفرضية البديلة $H_1: \mu \neq \mu_0$ بمستوى دلالة α.

القيم الحرجة , $Z_{\frac{\alpha}{2}} = -Z_{1-\frac{\alpha}{2}}$

$Z_{1-\frac{\alpha}{2}}$



مثال1:

تخضع أوزان عبوات احد مساحيق الغسيل لتوزيع طبيعي انحرافه المعياري7 غم ومعدله μ غم. عند مستوى دلالة 0.05 اختبر الفرضية : $H_0:\mu = 50$ ، $H_1:\mu \neq 50$ ، $\alpha=0.05$

إحصاء الاختبار هو: تحت فرض أن H_0 صحيحة فإن : $z=\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ يخضع لتوزيع طبيعي معياري.

المنطقة الحرجة: بما أن الفرضية البديلة لا تحدد اتجاها واحدا, فإن الاختبار المناسب ذو طرفين،

والقيم الحرجة تكون:

$Z_{0.975} = 1.96$ ، $Z_{0.025}=-1.96$ أي نرفض H_0 إذا كان $|z| > 1.96$

Z المحسوبة: $z=\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ ، $Z=\frac{56-50}{7/\sqrt{12}}=2.97$

المقارنة: نقارن قيمة Z بالقيمة الحرجة: من الواضح أن 2.97>1,96 أي أن قيمة Z تقع في المنطقة الحرجة , لذلك ارفض H_0 لصالح $\mu > 50$ ولاحظ أن النتيجة كانت لصالح $\mu > 50$ لأن Z وقعت في منطقة الرفض اليمنى , أي على يمين $Z_{1-\alpha/2}$.

مثال2:

أخذت عينة من *64* طالب من مدارسة فوجدنا متوسط الطول *155 cm*. فإذا كان الانحراف المعياري للمجتمع يساوي *5cm*. اختبري الفرض القائل: $H_1: \mu \neq 160$، $H_0: \mu = 160$ وذلك عند مستوى معنوية : *0.05*

الحل: $Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ حيث: $n = 64$ ، $\sigma = 5$, $\mu_0 = 160$, $\bar{x} = 155$ ، $\alpha = 0.05$

وهذه تمثل قيمة Z . المحسوبة $\rightarrow Z = \frac{155-160}{5/\sqrt{64}} = -8$

عند مستوى معنوية $(\alpha = 0.05)$ وبما أن الاختبار من طرفين إذا قيم Z الجدولية سوف تكون كالتالي :

$Z_{\alpha/2} = Z_{0.05/2} = Z_{0.025} = 1.96$ وعليه فإن قيمة $-Z_{\alpha/2}$ ستكون :

$-Z_{\alpha/2} = -1.96$

ومن هنا نجد أن قيمة Z المحسوبة أصغر من قيمة Z الجدولية , أي أن قيمة تقع في منطقة الرفض, ومن هنا يتم رفض الفرض العدمي بأن $H_0: \mu = 160$.

9-2- اختبار الفرضيات المتعلقة بالوسط لمجتمع تباينه مجهول وحجم العينة كبير

أن المجتمع في هذه الحالة لم يفترض أنة طبيعي، ولكن حجم العينة كبير $n \geq 30$.

مثال:

على مستوى 5% اختبر الفرضية $H_0: \mu = 10$ مقابل الفرضية البديلة $H_1: \mu \neq 10$

إذا أعطت عينة حجمها *n=42* وسطاً حسابيا $\bar{x}=11.5$ وانحرافاً معيارياً *S=3.3*

الحل: H0:μ=10 ، μH1: ≠ 10 ، α=0.05 مستوى الدلالة .

إحصاء الاختبار :بافتراض أن H_0صحيحة فإن $Z=\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ يخضع لتوزيع طبيعي معياري تقريباً

المنطقة الحرجة : الفرضية البديلة ذات طرفين ،

إذن ، فالقيم الحرجة هي: $Z_{0.975}=1.96$, $Z_{0.025}=-1.96$

Z المحسوبة : $Z=\frac{X-\mu_0}{\frac{S}{\sqrt{n}}}$ ، $Z=\frac{11.5-10}{3.3/\sqrt{4.2}}=2.94$ ،

وذلك باستعمال S بدلاً من σ لأن حجم العينة أكبر من 30

المقارنة : من الواضح أن 1.96< 2.94 إذن ارفض H_0 لصالح μ>10 لأن Z وقعت في المنطقة الحرجة إلى يمين القيمة الحرجة اليمنى .

9 – 3 – اختبار الفرضيات المتعلقة بالوسط لمجتمع طبيعي تباينه مجهول العينة صغيرة

تباين المجتمع σ^2 مجهول وحجم العينة صغير : $(n<30)$

هذه الحالة تختلف عن الحالة الأولى من حيث التباين σ^2 غير معلومة وأنها تختلف عن الحالة الثانية من حيث حجم العينة صغير . ففي هذه الحالة نستعمل الإحصائية t التي يكون لها الشكل التالي:

$\boldsymbol{t}=\frac{\bar{x}-\mu_0}{s/\sqrt{n}}$ ، والإحصائية t تتبع توزيع t وذلك بدرجات حرية $(v=n-1)$ ومستوى معنوية (مستوى الدلالة) α أو $\alpha/2$ حسب نوع الاختبار.

وخطوات الاختبار حول μ هي : $H_0: \mu=\mu_0$ مقابل الفرضية البديلة :

$(i) H_1: \mu>\mu_0$, $(ii) H_1: \mu<\mu_0$, $(iii) H_1: \mu \neq \mu_0$

إحصائية الاختبار: بافتراض أن H_0 صحيحه ، أي $\mu = \mu_0$

فإن $T=\frac{\bar{x}-\mu_0}{S/\sqrt{n}}$ تخضع لتوزيع t ذي درجات الحرية $(n-1)$

تتحدد النقاط الحرجة حسب الحالات الثلاث في الفرضية البديلة فتكون في الحالة :

الحالة (i) النقطة الحرجة $t_\alpha = t_{(1-\alpha\ ,\ n-1)}$ و منطقة الرفض: ارفض H_0 إذا كان $t > t_\alpha$.

الحالة (ii) النقطة الحرجة $t_\alpha = -t_{(1-\alpha\ ,\ n-1)}$ ومنطقة الرفض: ارفض H_0 إذا كان $t > -t_\alpha$

الحالة (iii) النقطتان الحرجتان $t_{\frac{\alpha}{2}}, -t_{\frac{\alpha}{2}}$ حيث $t_{\frac{\alpha}{2}} = t_{(1-\frac{\alpha}{2},\ n-1)}$

ومنطقة الرفض هي نرفض H_0 إذا كان $t<-t$ أو $t > t_{\frac{\alpha}{2}}$ بعبارة أخرى إذا كان $|t| > t_{\frac{\alpha}{2}}$

مثال1:

أظهرت سجلات إحدى المدارس أن معدل تحصيل الطلبة في امتحان اللغة 410 . بدأت المدرسة بإعطاء دورات تقوية للطلبة . اختبر (عند مستوى دلالة 1%) فرضية أن هذا المعدل قد تحسن إذا أعطت نتائج 14طالباً وسطاً حسابياً $\bar{x}$=418 بانحراف معياري S=21 اعتبر أن نتائج طلبة المدارس تخضع لتوزيع طبيعي.

الحل: $H0:\mu=410$ ، $H1 : \mu>410$ ، مستوى الدلالة $\alpha=0.01$

نستعمل توزيع t لأن المجتمع طبيعي، تباينه غير معلوم، وحجم العينة 14 أي أقل من 30 .

إحصاء الاختبار هو $T=\frac{\bar{X}-410}{S/\sqrt{n}}$ يخضع لتوزيع t ذي درجات حرية $(n-1)$

بما أن الفرضية البديلة لها اتجاه واحد فإن القيمة الحرجة $t_{0.01}$ تحت درجات حرية 13.

$t_{0.01} = 2.65$ من جداول توزيع t تحت درجة 13.

t المحسوبة : $t=\frac{\bar{x}-\mu}{S/\sqrt{n}}$ ، $t = \frac{418-410}{21/\sqrt{14}} = \frac{8x3.742}{21} = 1.42$

نقارن بين قيمة t والقيمة الحرجة $t_{0.01}$ من الواضح أن $2.65 > 1.42$ أي أن t لا تقع في المنطقة الحرجة ، وبالتالي لا ترفض H_0 أي لا تستطيع استنتاج أن معدل تحصيل الطلبة قد تحسن .

مثال2:

إذا كان متوسط ربح سهم إحدى الشركات هو *15* د في العام الماضي. تم أخذ عينة من *7* مساهمين فوجدنا متوسط ربح السهم العام الحالي *17* د بانحراف معياري *2*. هل توافق المساهمين في تقديرهم لارتفاع متوسط ربح السهم هذا العام ؟ وذلك بمستوى معنوية *5*%.

<u>الحل</u>: فروض الاختبار وهي كالتالي: $H_0 : \mu = 15$ ، $H_1 : \mu > 15$

t المحسوبة :

$$t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}} = \frac{17-15}{2/\sqrt{7}} = 2.65$$

t الجدولية : عند مستوى معنوية $(\alpha = 0.05)$ وبما أن الاختبار من طرف واحد إذا قيمة t الجدولية سوف تكون كالتالي: $t_{(n-1,\alpha)} = t_{(7-1\,,0.05)} = t_{(6\,,0.05)} = 1.943$

قيمة t المحسوبة أكبر من قيمة t الجدولية,أي أن قيمة t تقع في منطقة الرفض , ومن هنا يمكن القول بأننا نوافق المساهمين على توقعاتهم بارتفاع متوسط الربح للسهم هذا العام.

10- اختبار الفرضيات حول تباين المجتمع:

عند الرغبة في اختبار الفرض أن التباين لمجتمع طبيعي يساوي قيمة معينة σ_0^2 ضد الفرض البديل ذي جانبين أن التباين لا يساوي σ_0^2 .. أي أننا نختبر الفرض:

$$H_0: \sigma^2 = \sigma_0^2 \quad \text{ضد} \quad H_1: \sigma^2 \neq \sigma_0^2$$

نختار عينة عشوائية من الحجم n من المجتمع موضع الدراسة ونحسب تباين العينة S^2

وعلى ذلك $X^2 = \frac{(n-1)S^2}{\sigma_0^2}$ قيمة للمتغير X^2 ..

المقارنة بين تبايني متجمعين:

إن مقارنة تبايني مجتمعين ضرورية من أجل مقارنة مجتمعين وخاصة لما نريد مقارنة متوسطي المجتمعين. ولإجراء المقارنة بين تبايني مجتمعين نختبر النسبة بين التباينين:

نأخذ عينة عشوائية حجمها n_1 من مجتمع طبيعي $N(\mu_1\ ,\ \sigma_1^2)$ ونخذ عينة مستقلة عن الأولى حجمها n_2 من مجتمع طبيعي $N(\mu_2\ ,\ \sigma_2^2)$ وكان s_1^2 تباين العينة الأولى s_2^2 تباين العينة الثانية:

الفرضية الصفرية $H_0 = \sigma_1^2 = \sigma_2^2$ مقابل الفرضية البديلة

$$(i) H_1: \sigma_1^2 > \sigma_2^2\ , \quad (ii) H_1: \sigma_1^2 < \sigma_2^2\ , \quad (iii) H_1: \sigma_1^2 \neq \sigma_2^2$$

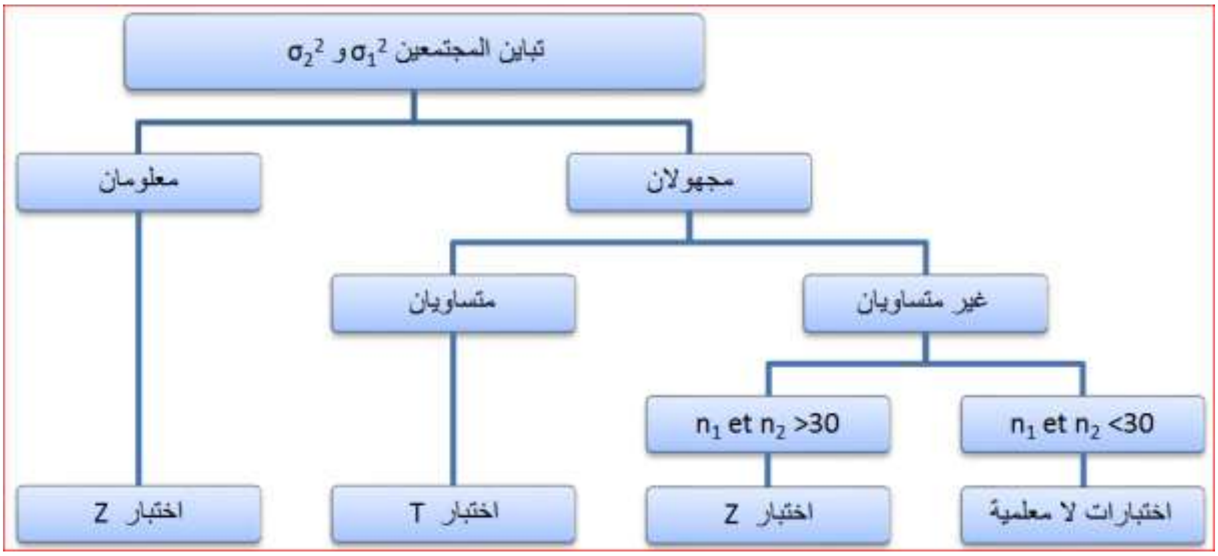
إحصائية الاختبار: $\frac{\left(\frac{n_1}{n_1-1}\right)s_1^2}{\left(\frac{n_2}{n_2-1}\right)s_2^2}$ هذه الإحصائية تخضع تحت فرض H_0 صحيحة لتوزيع F ذي درجات الحرية (n_2-1) في البسط (n_1-1) في المقام. لا بد أن تكون S_1 أكبر من S_2 لأن F>1 دائما.

F الجدولية: نستخرج F الجدولية بدرجة حرية n_1-1 و n_2-1 و مستوى الدلالة α.

القرار: إذا كانت F المحسوبة أكبر من F الجدولية نرفض H_0.

11- اختبار الفرضيات المتعلقة بالفرق بين وسطين (مقارنة متوسطي مجتمعين):

سنبحث هذا النوع من الاختبارات حسب كون المجتمعات تتبع توزيع طبيعياً أم لا آخذين بعين الاعتبار معلومية التباينات أم عدمها وكبر حجوم العينات أو صغرها (الشكل الموالي يلخص هذه الحالات):



11 – 1- مجتمعان طبيعيان تبايناهما معلومان: ليكن $\bar{X}$ و$\bar{Y}$ متوسط العينتين،

نعلم من توزيع المعاينة للوسط الحسابي المدروس سابقا أن:

$\bar{X}$ يتبع التوزيع الطبيعي $N(\mu_1, {}^{\sigma_1}/_{\sqrt{n_1}})$ و $\bar{Y}$ يتبع التوزيع الطبيعي $N(\mu_2, {}^{\sigma_2}/_{\sqrt{n_2}})$

ومنه الفرق $\bar{X} - \bar{Y}$ يتبع التوزيع الطبيعي $N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$

يمكن الآن حساب القيمة المعيارية Z كالآتي:

$$z = \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \quad ---\rightarrow N(0,1)$$

نظرية 1:

إذا أخذت عينة عشوائية حجمها n من مجتمع $N(\mu_1 , \sigma_1)$ وأخذت عينة عشوائية ثانية ومستقلة عن الأولى حجمها n_2 من مجتمع $N(\mu_2 , \sigma_2)$ وكان σ_1^2 , σ_2^2 معلومتين فإن إحصاء الاختبار للفرضية $H_0:\mu_1 = \mu_2$ هو :

$$Z = \frac{(\bar{X}-\bar{Y})}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$$

الذي توزيعه $N(0,1)$ حيث $\bar{X}, \bar{Y}$ هما وسطا العينتين على التوالي .

من هذه النظرية نرى أن خطوات اختبار الفرضية $H_0: \mu_1 = \mu_2$ هي نفسها الخطوات السابقة مع الأخذ بعين الاعتبار إحصاء الاختبار الجديد ..

أما إذا كانت المجتمعات لا تخضع للتوزيع الطبيعي فإننا نطبق نظرية تقارب التوزيعات (النهاية المركزية) إذا كانت حجوم العينات كبيرة.

11 – 2- مجتمعان (لا ندري بأنهما طبيعيين) تبايناهما معلومان وحجم العينتين كبير:

نظرية 2:

إذا أخذت عينتان عشوائيتان مستقلتان من مجتمعين متوسط الأول μ_1 وتباينه σ_1^2 معلوم ومتوسط الثاني μ_2 وتباينه σ_2^2 معلوم وكان حجم الأولى n_1 كبيراً وحجم الثانية n_2 كبيراً فإن إحصاء الاختبار للفرضية $H_0: \mu_1 = \mu_2$ هو :

$$Z = \frac{(\bar{X}-\bar{Y})}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$$

الذي يخضع لتوزيع طبيعي معياري $N(0,1)$ تقريباً .

إن خطوات اختبار الفرضية $H_0: \mu_1 = \mu_2$ هي نفسها الخطوات السابقة مع الأخذ بعين الاعتبار إحصاء الاختبار الجديد.

ونلاحظ أيضا في هاتين النظريتين:

(1) إذا لم تكن التباينات معلومة فإننا نستطيع استعمال تباينات العينات S_1^2 ، S_2^2 ويبقى إحصاء الاختبار نفسه شريطة أن n_1, n_2 كبيرتان(كل منهما اكبر من 30)

(2) إذا كانت الفرضية الصفرية $\boldsymbol{H_0: \mu_1 = \mu_2 = d}$ حيث d عدد ثابت بدلاً من:
$H_0: \mu_1 = \mu_2$
أي $\mu_1 = \mu_2 = 0$) فإن إحصاء الاختبار في كلتي النظريتين يصبح :

$$Z = \frac{(\bar{X}-\bar{Y})-d}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$$

مثال (1):

أخذت عينة حجمها 22 من المجتمع $\mu(\mu_1 , 110)$ فأعطت $\bar{x}= 83$ وأخذت عينة ثانية مستقلة عن الأولى وحجمها 27 من المجتمع $N(\mu_2, 81)$ فأعطت الوسط الحسابي $\bar{Y}=69$

أ) اختبر $H_0: \mu_1=\mu$ مقابل $H_1: \mu_1>\mu_2$ على مستوى دلالة $\alpha=0.05$

ب) اختبر الفرضية $H_0: \mu_1-\mu_2=10$ مقابل $H_1: \mu_1-\mu_2>10$ على مستوى دلالة $\alpha = 0.05$

الحل:

(أ) الفرضية الصفرية $H_0: \mu_1 = \mu$ ضد الفرضية البديلة $H_1: \mu_1 = \mu$ ، $\alpha = 0.05$

إحصاء الاختبار تحت فرض أن H_0 صحيحة فإن الإحصائية : $Z=\frac{(X-Y)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$

تتبع $N(0,1)$

النقاط الحرجة ومنطقة الرفض: بما أن الاختبار ذو طرف واحد فالنقطة الحرجة :

$Z_{0.95}=1.645$

ومنطقة الرفض: نرفض H_0 إذا كان $Z>1.645$

Z المحسوبة : $Z=\frac{83-69}{\sqrt{\frac{110}{22}+\frac{81}{27}}}=\frac{14}{\sqrt{8}}=4$

المقارنة : بما أن $4.95>1.645$ تقع في منطقة الرفض فإننا نرفض H_0 لصالح H_1

ب) $H_0: \mu_1-\mu_2=10$ ، $H_1: \mu_1-\mu_2>10$ ، $\alpha=0.05$

إحصاء الاختبار تحت فرض H_0 صحيحة فإن الإحصائية: $Z=\frac{(X-Y)-10}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$

تتبع $N(0,1)$ (التوزيع الطبيعي المعياري)

منطقة الرفض تبقى نفسها , ارفض H_0 إذا كان : $Z>1.645$

Z المحسوبة : $Z=\frac{(83-69)-10}{\sqrt{\frac{110}{22}+\frac{81}{27}}}=\frac{4}{\sqrt{8}}=1.41$

المقارنة : من الواضح أن $1.41<1.645$ إذا لا ترفض H_0 . أي أنه لا يوجد دلالة بأن الفرق بين الوسطين يزيد عن 10..

11 – 3– مجتمعان تبايناهما مجهولان ومتساويان:

نتحقق من تساوي التباينين باستعمال اختبار تساوي التباينين المذكور سابقا (المبحث 4 من هذا الفصل).

نظرية (1) :

أخذت العينة العشوائية التي حجمها n_1 من مجتمع $N(\mu_1, \sigma_1^2)$ فأعطيت الوسط الحسابي $\bar{x}$ والتباين S_1^2 و أخذت العينة العشوائية التي حجمها n_2 من مجتمع طبيعي $N(\mu_2, \sigma_2^2)$ مستقل عن الاول فأعطت الوسط الحسابي $\bar{Y}$ والتباين S_2^2 فإذا كان $\sigma_1^2 = \sigma_2^2$ و كانت n_1 , n_2 صغيرتين يكون إحصاء اختبار الفرضية $H_0 : \mu_1 = \mu_2$ هو :

$$T = \frac{(\bar{X} - \bar{Y})}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

يخضع لتوزيع t ذي درجات الحرية $(n_1 + n_2 - 2)$ حيث :

التباين المشترك (التباينين متساويين ويأخذان هذه القيمة)

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

لاحظ انه بسبب صغر الحجمين n_1 , n_2 فلا نستطيع استعمال نظرية النهاية المركزية .

مثال (1) :

أخذت عينة عشوائية حجمها 9 من أوزان الأطفال الذكور حديثي الولادة في احد المستشفيات فأعط (بالكغ) $\bar{x} = 3.2$, $S_1 = 1.3$, واخذت عينة عشوائية حجمها 15 من أوزان الإناث حديثات الولادة في المستشفى فأعطت $\bar{Y} = 2.8$, $S_2 = 1.2$.

على فرض أن كلاً من أوزان الذكور وأوزان الإناث يخضع لتوزيع طبيعي ذي التباين نفسه ,

اختبر الفرضية:

أ) $H_0: \mu_1 = \mu_2$ مقابل $H_1: \mu_1 > \mu_2$ على مستوى $\alpha = 0.05$ حيث μ_1 معدل أوزان الذكور , μ_2 معدل أوزان الإناث .

الحـل : شروط النظرية متحققة . $H_0: \mu_1 = \mu_2$ مقابل الفرضية البديلة $H_1: \mu_1 > \mu_2$ ، $\alpha = 0.05$.

إحصاء الاختبار : $T = \frac{(\bar{X}-\bar{Y})}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ يخضع لتوزيع t ذي درجات الحرية $(n_1 + n_2 - 2)$

منطقة الرفض: بما أن الاختبار ذو طرف واحد فان النقطة الحرجة هي:

$$t_{0.05} = t\,(0.95\ ,22) = 1.717$$

ومنطقة الرفض: نرفض H_0 إذا كان $T > 1.717$

t المحسوبة:

$$S_p^2 = \frac{8\times(1.3)^2 + 14\times(1.2)^2}{9+15-2} = \frac{33.68}{22} = 1.53 \rightarrow S_p = \sqrt{1.53} = 1.24$$

$$T = \frac{3.2-2.8}{1.24\sqrt{\frac{1}{9}+\frac{1}{15}}} = \frac{0.4}{1.24\sqrt{0.1778}} = \frac{0.4}{0.523} = 0.745$$

المقارنة : بما أن $1.717 > 0.745$ اذا لا نرفض H_0 أي لا يوجد دلالة كافية أن

معدل أوزان الأطفال الذكور حديثي الولادة أكبر من معدل أوزان الإناث حديثي الولادة .

11 – 4 – مجتمعان تبايناهما مجهولان وغير متساويان:

إذا تبين لنا بعد التحقق أنت تباين المجتمعين هو غير متساو، نكون أمام حالتين حسب حجم العينتين.

11-4-1- n_1 et $n_2>30$: نستعمل نفس احصائية الاختبار لحالة التباينان معلومان:

$$Z = \frac{(\bar{X}-\bar{Y})}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \rightarrow N(0,1)$$

نستبدل قيم التباينين بتقديراتهما المحسوبة من العينتين: $\hat{\sigma}_1^2 = \frac{n_1}{n_1-1}S_1^2$ et $\hat{\sigma}_2^2 = \frac{n_2}{n_2-1}S_2^2$

تصبح إذن إحصائية الاختبار كالتالي: $$Z = \frac{(\bar{X}-\bar{Y})}{\sqrt{\frac{S_1^2}{n^1-1}+\frac{S_2^2}{n^2-1}}} \rightarrow N(0,1)$$

11-4-2- n_1 et $n_2<30$: نستعمل في هذه الحالة الاختبارات اللامعلمية. Testes non paramétriques

12- اختبار الفرضيات المتعلقة بالفرق بين نسبتين

(مقارنة نسبتي مجتمعين):

إن توزيع المعاينة للنسبة المحسوبة من العينة : f=k/n تتبع التوزيع الطبيعي $N\left(P, \sqrt{\frac{P(1-P)}{n}}\right)$ وبنفس الطريقة إذا كان لدينا مجتمعين فيهما النسب P_1 و P_2 وأخذنا منهما عينتين عشوائيتين حجمهما n_1 و n_2 والنسبة فيهما f_1 و f_2 . يكون الفرق $f_2 - f_1$ يتبع التوزيع الطبيعي: $N\left(P_1 - P_2, \sqrt{\frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}}\right)$

يمكن استنتاج القيمة المعيارية : $Z = \frac{(f_1-f_2)-(P_1-P_2)}{\sqrt{\frac{P_1(1-P_1)}{n_1}+\frac{P_2(1-P_2)}{n_2}}}$ وإذا كانت H_0 صحيحة تصبح قيمة Z كالآتي:

$$Z = \frac{|f_1-f_2|}{\sqrt{\hat{P}(1-\hat{P})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \dashrightarrow N(0,1)\ , \quad \text{avec} \quad \hat{P} = \frac{n_1f_1+n_2f_2}{n_1+n_2} = \frac{k_1+k_2}{n_1+n_2}$$

نستخرج قيمة Z الجدولية ونقارنها مع قيمة Z المحسوبة. إذا كانت المحسوبة أكبر من الجدولية نرفض H_0

13 – اختبار الاستقلالية باستعمال كاي تربيع khi2 χ^2

13 – 1 – مقدمة:

هناك العديد من الاستخدامات لتوزيع كاي تربيع، من بينها اختبار الفروض حول البيانات النوعية وهي بيانات لا تقبل العمليات الحسابية المعروفة فإذا كان لدينا عينة من الأشخاص يمكننا إيجاد عدد المتزوجين وكذا عدد الأشخاص الذين يفضلون مشاهدة القناة الجزائرية في نفس العينة وبإتباع بعض الخطوات يمكننا اختبار استقلال (أو عدم استقلال) الحالة الاجتماعية عن تفضيل القناة الجزائرية، حيث يمكن إيجاد التكرارات المتوقعة بافتراض استقلال المتغيرين ثم المقارنة بين التكرارات الفعلية وبين التكرارات المتوقعة في تلك العينة. فإذا كان الفرق كبيرا نقول أن هناك علاقة بين المتغيرين وإذا كان الفرق صغيرا نقبل افتراض أن المتغيرين مستقلين.

13 – 2 – شرح اختبار كاي تربيع:

نرجع لمثال علاقة الحالة الاجتماعية بتفضيل القناة الجزائرية:

نضع الحدث A : تفضيل القناة الجزائرية (نعم)

نضع الحدث B : الحالة الاجتماعية أعزب

$$P(A)=\frac{\text{المحققة}}{\text{الكلية}}=\frac{55}{100}\quad;\quad P(B)=\frac{\text{المحققة}}{\text{الكلية}}=\frac{35}{100}\quad;\quad P(A\cap B)=\frac{55\times 35}{100\times 100}$$

ولإيجاد التكرارات المتوقعة نضرب الاحتمال المتحصل عليه في المجموع الكلي:

$$e_{ij}=\sum o_{ij}\times P(A\cap B)=100\times\frac{55\times 35}{100\times 100}=\frac{55\times 35}{100}=\frac{\sum \text{Ligne}_i\times\sum \text{Colonne}_j}{\sum \text{Observations}}=19.25$$

القيم المشاهدة O_{ij}:

نقوم بنفس العملية على كل خلايا الجدول لإيجاد التكرارات المتوقعة

		أعزب	متزوج
تفضيل القناة ج	نعم	20	35
	لا	15	30

القيم المتوقعة e_{ij}:

$$e_{ij} = \frac{\sum \text{Ligne}_i \times \sum \text{Colonne}_j}{\sum \text{Observations}}$$

		أعزب	متزوج
تفضيل القناة ج	نعم	19.25	35.75
	لا	15.75	29.25

– حساب الفروق: إذا كانت الأرقام في الجدول الثاني هي نفسها الموجودة في الجدول الأول، نقول أن افتراضنا السابق (المتغيران مستقلان) صحيح تماما. أما إذا لم تكن الأرقام نفسها (وهذا هو الغالب) فإننا سنبني قرارنا على كمية الفرق الموجود بين التكرارات المشاهدة والتكرارات المتوقعة.

كيف نحسب كمية الفروق الموجودة بين التكرارات المشاهدة والمتوقعة ؟

أ) إذا حسبنا مباشرة مجموع الفروق الموجودة $\sum(e_{ij} - o_{ij})$ فإن بعض الفروق السالبة ستذهب مع الفروق الموجبة وبالتالي هذا المجموع لا يعطي فعلا مجموع الفروق الموجودة.

ب) من أجل التخلص من الإشارة يمكن استعمال القيمة المطلقة $\sum|e_{ij} - o_{ij}|$ لكننا سنتحصل على قيمة لمجموع الفروق لها علاقة بوحدة القياس (مئات الكيلوغرامات، أعشار السنتيم، ملايين الحبات، . . .) وفي هذه الحالة يصعب إيجاد القيم الحرجة التي يمكننا المقارنة بها لقول أن الفروق كبيرة أم لا.

ج) من أجل ما سبق (من أجل التخلص من الإشارة وكذا التخلص من وحدة القياس)

نستعمل العلاقة التالية التي تعطينا مجموع الفروقات الموجودة بين التكرارات المشاهدة والتكرارات المتوقعة بدون وحدة قياس:

نسمي المجموع المتحصل عليه في هذه العلاقة بكاي تربيع المحسوبة :

$$X^2 = \sum \frac{(e_{ij}-o_{ij})^2}{e_{ij}}$$

نطبق هذه العلاقة على المعطيات السابقة :

$$X^2 = \frac{(19.25-20)^2}{19.25} + \frac{(35.75-35)^2}{35.75} + \frac{(15.75-15)^2}{15.75} + \frac{(29.25-30)^2}{29.25} = 0.083$$

نجد X^2 المحسوبة

13 – 3 – جدول كاي تربيع:

جدول كاي تربيع موجود في الملاحق، وهو يقدم قيمة كاي تربيع الحرجة التي نقارنها مع كاي تربيع المحسوبة.

ولكي نستخرج القيمة الحرجة من الجدول لا بد من معرفة:

– العمود: نسبة الثقة التي نستعملها في الاختبار (مثلا: نختار العمود P=0.95 من أجل ثقة 95%)

– السطر: درجة الحرية في الجدول : (عدد الأسطر–1)x(عدد الأعمدة–1)

ddl=(nombre_ligne−1)x(nombre_colonne−1)

من أجل مثالنا السابق ومن أجل نسبة 95% ثقة ودرجة حرية: ddl=(2−1)x(2−1)=1

نجد: $X^2 = 3.84$

13 – 4 – مقارنة المحسوبة بالقيمة الحرجة أو لجدولية:

إذا كانت الفروقات بين التكرارات المشاهدة والتكرارات المتوقعة صغيرة ويمكننا إهمالها، في هذه الحالة نقبل الافتراض الأول (المتغيران مستقلان). والذي يجعلنا نقول أن الفروق كبيرة أم صغيرة هو مقارنة مجموع الفروق (X2 المحسوبة) من القيمة الحرجة ونستعمل من أجل اتخاذ القرار القاعدة التالية:

قاعدة لاتخاذ القرار: إذا كانت X2 المحسوبة أكبر من X2 الجدولية نرفض الفرضية القائلة باستقلالية المتغيرين.

في مثالنا نلاحظ أن المحسوبة أقل بكثير من الجدولية وبالتالي لا يمكننا أن نرفض الفرضية الأولى (الصفرية) التي وضعناها في (أ) أي أن المتغيران فعلا مستقلان ، وبتعبير آخر نقول أنه لا علاقة بين الحالة الاجتماعية للفرد وتفضيله أم لا القناة الجزائرية وذلك في حدود ثقة 95%.

3 – ملخص خطوات الاختبار: نتبع الخطوات التالية لإجراء اختبار X2

1 – وضع الفرضية الصفرية H0 ونفترض فيها بأنه ليس هناك علاقة بين المتغيرين.

2 – حساب التكرارات المتوقعة e . . : $\longleftarrow$ $e_{ij} = \dfrac{(somme\ \ ligne)(somme\ \ colonne)}{(total)}$

3 – حساب X2 . . . . :

$$X^2 = \sum \frac{(e_{ij} - o_{ij})^2}{e_{ij}}$$

4 – استخراج X2 الجدولية (من الجدول)

5 – اتخاذ القرار حول قبول أو رفض H0

13 – 5 – تمرين تطبيقي مع الحل:

لدراسة العلاقة بين مستوى الدخل ونوع السكن حصلنا على عينة من 100 شخص ثم قمنا بتصنيفهم حسب ظاهرتين مستوى الدخل (مرتفع منخفض) ونوع السكن (فيلا أو شقة)، فحصلنا على التكرارات الفعلية الموضحة في الجدول التالي:

	مرتفع	منخفض	المجموع
شقة	21	49	70
فيلا	19	11	30
المجموع	40	60	100

الإشكال: هل هناك علاقة بين نوع السكن ومستوى الدخل في حدود ثقة 95% ؟

1- الفرضية الصفرية: مستوى الدخل مستقل عن نوع السكن = H0

2 – حساب التكرارات المتوقعة: من دراستنا للاحتمالات نعلم انه إذا كان الحادثين العشوائيين مستقلين عن بعضهما البعض فان احتمال وقوعهما معا يساوي حاصل ضرب احتمال وقوع كل منهما، بناءا على ذلك بافتراض استقلال الدخل عن نوع السكن نستطيع إيجاد التكرار المتوقع للأشخاص ذوي الدخل المنخفض الذين يقطنون في فلل، وذلك بضرب احتمال التقاطع في عدد حجم العينة. أي :

$$\frac{60 \times 30}{100} = 18$$

وبنفس الطريقة نحسب باقي القيم ونحصل على جدول الاقتران (التكرارات المتوقعة) :

	مرتفع	منخفض
شقة	28	42
فيلا	12	18

3 – كاي تربيع المحسوبة:

$$\chi^2=\sum\frac{(e_{ij}-o_{ij})^2}{e_{ij}}=\frac{(21-28)^2}{28}+\frac{(19-12)^2}{12}+\frac{(11-18)^2}{18}+\frac{(49-42)^2}{42}=9.72$$

4 – كاي تربيع الجدولية: عند مستوى معنوية α = 0.05 (ثقة 95%) $\chi^2=3.841$

ودرجة حرية=ddl=(عدد الأسطر–1)x(عدد الأعمدة–1)=1

5 – القرار: X2 المحسوبة اكبر من X2 الجدولية. لذلك نرفض H0 ونستنتج وجود علاقة

بين مستوى الدخل وبين نوع السكن.

14 – تحليل التباين:

يعتبر اختبار التوزيع الطبيعي من أقوى الاختبارات الإحصائية للمقارنة بين متوسطي مجتمعين معلومي التباين، ويعتبراختبار t هو الأقوى عندما يكون تباين المجتمعين مجهولا ولكنهما متساويين والمجتمعان طبيعيين. ولكن إذا زاد عدد المقارنات عن 2 فإننا نستعمل طريقة تحليل التباين.

وتحليل التباين يجزء التباين الكلي إلى عدة عناصر لكل منها مصدره ثم يعطي مقدار مشاركة كل مصدر من هذه المصادر في إجمالي التباين.

1 – تحليل التباين الأحادي:

التحليل التباين الأحادي به مصدر واحد للاختلاف (المعالجات)، والنموذج الرياضي له هو:

$$Yij = \mu + Yi + Yij \qquad (\mu i = \mu + Yi)$$

حيث: Yij : المفردة j الناتجة من المعالجة i

μ : المتوسط الكلي ، μi : متوسط المجموعة i

ei : تأثير المعالجة ، eij : الخطأ العشوائي

نتائج القياسات الخاصة بتحليل التباين الأحادي تكون كما يلي:

	المعالجات						
	1	2	…	i	…	t	
	y11 y12 … y1n	y21 y22 … y2n	… … … …	yi1 yi2 … yin	… … … …	yt1 yt2 … ytn	
الحجم	n1	n2	…	ni	…	nt	n
المجموع	y1.	y2.	…	yi.	…	yt.	y..
المتوسط	$\overline{X}_{1.}$	$\overline{X}_{2.}$	…	$\overline{X}_{1.}$	…	$\overline{X}_{t.}$	$\overline{X}_{..}$
المربع	y1.2	y2.2	…	yi.2	…	yt.2	y..2

أ) – كتابة الفرضيات:

H0 : $\mu 1 = \mu 2 = \ldots = \mu t$

H1 : على الأقل متوسطين غير متساويين

ب) – تجزئة التباين الكلي: SST = SSTRT + SSE

مجموع المربعات

$$SST = \sum_{i=1}^{t}\sum_{i=1}^{n} y_{ij}^2 - \frac{y_{..}^2}{N}$$

مجموع المربعات

$$SSTRT = \sum_{i=1}^{t} \frac{y_{i.}^2}{n_i} - \frac{y_{..}^2}{N}$$

مجموع مربعات الخطأ العشوائي: SSE = SST – SSTRT

ج) – جدول تحليل التباين: ANOVA:

مصدر الاختلاف	مجموع المربعات	درجات الحرية	متوسط المربعات	F المحسوبة
SV	SS	ddl	MS	
المعالجات TRT	SSTRT	t–1	SSTRT/(t–1)	MSTRT
الخطأ E	SSE	N–t	SSE/ (N–t)	MSE
المجموع T	SST	N–1		

د) – القرار: من جدول F وبدرجات حرية ($v1=t-1$) و($v2=n(t-1)$) ومستوى معنوية (α =5%) أو (α =1%) نستخرج قيمة $F\alpha,v1,v1$ الجدولية.

إذا كانت F المحسوبة أقل من $F\alpha,v1,v1$ الجدولية نقبل H0، وإذا كانت F المحسوبة أكبر من $F\alpha,v1,v1$ الجدولية نقوم بمقارنة متوسطات المجموعات (μi) مثنى مثنى لمعرفة من متساوية منهن ومن غير متساوية مع الأخرى.

مثال: 3 مجموعات من الطلبة، تدرس كل مجموعة عند أستاذ، وتحصلوا على النقاط التالية، هل يمكن القول (عند مستوى معنوية 5%) أن هناك فروق معنوية بين متوسطات الدرجات المعطاة من قبل الأساتذة الثلاثة؟

	المعالجات (الأساتذة)			
	1	2	3	
	95 , 32 , 47 75 , 83 , 84 73 , 68	85 , 90 , 79 50 , 32 , 84 78 , 95 , 65 80	79 , 92 , 63 68 , 76 , 20 37 , 75 , 86	t = 3
ni	8	10	9	N=27
yi.	557	738	596	y..=1891
$\overline{X}_{1.}$	69.6	73.8	66.22	$\overline{X}_{..}$=70.04
$yi.^{2}$	310249	544644	355216	$\frac{y_{..}^2}{N}$=132440
$\frac{Y_{i.}^2}{n_i}$	38781,13	54464,4	39468.44	132714

$$\sum\sum y_{ij}^2 = 143525 \quad \rightarrow \text{SST} = 143525 - 132440 = 11085$$

$$\sum\sum \frac{Y_{i.}^2}{n_i} = 132714 \rightarrow \text{SSTRT} = 132714 - 132440 = 274$$

يمكننا الآن رسم جدول تحليل التباين التالي:

SV	SS	ddl	MS	F
المعالجات TRT	274	2	137	0.304
الخطأ E	10811	24	450.46	
المجموع T	11085	26		

القرار: من جدول F وبدرجات حرية 2 و 24 و α=0.05، نجد أن F0.05,2,24=3.4، وحيث أن F المحسوبة أقل من F الجدولية إذن لا توجد معلومات كافية لرفض H0.

ملاحظة: يمكن الحصول على قيمة F الجدولية باستعمال: =INVERSE.LOI.F(0,05; 2; 24) في برنامج Excel

المقارنات المتعددة:

إذا كانت نتيجة اختبار تحليل التباين تشير إلى وجود فروق جوهرية بين متوسطات المعالجات (حالة F المحسوبة تكون أكبر من F الجدولية) يكون في هذه الحالة السؤال الذي يجب الإجابة عليه : بين أي من المتوسطات تكون الفروق أو الاختلافات؟

وللإجابة على هذا السؤال يجب إجراء عدة مقارنات بين متوسطات المعالجات لتحديد أي المتوسطات تختلف عن الأخرى. وهذا ما يطلق عليه تسمية المقارنات المتعددة. وهناك العديد من الطرق لإجراء هذه المقارنات لكننا سنشرح طريقة فيشر لأقل فرق معنوي.

اختبار فيشر لأقل فرق معنوي Fisher's least significant difference (LSD) :

يستخدم هذا الاختبار للمقارنة بين متوسطي كل معالجتين على حده. ويوصي فيشر بعدم استخدام هذا الاختبار إلا إذا كان اختبار F معنوي.

وصيغة أقل فرق معنوي تكون كما يلي:

$$LSD = (t_{\frac{\alpha}{2},N-t})\sqrt{MSE\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

حيث t تمثل عدد المعالجات و N العدد الكلي للمفردات في التجربة. وسوف نرفض H0 :μi= μj إذا كانت : $|\bar{y}_i - \bar{y}_j| > LSD$.

مثال:

استخدم اختبار فيشر لتحديد أي المتوسطات تختلف عن بعضها البعض في المثال التالي:

جربنا 4 فيتامينات على 20 بقرة ولاحظنا الزيادة في الأوزان التالية:

A	B	C	D
46	65	27	15
53	59	37	30
54	43	25	27
39	49	35	28
33	37	43	40
yi.=225	253	167	140
$\bar{y}_{i.} = 45$	50.6	33.4	28

جدول تحليل التباين هو الآتي:

المصدر	درجات الحرية	مجموع المربعات	متوسط المربعات	F المحسوبة	F الجدولية
المعالجات TRT	3	SST=1613.35	MST=537.78	6.21	3.24
الخطأ ERR	16	SSE=1386.40	MSE=86.65		F0.05 ; 3 ; 16
المجموع Total	19	SST=2999.75			

النتيجة : ليس للفيتامينات الأربعة نفس التأثير عند مستوى معنوية 5%.

الحل: α=0.05 ، N−t =16 وعليه فإن : $t_{\frac{\alpha}{2},N-t} = t_{0.025,16} = 2.12$ وحيث أن حجم العينة متساوي في جميع المعالجات فإن قيمة LSD ستكون متساوية في جميع المقارنات:

$$\mathrm{LSD} = \left(t_{\frac{\alpha}{2},N-t}\right)\sqrt{\mathrm{MSE}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)} = (2.12)\sqrt{86.65\left(\frac{1}{5}+\frac{1}{5}\right)} = 12.48$$

سيكون إذن اختبار LSD كما يلي:

الفرضية	الفرق المطلق بين المتوسطين	LSD	القرار
H0 :μA= μB	5.6	12.48	لا نرفض H0
H0 :μB= μC	17.2	12.48	نرفض H0
H0 :μB= μD	22.6	12.48	نرفض H0
H0 :μA= μC	11.6	12.48	لا نرفض H0
H0 :μA= μD	17	12.48	نرفض H0
H0 :μC= μD	5.4	12.48	لا نرفض H0

النتيجة: المتوسطات المتساوية والمختلفة مختصرة في الشكل التالي: ———

$\mu_B \quad \mu_A \quad \mu_C \quad \mu_D$

أمثلة و تمارين

1 - العينات:

مسألة محلولة:

في منطقة من مناطق هذا الوطن الحبيب، اشتكى الأهالي من عزوف أبنائهم عن المطالعة والمذاكرة في المنزل وبالخصوص تلاميذ التعليم المتوسط. فكلفوك بدراسة هذه الظاهرة والبحث عن أسبابها واقتراح الحلول لها. فبعد فهمك للإشكالية وضعت فرضيات وأعددت استمارة من مجموعة أسئلة للإجابة على الفرضيات.

- ما هو المجتمع المدروس؟

- إذا لم تكن لديك الإمكانيات والوقت الكافي لدراسة كل مفردات المجتمع، ماذا تفعل؟

- ما نوع العينة المختارة؟ وكيف يتم اختيار مفرداتها؟

إذا علمت أن هناك فرق بين مطالعة الذكور ومطالعة الإناث (يعني هناك تباين بين الإناث والذكور في ما يخص المطالعة في المنزل):

- ما نوع العينة المختارة؟ وكيف يتم اختيار مفرداتها؟

إذا علمت أن عدد التلاميذ في التعليم المتوسط في هذه المنطقة هو 450 بنت و550 ذكر، وأردت استخراج عينة متكونة من 200 تلميذ :

- كم يكون نصيب كل جنس في العينة؟

- استعمل جدول الأرقام العشوائية واختر فقط 10 بنات من بين ال450 بنت. (أذكر الخطوات)

الحل النموذجي:

- المجتمع المدروس: كل التلاميذ في التعليم المتوسط الساكنين في المنطقة المذكورة.

- إذا لم تكن لدي الإمكانيات والوقت الكافي لدراسة كل مفردات المجتمع، أستعمل العينة.

- العينة المختارة هي عينة عشوائية بسيطة ويتم اختيار مفرداتها عن طريق القرعة أو بواسطة جدول الأرقام العشوائية وذلك بعد الحصول على قائمة اسمية لكل التلاميذ.

– إذا كان هناك تباين بين الإناث والذكور في ما يخص المطالعة في المنزل فالعينة المختارة هي عينة عشوائية طبقية ويتم اختيار مفرداتها بحساب عدد البنات وعدد الذكور في العينة ثم نسحب عينة عشوائية بسيطة من كل فئة، وذلك بعد الحصول على قائمة اسمية لكل التلاميذ بنات وذكور.

– إذا كان عدد التلاميذ 450 بنت و550 ذكر، وأردت استخراج عينة متكونة من 200 تلميذ :

نصيب البنات في العينة: $\frac{200}{1000}x450 = 90$ ، نصيب الذكور في العينة: $\frac{200}{1000}x550 = 110$

– بجدول الأرقام العشوائية نختار 10 بنات من بين ال450 بنت. مثلا، من العمود الأول، نقرأ 3 أرقام على اليمين نسجل كل رقم أقل من 450 وبدون تكرار حتى نحصل على 10 أرقام.

تمرين غير محلول:

نريد اختيار عينة متكونة من 30 شخص من بين 565 شخصا .

أ) – ما نوع العينة المستعملة إذا علمت أننا لدينا قائمة اسمية لكل الأشخاص ؟ استخرج هذه العينة باستعمال جدول الأرقام العشوائية.

ب) – ما نوع العينة المستعملة إذا علمت أنه لدينا قائمتين ، الأولى بها 350 امرأة و الثانية بها 215 رجلا؟ استخرج هذه العينة باستعمال جدول الأرقام العشوائية.

ج) – ما نوع العينة المستعملة إذا علمت أنه ليس لدينا قائمة ولكننا نعلم أن الأشخاص كلهم سيخرجون من باب القاعة التي هم متواجدون بها؟ استخرج هذه العينة .

فكرة على الحل: العينة الأولى عشوائية بسيطة، الثانية عشوائية طبقية والثالثة منتظمة.

تمارين أخرى حول التقديرات:

(1) – مجتمع حجمه 800 أخذت منه عينة حجمها 64 متوسطها 50 وانحرافها المعياري 20. أوجد مجال الثقة لمتوسط المجتمع في حدود ثقة 95 % . z [54.7 – 45.3]

(2) – مجتمع توزيعه طبيعي حجمه 1000 و انحرافه المعياري 30 أخذت منه عينة حجمها 25 ومتوسطها 80 . أوجد مجال الثقة لمتوسط المجتمع في حدود ثقة 90 % ، 95 % ، 99 % .

z [95.48 – 64.52] ، [91.76 – 68.24] ، [89.84 – 70.16]

(3) – مجتمع حجمه 500 و انحرافه المعياري 40 أخذت منه عينة حجمها 36 ومتوسطها 380. أوجد مجال الثقة لمتوسط المجتمع في حدود ثقة 95 % . z [392.54 – 367.46]

(4) – مجتمع حجمه 1200 عامل أخذت منه عينة حجمها 100 عامل، 70 منهم يفضلون طريقة الدوام الجديدة. أوجد مجال الثقة لنسبة العمال الذين يفضلون هذه الطريقة بثقة 95 % . [0.79 – z 0.61]

(5) – في مدرسة بها 1000 تلميذ، أخذنا عينة من 35 تلميذ ووجدنا أن 7 منهم مصابين بأعراض مرض معين، أوجد مجال الثقة لنسبة المصابين بهذه الأعراض في كل المدرسة بدرجة ثقة 95 %.

z [0.332 – 0.068]

(6) – مجتمع حجمه 1000 يتبع التوزيع الطبيعي، أخذت منه عينة حجمها 25 مفردة متوسطها 80. أوجد مجال أوجد مجال الثقة لمتوسط المجتمع في حدود ثقة 90 % ، 95 % ، 99 % . قارن مع نتائج التمرين (2) [90.27 – 69.73] ، [92.29 – 67.62] ، t [96.78 – 63.22]

(7) – أخذت عينة من 9 مصابيح وكان متوسط عمرها 300 ساعة بانحراف معياري s=45، إذا افترضنا أن عمر المصابيح يتبع التوزيع الطبيعي، أوجد مجال الثقة لمتوسط عمر المصابيح في حدود 90 % . t [328 – 272]

(8) – يرغب صاحب مصنع مصابيح كهربائية تقدير نسبة المصابيح المعيبة بدقة ±0.1 بدرجة ثقة 95 % ما هو الحد الأدنى لحجم العينة المطلوب إذا كان يعلم أن هذه النسبة عادة هي 0.2؟.
[62]

ما هو هذا الحجم إذا كان صاحب المصنع لا يعلم أن هذه النسبة عادة هي 0.2؟. [97]

9– اختبار كاي تربيع:

تمرين محلول:

الجدول الموالي يمثل توزيع الموظفين حسب الجنس على مديريات مؤسسة ما.

المديريات	المديرية (أ)	المديرية (ب)	المديرية (ج)	المديرية (د)
عدد الإناث	20	40	10	30
عدد الذكور	30	20	25	25

يدعي مدير هذه المؤسسة أنه لا يفرق بين الرجال والنساء في المناصب داخل المديريات. تأكد من ذلك في حدود 95 % ثقة.

الحل:

H0 ليس هناك فرق بين الذكور والإناث – ليس للجنس علاقة بتوزيع الموظفين على المديريات

بعد حساب القيم المتوقعة، نجد:

X2 المحسوبة : 15,5498 ،

درجة الحرية : 3 = (4−1)(2−1) = ddl ، X2 الجدولية : 7,815

القرار: المحسوبة أكبر من الجدولية نرفض 0H (إدعاء المدير خاطئ)

10- تمرين محلول على تحليل التباين:

3 مجموعات من الطلبة، تدرس كل مجموعة عند أستاذ، وتحصلوا على النقاط التالية،

(الأساتذة)		
1	2	3
95 , 32 , 47 , 75 , 83 , 84 , 73 , 68	85 , 90 , 79 , 50 , 32 , 84 , 78 , 95 , 65 , 80	79 , 92 , 63 , 68 , 76 , 20 , 37 , 75 , 86

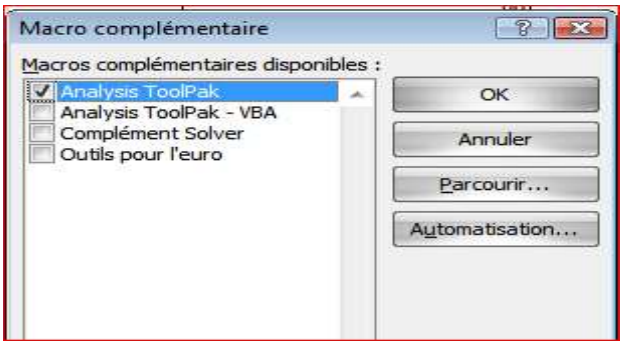
هل يمكن القول (عند مستوى معنوية 5%) أن هناك فروق معنوية بين متوسطات الدرجات المعطاة من قبل الأساتذة الثلاثة؟

هذا التمرين محلول في الدرس الخاص بتحليل التباين، وكل الحسابات موضحة في الحل المعطى، المطلوب هنا إعادة الإجابة على السؤال باستعمال برنامج تحليل التباين الموجود في Excel.

الحل:

أولا: إضافة برنامج تحليل المعطيات الإحصائية إلى Excel إذا لم يكن قد أضيف بعد وهذا عن طريق:

Menu Fichier / Options / Compléments / Analysis ToolPak (Atteindre)



ثانيا: اختر : Menu Données / Utilitaire d'analyse / Analyse de variance : un facteur (Ok)

ثالثا: في النافذة التي تظهر حدد جدول المعطيات في (Plage d'entrée) ومكان حصولك على النتائج في (Plage de sortie) وستتحصل على النتائج التالية:

Analyse de variance: un facteur

RAPPORT DÉTAILLÉ

Groupes	Nombre d'échantillons	Somme	Moyenne	Variance
Colonne 1	9	596	66,22	549,44
Colonne 2	10	738	73,8	379,51
Colonne 3	8	557	69,625	428,55

ANALYSE DE VARIANCE

Source des variations	Somme des carrés	Degré de liberté	Moyenne des carrés	F	Probabilité	Valeur critique pour F
Entre Groupes	273,93	2	136,97	0,304	0,7406	3,403
A l'intérieur des groupes	10811,03	24	450,46			
Total	11084,96	26				

كما نرى تحصلنا على نفس النتيجة لكن هذه المرة بدون إجراء أي حسابات.

القرار (أنظر حل التمرين في درس تحليل التباين)

تمرين محلول حول معامل سبيرمن لارتباط الرتب:

معامل سبيرمان لارتباط الرتب بين متغيرين X و Y: هو $r_s = 1 - \frac{6(\sum d^2)}{n(n^2-1)}$

حيث: $\sum d^2$ هو مجموع مربعات الفروق بين رتب المتغيرين و n هي عدد أزواج القيم.

أن قيمة معامل ارتباط الرتب تنحصر بين –1، +1 فإذا كان معامل ارتباط الرتب يساوي + 1 (ارتباط طردي تام بين الرتب). وإذا كان معامل ارتباط الرتب يساوي – 1 (ارتباط عكسي تام بين الرتب).

(مثال)

البيانات التالية تمثل أعداد الساعات التي ذاكرها عشرة طلاب والدرجات التي حصلوا عليها:

عدد الساعات X	10	6	12	14	11	6	19	16	3	9
الدرجات y	60	48	83	76	74	58	98	89	37	69

أحسب معامل سبيرمان لارتباط الرتب.

(الحل)

ننظم الحل في الجدول التالي مع ملاحظة أن n = 10

لحساب معامل سبيرمان لارتباط الرتب نقوم بترتيب كل من المتغيرين ترتيباً تصاعديا. نأخذ أقل قيمة من قيم المتغير الرتبة رقم 1، والقيمة الأعلى منها مباشرة الرتبة رقم 2 وهكذا (بالنسبة لكل من المتغيرين). وعند تساوي قيمتين (أو أكثر) من قيم المتغير نعطي كل قيمة رتبة مختلفة (كما لو كانت القيم غير متساوية) ثم نحسب متوسط هذه الرتب، ويعطى هذا المتوسط لكل من هذه القيم المتساوية.

عدد الساعات X	الدرجات Y	رتب X	رتب Y	الفروق d	d2
10	60	6	7	− 1	1.00
6	48	8.5	9	− 0.5	0.05
12	83	4	3	1	1.00
14	76	3	4	− 1	1.00
11	74	5	5	0	0
6	58	8.5	8	0.5	0.25
19	98	1	1	0	0
16	89	2	2	0	0
3	37	10	10	0	0
9	69	7	6	1	1.00
المجموع				Zero	4.50

وبالتعويض في القانون حيث $\sum d^2 = 4.5$ ، $n = 10$ نحصل على:

$$r_s = 1 - \frac{6(\sum d^2)}{n(n^2-1)} = 1 - \frac{6(4.5)}{10(1-99)} = 1 - \frac{27}{990} = 1 - 0.027 \qquad r_s = 0.973$$

مما يعني أننا أمام علاقة طردية قوية بين المتغيرين. فكلما زادت عدد الساعات التي يدرسها الطالب في هذا المثال، كلما زادت درجاته في الامتحانات قوة وذلك بما نسبته 97 %.

تمرين محلول حول معامل بيرسون للارتباط الخطي البسيط:

معامل بيرسون للارتباط الخطي البسيط بين متغيرين X و Y:

هو "متوسط حاصل ضرب الوحدات المعيارية للمتغيرين" $r = \frac{1}{n} \sum \left(\frac{x - \bar{x}}{S_x}\right)\left(\frac{y - \bar{y}}{S_y}\right)$

حيث $\bar{x} ; \bar{y}$ هما المتوسط الحسابي للمتغيرين و $S_x ; S_y$ انحرافهما المعياري.

الصيغة المختصرة والتي تعطي النتائج نفسها: $r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{n \sum x^2 - (\sum x)^2} \sqrt{n \sum y^2 - (\sum y)^2}}$

(مثال):

البيانات التالية تمثل أعمار ثمانية من الناخبين ودخولهم اليومية بالدولار، والمطلوب حساب معامل بيرسون للارتباط الخطي بين الأعمار والدخول.

الأعمار x : 25 32 29 43 38 51 47 35

الدخول y : 10 18 15 35 40 62 100 50

الحل: لحساب معامل بيرسون للارتباط الخطي يلزم حساب المجاميع:

$\sum x, \sum y, \sum xy, \sum x^2, \sum y^2$ لذلك يتم تنظيم حساب هذه المجاميع كما يلي:

الأعمار x	الدخول y	xy	x2	y2
25	10	250	625	100
32	18	576	1024	324
29	15	435	841	225
43	35	1505	1849	1225
38	40	1520	1444	1600
51	62	3162	2601	3844
47	100	4700	2209	10000
35	50	1750	1225	2500
300	330	13898	11818	19818

ثم نطبق في الصيغة المختصرة لمعامل الارتباط حيث $n = 8$:

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

$$= \frac{8(13898) - (300)(330)}{\sqrt{8(11818) - (300)^2}\sqrt{8(19818) - (330)^2}}$$

$$= \frac{111184 - 99000}{\sqrt{94544 - 9000}\sqrt{158544 - 108900}} = \frac{12184}{\sqrt{4544}\sqrt{49644}} = \frac{12184}{15019.6} \Rightarrow r = 0.81$$

معامل بيرسون للارتباط الخطي بين أعمار الناخبين ودخولهم اليومية يساوي 0.81 وهو ارتباط طردي (لأن إشارته موجبة) وقوى (لأنه قريب من الواحد). بمعنى أخر، إن هناك علاقة طردية قوية بين عمر الناخب ودخله مقدارها 81 %. فمع زيادة عمر الناخب يزيد دخله، والعكس صحيح.

المراجع:

أهم المراجع باللغة العربية:

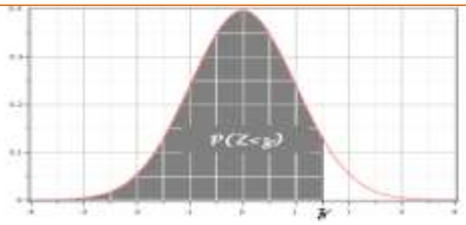
- أحمد عبد السميع طبية - مبادئ الإحصاء – الطبعة الأولى - دار البداية - عمان - 2008
- ثروت محمد عبد المنعم - مدخل حديث للإحصاء والاحتمالات - (مكتبة العبيكان)
- جلال الصياد، عبد الحميد محمد ربيع - مبادئ الطرق الاحصائية - الطبعة الأولى - المملكة العربية السعودية - جدة - 1983
- رابح حمودي – الرفيق في علم الاحصاء – دار المعرفة – الجزائر 2002
- عبد الكريم بوحفص – الاحصاء المطبق في العلوم الاجتماعية والانسانية، مدعم بتطبيقات وتمارين محلولة – الطبعة الثالثة منقحة – ديوان المطبوعات الجامعية – مارس 2011
- علي عبد السلام العماري، د. علي حسين العجيلي – الاحصاء والاحتمالات النظرية والتطبيق – مالطا – شركة ELGA – 2000
- محمود محمد إبراهيم هندي، د. خلف سلمان سلطان سلمان - مفاهيم لطرق التحليل الإحصائي - (مكتبة الرشد)
- موساوي عبد النور، بركان يوسف – الإحصاء Statistique 2 - دار العلوم للنشر والتوزيع – عنابة – 2010

أهم المراجع باللغة الفرنسية:

- Abdeldjellil Bezzaoucha – Les fondements de la biostatistique et de l’épidémiologie en sciences médicales – OPU Alger – 2009
- Admane O, Hoang-ky, Ouakli N. – Statistique cours et exercices pour les étudiants du tronc commun Bio-Médical – OPU – Alger - 2006
- Didier Cornuel, Paul Derreumaux – Calcul des probabilités, exercices corrigés et rappels de cours – Paris – éditions Sirey – 1976
- François Dress - Les probabilités et statistique de A à Z - Dunod
- J.F.Boisvieux, J.I.Golmard, A.Mallet, V.Morice - Statistiques PCEM1 - Paris VI - 2002
- Khaled Khaldi – Méthodes statistiques et probabilités – Casbah éditions – Alger – 2000
- Les Fichiers Vuibert – Sciences économiques et sociales – Terminale ES – Paris – Vuibert – Août 1997
- Murray R.Spiegel – Mini Schaum’s, Statistique – EdiScience – 2002
- Nadia Benhaddou Bakkioui - Resumé théorique et guide de travaux pratiqued, Module: Statistiqued - Maroc
- Renée Veysseyre - Aide mémoire Statistique et probabilités pour l'ingénieur - 2e édition - Dunod - Paris - 2006
- Yves Tillé - Résumé du cours de statistique descriptive - Janvier 2008

جدول الأعداد العشوائية

58362	26638	63247	12561	90699	37091	46929	89058	89610	51155
98218	28245	51679	85910	85872	74086	18659	47180	54992	17540
15732	21196	18489	39664	91103	16495	22475	31470	25887	57744
66418	44470	40567	66220	89760	43861	21417	92577	87327	42997
77016	72343	41997	84354	50722	28276	39730	83148	51918	86105
72571	68100	96023	54214	97040	49810	58539	81743	93514	28158
15396	34117	50800	44175	61661	13394	67971	32081	46658	11831
72660	18101	92493	49529	64960	53793	60466	13886	92584	77709
89980	27484	18407	45849	28613	44056	38340	34868	48535	91434
66104	85700	65537	54254	29978	11045	45281	81300	67977	18530
20747	86658	63549	38522	29848	26191	53859	99963	69595	98387
42529	18397	46883	87886	87707	13093	64668	54970	72121	78620
13860	23248	93968	97952	69060	88347	55367	86856	92500	77404
63267	75097	46642	48001	28556	63489	99021	71159	85154	24747
45736	19918	55030	12175	42267	85378	42570	50322	61430	11562
34180	87952	83290	89719	83785	42648	38372	51542	14623	37126
50342	59810	96204	34747	68700	47719	73067	70790	96113	37226
66883	18649	52127	93585	72948	78988	55721	47169	86907	40575
26139	41194	73124	22605	42863	92381	12873	93100	29315	76280
77042	32345	55394	94605	81329	14060	73681	80074	63490	65926
16787	30013	17790	24083	30766	50660	55964	59428	80788	42585
15946	74386	83708	97802	83343	87782	77189	59562	64515	78629
58888	74392	99248	90890	49785	60224	86308	82544	20073	42222
22700	85754	17948	93902	17395	83096	21945	16713	17395	47793
53849	41791	22629	58474	12145	56640	50289	41584	80671	74220
37256	30678	21309	44727	86736	72571	47452	82650	65544	95852
73968	99793	19726	44443	61605	83195	19667	34875	97671	28130
14524	76227	26805	39780	92485	10790	54805	11272	89504	34882
98638	29455	87573	18169	18098	87037	54140	75683	93480	55400
32479	28112	47159	47345	86871	82843	93191	93691	22547	34637
61864	98060	41000	71756	79732	16162	11037	81626	91382	21126
59737	81657	87086	77072	18733	25328	11408	13020	29882	40117
48086	48229	98703	97314	70864	69432	77501	87243	64487	86736
21918	77948	71201	15999	64839	50173	34978	31808	83796	86997
51846	36998	15248	12168	47547	15428	68276	12420	61040	25521
60123	29738	71828	80670	87560	75183	49500	63687	62663	31007
87471	24372	19352	52613	94976	24917	98435	45238	27206	81354
83811	89596	27702	44318	20442	92130	18974	34182	90540	33156
80029	69629	31397	25848	15185	30538	89428	81827	55364	69420
72168	67120	81817	65244	33192	75973	61417	68988	24308	14673
94875	56371	34951	24437	39438	73508	65651	39156	32957	19300
55977	26809	62817	85878	49805	93932	44136	91096	18501	70216
26160	26491	23549	32708	83531	61620	82798	46036	82136	34614
57319	11365	48082	27938	48150	56151	44061	56116	32650	44235
79768	80482	60126	98360	98664	58921	85423	95442	59063	65454
66666	48821	63191	81974	39171	13070	66229	58935	75991	57238
44055	26938	21844	56975	31330	60606	69857	70929	90870	42320
15584	69540	77840	37583	14419	87421	15447	11377	28004	77794
81248	78155	63365	26679	75966	73742	22584	78427	77341	62805
84990	68407	31600	60990	13361	58271	45605	36345	59066	92239



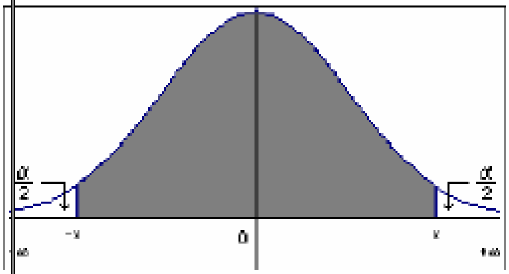
جدول التوزيع الطبيعي المعياري

Exemple :

$P(z < 1.96) = 0.975$

Ligne : 1.9 Colonne : 0.06

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983
3.6	0.99984	0.99985	0.99985	0.99986	0.99986	0.99987	0.99987	0.99988	0.99988	0.99989
3.7	0.99989	0.99990	0.99990	0.99990	0.99991	0.99991	0.99992	0.99992	0.99992	0.99992
3.8	0.99993	0.99993	0.99993	0.99994	0.99994	0.99994	0.99994	0.99995	0.99995	0.99995
3.9	0.99995	0.99995	0.99996	0.99996	0.99996	0.99996	0.99996	0.99996	0.99997	0.99997

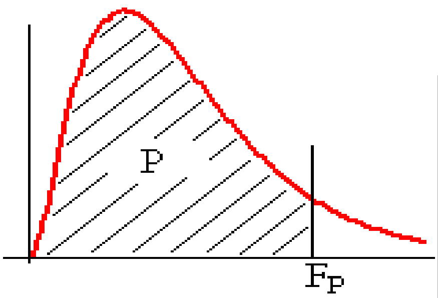


جدول توزيع ستودنت

Cette table donne les fractiles de la loi de Student à ν degrés de liberté : valeur t ayant la probabilité α d'être dépassée en valeur absolue : $P(-t < T < t) = 1 - \alpha$.
Ou : $P(T < -t) = \alpha/2 = P(T > t)$

<u>Exemple</u>: α = 5%=0,05 et ddl=2 t=4.3027

α	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0.05	0.02	0.01	0.005	0.001
1- α	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95	0.975	0.99	0.995	0.9975	0.9995
ddl														
1	0.1584	0.3249	0.5095	0.7265	1	1.3764	1.9626	3.0777	6.3137	12.706	31.821	63.656	127.32	636.58
2	0.1421	0.2887	0.4447	0.6172	0.8165	1.0607	1.3862	1.8856	2.92	4.3027	6.9645	9.925	14.089	31.6
3	0.1366	0.2767	0.4242	0.5844	0.7649	0.9785	1.2498	1.6377	2.3534	3.1824	4.5407	5.8408	7.4532	12.924
4	0.1338	0.2707	0.4142	0.5686	0.7407	0.941	1.1896	1.5332	2.1318	2.7765	3.7469	4.6041	5.5975	8.6101
5	0.1322	0.2672	0.4082	0.5594	0.7267	0.9195	1.1558	1.4759	2.015	2.5706	3.3649	4.0321	4.7733	6.8685
6	0.1311	0.2648	0.4043	0.5534	0.7176	0.9057	1.1342	1.4398	1.9432	2.4469	3.1427	3.7074	4.3168	5.9587
7	0.1303	0.2632	0.4015	0.5491	0.7111	0.896	1.1192	1.4149	1.8946	2.3646	2.9979	3.4995	4.0294	5.4081
8	0.1297	0.2619	0.3995	0.5459	0.7064	0.8889	1.1081	1.3968	1.8595	2.306	2.8965	3.3554	3.8325	5.0414
9	0.1293	0.261	0.3979	0.5435	0.7027	0.8834	1.0997	1.383	1.8331	2.2622	2.8214	3.2498	3.6896	4.7809
10	0.1289	0.2602	0.3966	0.5415	0.6998	0.8791	1.0931	1.3722	1.8125	2.2281	2.7638	3.1693	3.5814	4.5868
11	0.1286	0.2596	0.3956	0.5399	0.6974	0.8755	1.0877	1.3634	1.7959	2.201	2.7181	3.1058	3.4966	4.4369
12	0.1283	0.259	0.3947	0.5386	0.6955	0.8726	1.0832	1.3562	1.7823	2.1788	2.681	3.0545	3.4284	4.3178
13	0.1281	0.2586	0.394	0.5375	0.6938	0.8702	1.0795	1.3502	1.7709	2.1604	2.6503	3.0123	3.3725	4.2209
14	0.128	0.2582	0.3933	0.5366	0.6924	0.8681	1.0763	1.345	1.7613	2.1448	2.6245	2.9768	3.3257	4.1403
15	0.1278	0.2579	0.3928	0.5357	0.6912	0.8662	1.0735	1.3406	1.7531	2.1315	2.6025	2.9467	3.286	4.0728
16	0.1277	0.2576	0.3923	0.535	0.6901	0.8647	1.0711	1.3368	1.7459	2.1199	2.5835	2.9208	3.252	4.0149
17	0.1276	0.2573	0.3919	0.5344	0.6892	0.8633	1.069	1.3334	1.7396	2.1098	2.5669	2.8982	3.2224	3.9651
18	0.1274	0.2571	0.3915	0.5338	0.6884	0.862	1.0672	1.3304	1.7341	2.1009	2.5524	2.8784	3.1966	3.9217
19	0.1274	0.2569	0.3912	0.5333	0.6876	0.861	1.0655	1.3277	1.7291	2.093	2.5395	2.8609	3.1737	3.8833
20	0.1273	0.2567	0.3909	0.5329	0.687	0.86	1.064	1.3253	1.7247	2.086	2.528	2.8453	3.1534	3.8496
21	0.1272	0.2566	0.3906	0.5325	0.6864	0.8591	1.0627	1.3232	1.7207	2.0796	2.5176	2.8314	3.1352	3.8193
22	0.1271	0.2564	0.3904	0.5321	0.6858	0.8583	1.0614	1.3212	1.7171	2.0739	2.5083	2.8188	3.1188	3.7922
23	0.1271	0.2563	0.3902	0.5317	0.6853	0.8575	1.0603	1.3195	1.7139	2.0687	2.4999	2.8073	3.104	3.7676
24	0.127	0.2562	0.39	0.5314	0.6848	0.8569	1.0593	1.3178	1.7109	2.0639	2.4922	2.797	3.0905	3.7454
25	0.1269	0.2561	0.3898	0.5312	0.6844	0.8562	1.0584	1.3163	1.7081	2.0595	2.4851	2.7874	3.0782	3.7251
26	0.1269	0.256	0.3896	0.5309	0.684	0.8557	1.0575	1.315	1.7056	2.0555	2.4786	2.7787	3.0669	3.7067
27	0.1268	0.2559	0.3894	0.5306	0.6837	0.8551	1.0567	1.3137	1.7033	2.0518	2.4727	2.7707	3.0565	3.6895
28	0.1268	0.2558	0.3893	0.5304	0.6834	0.8546	1.056	1.3125	1.7011	2.0484	2.4671	2.7633	3.047	3.6739
29	0.1268	0.2557	0.3892	0.5302	0.683	0.8542	1.0553	1.3114	1.6991	2.0452	2.462	2.7564	3.038	3.6595
30	0.1267	0.2556	0.389	0.53	0.6828	0.8538	1.0547	1.3104	1.6973	2.0423	2.4573	2.75	3.0298	3.646
∞	0.1257	0.2533	0.3853	0.5244	0.6744	0.8416	1.0364	1.2816	1.6449	1.96	2.3264	2.5759	2.8072	3.2908



جدول توزيع كاي تربيع χ^2

Cette table donne les fractiles F_P de la loi de khi-deux

à ddl degrés de liberté : P = Probabilité ($\chi^2 < F_P$)

(degré de signification $\alpha = 1 - P$)

Exemple: α = 5%=0,05 (P=0.95) et ddl=2 χ^2=5.991

ddl \ P	0.5	0.6	0.7	0.75	0.8	0.85	0.9	0.95	0.98	0.99	0.999
1	0.455	0.708	1.074	1.323	1.642	2.072	2.706	3.841	5.412	6.635	10.827
2	1.386	1.833	2.408	2.773	3.219	3.794	4.605	5.991	7.824	9.210	13.815
3	2.366	2.946	3.665	4.108	4.642	5.317	6.251	7.815	9.837	11.345	16.266
4	3.357	4.045	4.878	5.385	5.989	6.745	7.779	9.488	11.668	13.277	18.466
5	4.351	5.132	6.064	6.626	7.289	8.115	9.236	11.070	13.388	15.086	20.515
6	5.348	6.211	7.231	7.841	8.558	9.446	10.645	12.592	15.033	16.812	22.457
7	6.346	7.283	8.383	9.037	9.803	10.748	12.017	14.067	16.622	18.475	24.321
8	7.344	8.351	9.524	10.219	11.030	12.027	13.362	15.507	18.168	20.090	26.124
9	8.343	9.414	10.656	11.389	12.242	13.288	14.684	16.919	19.679	21.666	27.877
10	9.342	10.473	11.781	12.549	13.442	14.534	15.987	18.307	21.161	23.209	29.588
11	10.341	11.530	12.899	13.701	14.631	15.767	17.275	19.675	22.618	24.725	31.264
12	11.340	12.584	14.011	14.845	15.812	16.989	18.549	21.026	24.054	26.217	32.909
13	12.340	13.636	15.119	15.984	16.985	18.202	19.812	22.362	25.471	27.688	34.527
14	13.339	14.685	16.222	17.117	18.151	19.406	21.064	23.685	26.873	29.141	36.124
15	14.339	15.733	17.322	18.245	19.311	20.603	22.307	24.996	28.259	30.578	37.698
16	15.338	16.780	18.418	19.369	20.465	21.793	23.542	26.296	29.633	32.000	39.252
17	16.338	17.824	19.511	20.489	21.615	22.977	24.769	27.587	30.995	33.409	40.791
18	17.338	18.868	20.601	21.605	22.760	24.155	25.989	28.869	32.346	34.805	42.312
19	18.338	19.910	21.689	22.718	23.900	25.329	27.204	30.144	33.687	36.191	43.819
20	19.337	20.951	22.775	23.828	25.038	26.498	28.412	31.410	35.020	37.566	45.314
21	20.337	21.992	23.858	24.935	26.171	27.662	29.615	32.671	36.343	38.932	46.796
22	21.337	23.031	24.939	26.039	27.301	28.822	30.813	33.924	37.659	40.289	48.268
23	22.337	24.069	26.018	27.141	28.429	29.979	32.007	35.172	38.968	41.638	49.728
24	23.337	25.106	27.096	28.241	29.553	31.132	33.196	36.415	40.270	42.980	51.179
25	24.337	26.143	28.172	29.339	30.675	32.282	34.382	37.652	41.566	44.314	52.619
26	25.336	27.179	29.246	30.435	31.795	33.429	35.563	38.885	42.856	45.642	54.051
27	26.336	28.214	30.319	31.528	32.912	34.574	36.741	40.113	44.140	46.963	55.475
28	27.336	29.249	31.391	32.620	34.027	35.715	37.916	41.337	45.419	48.278	56.892
29	28.336	30.283	32.461	33.711	35.139	36.854	39.087	42.557	46.693	49.588	58.301
30	29.336	31.316	33.530	34.800	36.250	37.990	40.256	43.773	47.962	50.892	59.702

Pour $\nu > 30$, La loi du χ^2 peut –être approximée par la loi normale $N(\nu, \sqrt{\nu})$

جدول توزيع فيشر

$F_{0,05\,,\,v1\,,\,v2}$

v1 : le numérateur des degrés de liberté, v2 : le dénominateur des degrés de liberté.

v1	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**12**	**15**	**20**	**24**	**30**	**40**	**60**	**120**
v2																		
1	161	199	215	224	230	234	236	238	240	241	243	245	248	249	250	251	252	253
2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,5	19,5	19,5	19,5	19,5
3	10,1	9,6	9,3	9,1	9,0	8,9	8,9	8,8	8,8	8,8	8,7	8,7	8,7	8,6	8,6	8,6	8,6	8,5
4	7,7	6,9	6,6	6,4	6,3	6,2	6,1	6,0	6,0	6,0	5,9	5,9	5,8	5,8	5,7	5,7	5,7	5,7
5	6,6	5,8	5,4	5,2	5,1	5,0	4,9	4,8	4,8	4,7	4,7	4,6	4,6	4,5	4,5	4,5	4,4	4,4
6	6,0	5,1	4,8	4,5	4,4	4,3	4,2	4,1	4,1	4,1	4,0	3,9	3,9	3,8	3,8	3,8	3,7	3,7
7	5,6	4,7	4,3	4,1	4,0	3,9	3,8	3,7	3,7	3,6	3,6	3,5	3,4	3,4	3,4	3,3	3,3	3,3
8	5,3	4,5	4,1	3,8	3,7	3,6	3,5	3,4	3,4	3,3	3,3	3,2	3,2	3,1	3,1	3,0	3,0	3,0
9	5,1	4,3	3,9	3,6	3,5	3,4	3,3	3,2	3,2	3,1	3,1	3,0	2,9	2,9	2,9	2,8	2,8	2,7
10	5,0	4,1	3,7	3,5	3,3	3,2	3,1	3,1	3,0	3,0	2,9	2,8	2,8	2,7	2,7	2,7	2,6	2,6
11	4,8	4,0	3,6	3,4	3,2	3,1	3,0	2,9	2,9	2,9	2,8	2,7	2,6	2,6	2,6	2,5	2,5	2,4
12	4,7	3,9	3,5	3,3	3,1	3,0	2,9	2,8	2,8	2,8	2,7	2,6	2,5	2,5	2,5	2,4	2,4	2,3
13	4,7	3,8	3,4	3,2	3,0	2,9	2,8	2,8	2,7	2,7	2,6	2,5	2,5	2,4	2,4	2,3	2,3	2,3
14	4,6	3,7	3,3	3,1	3,0	2,8	2,8	2,7	2,6	2,6	2,5	2,5	2,4	2,3	2,3	2,3	2,2	2,2
15	4,5	3,7	3,3	3,1	2,9	2,8	2,7	2,6	2,6	2,5	2,5	2,4	2,3	2,3	2,2	2,2	2,2	2,1
16	4,5	3,6	3,2	3,0	2,9	2,7	2,7	2,6	2,5	2,5	2,4	2,4	2,3	2,2	2,2	2,2	2,1	2,1
17	4,5	3,6	3,2	3,0	2,8	2,7	2,6	2,5	2,5	2,4	2,4	2,3	2,2	2,2	2,1	2,1	2,1	2,0
18	4,4	3,6	3,2	2,9	2,8	2,7	2,6	2,5	2,5	2,4	2,3	2,3	2,2	2,1	2,1	2,1	2,0	2,0
19	4,4	3,5	3,1	2,9	2,7	2,6	2,5	2,5	2,4	2,4	2,3	2,2	2,2	2,1	2,1	2,0	2,0	1,9
20	4,4	3,5	3,1	2,9	2,7	2,6	2,5	2,4	2,4	2,3	2,3	2,2	2,1	2,1	2,0	2,0	1,9	1,9
21	4,3	3,5	3,1	2,8	2,7	2,6	2,5	2,4	2,4	2,3	2,3	2,2	2,1	2,1	2,0	2,0	1,9	1,9
22	4,3	3,4	3,0	2,8	2,7	2,5	2,5	2,4	2,3	2,3	2,2	2,2	2,1	2,0	2,0	1,9	1,9	1,8
23	4,3	3,4	3,0	2,8	2,6	2,5	2,4	2,4	2,3	2,3	2,2	2,1	2,0	2,0	2,0	1,9	1,9	1,8
24	4,3	3,4	3,0	2,8	2,6	2,5	2,4	2,4	2,3	2,3	2,2	2,1	2,0	2,0	1,9	1,9	1,8	1,8
25	4,2	3,4	3,0	2,8	2,6	2,5	2,4	2,3	2,3	2,2	2,2	2,1	2,0	2,0	1,9	1,9	1,8	1,8
26	4,2	3,4	3,0	2,7	2,6	2,5	2,4	2,3	2,3	2,2	2,1	2,1	2,0	1,9	1,9	1,9	1,8	1,7
27	4,2	3,4	3,0	2,7	2,6	2,5	2,4	2,3	2,3	2,2	2,1	2,1	2,0	1,9	1,9	1,8	1,8	1,7
28	4,2	3,3	2,9	2,7	2,6	2,4	2,4	2,3	2,2	2,2	2,1	2,0	2,0	1,9	1,9	1,8	1,8	1,7
29	4,2	3,3	2,9	2,7	2,5	2,4	2,3	2,3	2,2	2,2	2,1	2,0	1,9	1,9	1,9	1,8	1,8	1,7
30	4,2	3,3	2,9	2,7	2,5	2,4	2,3	2,3	2,2	2,2	2,1	2,0	1,9	1,9	1,8	1,8	1,7	1,7
40	4,1	3,2	2,8	2,6	2,4	2,3	2,2	2,2	2,1	2,1	2,0	1,9	1,8	1,8	1,7	1,7	1,6	1,6
60	4,0	3,2	2,8	2,5	2,4	2,3	2,2	2,1	2,0	2,0	1,9	1,8	1,7	1,7	1,6	1,6	1,5	1,5
120	3,9	3,1	2,7	2,4	2,3	2,2	2,1	2,0	2,0	1,9	1,8	1,8	1,7	1,6	1,6	1,5	1,4	1,4

جدول توزيع فيشر (تابع)

$F_{0,01\ ,\ v1\ ,\ v2}$

v1 : le numérateur des degrés de liberté, v2 : le dénominateur des degrés de liberté.

v1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120
v2																		
1	4052	4999	5403	5624	5763	5859	5928	5981	6022	6055	6106	6157	6208	6234	6260	6286	6313	6339
2	98,5	99,0	99,2	99,2	99,3	99,3	99,4	99,4	99,4	99,4	99,4	99,4	99,4	99,5	99,5	99,5	99,5	99,5
3	34,1	30,8	29,5	28,7	28,2	27,9	27,7	27,5	27,3	27,2	27,1	26,9	26,7	26,6	26,5	26,4	26,3	26,2
4	21,2	18,0	16,7	16,0	15,5	15,2	15,0	14,8	14,7	14,5	14,4	14,2	14,0	13,9	13,8	13,7	13,7	13,6
5	16,3	13,3	12,1	11,4	11,0	10,7	10,5	10,3	10,2	10,1	9,9	9,7	9,6	9,5	9,4	9,3	9,2	9,1
6	13,7	10,9	9,8	9,1	8,7	8,5	8,3	8,1	8,0	7,9	7,7	7,6	7,4	7,3	7,2	7,1	7,1	7,0
7	12,2	9,5	8,5	7,8	7,5	7,2	7,0	6,8	6,7	6,6	6,5	6,3	6,2	6,1	6,0	5,9	5,8	5,7
8	11,3	8,6	7,6	7,0	6,6	6,4	6,2	6,0	5,9	5,8	5,7	5,5	5,4	5,3	5,2	5,1	5,0	4,9
9	10,6	8,0	7,0	6,4	6,1	5,8	5,6	5,5	5,4	5,3	5,1	5,0	4,8	4,7	4,6	4,6	4,5	4,4
10	10,0	7,6	6,6	6,0	5,6	5,4	5,2	5,1	4,9	4,8	4,7	4,6	4,4	4,3	4,2	4,2	4,1	4,0
11	9,6	7,2	6,2	5,7	5,3	5,1	4,9	4,7	4,6	4,5	4,4	4,3	4,1	4,0	3,9	3,9	3,8	3,7
12	9,3	6,9	6,0	5,4	5,1	4,8	4,6	4,5	4,4	4,3	4,2	4,0	3,9	3,8	3,7	3,6	3,5	3,4
13	9,1	6,7	5,7	5,2	4,9	4,6	4,4	4,3	4,2	4,1	4,0	3,8	3,7	3,6	3,5	3,4	3,3	3,3
14	8,9	6,5	5,6	5,0	4,7	4,5	4,3	4,1	4,0	3,9	3,8	3,7	3,5	3,4	3,3	3,3	3,2	3,1
15	8,7	6,4	5,4	4,9	4,6	4,3	4,1	4,0	3,9	3,8	3,7	3,5	3,4	3,3	3,2	3,1	3,0	3,0
16	8,5	6,2	5,3	4,8	4,4	4,2	4,0	3,9	3,8	3,7	3,6	3,4	3,3	3,2	3,1	3,0	2,9	2,8
17	8,4	6,1	5,2	4,7	4,3	4,1	3,9	3,8	3,7	3,6	3,5	3,3	3,2	3,1	3,0	2,9	2,8	2,7
18	8,3	6,0	5,1	4,6	4,2	4,0	3,8	3,7	3,6	3,5	3,4	3,2	3,1	3,0	2,9	2,8	2,7	2,7
19	8,2	5,9	5,0	4,5	4,2	3,9	3,8	3,6	3,5	3,4	3,3	3,2	3,0	2,9	2,8	2,8	2,7	2,6
20	8,1	5,8	4,9	4,4	4,1	3,9	3,7	3,6	3,5	3,4	3,2	3,1	2,9	2,9	2,8	2,7	2,6	2,5
21	8,0	5,8	4,9	4,4	4,0	3,8	3,6	3,5	3,4	3,3	3,2	3,0	2,9	2,8	2,7	2,6	2,5	2,5
22	7,9	5,7	4,8	4,3	4,0	3,8	3,6	3,5	3,3	3,3	3,1	3,0	2,8	2,7	2,7	2,6	2,5	2,4
23	7,9	5,7	4,8	4,3	3,9	3,7	3,5	3,4	3,3	3,2	3,1	2,9	2,8	2,7	2,6	2,5	2,4	2,4
24	7,8	5,6	4,7	4,2	3,9	3,7	3,5	3,4	3,3	3,2	3,0	2,9	2,7	2,7	2,6	2,5	2,4	2,3
25	7,8	5,6	4,7	4,2	3,9	3,6	3,5	3,3	3,2	3,1	3,0	2,9	2,7	2,6	2,5	2,5	2,4	2,3
26	7,7	5,5	4,6	4,1	3,8	3,6	3,4	3,3	3,2	3,1	3,0	2,8	2,7	2,6	2,5	2,4	2,3	2,2
27	7,7	5,5	4,6	4,1	3,8	3,6	3,4	3,3	3,1	3,1	2,9	2,8	2,6	2,6	2,5	2,4	2,3	2,2
28	7,6	5,5	4,6	4,1	3,8	3,5	3,4	3,2	3,1	3,0	2,9	2,8	2,6	2,5	2,4	2,4	2,3	2,2
29	7,6	5,4	4,5	4,0	3,7	3,5	3,3	3,2	3,1	3,0	2,9	2,7	2,6	2,5	2,4	2,3	2,2	2,1
30	7,6	5,4	4,5	4,0	3,7	3,5	3,3	3,2	3,1	3,0	2,8	2,7	2,5	2,5	2,4	2,3	2,2	2,1
40	7,3	5,2	4,3	3,8	3,5	3,3	3,1	3,0	2,9	2,8	2,7	2,5	2,4	2,3	2,2	2,1	2,0	1,9
60	7,1	5,0	4,1	3,6	3,3	3,1	3,0	2,8	2,7	2,6	2,5	2,4	2,2	2,1	2,0	1,9	1,8	1,7
120	6,9	4,8	3,9	3,5	3,2	3,0	2,8	2,7	2,6	2,5	2,3	2,2	2,0	2,0	1,9	1,8	1,7	1,5

مفهوم وحساب درجة الحرية degré de liberté (ddl)

لشرح معنى درجة الحرية نضرب مثال بمدرب كرة القدم لديه أحد عشرة (11) لاعبا سيدخلون الملعب الواحد تلو الآخر ويقوم هذا المدرب بوضع كل لاعب في المكان الذي يراه مناسبا له على الملعب. بما أن هناك 11 مكانا في لعبة كرة القدم، فإن للمدرب 11 اختيارا بالنسبة للاعب الأول الذي يدخل عليه، و بعد أن يختار له مكانا يبقى لديه 10 اختيارات للاعب الموالي و هكذا حتى يملأ عشرة أماكن ويبقى لديه مكان شاغر واحد فقط يتوجب عليه إعطاؤه للاعب الأخير دون إمكانية الاختيار.

في هذا المثال، كان للمدرب الحرية في اختيار 10 أماكن للاعبين أما المكان الأخير فلا خيار فيه. يعني أن للمدرب درجة الحرية = عدد اللاعبين – 1 = 11-1 = 10.

بنفس الطريقة في الاحصاء، إذا كان لدينا خمسة (5) أرقام ونعلم أن مجموعهم هو 20 فلدينا الحرية في اختيار أربعة (4) أرقام (مثلا: 1 و 7 و 2 و 4) لكن الرقم الخامس فلا يمكننا اختياره لأن قيمته تكون من طرح الأرقام الأربعة من المجموع: الرقم الخامس = 20 – (1+7+2+4) = 20- 14 = 6، إذن قيمة الرقم الخامس لا بد وأن تكون 6. عدد درجات الحرية هنا هي عدد القيم – 1.

إذا كان لدينا جدولا يتكون من (r) من الأسطر و من (c) من الأعمدة ولدينا مجاميع الأسطر والأعمدة فإن درجة الحرية (كم رقما يمكننا وضعه بحرية داخل هذا الجدول حتى تبقى المجاميع كما هي) تحسب كما يلي: ddl=(r-1)(c-1) ولنوضح هذا بمثال:

ليكن لدينا الجدول التالي:

مجاميع الأسطر				
70				
30				
	20	**30**	**50**	مجاميع الأعمدة

يجب أن نلاحظ أن هذا الجول يتكون من سطرين وثلاثة أعمدة يعني (r=2) و (c=3) وليس كما قد يتوهم البعض بأن للجدول أربعة أسطر و خمسة أعمدة، فلا نحسب أعمدة وأسطر العناوين و المجاميع.

إذن للجدول سطرين وثلاثة أعمدة يعني أن هناك 2*3 = 6 قيم يجب وضعها في خانات الجدول. لكن حتى نحافظ على المجاميع دون تغيير فلدينا الحرية في اختيار قيمتين فقط من بين 6 . يعني:
ddl=(r-1)(c-1)= (2-1)(3-1)=2

مجاميع الأسطر				
70	10	20	**40**	
30	10	**10**	10	
	20	**30**	**50**	مجاميع الأعمدة

إذا وضعنا الرقمين 40 و 10 مثلا في الجدول (الرقمين بالبنط العريض) فإن الأرقام الأخرى ستنتج أوتوماتيكيا من معرفتنا للمجاميع (الأرقام الصغيرة).