وزارة التعليم العالي والبحث العلمي

جامـعة الجزائر 3

كليـة العلــوم الاقتــصادية والعــلوم التــجاريــة وعـلوم التســيير

مطبوعة خاصة بمقياس:

الاحصاء 3

موجهة لطلبة السنة الثانية ليسانس جميع الشعب

من إعداد:

الدكتور طويطو محمد

أستاذ محاضر أ

السنة الجامعية 2017/2016

مقدمة

الإحصاءعلم يهتم بالمعلومات والبيانات – ويهدف إلى تجميعها وتبوبيها وتنظميها وتحليلها واستخلاص النتائج منها بلوتعميم نتائجها – واستخدامها في اتخاذ القرارات، وأدى التقدم المذهل في تكنولوجيا المعلومات واستخدام الحاسبات الآلية إلى مساعدة الدارسين والباحثين ومتخذي القراراتفي الوصول إلى درجات عالية ومستويات متقدمةمن التحليل ووصف الواقع ومتابعته ثم إلى التنبؤ بالمستقبل .

ولمتعد البحوث الاقتصادية والاجتماعية والإدارية وغيرها في وقتنا المعاصر ، وفي ظل التقدم التكنولوجي الهائل في كافة ميادين حياتنا اليومية، تكتفي بمجرد عرض المشاكل ودراسة الظواهر وتحديد الأسباب واستخلاص النتائج واتخاذ القرارات بطريقة سطحية مجردة عن أسلوبالإقناع والتقدير والقياس.

ولقدأصبح الاتجاه العام في مثل هذه البحوث والدراسات هو استخدام طرق القياس الكمية ووسائلالإقناعالإحصائية وذلك لتحديد الخصائص وإبراز الاتجاهات العامة في الظواهر الاجتماعيةوالإدارية،وتحليل العلاقات المتشابكة والمتبادلة بين الظواهر على أساس موضوع غير متميز.

وعلمالإحصاء يعطي للباحثين في مجال العلوم الاقتصادية والاجتماعية والإدارية،العديد من الطرق والأساليب اللازمة لضرورة القيام بالدراسات والبحوث الاقتصادية والاجتماعية والإدارية والجغرافية علي أساس من القياس لحركة العديد من المتغيراتالمحددة للظواهر موضوع الدراسة.

وتستخدم كلمة الإحصاء لتشير إلى عملية جمع البيانات الكمية والأساليب المستعملة في معالجة تلك البيانات، وقد نعني بهذه الكلمة أيضا عملية استخلاص بعض الاستنتاجات من دراسة عينة صغيرة لصياغة تعميمات يمكن تطبيقها علي مجتمعات اكبر حجما.

فبحوثالرأي العام على سبيل المثال تقوم على مقابلة ودراسة عينة صغيرة من أفراد المجتمع ولكن نتائجها تستخدم في الاستدلال علىاتجاهاتالرأي العام في المجتمع ككل،وبذلك يمكن القول بان الإحصاء يشير إلي طرق تنظيم وتلخيص البيانات وإلىالأساليب التي تستخدم في تحليل وتفسير النتائج واستخلاصاتها والتييمكن تعميمها على مجتمع الدراسة.

فالإحصاء هو علم يبحث في طريق جمع الحقائق الخاصة بالظواهر العلمية الاجتماعية التي تتمثل في حالات أو مشاهدات متعددة، وفى كيفية تسجيل هذه الحقائق في صورة قياسية رقمية، وتلخيصها بطريقة يسهل بها معرفة اتجاهات الظواهر وعلاقات بعضها ببعض، ويبحث أيضاً في دراسة هذه العلاقات والاتجاهات واستخدامها في تفهم حقيقة الظواهر ومعرفة القوانين التي تسير تبعاً لها .

ومن هنا يتضح أن الإحصاء لا غنى عنه لأي باحث في شتى المجالات المختلفة إذ اعتمد في بحثه على الأسلوب العلمي. أي أن الإحصاء هو عصا الباحث التي تقوده إلى الطريق الصحيح، وهى الأداة التي تساعده على تفسير الظواهر التي يدرسها وتوضيح النتائج التي يحصل عليها ودلالات البيانات والأرقام التي يحصل عليها.

يهدف هذا المقرر إلى تعريف طلاب قسم الاقتصاد بعلم الإحصاء (3) وأهميته ودوره في تسهيل عمل الباحث الاقتصادي في التعامل مع مجتمع البحث بدءا من إيجاد توزيع المعاينة للعينات وتقديراتها ومجالات الثقة الخاصة بها واختبارات الفروض ودلالتها كاختبار (الطبيعي N، ستيودنت t، فيشر F، χ^2 ...الخ)، وذلك بهدف إكساب الطالب مجموعة من الخبرات في مجال الإحصاء الاقتصادي كي تساعده في عرض نتائج البحوث الاقتصادية الكيفية بصورة كمية محدده وواضحة ومختصره ودقيقة.

1-الاطار المفاهيمي.

يستند هذا القسم من الأساليب الإحصائية إلى مجموعة من النظريات الإحصائية لعل أهمها نظرية الاحتمالات ونظرية توزيع المعاينة اللتان تمثلان حلقة الوصل بين الإحصاء الوصفيوالإحصاء (3). ويسعى هذا النوع من الأساليب الإحصائية إلى الوصول إلى تقديرات لمعالم وخصائص مجتمعات الدراسة من خلال ما هو متوفر من معلومات عن العينات

المختارة من تلك المجتمعات، فضلا عن اختبار الفروض الإحصائية عن مجتمع البحث على أساس البيانات المتاحة عن عينات الدراسة. ويطلق على هذا النوع من الأساليب أكثر من تسمية تؤدى جميعها إلى نفس المعنى فأحيانا يسمى بالإحصاء (3)أو الإحصاء الاستدلالي، أو الاستنباطي أو التعميمي حيث يهدف إلى الوصول إلى تعميمات عن مجتمع الدراسة من خلال العينة المسحوبة من هذا المجتمع. ويشمل هذا النوع من الأساليب الإحصائية، الاحتمالات، توزيع المعاينة، مجالات الثقة،اختبار الفروض.

1-1- <u>مفهوم الاستدلال:</u> ويقصد بوظيفة الاستدلال اشتقاق النتائج من دراسة وفحص المقدمات والبيانات المتوافرة عن ظاهرة معينة. ولهذا يطلق علىالعملية الإحصائية التي تستخدم الاستدلالي علىأساس المنطق الاستدلالي المبني على نظرية الاحتمالات الرياضية فمن عينة محددة من أعمالأحد المصانع وباستخدام الأسلوبالإحصاء الاستدلالي يكون من الممكن التنبؤ بمعدلات الزيادة في الإنتاج ومقدار التغير في نسبة الغياب وفي هذه الحالة نجد أن الدقة في التنبؤ تعتمد علي عوامل كثيرة من أهمها ملائمة الأدواتالإحصائية المستخدمة وحجم العينة محل الدارسة والإجراءاتالإحصائية اتخذت عند اختيارها.

وتعتبر وظيفة الاستدلال أو الاستقراء من الأهمية بمكان في البحث العلمي فمثلا:

إذا كانت الظاهرة موضوع الدراسة والتحليل ممثلة للمجتمع الذيتنتميإليه فانه يمكن الحصول على نتائج معنوية عن المجتمع بتحليل بيانات هذه الظاهرة وهو ما يعرف بالاستدلال ويعتمد هذا الأسلوبفي البحث على الشروط التي يجب توافرها حتى يكون هذا الاستدلال سليما، وبما أن الاستدلال لا يمكن أن يكون مؤكد فان لغة الاحتمال تستخدم عند عرض النتائج.

وتعتبر وظيفة الاستقراء لها أهمية كبيرة فهي تمكن الباحث من الوصول إلى تعميمات عن المجتمع على أساس المعلومات المتاحة من عينة منه، وفى هذه الحالة فان أساليب ومقاييس الوصف يقتصر وصفها على ذلك الجزء (العينة) فقط من المجتمع، ومن هنا تأتيأهمية وظيفة الاستقراء فهي تمكننا من وصف المجتمع (التعميم) باستخدام بيانات العينة .

إن القوانين في العلوم الطبيعية والاجتماعية تجد برهانها عند الوقائع والحقائق الإحصائية ولذا يعد الاستقراء الإحصائيأساسا لتطور المعرفة العلمية باعتباره البرهان لهذه القوانين. ووظيفة الاستقراء تحقق مطلبين أساسيينفي البحث: الأول تقدير خواص المجتمع والثاني اختبارات الفروض حول هذهالخواص. ولا تقتصر هذه الوظيفة على مجرد الاستقراء بل تقدم لنا تقييما عن مدى دقة هذا الاستقراء وأكثر من ذلك فهي تمكننا من التحكم في مستوى الدقة وذلك بعدة طرق منها استخدام الأسلوب المناسب للمعاينة والحجم المناسب للعينة. وباختصار فان هذه الوظيفة للإحصاء تمدنا بالاستقراء المنطقي وتختلف الأساليب المتبعة في الاستقراء حسب طبيعة محل الاستقراء.

* الإحصاء الاستدلالي (الاستقرائي) بمعنى استخلاص النتائج العامة من النتائج الجزئية.
* التحليل العاملي، ويهتم في قياس العوامل الكامنة وراء الظواهر من اجل صياغة النتائج بصورة نظريات علمية .

1-2- <u>مفهوم المتغيرات :</u>

هناك حقيقة، تقول : إن مفهوم المتغيرات يتصل بالبيانات الكمية. ويعرف المتغير بـ "انه كمية تتغير(أي تختلف قيمتها) من مفردة لأخرى". ويقصد أيضا بالمتغير أيةخاصية تأخذ قيما باختلافالمتغيرات(الأفراد، أو ألاماكن، أو الأشياء)في العمل الإحصائي بأشكال متعددة ، منها:

1-2-1- <u>المتغيرات الكمية</u>:وهي تلك الصفات التي يتملكها الأفراد أو الأشياء،والتي بالإمكان قياسها..

1-2-2- <u>المتغيرات النوعية</u>: هناك العديد من الخواص لا يمكن قياسها (كما هو الحال عند قياس الأطوال أو الأوزان أو الأعمار) بل بالإمكان تجزئتها إلى نوعين (نعم ، لا) .. فعلى سبيل المثالعندما يراد معرفة مدى ممارسة طلبة الجامعة للأنشطة الرياضية يكونوا إما (ممارسين لها) أو (غير ممارسين لها) ..

1-2-3- المتغيرات العشوائية: من خلال عملية قياس أي متغير مبحوث نصل إلى قيمة ذلك المتغير(أي انه صدفة ظهر طول اللاعب بهذه القيمة، بمعنى لا تعرف الأسباب) فان المتغير في هذه الحالة يسمى بـ (المتغير العشوائي)، والمتغيرات العشوائية، أنواع، منها:

- **المتغير العشوائي المتقطع(المنفصل):**

فالمتغير العشوائي المتقطع يتميز بتقطعات في القيم التي يأخذها (أي إن هذه القيم تكون منفصلة الواحدة عن الأخرى) ... وان هذه الفواصل أو التقطعات تدل على عدم وجود قيم واقعة بين قيم أخرى معلومة يمكن للمتغير أن يأخذها.

- **المتغير العشوائي المستمر:**

يقصد بالمتغير العشوائي المستمر، هو المتغير الذي لا يخضع لخاصية التقطعات (القيم المنفصلة) كما هو الحال عند المتغير العشوائي المتقطع ... والمتغير العشوائي المستمر، يتضمن مختلف القياسات التي بالإمكان إجراؤها على الأفراد أو الأشياء مثل الطول أو الوزن أو العمر ... وغيرها .

1-3- مفهوم العينات والمجتمع الإحصائي:

1-3-1 مفهوم المجتمع Population: المقصود بالمجتمع هو مجموعة من الأفراد كالمجتمع العربي الذي نقصد به مجموعة من أفراد ذوي خصائص معينة أو نقول المجتمع الجزائريأو مجتمع ولاية جيجل، وهنا يقصد المجتمع كافة الأفراد الذين يسكنون في منطقة جغرافية معينة في وقت معين.

يعرف المجتمع بأنهعبارة عن جميع المفردات التي يمكن أن يأخذها المتغير، أما في الإحصاء فان مفهوم المجتمع يستخدم في مجالات أوسع فهو لا يشمل مجتمعات فحسب بل يشمل المجموعات المختلفة للموضوعات المختلفة من ظواهر طبقية وأشياء مهما كانت ذات خصائص مشتركة.

ولهذا يمكن للإحصائي أن يعرف المجتمع تبعا لأغراضه الخاصة بأنه مجموعة معينة من الحيوانات أو الأشجار أو الأفراد، ويمكن أن يكون المجتمع لباحث تربوي مجموعة معينة من الطلاب كأن يكون طلبة كلية الاقتصاد في جامعة جيجل أو طلاب أي كلية أخرى أوطلاب السنة الثانية علوم اقتصادية في كلية الاقتصاد بجيجل وهكذا.
ويمكن تصنيف المجتمعات إلى نوعين :

<u>المجتمع المحدود:</u> وهو الذي يمكن حساب أعداد أفراده كما في حالة عدد الطلابأو عدد أفراد الشعب الجزائري.

<u>المجتمع غير المحدود:</u> كما في حالة عدد الملاحظات أو التجارب العلمية أو عدد المحاضرات التي تلقى في الجامعات في كافة أنحاء العالم.

1–3–2 <u>**معالم المجتمع** Paramètre d’une population:</u>

نقصد بمعالم المجتمع مجموعة من خصائصه مثل المتوسط، التباين، معامل التماثل، ... من خصائص المجتمع أيضا طبيعة توزيعه الاحتمالي $f(x)$ كأن يكون طبيعيا أو غيره.

♦ مميزات استخدام المجتمع (الحصر الشامل) :

– دقة النتائج المتحصل عليها والوثوق في كفاءتها نظرا لجمع البيانات من كل فرد شمله البحث من دون ترك مفردة أو حالة.

- تجنب أخطاء التعميم التي تنتج من استخدام بيانات مأخوذة من عينة محددة من المجتمع وتطبيق نتائجها على المجتمع كله.

- تتفادى هذه الطريقة الأخطاء الشائعة والناجمة في غيرها من الطرائق (طريقة العينة) خاصة خطأ التحيز وخطأ الصدفة .

♦ عيوب استخدام المجتمع (الحصر الشامل):

-عالية التكاليف ويحتاج إلى إمكانيات كبيرة.

- يستغرق وقتا طويلا وتبذل فيه جهود كبيرة في جمع البيانات وتصنيفها.

- يحتاج إلى جهاز إداري وفني ضخم ومدرب للقيام به.

1-3-3- <u>مفهوم العينات:</u> تعرف العينة بأنها ذلك الجزء من المجتمع الذي يجري اختيارها على وفق قواعد وطرائق علمية بحيث تمثل المجتمع تمثيلا صحيحا.

- مميزات استخدام العينات في البحوث ما يلي :

- العينات تكتفي بعدد محدود من المفردات وليس جميعها،وذلك اقتصادا في الجهد والنفقات.

-أنها سريعة في إعطاء نتائج البحوث مقارنة بأسلوب الحصر الشامل.

- تتيح للباحث التعميق في مصادر الأحكام واتخاذ القرارات.

- تستخدم لأنها اقل عرضة للأخطاء مع الأساليب الأخرى.

- يعد استخدامها (العينات) من الوسائل المعنية بإثراء البحوث العلمية الأصلية.

-أنها طريقة مناسبة، حيث إمكانية تحديد مدى الثقة في نتائجها، وكذا نسبة تمثيلها للمجتمع.

- عيوب استخدامالعينة (أخطاء المعاينة):

– اخذ عينة من مصدر خاطئ، كأن تستخدم دليل الهاتف للحصول على عينة تمثل الرأي العام.

-التحيز الشخصي، ويحدث ذلك حينما يأخذ الباحث عينته المختارة من فئة معينة لها خصائص مميزة عن المجتمع الكلي.

– جمع بيانات ناقصة، فمثلا إهمال العامل الجغرافي عند دراسة المستوى الاقتصادي للسكان بتقسيم الأسر المبحوثة حسب دخولها.

– خطأ الصدفة، يزداد احتمال ورود هذا الخطأ كلما صغر حجم العينة.

1-3-4- أنواع العينات:

– العينات غير الاحتمالية: وهي تلك العينات التي يتم اختيارها بطريقة غير عشوائية، أي التي لا تعتمد على نظرية الاحتمالات، ومن عيوبها أنها لا تمثل مجتمع البحث تمثيلا دقيقا، ومن ثم فان نتائجها لا تصلح للتعميم على المجتمع كله، ومن أمثلة هذا النوع من العينات أن يختار الباحث عينة يرى أنها تمثل المجتمع الأصلي الذي يقوم بدراسته تمثيلا صادقا.

– العينات الاحتمالية:

– العينة العشوائية البسيطة:هي العينة التي تختار وحدتها من الإطار الخاص بها، على أساس يهيئ فرص انتقاء متكافئة لجميع وحدات المجتمع المسحوبة منها.

– العينة العشوائية الطبقية:في هذه الحالة ينبغي تقسيم المجتمع إلى أقسام أو طبقات مختلفة ثم يأخذ من كل قسم أو طبقة عينة متجانسة بطريقة عشوائية، على أن يكون حجم كل طبقة في العينة متناسبة مع حجم الطبقة المناظرة لها في المجتمع الأصلي.

– العينة العشوائية المنتظمة:يتم اختيار وحداتها بحيث تكون المسافة أو المدة بين كل وحدة وأخرى ثابتة لجميع وحدات العينة.

– العينة العشوائية العنقودية:وهي عينة تختار عن طريق استخدام تجمعات (عناقيد) تختار من المجتمع الأصلي بدلا من انتقاء المفردات بصفة مباشرة من هذا المجتمع.

1–4– <u>**مفهوم العينة النفادية والعينة غير النفادية** Echantillon exhaustif et non exhaustif:</u>

عندما يكون السحب بالإرجاع حيث يمكن أن تظهر المفردة أكثر من مرة في العينة، نسمي هذه المعاينة غير نفادية لأن تكرار العملية لا يؤدي إلى تقليص عدد المفردات في المجتمع، والعكس نسمي المعاينة بدون إرجاع معاينة نفادية. هناك فرضيتان تتكرران في عدد من العلاقات الرياضية التي سنراها لاحقا، هما فرضية أن قيم مفردات العينة مستقلة والمجتمع لانهائي. يتحقق شرط الاستقلال إذا كانت المعاينة غير نفادية، وإذا كانت كذلك، يمكن اعتبار المجتمع مجتمعا غير محدود.

1–5–<u>**مفهوم إحصائية المعاينة** Statistique de l’échantillonnage:</u>

لتقدير معالم المجتمع (متوسط المجتمع μ، تباين المجتمعσ^2 النسبةp ...) ننطلق من بيانات العينة، حيث نحتاج إلى حساب معالم مثل متوسط العينة $\bar{x}$، تباين العينة S^2، النسبة في العينةp'. بصفة عامة، نسمي كل قيمة تحسب انطلاقا من بيانات العينة من أجل تقدير قيمة معالم المجتمع إحصائية المعاينة. نظريا (رياضيا) إحصائية المعاينة هي كل دالة في المتغيرات العشوائية التي تمثل القيم المحصل عليها في العينة.

2 -المتغيرات العشوائية والتوزيعات الاحتمالية

Random Variables and Probability Distributions

مقدمة

يهتم هذا الفصل بدراسة المتغيرات العشوائية، من حيث تعريفها، وأنواعها، والتوزيعات الاحتمالية لها، وخصائص هذه التوزيعات، والتوزيعات الاحتمالية للمتغيرات العشوائية الخاصة.

المتغير العشوائي Random Variable:

المتغير العشوائي هو الذي يأخذ قيما حقيقية مختلفة تعبر عن نتائج فراغ العينة، ومن ثم مجال هذا المتغير، يشمل كل القيم الممكنة له، ويكون لكل قيمة من القيم التي يأخذها المتغير احتمال معين، وينقسم المتغير العشوائي إلى قسمين هما:

1- المتغيرات العشوائية المنفصلة Discrete Random Variables

2- المتغيرات العشوائية المتصلة(المستمرة) Continuous Random Variables

2-1-المتغيرات العشوائية المنفصلة

المتغير العشوائي المنفصل هو الذي يأخذ قيم بينية، ومتباعدة، ويرمز للمتغير العشوائي بشكل عام بحرف من الحروف الأبجدية الكبيرة.... $X, Y, Z,$ ويرمز للقيم التي يأخذها المتغير بالحروف الأبجدية الصغيرة، $x, y, z,$ ...، ومن أمثلة هذه المتغيرات:

1- عدد الأولاد الذكور في الأسرة المكونة من أربع أولاد X، $X:\{x=0,1,2,3,4\}$.

2- عدد العملاء الذين يتم إنهاء خدمتهم البنكية كل 10 دقائق Y، $Y:\{y=0,1,2,3,\ldots\}$.

3- عدد مرات استخدام نوع معين من الأسمدة خلال الدورة الزراعية.

4- عدد الوحدات التالفة من إنتاج مزرعة معينة تنتج 200 وحدة كل موسم.

5- عدد الوحدات التي تستهلكها الأسرة من سلعة معينة خلال الشهر.

وهكذا..... الأمثلة كثيرة

2-1-1 التوزيع الاحتمالي للمتغير العشوائي المنفصل

التوزيع الاحتمالي، هو الذي يبين احتمالات حدوث القيم التي يمكن يأخذها المتغير، والتي ترتبط باحتمالات النتائج الممكنة في فراغ العينة، وبمعنى آخر هو التكراري النسبي للقيم التي يمكن أن يأخذها المتغير.

فإذا كان المتغير العشوائي المنفصل X يأخذ القيم، $X:\{x=x_1,x_2,...,x_n\}$، وكان $P(X=x_i)=f(x_i)$ هو احتمال أن المتغير العشوائي يأخذ القيمة x_i، فإنه، يمكن تكوين جدول التوزيع الاحتمالي للمتغير العشوائي X، وهو جدول مكون من عمودين، الأول به القيم الممكنة للمتغير $X:\{x=x_1,x_2,...,x_n\}$، والثاني به القيم الاحتمالية لهذا المتغير $P(X=x_i)=f(x_i)$، أي أن:

جدول (1)جدول التوزيع الاحتمالي للمتغير العشوائي المنفصل

$f(x_i)$	x_i
$f(x_1)$	x_1
$f(x_2)$	x_2
$\vdots$	$\vdots$
$f(x_n)$	x_n
1	Σ

وتسمى الدالة $f(x_i)$ بدالة الاحتمال، ومن خصائص هذه الدالة ما يلي:

$$\boxed{\begin{array}{l} 1\text{-}\quad 0 < f(x_i) < 1 \\ 2\text{-}\quad \sum f(x_i) = 1 \end{array}}$$

مثال (1)

إذا كان من المعلوم أن نسبة مبيعات أحد المراكز التجارية من التفاح الأمريكي 0.60 ، بينما يكون نسبة مبيعاته من الأنواع الأخرى للتفاح 0.40، اشترى أحد العملاء عبوتين، والمطلوب:

1- كون فراغ العينة.

2- إذا عرف المتغير العشوائي X بأنه عدد العبوات المشتراة من التفاح الأمريكي، فأوجد الآتي:

- التوزيع الاحتمالي للمتغير العشوائي X .
- ارسم دالة الاحتمال لهذا المتغير.
- كون التوزيع الاحتمالي التجميعي.
- ما هو احتمال $P(X=1)$، $P(X\leq 1)$، $P(X=1.5)$، $P(X\leq 1.5)$
- حدد قيمة الوسيط، والمنوال لعدد العبوات المشتراة.

الحل:

تكوين فراغ العينة:

التجربة هنا هو شراء وحدتين من عبوات التفاح، ومن ثم فراغ العينة يتكون من أربع نتائج، هي:

- التوزيع الاحتمالي لعدد العبوات المشتراة من التفاح الأمريكي X

من المعلوم أن العميل اشترى عبوتين، وأن المتغير العشوائي هو عدد العبوات المشتراة من التفاح الأمريكي، لذا تكون القيم الممكنة للمتغير العشوائي هي:

$x=0$ إذا كانت العبوتين من النوع الآخر، أى إذا كانت نتيجة التجربة (آخر، آخر)

$x=1$ إذا كان أحد العبوتين من النوع الأمريكي، أي إذا كانت نتيجة التجربة (آخر، أمريكي) أو (أمريكي،آخر)

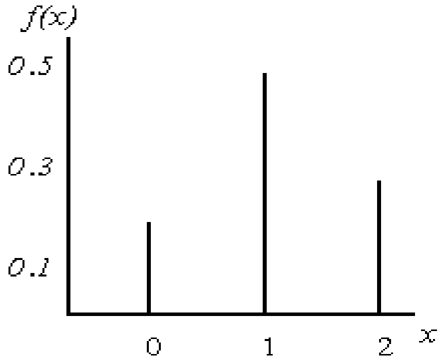
$x=2$ إذا كان العبوتين من النوع الأمريكي، أي إذا كانت نتيجة التجربة (أمريكي ، أمريكي)

ومن ثم يأخذ المتغير القيم: $X:\{x=0,1,2\}$ ، ويرتبط احتمالات هذه القيم باحتمالات نتائج التجربة المناظرة لها كما هو مبين أعلاه، ومن ثم يكون التوزيع الاحتمالي للمتغير العشوائي X هو:

جدول التوزيع الاحتمالي لعدد العبوات المشتراة من التفاح الأمريكي

x_i	$f(x_i)$
0	0.16
1	0.48
2	0.36
Σ	1

- رسم دالة الاحتمال $f(x)$:



- تكوين التوزيع الاحتمالي التجميعي:

التوزيع التجميعي، هو جدول يشمل الاحتمالات الناتجة من حساب الاحتمال $P(X \le x)$ ، ويرمز له بالرمز $F(x)$ ، أي أن دالة التوزيع الاحتمالي التجميعي تأخذ الصورة التالية:

$$\boxed{F(x) = P(X \le x)}$$

ومن ثم يمكن تكوين جدول التوزيع الاحتمالي التجميعي لعدد الوحدات المشتراة من التفاح الأمريكي كما يلي:

جدول التوزيع الاحتمالي، والتوزيع التجميعي لعدد العبوات المشتراه من التفاح الأمريكي

x_i	$f(x_i)$	$F(x_i)$
0	0.16	$F(0)=P(X\le 0)=0.16$
1	0.48	$F(1)=P(X\le 1)=0.16+0.48=0.64$
2	0.36	$F(2)=P(X\le 2)=0.64+0.36=1.00$
Σ	1	

- حساب الاحتمالات:- $P(X=1)$، $P(X\le 1)$، $P(X=1.5)$، $P(X\le 1.5)$

$$P(X=1)=f(1)=0.48$$
$$P(X\le 1)=F(1)=0.64$$
$$P(X=1.5)=f(1.5)=0$$
$$P(X\le 1.5)=F(1.5)=F(1)=0.64$$

- تحديد قيمة الوسيط، والمنوال.

الوسيط:- رتبة الوسيط هو 0.50 ، إذا الوسيط M هو القيمة التي تحقق الاحتمال:
$P(X\le M)=F(M)=0.50$، وهذا الاحتمال يقع بين القيمتين $(1,0)$ كما هو مبين بالرسم التالي:

	x_i	$F(x_i)$	
	0	0.16	
M			$F(M)=0.50$
	1	0.64	
	2	1.00	

إذا الوسيط قيمته هي:

$$M=0+\frac{0.5-0.16}{0.64-0.16}\times(1-0)=0.71$$

حساب المنوال:

المنوال $Mode$ = القيمة x_i المناظرة لأكبر قيمة احتمالية.

إذا المنوال هو: $Mode = 1$ حيث أنه يناظر أكبر قيمة احتمالية هي: $f(1) = 0.48$.

2-1-2 الوسط الحسابي والتباين للمتغير العشوائي المنفصل

أ- يرمز للوسط الحسابي للمتغير العشوائي بالرمز μ (ميو)، ويحسب بتطبيق المعادلة التالية:

$$\boxed{\mu = \sum x_i \; f(x_i)}$$

ب- وأما التباين ويرمز له بالرمز σ^2 (سيجما)، فيحسب بتطبيق المعادلة التالية:

$$\boxed{\begin{aligned}\sigma^2 &= \sum (x_i - \mu)^2 \; f(x_i) \\ &= \sum x_i^2 \; f(x_i) - \mu^2\end{aligned}}$$

مثال (2)

في المثال السابق احسب ما يلي:

أ- الوسط الحسابي لعدد العبوات المشتراة من النوع الأمريكي:

ب- احسب الانحراف المعياري لعدد العبوات المشتراة من النوع الأمريكي.

ت- أوجد معامل الاختلاف النسبي:

الحل

أ- الوسط الحسابي لعدد العبوات من النوع الأمريكي:

لحساب الوسط الحسابي والانحراف المعياري يتم استخدام المعادلة (3)، (4) وهذا يتطلب تكوين جدول يشمل المجاميع التالية: $\sum x_i^2 \; f(x_i)$, $\sum x_i \; f(x_i)$، وذلك كما يلي:

x_i	$f(x_i)$	$x_i f(x_i)$	$x_i^2 f(x_i)$
0	0.16	0	0
1	0.48	0.48	0.48
2	0.36	0.72	1.44
Σ	1	1.20	1.92

إذا الوسط الحسابي هو: $\mu = \sum x_i \ f(x_i) = 1.20$

ب- ولحساب الانحراف المعياري يجب أولا حساب التباين وهو:

$$\sigma^2 = \sum x_i^2 \ f(x_i) - \mu^2 = 1.92 - (1.20)^2 = 0.48$$

إذا الانحراف المعياري قيمته هي:

$$\sigma = \sqrt{\sigma^2} = \sqrt{0.48} = 0.693$$

ت- معامل الاختلاف النسبي هو:

$$C.V = \frac{\sigma}{\mu} \times 100 = \frac{0.693}{1.2} \times 100 = 57.7$$

تمرين للحل:-فيما يلي التوزيع الاحتمالي لعدد الوحدات التي تستهلكها الأسرة من أحد مساحيق النظافة خلال الشهر X، $X: \{x = 0,1,2,3,4,5\}$

x (عدد الوحدات التي تستهلكها الأسرة)	0	1	2	3	4	5
$f(x)$	0.15	0.30	0.25	0.23	0.05	0.02

والمطلوب:

1- حدد نوع هذا المتغير (عدد الوحدات التي تستهلكها الأسرة)

2- احسب الوسط والوسيط والمنوال والانحراف المعياري لعدد الوحدات المستهلكة.

3- كون جدول التوزيع التجميعي $F(x)$ ثم أوجد الآتي:

أ- نسبة الأسر التي يقل استهلاكها عن وحدتين

ب- نسبة الأسر التي يزيد استهلاكها عن 3 وحدات

ت- إذا كان لدينا 500 أسرة، فما هو عدد الأسر المتوقع أن يكون استهلاكها على الأقل 3 وحدات؟

4- احسب معامل الالتواء، وكذلك معامل الاختلاف النسبي، وعلق على النتائج.

2-2 التوزيعات الاحتمالية المنفصلة الخاصة

في كثير من النواحي التطبيقية، تتبع بعض الظواهر توزيعات احتمالية خاصة، وهي التوزيعات التي يمكن حساب احتمالات قيم المتغير عن طريق معادلة رياضية، تسمى بدالة الاحتمال $f(x)$ ، وهذه المعادلة لها معالم معينة، تسمى بمعالم المجتمع الذي ينسب له هذا التوزيع، وهذه المعالم ما هي إلا حقائق ثابتة مجهولة، وهي الأساس في حساب القيم الاحتمالية للتوزيع الاحتمالي للمجتمع محل الدراسة.

ومن أهم التوزيعات التي سيتم دراستها في هذا المقرر، توزيع ثنائي الحدين، والتوزيع البواسون.

2-2-1 التوزيع ثنائي الحدين The Binomial Distribution

يستخدم هذا التوزيع في الحالات التي يكون للظاهرة محل الدراسة نتيجتان فقط متنافيتان، النتيجة محل الاهتمام وتسمى بحالة النجاح، والأخرى تسمى بحالة الفشل، ومن أمثلة ذلك:

- عند إعطاء مريض نوع معين من الأدوية، لها نتيجتان: (استجابة للدواء، أو عدم استجابة)
- عند فحص عبوة بداخلها نوع معين من الفاكهة، لها نتيجتان (الوحدة إما أن تكون سليمة، أو تكون معيبة)
- عند إلقاء قطعة عملة، لها نتيجتان (ظهور الوجه الذي يحمل الصورة، أو الوجه الذي يحمل الكتابة)
- نتيجة الطالب في الاختبار (نجاح، رسوب)
- استخدام المزارع لبرنامج معين في الزراعة (يستخدم، أو لا يستخدم).

شكل التوزيع الاحتمالي ثنائي الحدين

إذا كررت محاولة n من المرات، بحيث أن كل محاولة لها نتيجتان فقط متنافيتان هما:

- النتيجة محل الاهتمام " حالة نجاح " وتتم باحتمال ثابت في كل محاولة هو p
- النتيجة الأخرى " حالة فشل " وتتم باحتمال ثابت أيضا هو $q = 1 - p$

وبافتراض أن هذه المحاولات مستقلة، بمعنى أن نتيجة كل محاولة ليس لها علاقة بنتيجة المحاولة الأخرى، وإذا كان المتغير العشوائي X يعبر عن عدد حالات النجاح "عدد النتائج محل الاهتمام" في الـ n محاولة، فإن مدي المتغير العشوائي X والذي يعبر عن عدد حالات النجاح هو: $X : \{x = 0,1,2,...,n\}$ ، ومن ثم يحسب الاحتمال $P(X = x) = f(x)$ بتطبيق المعادلة التالية:

$$\boxed{f(x) = \binom{n}{x} p^x q^{n-x} \quad , \quad x = 0,1,2,\ldots n}$$

حيث أن $\binom{n}{x}$ هي عدد طرق اختيار x من n مع إهمال الترتيب، وتحسب كما يلي:

$$\boxed{\binom{n}{x} = \frac{n(n-1)(n-2)\ldots(n-x+1)}{x(x-1)(x-2)\ldots 3\times 2\times 1}}$$

$$\binom{7}{3} = \frac{7\times 6\times 5}{3\times 2\times 1} = 35 \;=\; \binom{7}{4}$$

$$\binom{7}{0} = \binom{7}{7} = 1$$

مثال (3)

إذا كان من المعلوم أن نسبة الشفاء من مرض معين باستخدام نوع معين من العقاقير الطبية هو 0.60، إذا تناول هذا العقار 5 مصابين بهذا المرض. إذا عرف المتغير العشوائي X بأنه عدد الذين المستجيبين (حالات الشفاء) لهذا العقار.

المطلوب:

أ- ما هو نوع المتغير؟

ب- اكتب شكل دالة الاحتمال $f(x)$ لهذا المتغير.

ت- احسب الاحتمالات التالية:

- ما احتمال استجابة 3 مرضى لهذا العقار؟
- ما هو احتمال استجابة مريض واحد على الأقل؟
- ما هو احتمال استجابة 2 مرضى على الأكثر؟

ث- احسب الوسط الحسابي، والانحراف المعياري لعدد حالات الاستجابة.

ج- حدد شكل التوزيع.

الحل:

أ- عدد حالات الاستجابة X متغير كميمنفصل ، ومدى هذا المتغير في هذه الحالة هو: $X:\{x=0,1,2,3,4,5\}$.

ب- شكل دالة الاحتمال: $n=5$ ، $p=0.60$ ، $q=1-p=0.40$ إذا:

$$f(x) = \binom{n}{x}(p)^x(q)^{n-x}$$

$$= \binom{5}{x}(0.6)^x(0.4)^{5-x} \quad , \quad x = 0,1,2,3,4,5$$

ت- حساب الاحتمالات:

- حساب احتمال استجابة 3 مرضى لهذا الدواء: $P(x=3) = f(3)$

$$f(3) = \binom{5}{3}(0.6)^3(0.4)^{5-3} = \frac{5\times4\times3}{3\times2\times1}\times0.216\times0.16 = 10\times0.03456$$
$$= 0.3456$$

- حساب احتمال استجابة مريض واحد على الأقل: $P(x\geq1)$

$$P(x\geq1) = f(1)+f(2)+f(3)+f(4)+f(5) = 1-f(0)$$
$$= 1-\left[\binom{5}{0}(0.6)^0(0.4)^5\right] = 1-1\times1\times0.01024 = 0.98976$$

- حساب احتمال استجابة 2 مرضى على الأكثر: $P(x\leq2)$:

$$P(x\leq2) = f(2)+f(1)+f(0)$$

$$= \binom{5}{2}(0.6)^2(0.4)^3+\binom{5}{1}(0.6)^1(0.4)^4+\binom{5}{0}(0.6)^0(0.4)^5$$
$$= \frac{5\times4}{2\times1}(0.36)(0.064)+\frac{5}{1}(0.6)(0.0256)+1(1)(0.01024$$
$$= 0.2304+0.0768+0.01024 = 0.31744$$

ث- حساب الوسط الحسابي، والانحراف المعياري لعدد حالات الاستجابة:

- الوسط الحسابي (μ) في حالة التوزيع ثنائي الحدين يحسب بتطبيق المعادلة (8-3)، وباستخدام العمليات الرياضية يمكن الوصول إلى النتيجة التالية:

$$\boxed{\mu = \sum x\ f(x) = np}$$

إذا الوسط الحسابي هو:

$$\mu = np = 5(0.60) = 3$$

- الانحراف المعياري هو الجذر التربيعي الموجب للتباين، ولحساب التباين في التوزيع ثنائي الحدين يتم تطبيق المعادلة (8-4)، ومنها يمكن التوصل إلى الصورة التالية:

$$\boxed{\sigma^2 = npq}$$

إذا تباين عدد حالات الاستجابة هو:

$$\sigma^2 = npq$$
$$= 5(0.60)(0.40) = 1.2$$

ومن ثم يأخذ الانحراف المعياري الصورة التالية:

$$\sigma = \sqrt{npq}$$
$$= \sqrt{1.2} = 1.095$$

ويمكن حساب معامل الاختلاف النسبي، بتطبيق المعادلة التالية:

$$V.C = \frac{\sigma}{\mu} \times 100 = \frac{1.095}{3} \times 100 = 36.5\%$$

ج- تحديد شكل التوزيع:

يتحدد شكل التوزيع ثنائي الحدين وفقا لقيمة احتمال النجاح p كما يلي:

إذا كان $p = 0.5$ فإن التوزيع الاحتمالي ثنائي الحدين يكون متماثل.

إذا كان $p < 0.5$ فإن التوزيع الاحتمالي ثنائي الحدين يكون موجب الالتواء.

إذا كان $p > 0.5$ فإن التوزيع الاحتمالي ثنائي الحدين يكون سالب الالتواء.

وحيث أن $p = 0.6 > 0.5$ فإن توزيع عدد حالات الاستجابة سالب الالتواء.

2-2-2 التوزيع البواسوني Poisson Distribution

يكثر استخدام هذا التوزيع في الحالات التي تقع فيها الأحداث وفقا لمعدلات زمنية، وكذلك في حالة الأحداث نادرة الوقوع، ومن أمثلة ذلك:

- عدد الوحدات التي تستهلكها الأسرة من سلعة معينة خلال الشهر. $X : \{x = 0,1,2,...\}$
- عدد مرات ري نوع معين من المحاصيل الزراعية خلال الموسم. $X : \{x = 0,1,2,...\}$
- عدد العملاء الذين يتم خدمتهم البنكية كل 10 دقائق. $X : \{x = 0,1,2,...\}$
- عدد مرات زيارة المريض للطبيب كل سنة. $X : \{x = 0,1,2,...\}$
- عدد مرات تناول الأسرة للحوم الحمراء خلال الأسبوع. $X : \{x = 0,1,2,...\}$
- عدد أخطاء الطباعة لكل صفحة من صفحات الكتاب. $X : \{x = 0,1,2,...\}$

وهكذا الأمثلة كثيرة

شكل التوزيع الاحتمالي البواسوني

إذا كان متوسط عدد مرات وقوع حادث وفقا لمعدل زمني معين هو μ، وكان المتغير العشوائي X يعبر عن عدد مرات وقوع الحادث وفقا لهذا المعدل، فإن مدي المتغير العشوائي X هو: $X : \{x = 0,1,2,...\}$، وهذا المدى

عبارة عن فئة مفتوحة من اليمين، فإن الاحتمال $P(X=x)=f(x)$ والذي يعبر عن احتمال وقوع الحادث عدد x من المرات وفقا لهذا المعدل، يحسب بتطبيق المعادلة التالية:

$$\boxed{f(x)=\frac{e^{-\mu}\mu^{x}}{x!}\ ,\ \ x=0,1,2,\dots}$$

حيث أن e هي أساس اللوغاريتم الطبيعي، وتوجد في بعض الآلات الحاسبة، وقيمتها هي: $e=2.718$ تقريبا، ويمكن حساب قيمتها باستخدام الآلة الحاسبة باتباع الخطوات التالية من الشمال إلى اليمين:

مثلا إيجاد $e^{-1.5}$

SHIFT e^{x} (-) 1.5 = 0.22323016 — النتيجة

وأما $x!$ فتسمى "مضروب العدد x " ويساوي: $x!=x(x-1)(x-2)\dots3\times2\times1$

مثال (4)

إذا كان من المعلوم أن عدد الوحدات التي تستهلكها الأسرة من سلعة معينة خلال الشهر تتبع توزيع بواسون بمتوسط 3 وحدات شهريا، إذا عرف المتغير العشوائي X بأنه عدد الوحدات التي تستهلكها الأسرة خلال الشهر من هذه السلعة.

المطلوب:

أ- ما هو نوع المتغير العشوائي؟

ب- اكتب شكل دالة الاحتمال $f(x)$ لهذا المتغير.

ح- احسب الاحتمالات التالية:

- احتمال أن الأسرة تستهلك وحدتين خلال الشهر؟
- احتمال أن أسرة ما تستهلك وحدة واحد على الأقل خلال الشهر؟
- احتمال أن أسرة ما تستهلك 3 وحدات على الأكثر خلال الشهر؟

خ- احسب الوسط الحسابي، والانحراف المعياري لعدد الوحدات المستهلكة.

د- حدد شكل التوزيع.

الحل:

أ- عدد الوحدات التي تستهلكها الأسرة X متغير كميمنفصل ، ومدى هذا المتغير في هذه الحالة هو: $X:\{x=0,1,2,3,...\}$.

ب- شكل دالة الاحتمال:

بما أن متوسط عدد الوحدات التي تستهلكها الأسرة خلال الشهر هو: $\mu=3$، إذا دالة الاحتمال هي:

$$f(x)=\frac{e^{-\mu}\mu^{x}}{x!}$$

$$=\frac{e^{-3}3^{x}}{x!}\quad,\quad x=0,1,2,...$$

ح- حساب الاحتمالات:

- حساب احتمال أن أسرة ما تستهلك وحدتين خلال الشهر، $f(2)$

$$f(2)=\frac{e^{-3}3^{2}}{2!}=\frac{0.0498(9)}{2\times1}=0.22404$$

- احتمال أن أسرة ما تستهلك وحدة واحد على الأقل خلال الشهر هو:

$$P(X\geq1)=f(1)+f(2)+....$$

$$=1-f(0)=1-\frac{e^{-3}3^{0}}{0!}=\frac{0.0498}{1}=1-0.0498=0.9502$$

- احتمال أن أسرة ما تستهلك 3 وحدات على الأكثر خلال الشهر هو:

$$P(X\leq3)=f(3)+f(2)+f(1)+f(0)$$

$$=\frac{e^{-3}3^{3}}{3!}+\frac{e^{-3}3^{2}}{2!}+\frac{e^{-3}3^{1}}{1!}+\frac{e^{-3}3^{0}}{0!}\frac{0.0498}{1}$$

$$=0.0498\left(\frac{27}{6}+\frac{9}{2}+\frac{3}{1}+\frac{1}{1}\right)=0.0498(13)=0.6474$$

خ- حساب الوسط الحسابي، والانحراف المعياري لعدد حالات الاستجابة:

- الوسط الحسابي (μ) في حالة التوزيع البواسون هو معلمة معطاة هي:

$$\mu=3$$

في هذا التوزيع، فإن التباين يساوي الوسط الحسابي:

أي أن: $\sigma^{2}=\mu=3$

ومن ثم يكون الانحراف المعياري هو:

$$\sigma=\sqrt{\mu}=\sqrt{3}=1.732$$

ويمكن حساب معامل الاختلاف النسبي، بتطبيق المعادلة التي سبق استخدامها في الفصل السابق، وهو:

$$V.C = \frac{\sigma}{\mu} \times 100 = \frac{1.732}{3} \times 100 = 57.7\%$$

د- تحديد شكل التوزيع:

دائما التوزيع البواسون موجب الالتواء.

2-2-3- التوزيع الهندسي الزائد: Distribution hyper géométrique

(أ) استنتاج صيغة قانون التوزيع الهندسي الزائد:

مثال 5: صندوق به 6 كريات منها 4 بيضاء و2 حمراء. نسحب بدون إرجاع 3 كريات. احسب احتمال الحصول على كرتين بيضاوين، 3 كريات بيضاء، كرية واحدة بيضاء، ولا كرية بيضاء.

نفترض أننا نسحب من صندوق كريات **بدون إرجاع** عددها n، إذا كان الصندوق يحتوي على N كرية منها b بيضاء وr حمراء $(N = b + r)$ فإن احتمال الحصول على عدد معين $x \leq b$ من الكريات البيضاء يمكن أن نحصل عليه من خلال القانون الكلاسيكي للاحتمالات (ع الحالات الملائمة / ع الحالات الممكنة) وذلك باستخدام التوفيقات:

$$\boxed{P(X = x) = \frac{C_b^x \cdot C_r^{n-x}}{C_N^n}}$$

تسمى هذه الصيغة: قانون التوزيع الهندسي الزائد ونكتب $X \sim H(N, b, p)$ حيث:

$p = b/N$ و $q = 1-p = r/N$

يمكن الآن الإجابة على أسئلة المثال كما يلي:

$P(X=2) = C_4^2 \cdot C_2^1 / C_6^3 = 12/20$, $P(x = 3) = C_4^3 \cdot C_2^0 / C_6^3 = 1/5$, …

(ب) خصائص التوزيع الهندسي الزائد

$$\mu = np, \qquad \sigma^2 = npq\left(\frac{N-n}{N-1}\right)$$

ملاحظة: للتوزيع الهندسي الزائد علاقة بالتوزيع الثنائي سنذكرها عندما نتطرق لهذا الأخير.

2-2-4- التوزيع الهندسي الزائد المتعدد: Distribution Multi-hypergéométrique

أ. استنتاج صيغة قانون التوزيع الهندسي الزائد المتعدد

يمكن بسهولة تعميم القانون السابق على حالة وجود أكثر من صنفين (k صنف)، حيث من كل صنف لدينا Ni كرية، $(\Sigma Ni = N)$ ، ولحساب احتمال نتيجة معينة؛ مثلا 2 كريات بيضاء $(X_1 = 2)$ ، 5 حمراء، 1 زرقاء، . . . يمكن حساب عدد الحالات الملائمة والممكنة من خلال التوفيقات كما يلي:

$$P(X_1 = x_1, X_2 = x_2, \ldots X_k = x_k) = \frac{C_{N_1}^{x_1} C_{N_2}^{x_2} \ldots C_{N_k}^{x_k}}{C_N^n} \qquad \sum_1^k N_i = N, \sum_1^k x_i = n$$

ب. خصائص التوزيع الهندسي الزائد المتعدد

$$E(X_i) = n\frac{N_i}{N} = np_i$$

ملاحظة : للتوزيع الهندسي الزائد علاقة بالتوزيع الثنائي سنذكرها عندما نتطرق لهذا الأخير.

2-2-5-توزيع برنولي Distribution de Bernoulli

أ. استنتاج صيغة قانون برنولي

نقول عن تجربة أنها "برنولية" إذا كانت تحتمل نتيجتين (حدثين) متنافيتين A و'A. نسمي A نجاح و'A فشل.

نعتبر المتغيرة ع X التي تمثل **عدد مرات النجاح**، تأخذ X القيمة 1 عند تحقق الحدث A و0 في الحالة المعاكسة.

نرمز عادة ب p "احتمال النجاح" لاحتمال تحقق الحدث A وp - q = 1 احتمال الحدث المعاكس (الفشل). يعين توزيع برنولي كما يلي :

$$\boxed{P(X=1) = p, \qquad P(X=0) = q, \qquad X = 0, 1.}$$

ونكتب **X ~ B(1, p)**

ب. خصائص توزيع برنولي

$E(X) = \Sigma x_i p_i$ $\qquad E(X) = p. \quad <= \quad = 1.p + 0.q = p$

$V(X) = E(X^2) – E(X)^2 = (1^2.p + 0^2.q) – p^2 = p – p^2 = p(1- p) = pq$

$=> \mathbf{V(X) = qp.}$

الدالة المتجددة للعزوم $M(t) = E(e^{xt}) = e^{0t}q + e^{1t}p => \mathbf{M(t) = q + pe^t}.$

معامل التماثل $\mu_3 = \sum (x-\mu)^3 p(x) = (0-p)^3 q + (1-p)^3 p = pq^3 - qp^3 = qp(q^2 - p^2)$

$$\alpha_3 = \frac{\mu_3}{\sigma^3} = \frac{qp(q^2-p^2)}{qp\sqrt{qp}} \qquad = \frac{q^2-p^2}{\sqrt{qp}}$$

2–2–6- التوزيع الثنائي السالب (باسكال) Distribution binomiale négative

أ. استنتاج صيغة قانون التوزيع الثنائي السالب:

مثال: نرمي قطعة نقود إلى غاية الحصول على 3 مرات صورة (متتالية أو لا). أحسب احتمال أن نحصل على ذلك بعد 5 رميات، 4 رميات، 3 رميات، توقع عدد الرميات اللازمة وأحسب التباين.

من جديد ليكن لدينا تجربة برنولية (نتيجتين نجاح وفشل) مكررة، لكن هذه المرة إلى غاية الحصول على عدد معين (r) من النجاحات. X في هذه الحالة هي عدد مرات تكرار التجربة إلى غاية الحصول على r نجاح.

كيف يحسب الاحتمال ؟ نعلم أن تحقق النجاح r مرة احتماله p^r واحتمال الفشل x-r مرة يساوي q^{x-r} . إذا الاحتمال المطلوب يتضمن جداء هذين الاحتمالين $p^r q^{x-r}$. لكن هناك عددا من الطرق الملائمة لتحقيق r نجاح من بين X تجربة مع العلم أن آخر تجربة هي نجاح. هذا العدد يساوي إذا عدد الطرق الملائمة لاختيار r-1 نجاح من بين x-1 تجربة C^{r-1}_{x-1} (التجربة الأخيرة معلومة النتيجة).

$$\boxed{P(X=x)=C_{x-1}^{r-1}p^r q^{x-r}, \qquad X=r, r+1, r+2, \ldots +\infty, \qquad r=1,2,3,\ldots,+\infty}$$

يسمى هذا التوزيع توزيع باسكال أو الثنائي السالب ونكتب: X~B (N, r, p)

يمكن إذا الإجابة على أسئلة المثال السابق بما يلي:

$$P(X=5)=C^{3-1}_{5-1}\ p^3 q^{5-3}=C^2_4\ (½)^3 (½)^2 = 6\ (1/8)\ (1/4)=9/32$$

$$\mu = r/p = 3/(1/2) = 6\ ,\ \sigma^2 = rq/p^2 = 3\ (1/2)\ /\ (1/2)^2 = 12/2 = 6$$

ب. خصائص التوزيع الثنائي السالب

$$\mu=\frac{r}{p}, \qquad \sigma^2=\frac{rq}{p^2}, \qquad M(t)=p\frac{e^t}{\left(1-qe^t\right)^r}$$

$$\alpha_3=\frac{q+1}{\sqrt{q}}\ ,\ \ \alpha_4=3+\frac{(q+2)^2+3(nq-1)}{nq}$$

2–2–7– التوزيع الهندسي Distribution géométrique

أ. استنتاج صيغة قانون التوزيع الهندسي

نرمي قطعة نقدية إلى أن نحصل على صورة. احتمال أن يتطلب ذلك 4 رميات هو: = P(X= 4) P(PPPF)

نعود من جديد إلى التجربة البرنولية وهذه المرة نكرر التجربة إلى غاية الحصول على النتيجة أو الحدث المطلوب (نجاح مرة واحدة). المتغيرة العشوائية X التي تمثل عدد مرات تكرار التجربة (بما فيها المرة التي حصل فيها النجاح) تتبع التوزيع الهندسي.

إذا رمزنا لاحتمال النجاح ب p ولاحتمال الفشل ب q فإن الاحتمال يمكن كتابته كما يلي: $P(X=4) = q^3p$

وبصفة عامة فإن احتمال أي قيمة لـX يعبر عنه كما يلي :

$$\boxed{P(X=x) = q^{x-1}p \qquad , \qquad X = 1, 2, 3, \dots}$$

ب. خصائص التوزيع الهندسي

$$\mu = \frac{1}{p}, \qquad \sigma^2 = \frac{q}{p^2}, \qquad M(t) = p\frac{e^t}{(1-qe^t)} \qquad \alpha_3 = \frac{q+1}{\sqrt{q}}, \qquad \alpha_4 = 12 + \frac{p^2}{q}$$

ملاحظة: التوزيع الهندسي ما هو إلا حالة خاصة من توزيع باسكال حيث $r = 1$

2-2-8- التوزيع المتعدد Distribution multinomiale

أ. استنتاج صيغة قانون التوزيع الثنائي المتعدد

مثال: نرمي قطعة نرد 4 مرات. أوجد احتمال الحصول على مرتين الرقم 6 ومرتين الرقم 1.

التوزيع المتعدد هو تعميم للتوزيع الثنائي، فبينما الأول يستعمل في حالة تجربة تقبل نتيجتين فقط، يستعمل التوزيع المتعدد للحالة العامة حيث يكون للتجربة عدد k من النتائج الممكنة. مع استقلالية التجارب عن بعضها. نرمز لهذه النتائج ب $A_1, A_2, \dots A_k$ ولاحتمالاتها ب $p_1, p_2, p_3, \dots p_k$. بما إن الأحداث (النتائج) Ai متنافية فإن:

$$p_1 + p_2 + p_3 + \dots + p_k = 1$$

إذا كررنا هذه التجربة متعددة النتائج عدد n من المرات فسيكون لدينا لكل حدث (نتيجة) متغيرة عشوائية تمثل عدد مرات وقوعه. نرمز لهذه المتغيرات ب$X_1, X_2, \dots X_k$ حيث $X_1 + X_2 + \dots + X_k = n$.

يحسب احتمال الحدث المركب: $X_1 = x_1, X_2 = x_2, \dots, X_k = x_k$ كما يلي :

$$\boxed{P(X_1 = x_1, X_2 = x_2, \dots, X_k = x_k) = \frac{n!}{x_1!x_2!\dots x_k!} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}}$$

ب. خصائص التوزيع المتعدد

$$E(X_1) = np_1, E(X_2) = np_2, \dots, \qquad E(X_k) = np_k$$
$$V(X_1) = np_1q_1, V(X_2) = np_2q_2, \dots \qquad V(X_k) = np_kq_k$$

ج. العلاقة مع التوزيع الهندسي الزائد المتعدد

في التوزيع الهندسي الزائد المتعدد، عندما $N \to \infty, Ni \to \infty, Ni/N \to pi$ ؛ يستخدم التوزيع المتعدد لحساب الاحتمالات.

مثال: إذا رمينا قطعة نرد42 مرة، أحسب احتمال أن يظهر كل رقم عدد من المرات يتناسب مع الرقم ذاته (الرقم 1 يظهر مرتين، الرقم 2 يظهر 4 مرات، الرقم 3 يظهر 6 مرات وهكذا).

$$P(X_1 = 2,\ X_2 = 4, \ldots X_6 = 12) = \frac{42!}{2!\ 4!\ 6!\ \ldots\ 12!}(1/6)^2(1/6)^4 \ldots (1/6)^{12}$$

مثال: نسحب من صندوق به 5 كريات مرقمة من 1 إلى 5 سبع مرات على التوالي كرية ثم نرجعها إلى الصندوق. أوجد احتمال: 3 كريات ذات رقم 1، كريتين ذات رقم 2 وكريتين ذات رقم 4.

- **خلاصة**

الجدول الملحق يلخص أهم النقاط حول التوزيعات المتقطعة الشهيرة.

توزيع	متى يستخدم	القيم الممكنة للمتغيرة	الاحتمال	التوقع والتباين
الهندسي الزائد **X~ H(N, b,p)**	سحب بدون إرجاع. كريات من صنفين.	X $=\{0,1,2,\ldots,b\}$, b $\le$ b + r = N	$P(X=x)=\dfrac{C_b^x \cdot C_r^{n-x}}{C_N^n}$	$\mu = np,$ $\sigma^2 = npq\left(\dfrac{N-n}{N-1}\right)$ q = r/N و p = b/N n عدد الكريات المسحوبة N العدد الكلي للكريات b عدد الكريات البيضاء r ع الكريات الحمراء
الهندسي الزائد المتعدد	نفس شروط ت الهندسي الزائد مع وجود أكثر من صنفين من الكريات.	Xi $=\{0,1,2,\ldots,$Ni$\}$, Σxi = n, ΣNi = N	$P(X_1=x_1,X_2=x_2,\ldots X_k=x_k)=$ $=\dfrac{C_{N1}^{x1}C_{N2}^{x2}C_{N3}^{x3}\cdots C_{Nk}^{xk}}{C_N^n}$	$E[X_i] = n\,(N_i/N) = np_i$
برنولي X~B(1, p)	تجربة واحدة (غير مكررة) تقبل نتيجتين.	X = $\{0,\ 1\}$	P(X = 1) = p, P(X = 0) = 1 − p = q	μ= p, σ^2= pq
الثنائي X~B(n, p)	تجارب ثنائية النتيجة، مكررة ومستقلة (p ثابت).	X = $\{0,1,2\ldots,$n$\}$	$P(X = x) = C^x_n\, p^x\, q^{n-x}$ $C_n^x = \dfrac{n!}{x!(n-x)!}$	μ= np, σ^2 = npq

باسكال (الثنائي السالب)	X هي عدد التجارب اللازمة للحصول على عدد r من النجاحات في تجارب برنولية مكررة.	X = {r, r +1, r +2, …, +∞}	$P(X = x) = C^{r-1}_{x-1}\, p^r\, q^{x-r}$	$\mu = r/p$, $\sigma^2 = rq/p^2$
الهندسي	X هي عدد التجارب اللازمة للحصول على النجاح الأول في تجارب برنولية مكررة.	X = {1,2,…,+∞}	$P(X = x) = q^{x-1}p$	$\mu = 1/p$, $\sigma^2 = q/p^2$
التوزيع المتعدد	هو تعميم للتوزيع الثنائي على تجربة مكررة متعددة النتائج.	$\forall i, 0 \le xi \le Ni,$ $\sum_{i=1}^{k} xi = n,$ $\sum_{i=1}^{k} Ni = N$	$P(X_1 = x_1, X_2 = x_2, ..., X_k = x_k) = \dfrac{n!}{x_1!x_2!\ldots x_k!} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k}$	$E(X_k) = np_k$, $V(X_k) = np_kq_k$
بواسون X~P(λ) **λ> 0**	X عدد النجاحات في عدد كبير جدا من التجارب البرنولية (عدد الوحدات التالفة في شحنة). أو أيضا عدد من الأحداث في فترة زمن.	X = {0,1,2,…,+∞}	$P(X = x) = \dfrac{\lambda^x e^{-\lambda}}{x!}$, $P(X = 0) = e^{-\lambda}$, $P(X \ge 1) = 1 - e^{-\lambda}$	$E(x) = V(x) = \lambda$

2-3- المتغيرات العشوائية المستمرة Continuous Random Variables

المتغير العشوائي المستمر، هو الذي يأخذ قيما متصلة، ويأخذ عدد لانهائي من القيم الممكنة له داخل مجاله، فإذا كان X متغير عشوائي مستمر، ويقع في المدى (a,b)، أي أن: $\{X = x : a < x < b\}$، فإن للمتغير

X عدد لانهائي من القيم تقع بين الحدين الأدنى والأعلى *(a,b)*، ومن الأمثلة على المتغيرات الكمية المستمرة ما يلي:

- كمية الألبان التي تنتجها البقرة في اليوم باللتر: $\{X = x : 10 < x < 40\}$
- المساحة المنزرعة بالأعلاف في الجزائر بالألف هكتار $\{X = x : 1000 < x < 15000\}$
- فترة صلاحية حفظ الدجاج المبرد بالأيام، $\{X = x : 1 < x < 5\}$
- وزن الجسم بالكيلوجرام للأعمار من *(40-30)*، $\{X = x : 55 < x < 80\}$

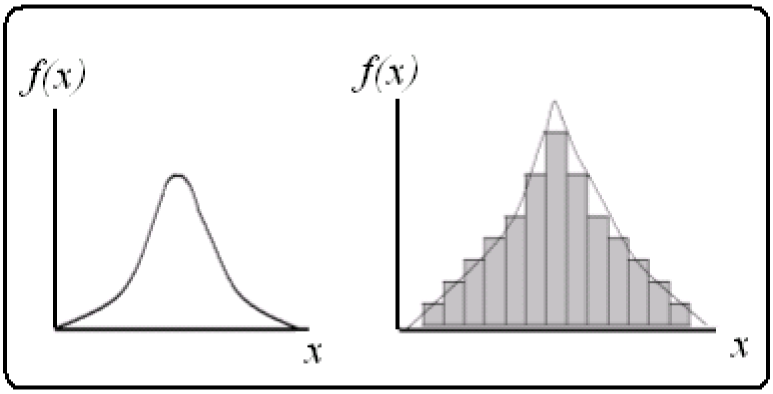
وهكذا الأمثلة على المتغير الكمي المستمر كثيرة.

2–3–1 التوزيع الاحتمالي للمتغير المستمر Continuous Probability

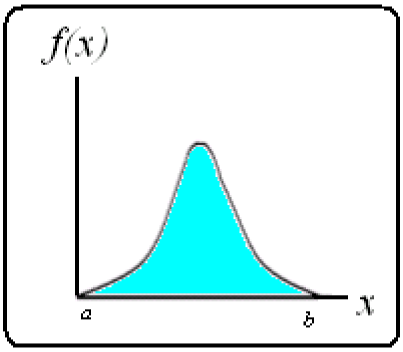
عند تمثيل بيانات المتغير الكمي المستمر في شكل مدرج تكراري النسبي، نجد أن شكل هذا المدرج هو أقرب وصف لمنحنى التوزيع الاحتمالي للمتغير المستمر، وكلما ضاقت الفترات بين مراكز الفئات، يمكن الحصول على رسم دقيق للمنحنى الخاص بدالة احتمال المتغير المستمر، كما هو مبين بالشكل التالي:

شكل (1)

شكل منحنى التوزيع الاحتمالي للمتغير العشوائي المستمر



والمساحة أسفل المنحنى تعبر عن مجموع الاحتمالات الكلية، ولذا تساوي هذه المساحة الواحد الصحيح، وتسمى الدالة *f(x)* بدالة كثافة الاحتمال Probability Distribution Function*(p.d.f)*، وبفرض المتغير العشوائي المستمر يقع في المدى: $X = \{x : a \prec x \prec b\}$، وأن منحنى هذه الدالة يأخذ الصورة التالية:



فإن من خصائص دالة كثافة الاحتمال $f(x)$ ما يلي:

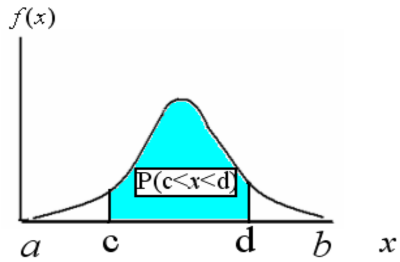
1- الدالة $f(x)$ موجبة داخل المدى (a,b) أي أن: $f(x) \succ 0$ ، $x \in (a,b)$

2- التكامل على حدود المتغير من الحد الأدنى a حتى الحد الأعلى b يعبر عن مجموع الاحتمالات الكلية، لذا يساوي الواحد الصحيح ، أي أن:

$$\boxed{\int_{x=a}^{x=b} f(x)\ dx = 1}$$

حيث أن الشكل الرياضي أعلاه يسمى بالتكامل المحدد من $x=a$ حتى $x=b$، وهذا يعني إيجاد المساحة أسفل المنحني بين (a,b).

3- لحساب احتمال أن المتغير العشوائي المستمر يقع في المدى (d,c) أي حساب الاحتمال $p(c<x<d)$ ، يجب حساب المساحة أسفل المنحني من $x=c$ حتى $x=d$ كما هي مبينة في الشكل البياني التالي:



ويتم ذلك بإيجاد التكامل المحدد في هذا المدى، كما يلي:

$$\boxed{p(c<x<d) = \int_{x=c}^{x=d} f(x)\ dx = [g(x)]_c^d = g(d) - g(c)}$$

4- في المتغير المستمر، يكون الاحتمال $p(x = value)$ مساويا للصفر، أي أن:

$$\boxed{p(x = value) = 0}$$

ولكي يمكننا حساب الاحتمالات، يجب عرض بعض قواعد التكامل التالية:

جدول (2)

بعض قواعد التكامل

(1) $\int x^n dx = \frac{x^{n+1}}{n+1}$ and $\int (a+bx)^n dx = \frac{(a+bx)^{n+1}}{b(n+1)}$ (2) $\int e^x dx = e^x$ and $\int e^{(a+bx)} dx = \frac{1}{b} e^{(a+bx)}$ (3) $\int \frac{1}{x} dx = \log_e(x)$ and $\int \frac{1}{(a+bx)} dx = \frac{1}{b}\log_e(a+bx)$	integration
(4) $\Gamma(n+1) = \int_0^\infty x^n e^{-x} dx = n! = n(n-1)(n-2)...3\times 2\times 1$	gamma
(5) $I\Gamma(n+1) = \int_0^a x^n e^{-x} dx = n!\left(1 - e^{-a}\sum_{i=0}^{n}\frac{a^i}{i!}\right)$	Incomplete gamma
(6) $B(m+1, n+1) = \int_0^1 x^n (1-x)^m dx = \frac{m!n!}{(m+n+1)!}$	Beta

مثال (6)

إذا كان الإنفاق الشهري للأسرة بالألف دج على المواد الغذائية له دالة كثافة احتمال تأخذ الصورة التالية:

$$f(x) = \begin{cases} cx(10-x) & , 0 < x < 10 \\ 0 & otherwise \end{cases}$$

والمطلوب:

1- حساب قيمة الثابت c

2- احسب احتمال أن إنفاق الأسرة يتراوح ما بين (8,5) ألف دج خلال الشهر.

3- إذا كان لدينا 600 أسرة، فما هو عدد الأسر المتوقع أن يقل إنفاقها عن 3 آلاف خلال الشهر؟

الحل

1- حساب قيمة c

من خصائص دالة كثافة الاحتمال:

$$\int_{x=a}^{x=b} f(x)\ dx = 1$$

إذا

$$\int_{x=0}^{x=10} cx(10-x)\ dx = c\int_{x=0}^{x=10}(10x-x^2)\ dx = c\left[10\left(\frac{x^2}{2}\right)-\frac{x^3}{3}\right]_0^{10}$$

$$= c\left[5x^2-\frac{x^3}{3}\right]_0^{10} = c\left[(5(100)-\frac{(1000)}{3})\right]-0$$

$$= \frac{500}{3}c = 1$$

$$c = 3/500 = 0.006$$

2- حساب أن إنفاق الأسرة يتراوح بين (8,5) ألف دج خلا الشهر هو.

$$p(5<x<8) = \int_{x=5}^{x=8} 0.006x(10-x)\ dx = 0.006\left[5x^2-\frac{x^3}{3}\right]_5^8$$

$$= 0.006\left[\left(5(8)^2-\frac{8^3}{3}\right)-\left(5(5)^2-\frac{5^3}{3}\right)\right] = 0.006[(149.3333)-(83.3333)]$$

$$= 0.006(66) = 0.396$$

3- إذا كان لدينا 600 أسرة، فإن عدد الأسر المتوقع أن يقل إنفاقها عن 3 آلاف خلال الشهر هو:

$$number\ \ of\ \ \ family = 600\ \ p(x<3)$$

$$= 600\int_0^3 0.006x(10-x)dx$$

$$= 3.6\left[5x^2-\frac{x^3}{3}\right]_0^3 = 3.6[45-9]-0 = 129.6 \approx 130$$

حوالي 130 أسرة.

2-3-2 المتوسط والتباين في التوزيع الاحتمالي المستمر

إذا كانت $f(x)$ هي دالة كثافة الاحتمال للمتغير العشوائي x، $a<x<b$ فإن التوقع الرياضي للدالة $h(x)$ تأخذ الصورة التالية:

$$\boxed{E\left(h(x)\right)=\int_a^b h(x)\ dx}$$

ومن ثم يمكن كتابة معادلة الوسط والتباين كما يلي.

$$\boxed{\begin{aligned}&\mu=E(x)=\int_a^b xf(x)dx\\&\sigma^2=E(x^2)-u^2\ ,\ E(x^2)=\int_a^b x^2f(x)\ dx\end{aligned}}$$

تابع مثال (6)

في المثال السابق أوجد المتوسط والانحراف المعياري ومعامل الاختلاف النسبي للإنفاق الشهري.

الحـل

1- المتوسط الحسابي

$$\mu=E(x)=xf(x)dx=\int_0^{10}x\left(0.006x(10-x)\right)=0.006\int_0^{10}(10x^2-x^3)dx$$

$$=0.006\left[10\frac{x^3}{3}-\frac{x^4}{4}\right]_0^{10}=0.006\left[\left(\frac{10000}{3}-\frac{10000}{4}\right)-(0)\right]$$

$$=60\left[\frac{1}{12}\right]=5$$

متوسط إنفاق الأسرة الشهري 5 آلاف دج.

2- الانحراف المعياري

$$\sigma^2=E(x^2)-u^2=E(x^2)-(5)^2$$

$$E(x^2)=\int_a^b x^2f(x)\ dx=0.006\int_0^{10}(10x^3-x^4)dx$$

$$=0.006\left[10\left(\frac{x^4}{4}\right)-\left(\frac{x^5}{5}\right)\right]_0^{10}=0.006\left[\frac{100000}{4}-\frac{100000}{5}\right]-0$$

$$=600\left(\frac{1}{20}\right)=30$$

إذا التباين هو : $\sigma^2 = 30 - 25 = 5$ ، ومن ثم يأخذ الانحراف المعياري القيمة التالية:

$$\sigma = \sqrt{variance} = \sqrt{5} = 2.236$$

3- معامل الاختلاف النسبي

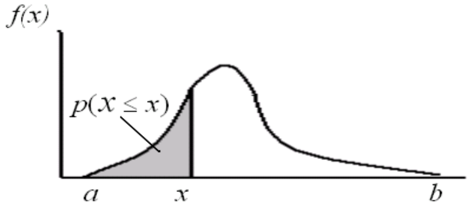
$$C.V = \frac{\sigma}{\mu} \times 100 = \frac{2.236}{5} \times 100 = 44.72\%$$

دالة التوزيع التجميعي Cumulative Distribution Function (C.D.F)

يرمز لهذه الدالة بالرمز $(C.D.F)=F(x)$ وتحسب بإيجاد الاحتمال:

$$\boxed{C.D.F = F(x) = P(X \le x) = \int_a^x f(x)dx}$$

ويمكن توضيحها بيانيا بالرسم التالي:



تابع مثال (6)

في المثال (6) أوجد دالة التوزيع التجميعي $C.D.F$، ثم استخدم هذه الدالة لحساب احتمال أن إنفاق الأسرة يقل عن 5 آلاف جنيه.

الحل

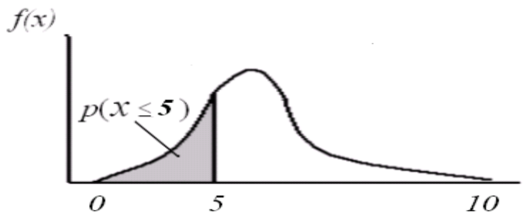
- إيجاد دالة التوزيع التجميعي $C.D.F$

$$F(x) = \int_0^x f(x)\ dx$$

$$= \int_0^x 0.006x(10-x)dx = 0.006\left[10\left(\frac{x^2}{2}\right) - \left(\frac{x^3}{3}\right)\right]_0^x$$

$$= 0.006\left[5x^2 - \left(\frac{x^3}{3}\right)\right]$$

- حساب الاحتمال المطلوب $F(5) = p(x \le 5)$ ، كما هو مبين بالرسم التالي:



ويمكن حساب هذا الاحتمال بالتعويض عن $x=5$ في الدالة $F(x)$ التي تم التوصل إليها، أي أن:

$$F(5)=P(x\leq 5)=$$
$$=0.006\left[5x^2-\frac{x^3}{3}\right]=0.006\left[125-\frac{125}{3}\right]$$
$$=0.006\left(\frac{250}{3}\right)=0.5$$

أي أن 50% من الأسر يقل إنفاقها عن 5 آلاف دج.

خصائص دالة التوزيع التجميعي

1- $F(x)\succ 0$ 2- $F(a)=0$ 3- $F(b)=1$ 4- $p(X\succ x)=1-F(x)$ 5- $f(x)=dF(x)/dx$

2-4- التوزيعات الاحتمالية المستمرة الخاصة Continuous Probability Distributions

هناك بعض التوزيعات الاحتمالية المستمرة الخاصة، ولها دوال كثافة احتمال محددة، وفيما يلي بعض هذه التوزيعات:

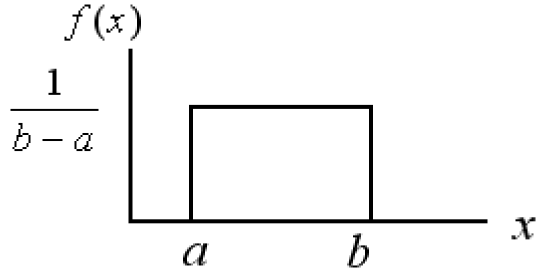
2-4-1 التوزيع المنتظم Uniform distribution

شكل دالة كثافة الاحتمال $p.d.f$

هو توزيع له دالة احتمال ثابتة، ويستخدم في حالة الظواهر التي يمكن أن تحدث بشكل منتظم، فإذا كان المتغير x متغير عشوائي له توزيع منتظم *Uniform*، مداه هو $a<x<b$ فإن دالة كثافة احتماله هي:

$$\boxed{f(x)=\frac{1}{b-a}\ ,\ a<x<b}$$

ويمكن تمثيل هذه الدالة بيانيا كما يلي:



معالم هذا التوزيع

توجد معلمتان لهذا التوزيع هما (b,a)، ولذا يكتب رمز لهذا التوزيع الصورة $x \sim U(a,b)$

خصائص التوزيع المستطيل

الوسط الحسابي μ ، والتباين σ^2 لهذا المتغير هما :

$$\mu = E(x) = \frac{a+b}{2} \quad , \quad \sigma^2 = \frac{(b-a)^2}{12}$$

على الطالب إثبات ذلك:

دالة التوزيع التجميعي $C.D.F$

تأخذ دالة التوزيع التجميعي $F(x)$ الشكل الآتي

$$\boxed{\begin{aligned} F(x) &= p(X \leq x) = \int_a^x f(x)dx = \frac{1}{b-a}\int_a^x dx \\ &= \frac{x-a}{b-a} \end{aligned}}$$

مثال (7)

استورد أحد المراكز التجارية 1500 طن بطاطس، ووضعها في مخزن، وقام ببيعها بكميات متساوية على مدار شهور السنة. إذا كانت الفترة الزمنية للبيع تتبع توزيع منتظم، فأوجد الآتي:

- دالة كثافة الاحتمال المعبرة عن الفترة الزمنية للبيع.
- بعد مرور سبعة أشهر من بداية البيع، ما هي الكمية الموجودة بالمخزن؟

الحـل

- دالة كثافة الاحتمال المعبرة عن الزمن:

بفرض أن المتغير x يعبر عن الفترة الزمنية للبيع مقاسة بالشهر، أي أن $0 < x < 12$، ومن ثم تأخذ دالة كثافة الاحتمال المعبرة عن الزمن الصورة التالية:

$$f(x) = \frac{1}{12-0} = \frac{1}{12} \;,\; 0 < x < 12$$

- حساب الكمية الموجودة بالمخزن بعد سبعة أشهر من بداية البيع.

بفرض أن Q هي كمية البطاطس المستوردة ، تكون الكية المتبقية بالمخزن بعد مرور سبعة أشهر من بداية البيع هي :

$$Q \times p(x > 7) = Q \times (1 - F(7)) = 1500(1 - \frac{7-0}{12-0}) = 625 \;\; Ton$$

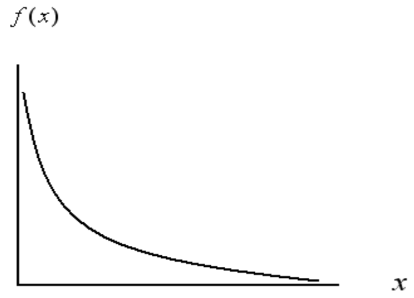
2–4–2 التوزيع الأسي السالب Negative Exponential distribution

شكل دالة كثافة الاحتمال $p.d.f$

إذا كان المتغير x متغير عشوائي له توزيع أسي سالب ، مداه هو $0 < x < \infty$ فإن دالة كثافة احتماله هي:

$$\boxed{f(x) = \theta\, e^{-\theta x} \,,\; 0 < x < \infty \;,\; \theta \succ 0}$$

ويمكن تمثيل هذه الدالة بيانيا كما يلي:



معالم هذا التوزيع

توجد معلمة واحدة هي (θ)

خصائص التوزيع الأسى السالب

الوسط الحسابي μ ، والتباين σ^2 لهذا المتغير هما:

$$\mu = E(x) = \frac{1}{\theta} \ , \quad \sigma^2 = \frac{1}{\theta^2}$$

دالة التوزيع التجميعي $C.D.F$

تأخذ دالة التوزيع التجميعي $F(x)$ الشكل الآتي

$$F(x) = p(X \le x) = \int_0^x f(x)dx = \left(1 - e^{-\theta x}\right)$$

مثال (8)

إذا كانت الفترة الزمنية لإنهاء خدمة العميل في البنك تتبع توزيع أسي بمتوسط 2 دقيقة، فأوجد ما يلي.

- دالة كثافة الاحتمال المعبرة عن الفترة الزمنية لإنهاء خدمة العميل.
- ما احتمال إنهاء خدمة العميل في أقل من دقيقة.

الحل

- دالة كثافة الاحتمال المعبرة عن الزمن:

بفرض أن المتغير x يعبر عن الفترة الزمنية لإنهاء خدمة العميل بالدقيقة، أي أن $0 < x < \infty$ ، فإن المتوسط $1/\theta = 2$ ، ومن ثم تصبح قيمة (θ) هي: $(\theta = 0.5)$ ، وتكتب دالة كثافة الاحتمال المعبرة عن الزمن على الصورة التالية:

$$f(x) = 0.5\ e^{-0.5\ x},\ 0 < x < \infty$$

- حساب احتمال إنهاء خدمة العميل في أقل من دقيقة.

$$p(x \le 1) = (1 - e^{-0.5x}) = (1 - e^{-0.5(1)}) = 0.3935$$

3-4-2 التوزيع الطبيعي The Normal Distribution

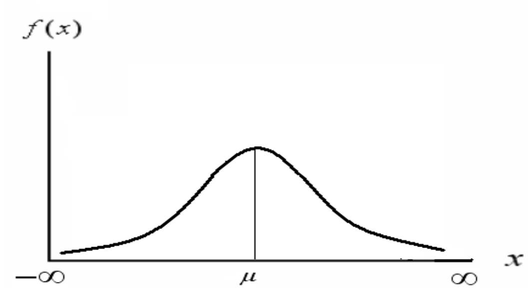
يعتبر هذا التوزيع من أكثر التوزيعات الاحتمالية استخداما في النواحي التطبيقية، ومنها الاستدلال الإحصائي شاملا التقدير، واختبارات الفروض، كما أن معظم التوزيعات يمكن تقريبها إلى هذا التوزيع، وفيما يلي عرض لهذا التوزيع.

شكل دالة كثافة الاحتمال $p.d.f$

إذا كان المتغير x متغير عشوائي له توزيع طبيعي ، مداه هو $-\infty < x < \infty$ فإن دالة كثافة احتماله هي:

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\ e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\ ,\ -\infty<x<\infty\ ,\ \pi=22/7$$

وهذا التوزيع له منحنى متماثل يأخذ الصورة التالية:



فهذا المنحنى متماثل على جانبي الوسط الحسابي μ.

معالم هذا التوزيع

توجد معلمتين لهذا التوزيع هما :

الوسط الحسابي : $E(x)=\mu$ والتباين : $\operatorname{var}(x)=\sigma^2$

ومن ثم يعبر عن توزيع المتغير x بالرموز : $x\sim N(\mu,\sigma^2)$ ويعني ذلك أن المتغير العشوائي x يتبع التوزيع الطبيعي بمتوسط μ ، وتباين σ^2 .

خصائص التوزيع الطبيعي

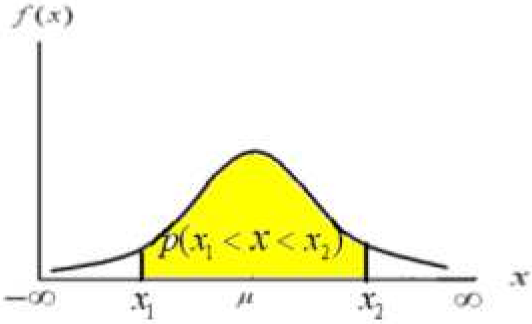
هذا التوزيع من أكثر التوزيعات الاحتمالية استخداما، بل يشتق منه كل التوزيعات الاحتمالية الأخرى المستخدمة في الاستدلال الإحصائي، ومن خصائص هذا التوزيع ما يلي:

1- الوسط الحسابي μ 2- والتباين σ^2

3- منحني هذا التوزيع متماثل على جانبي الوسط μ

كيفية حساب الاحتمالات $p(x_1<X<x_2)$

بفرض أن الاحتمال المطلوب حسابه هو $p(x_1 < X < x_2)$ ، وهذا الاحتمال يحدد بالمساحة التالية:



وحيث أن هذا التوزيع من التوزيعات المستمرة، فإن هذه المساحة (الاحتمال) تحسب بإيجاد التكامل التالي:

$$p(x_1 < X < x_2) = \int_{x_1}^{x_2} f(x)dx = \int_{x_1}^{x_2} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

وهذا التكامل يصعب حسابه، ومن ثم لجأ الإحصائيين إلى عمل تحويلة رياضية Transform، يمكن استخدام توزيعها الاحتمالي في حساب مثل هذه الاحتمالات، وهذه التحويلة هي:

$$z = \left(\frac{x-\mu}{\sigma}\right)$$

ويعرف المتغير الجديد z بالمتغير الطبيعي القياسي Standard Normal Variable ، وهذا المتغير له دالة كثافة احتمال تأخذ الصورة التالية:

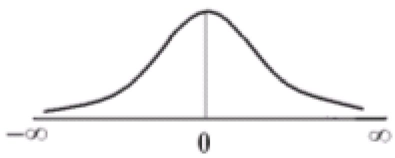
$$\boxed{f(Z) = \frac{1}{\sqrt{2\pi}}\ e^{-\frac{1}{2}z^2}\ ,\ -\infty < z < \infty\ ,\ \pi = 22/7}$$

ومن خصائص هذا التوزيع ما يلي:

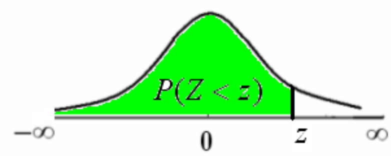
1- متوسطه هو: $E(z) = 0$ 2- تباينه هو: $\text{var}(z) = 1$

ومن ثم يعبر عن توزيع المتغير z بالرموز : $z \sim N(0,1)$ ويعني ذلك أن المتغير العشوائي x يتبع التوزيع الطبيعي القياسي بمتوسط (0) ، وتباين (1) .

3-يأخذ المنحنى الشكل الناقوس المتماثل على جانبي الصفر:



وصمم الإحصائيين جداول إحصائية لحساب دالة التوزيع التجميعي: $F(z)=P(Z<z)$، كما هو مبين بالرسم التالي:

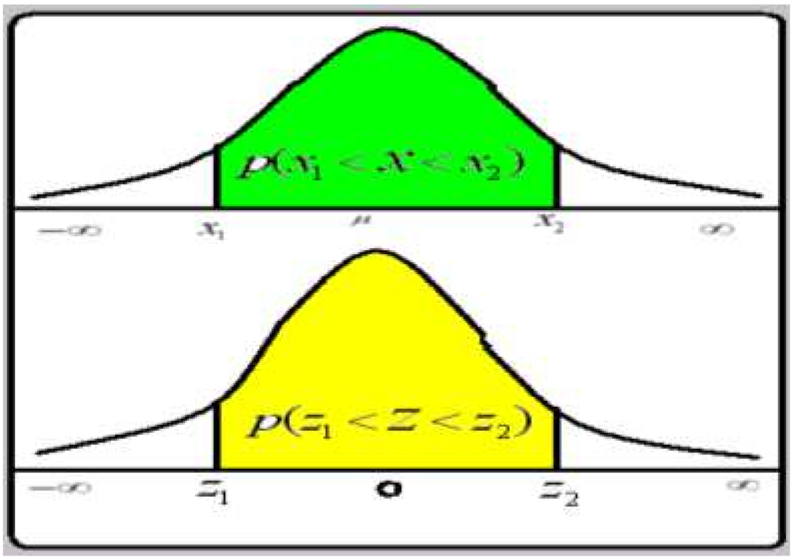


ونعود الآن إلى خطوات حساب الاحتمال $p(x_1<X<x_2)$ باستخدام التحويلة $z=(x-\mu)/\sigma$:

1- يتم تحويل القيم الطبيعية (x_1,x_2) إلى قيم طبيعية قياسية:

$$z_1=(x_1-\mu)/\sigma \;,\; z_2=(x_2-\mu)/\sigma$$

2- ومن ثم يكون الاحتمال: $p(x_1<X<x_2)=p(z_1<Z<z_2)$:



3- تستخدم جداول التوزيع الطبيعي القياسي، والذي يعطي المساحة الخاصة بالاحتمال

$$F(z)=P(Z<z)$$

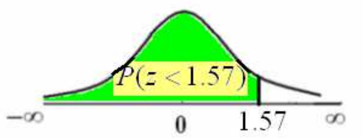
4- طريقة استخدام جدول التوزيع الطبيعي القياسي في حساب الاحتمالات

أوجد الاحتمالات التالية:

أ- $P(z<1.57)$ ب- $P(z<-2.33)$ ج- $P(z>1.96)$ د- $P(-2.01<z<1.28)$

الحل

أ- تحدد المساحة المعبرة عن الاحتمال $P(z<1.57)=F(1.57)$ أسفل المنحنى كما يلي



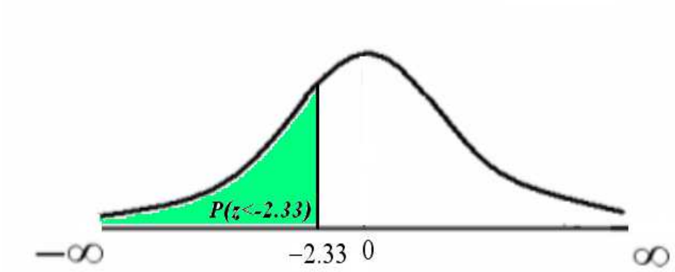
ويتم استخدام الجدول كما هو مبين :

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
⋮										
1.00										
1.10										
1.20										
1.30										
1.40										
1.50								0.9418		
⋮										

ويكون الاحتمال المطلوب هو: $P(z<1.57)=F(1.57)=0.9418$

ب- المساحة أسفل المنحنى المعبرة عن الاحتمال $P(z<-2.33)=F(-2.33)$ موضحة كالتالي:

P(z<-2.33)

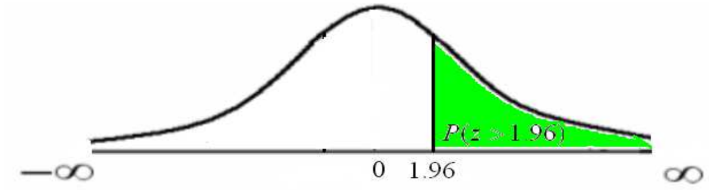


z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.										
.										
.										
.										
-2.70										
-2.60										
-2.50										

-2.40	
-2.30	0.0099
.	
.	
.	

ومن ثم يكون : $P(z < -2.33) = 0.0099$

ج- تحدد المساحة المعبرة عن الاحتمال $P(z > 1.96)$ كالتالي:



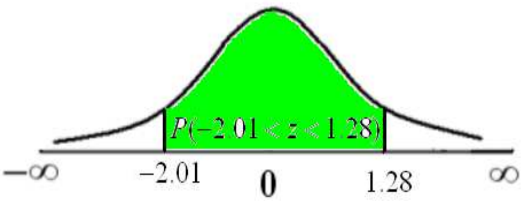
وهذا الاحتمال يحسب باستخدام خصائص دالة التوزيع التجميعي ، حيث أن :

$$P(z > 1.96) = 1 - p(z < 1.96) = 1 - F(1.96)$$

وبالكشف في الجدول بنفس الطريقة السابقة على القيمة 1.96 نجد أن : $p(z < 1.96) = 0.9750$ ،

ومن ثم يكون الاحتمال المطلوب هو: $P(z > 1.96) = 1 - 0.9750 = 0.0250$

د- المساحة أسفل المنحنى المعبرة عن الاحتمال $P(-2.01 < z < 1.28)$ هي:



وباستخدام أيضا خصائص دالة التوزيع التجميعي يمكن حساب هذا الاحتمال ، حيث أن :

$$P(-2.01 < z < 1.28) = F(1.28) - F(-2.01)$$

وبالكشف في الجدول عن هاتين القيمتين ، نجد أن:

$$P(-2.01 < z < 1.28) = 0.8997 - 0.0222 = 0.8775$$

مثال(9)

إذا كان الدخل السنوي للأسرة في أحد مناطق الجزائر يتبع توزيع طبيعي متوسطه 80 ألف دج، وتباينه 900. والمطلوب:

1- كتابة قيمة معالم التوزيع الاحتمالي للدخل السنوي.

2- كتابة شكل دالة كثافة الاحتمال.

3- ما هي نسبة الأسر التي يقل دخلها عن60 ألف دج ؟

4- ما هو الدخل الذي أقل منه 0.975 من الدخول؟

الحل

1- كتابة قيمة معالم التوزيع الاحتمالي للدخل السنوي.

بفرض أن x متغير عشوائي يعبر عن الدخل السنوي بالألف دج، وهو يتبع التوزيع الطبيعي، ومعالمه هي:

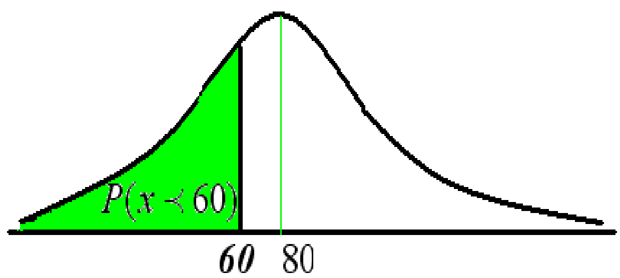
أ- المتوسط $E(x)=\mu=80$ ب- التباين هو: $Var(x)=\sigma^2=900$

أي أن : $x \sim N(80,900)$

2- شكل دالة كثافة الاحتمال

$$f(x)=\frac{1}{30\sqrt{2\pi}}\ e^{-\frac{1}{2}\left(\frac{x-80}{30}\right)^2}\ ,\ -\infty<x<\infty\ ,\ \pi=22/7$$

3- نسبة الأسر التي يقل دخلها عن60 ألف دج هي: $P(x \prec 60)$



ويتبع الخطوات المذكورة سابقا في حساب الاحتمال كما يلي:

$$P(x<60)=p\left(z<\frac{x-\mu}{\sigma}\right)$$
$$=P\left(z<\frac{60-80}{30}\right)=P(z<-0.67)=F(-0.67)$$

وبالكشف مباشرة عن هذه القيمة في جدول التوزيع الطبيعي القياسي ، نجد أن

$$P(x<60)=P(z<-0.67)=0.2514$$

4- الدخل الذي أقل منه 0.975 من الدخول: في هذه الحالة يبحث عن قيمة المتغير (x) الذي أقل منه 0.975 ، بفرض أن هذا المتغير هو (x_1) ، فإن :

$$P(x<x_1)=p\left(z<\frac{x_1-80}{30}\right)=0.975$$

بالكشف بطريقة عكسية ، حيث نبحث عن المساحة 0.9750 نجدها تقع عند تقاطع الصف 1.9 ، والعمود 06. أي أن قيمة $z=1.96$ ، ويكون :

$$1.96=\frac{x_1-80}{30}\quad,\ Then\quad x_1=30(1.96)+80=138.8$$

إذا الدخل هو 138.8 ألف دج في السنة.

2-4-4-التوزيع الأسي Distribution exponentielle

عادة ما يستخدم التوزيع الأسي في مسائل متعلقة بقياس الزمن. من ذلك مدة خدمة شباك البريد، مدة مكالمة هاتفية، مدة تفريغ باخرة شحن، مدة تصليح آلة، مدة انتظار زبون قبل الحصول على الخدمة...في العلوم الدقيقة يستخدم التوزيع الأسي لتمثيل مدة حياة الذرات المشعة (atomes radioactives) قبل أن تتفكك، حيث يعبر الوسيط عن اللحظة التي يبقى فيها نصف المجتمع الأصلي.

من الضروري فهم الآتي: كقاعدة عامة يستخدم التوزيع الأسي لتمثيل مدة حياة ظاهرة ما إذا كان لها متوسط ثابت $1/\lambda$ وكانت هذه الظاهرة لا تخضع للتقادم (vieillissement) أي أن مدة حياة الظاهرة بعد لحظة ما T لا تتبع اللحظة T؛ أي لا تتأثر بالمدة التي دامتها

الظاهرة من قبل. مثلا قد نستبعد استخدام التوزيع الأسي لتمثيل مدة حياة آلة عاملة قبل تعطلها لأن احتمال تعطلها في لحظة ليس مستقلا عن المدة التي عملتها الآلة من قبل، كذلك الأمر بالنسبة لمدة حياة الإنسان.

عمليا، نتحقق من دقة تمثيل التوزيع الأسي _أو أي توزيع آخر_ لظاهرة ما من خلال تقنيات اختبارات الفروض، وبالتحديد اختبار التجانس و التعديل.

نشير أخيرا إلى أن للتوزيع الأسي علاقة بالتوزيع بواسون، فإذا كان وقوع أحداث ما يتبع هذا التوزيع، فإن المدة بين وقوع حدثين تتبع التوزيع الأسي؛ كمثال على ذلك، إذا كان وصول الزبائن إلى مركز خدمة ما يتبع التوزيع بواسون فإن المدة الزمنية بين وصول زبون "أ" والزبون الموالي تتبع التوزيع الأسي. تتبين هذه العلاقة عند استنتاج صيغة القانون الأسي.

2-4-5-صيغة القانون الأسيأو دالة الكثافة و الدالة التجميعية للتوزيع.

بينت دراسة أن عدد حوادث العمل في معمل معين تتبع توزيع بواسون بمعدل λ حادث يوميا.

أوجد احتمال أن يسجل حادث على الأقل (حادث أو أكثر) في مدة t يوم.

$$P(X \geq 1) = 1-e^{-\lambda t} P(X \geq 1) = 1- P(0) = 1-[\lambda^{0t} * e^{-\lambda t}/0!] \quad =>$$

لنرمز ب T للزمن (باليوم) بين حادثينإذن سيكون لدينا $f(t)$ دالة الكثافة للزمن بين حادثين، و $F(t) = P(T \leq t)$ دالة التوزيع ل T.

لنحسب احتمال P أن يكون الزمن بين حادثين يوم أو أقل:

لدينا $P = P(T \leq t = 1)$ إذن:

$$P = F(t = 1) \ \ldots\ldots\ldots\ldots \ (1)$$

لاحظ من ناحية أخرى أن P هو معادل لاحتمال أن يسجل على الأقل حادث في يوم معين:

$P = P(X \geq 1) = 1 - e^{-\lambda t}$ ....(2)

من (1) و(2) نستنتج أن $\boxed{F(t) = 1 - e^{-\lambda t}}$ ....(3)

و منه $f(t) = F(t)' = (1 - e^{-\lambda t})'$

إذن $f(t) = \lambda e^{-\lambda t}$

قاعدة: إذا كان حدث عشوائي ما يتكرر في الزمن وفق توزيع بواسون:

$$\boxed{p_\tau(x) = \frac{(\lambda\tau)^x e^{-\lambda\tau}}{x!}}$$

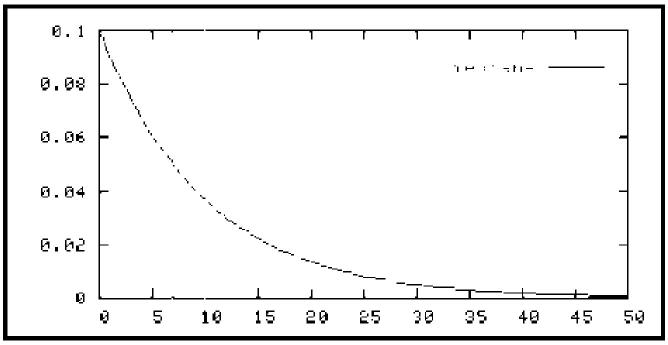
فإن الزمن T بين حادثين يتبع التوزيع التالي:

$$\boxed{f(\tau) = \begin{cases} \lambda e^{-\lambda\tau} & , \quad \tau > 0 \\ 0 & , \quad \tau \leq 0 \end{cases}}$$

حيث λ عدد حقيقي موجب.

و يسمى هذا التوزيع التوزيع الأسي ويسمى أيضا التوزيع الأسي السالب لعلاقته بتوزيع بواسون.

1. التمثيل البياني للتوزيع الأسي



رسم1 دالة الكثافة للتوزيع الأسي

2. خصائص التوزيع الأسي

$$\mu = 1/\lambda \quad , \quad \sigma^2 = 1/\lambda^2 \quad , \quad Med = \mu \ln(2) < \mu \quad , \quad M_x(t) = \frac{\lambda}{\lambda - t}$$

2-4-6-توزيع قاما Distribution gamma

توزيعي قاما و بيتا يمثلان مجموعة واسعة من التوزيعات ذات معلمتين تتميز بمرونة وقدرة على توليد توزيعات متعددة حسب قيم المعلمتين. ندرس هذين التوزيعين أيضا لعلاقتهما بالتوزيعات F، t، و ك2. يستخدم توزيع قاما لتمثيل بعض الظواهر مثل توزيع الدخل والادخار تحت شروط معينة[1].

1-صيغة القانون.

نقول عن متغيرة عشوائية أنها تتبع توزيع قاما إذا كانت دالة كثافتها كما يلي:

$$f(x) = \begin{cases} \dfrac{x^{\alpha-1} e^{-x/\beta}}{\beta^{\alpha} \Gamma(\alpha)}, & x > 0 \\ 0 \quad , & x \le 0 \end{cases} \quad , \alpha > 0, \beta > 0$$

حيث $\Gamma(\alpha)$ هي الدالة قاما: $\alpha > 0 \qquad \Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t} dt$

[1] أنظر: آيفازيان و آخرون، مبادئ النمذجة و المعالجة الأولية للبيانات، سلسلة : Mathématiques، Editions Mir، موسكو، 1983، ترجمه من الروسية إلى الفرنسية جيلالي مبارك، 1986. ص158.

ونكتب $X \sim \Gamma(\alpha,\beta)$

2-خصائص توزيع قاما

$\mu = \alpha\beta$, $\sigma^2 = \alpha\beta^2$, $M(t) = (1-\beta t)^{-\alpha}$

Pour $\alpha>1$: $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$ et si $\alpha\in N$: $\Gamma(\alpha) = (\alpha-1)!$,

$\Gamma(1/2) = \sqrt{\pi}$

من خصائص توزيع قاما علاقته بالتوزيع الأسي كما سنرى في السلسلة.

مثال. أحسب ما يلي:

$$\int_0^\infty t^4e^{-t}dt,\quad \int_0^\infty x^6e^{-x}dx,\quad \int_0^\infty \frac{e^{-x}}{x^{1/2}}dx,\quad \Gamma(7),\quad \Gamma(4.5),\quad \Gamma(2.5).$$

$$\int_0^\infty t^4e^{-t}dt = \Gamma(5) = 4! = 24, \int_0^\infty x^6e^{-x}dx = \Gamma(7) = 6! = 720, \int_0^\infty \frac{e^{-x}}{x^{1/2}}dx = \int_0^\infty x^{-1/2}e^{-x}dx = \Gamma(1/2) = \sqrt{\pi}$$

$$\Gamma(4.5) = \Gamma(3.5+1) = 3.5\Gamma(3.5) = 3.5(2.5)\Gamma(2.5) = 3.5(2.5)(1.5)(0.5)\Gamma(0.5) = 3.5(2.5)(1.5)(0.5)\sqrt{\pi}$$

$$\Gamma(7) = 6! = 720, \qquad \Gamma(2.5) = 1.5\Gamma(0.5) = 1.5\sqrt{\pi}$$

مثال. أحسب المتوسط والتباين للمتغيرات العشوائية X و Y و Z المعرفة كما يلي:

$$f(x) = \begin{cases} \dfrac{x^4e^{-x/2}}{2^5\Gamma(5)}, & x>0 \\ 0 \quad , & x\le 0 \end{cases},\quad f(y) = \begin{cases} \dfrac{y^3e^{-y/4}}{4^4(6)}, & y>0 \\ 0 \quad , & y\le 0 \end{cases},\quad f(z) = \begin{cases} \dfrac{z^2e^{-z}}{6}, & z>0 \\ 0 \quad , & z\le 0 \end{cases}$$

$$\mu_x = \alpha\beta = 5(2) = 10, \sigma^2{}_x = \alpha\beta^2 = 5(2^2) = 20, \mu_y = 4(4) = 16, \sigma^2{}_y = 4(4^2) = 64, \mu_z = 3(1) = 3, \sigma^2{}_z = 3$$

2-4-7-توزيع بيتا Distribution bêta

يتميز توزيع بيتا بمرونته الكبيرة تبعا لقيم معلمتيه (أنظر الرسم 14) حيث يستخدم لحساب توزيع t^2، F، التوزيع الثنائي، الثنائي السالب وغيرها[2]، وتستخدم لتمثيل بعض المتغيرات التي تتراوح بين 0 و 1، مثل نسبة ما كنسبة التالف أو المبيعات، إلخ.

1. صيغة القانون.

نقول عن متغيرة عشوائية أنها تتبع توزيع بيتا إذا كانت دالة كثافتها كما يلي:

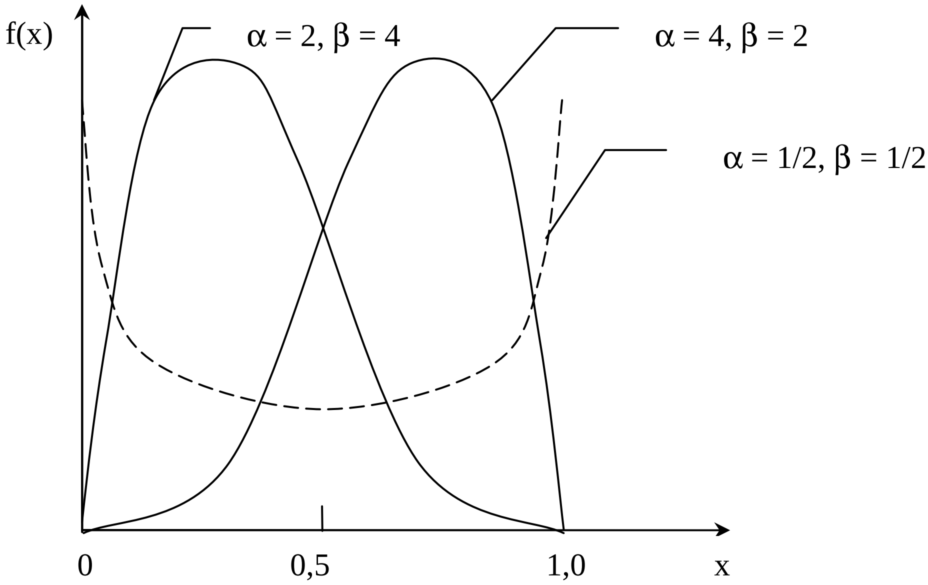
$$\boxed{f(x)=\begin{cases}\dfrac{x^{\alpha-1}(1-x)^{\beta-1}}{\beta(\alpha,\beta)} & 0<x<1 \\ 0 & ailleurs\end{cases} \qquad (\alpha,\beta>0)}$$

حيث $B(\alpha, \beta)$ هي الدالة بيتا: $\alpha,\beta>0 \qquad B(\alpha,\beta)=\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}du,$

و نكتب~ $X\ B(\alpha,\beta)$

$\alpha = 4,\ \beta = 2$

[2]المرجع السابق.



التمثيل البياني لدالة الكثافة للتوزيع بيتا من أجل قيم مختلفة للمعالم

2. خصائص توزيع بيتا

$$\alpha = \frac{\alpha}{\alpha+\beta}, \qquad \sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

3. العلاقة بين الدالتين قاما وبيتا:

$$B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

مثال. أحسب ما يلي:

$$B(3,4), \qquad B(1/2,1/2), \quad B(n,2), \qquad B(1,n), \qquad B(n,1) \quad (n \in N)$$

$$B(3,4) = \frac{(3-1)!(4-1)!}{(7-1)!} = \frac{2}{120} = \frac{1}{60}, \quad B(1/2,1/2) = \frac{\left(\sqrt{\pi}\right)\left(\sqrt{\pi}\right)}{(1/2)+(1/2)} = \pi,$$

$$B(n,\ 2) = \frac{(n-1)!1!}{(2+1)!} = \frac{1}{n(n+1)!}, \qquad B(1,n) = \frac{1(n-1)!}{n!} = \frac{(n-1)!}{n(n-1)!} = \frac{1}{n}, \qquad B(n,1) = \frac{(n-1)!1}{n!} = \frac{1}{n}$$

مثال. أحسب ما يلي: $B(3,2)$, $\int_0^1 x(1-x)dx$, $\int_0^1 x^4(1-x)^3\,dx,$

$$B(3,2)=1/[n(n+1)]=1/[3(4)]=1/12$$

$$\int_0^1 x^4(1-x)^3dx=B(5,4)=\frac{\Gamma(5)\Gamma(4)}{\Gamma(5+4)}=\frac{4!3!}{8!}=\frac{1}{280}, \qquad \int_0^1 x(1-x)dx=B(2,2)=\frac{1!1!}{3!}=1/6$$

وباستعمال العلاقة $B(\alpha,\beta)=\dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ نجد أن دالة الكثافة للتوزيع بيتا تكتب أيضا:

$$f(x)=\begin{cases}\dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}, & 0<x<1\\ 0 \quad , & ailleurs\end{cases}$$

للتوزيعين قاما وبيتا علاقة بعدد من التوزيعات المهمة كالتوزيع الأسي وتوزيع كاي تربيع، من ذلك مثلا أن التوزيع الأسي هو حالة خاصة من توزيع قاما عندما $\alpha=1,\ \beta=1/\lambda$.

مثال3. أحسب النسبة المتوقعة للإنتاج التالف والتباين، إذا كانت نسبة الإنتاج التالف تتبع التوزيع التالي:

$$f(x)=\begin{cases}6(1-x)^5, & 0<x<1\\ 0 \quad , & \text{sinon}\end{cases}$$

من المثال السابق لدينا: $B(1,n)=1/n \Rightarrow 6=1/B(1,6)$.

بوضع α و β يساويان 1 و 6 على التوالي، نجد أن $X \sim B(1,6)$

ومنه: $\mu=\dfrac{\alpha}{\alpha+\beta}=1/2, \qquad \sigma^2=\dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}=\dfrac{4}{16(5)}=1/20.$

مثال. نسبة الإنتاج المباع في مؤسسة تتبع التوزيع التالي. أحسب النسبة المتوقعة، واحتمال أن تبلغ النسبة أكثر من 35% .

$$f(x)=\begin{cases}12x^2(1-x), & 0<x<1\\ 0 & ,\ \text{sinon}\end{cases}$$

$$12=3*4\Rightarrow 12=1/B(3,2)\Rightarrow X\sim B(3,2)\Rightarrow \mu=\frac{\alpha}{\alpha+\beta}=\frac{3}{3+2}=3/5=60\%$$

$$P(X>0.35)=\int_{0,35}^{1}12x^2(1-x)dx.$$

$$soit: v=x^2dx\ et\, u=1-x\Rightarrow P(X>0.35)=12\left(\left[(1-x)\frac{x^3}{3}\right]_{0,35}^{1}+\int_{0,35}^{1}x^2dx\right)=0.3125$$

خلاصة

الجدول التالي يلخص خصائص التوزيعات الاحتمالية المستمرة الأكثر استخداما.

لتوزيع	دالة الكثافة ، التوقع والتباين		خصائص التوزيع
التوزيـــع الطبيعـــي المعياري X~N(0, 1)	$f(z)=\frac{1}{\sqrt{2\pi}}e^{-z^2/2} \qquad -\infty<z<\infty$ E(Z) = 0, V(Z) = 1		$P(Z \le -z) = 1- P(Z \le z) =$ $P(Z \ge z)$ $P(-\sigma \le X \le \sigma) =$ $P(-1 \le Z \le 1) = 0.6837,$ $P(-2\sigma \le X \le 2\sigma)=$ $P(-2 \le Z \le 2) = 0.9544,$ $P(-3\sigma \le X \le 3\sigma)=$ $P(-3 \le Z \le 3) = 0.9973.$
التوزيع الأسي	$\mu = 1/\lambda,$ $\sigma^2 = 1/\lambda^2$ $F(x) = 1-e^{-\lambda x}$	$f(\tau)=\begin{cases}\lambda e^{-\lambda\tau} & , \quad \tau>0\\ 0 & , \quad \tau\le 0\end{cases}$	$P(X \le \mu) = 0.63$
توزيع قاما X~Γ(α,β)	$\begin{cases}\frac{x^{\alpha-1}e^{-x/\beta}}{\beta^\alpha\Gamma(\alpha)}, & x>0\\ 0 \quad , & x\le 0\end{cases} \quad ,\alpha>0,\beta>0$ $\mu = \alpha\beta,$ $\sigma^2 = \alpha\beta^2$		$\Gamma(\alpha)=\int_0^\infty t^{\alpha-1}e^{-t}dt \qquad \alpha>0$

$B(\alpha,\beta)=\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}du, \qquad \alpha,\beta>0$ $B(\alpha,\beta)=\dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$	$f(x)=\begin{cases}\dfrac{x^{\alpha-1}(1-x)^{\beta-1}}{\beta(\alpha,\beta)} & 0<x<1 \\ 0 & ailleurs\end{cases} \qquad (\alpha,\beta>0)$ $\alpha=\dfrac{\alpha}{\alpha+\beta}, \quad \sigma^2=\dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	**توزيع بيتا** B(α,β)X~

3-نظرية توزيع المعاينة

يعتبر أسلوب المعاينة من أهم الأساليبالإحصائية التي نستخدمها لدراسة مجموعة كبيرة من المفردات (تسمى مجتمع) بقصد التعرف على خواصها عن طريق دراسة مجموعة صغيرة من هذه المفردات (تسمى عينة)، حيث نقوم بدراسة هذه العينة بدقة للتعرف على خواصها ومعرفة معالمها مثل الوسط الحسابي وغير ذلك من المقاييس الإحصائية ثم نقوم بتعميم النتائج التي نحصل عليها على المجتمع الأصلي للتعرف على خواصه ومعالمه وهذا هو الهدف الأساسي من الدراسة. وبالطبع لا يكون هذا التعميم من العينة إلى المجتمع له معنى أو قيمة علمية إلاإذا تم اختيار المفردات بطريقة نضمن بها أن تكون العينة ممثلة للمجتمع تمثيلا صادقا.

3-1- تعريف توزيع المعاينة وعدد العينات المسحوبة:

3-1-1- تعريف توزيع المعاينة: نفرض أن لدينا مجتمعا من المفردات يتبع توزيعا احتماليا معينا وأننا بصدد سحب عينة حجمها n من هذا المجتمع، بالطبع ليس معنى هذا أن هناك عينة واحدة يمكن سحبها ولكن يكون أمامنا عدد كبير من العينات التي يمكن سحبها من هذا المجتمع والتي حجم كل منها هو n من المفردات.

والآن نفرض أننا سحبنا عينة حجمها n من هذا المجتمع وحسبنا من هذه العينة مقياسا معينا (وليكن الوسط الحسابي) ثم سحبنا عينة ثانية لها نفس الحجم n وحسبنا منها نفس

المقياس ثم سحبنا عينة ثالثة وحسبنا منها نفس المقياس وهكذا بالنسبة لجميع العينات التي يمكن سحبها من هذا المجتمع، سنجد أمامنا عددا كبيرا من القيم لنفس المقياس ولا نتوقع أن تكون جميع القيم التي حصلنا عليها من العينات لهذا المقياس متساوية وإنما ستكون مختلفة عن بعضها وتكون مجتمعا أخر، عدد مفرداته أكبر بكثير من عدد مفردات المجتمع الأصلي.

وعلى ذلك يمكن النظر إلى هذا المقياس على أنه متغير عشوائي يأخذ قيما مختلفة (هي التي حصلنا عليها من هذه العينات) ويتبع توزيعا معينا ربما يختلف أو لا يختلف عن توزيع المجتمع الأصلي يسمى بتوزيع المعاينة لهذا المقياس سواء كان هذا المقياس هو الوسط الحسابي أو نسبة المفردات التي لها صفة معينة أو الانحراف المعياري أو غيره من المقاييس الإحصائية.

3-1-2- عدد العينات المسحوبة من المجتمع:

لإيجاد عدد العينات الممكنة والمسحوبة من المجتمع حجمه N هناك حالتين في حالة السحب بالإرجاع وفي حالة السحب بدون إرجاع:

- ♦ في حالة السحب بالإرجاع فان عدد العينات الممكنة يساوي N^n .
- ♦ في حالة السحب بدون إرجاع تنقسم إلى قسمين:

– حالة السحب بدون إرجاع والترتيب غير مهم: $C_N^n = \frac{N!}{(N-n)!(n!)}$

– حالة السحب بدون إرجاع والترتيب مهم: $\frac{N!}{(N-n)!}$.

ملاحظة:في حالة السحب بدون إرجاع ولم يتم ذكر نوعية الترتيب، فإننا نعتبر الترتيب غير مهم.

3-2- توزيعالمعاينة للمتوسطات:

نفرض أن لدينا مجتمعا ولتكن مفرداته هي: $X_1, X_2, X_3........$.

ونفرض أننا سحبنا من هذا المجتمع عينة حجمها n وحسبنا وسطها الحسابي فوجدناه $\bar{x}_1$ ثم سحبنا عينة أخرى لها نفس الحجم وحسبنا وسطها الحسابي فوجدناه $\bar{x}_2$، ثم عينة ثالثة لها نفس الحجم ووجدنا أن وسطها الحسابي $\bar{x}_3$ وهكذا بالنسبة لكل العينات التي حجمها n والتي يمكن سحبها من هذا المجتمع. سنجد في النهاية أننا سنحصل على مجموعة جديدة من المفردات هي المتوسطات الحسابية لهذه العينات وهي تكون مجتمعا جديدا يسمى مجتمع المتوسطات الحسابية للعينات التي حجمها n والتي يمكن سحبها من المجتمع الأصلي ويمكن كتابة مفردات المجتمع الجديد على النحو التالي: $\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{x}_4,............$.

وهذه المتوسطات تختلف عن بعضها كما أنها تكون مجتمعا جديدا له توزيع احتمالي يهمنا معرفته ودراسته، ومجتمع المتوسطات الحسابية $\bar{X}$ كأي مجتمع أخر له توزيع احتمالي يتمتع بجميع صفات وخواص التوزيعات الاحتمالية وبالطبع له متوسط وانحراف معياري.

3-2-1- متوسط توزيع المعاينة للمتوسطات:

إذا كانت متغيرة عشوائية تمثل مجتمع ما و $\bar{x}$ متغيرة عشوائية تمثل متوسط عينة مسحوبة من ذات المجتمع حيث حجم المجتمع يرمز له بالرمز N وحجم العينة يرمز لها بالرمز n، فإن القيمة المتوقعة لمتوسط العينة $E(\bar{X})$ تكتب كما يلي:

$$E(\bar{X}) = \mu_{\bar{x}} = \mu$$

البرهان : لنرمز ب Xi لقيم المتغيرة الأصليةX .

$$E(\overline{X}) = E\left(\frac{1}{n}\sum_i Xi\right) = \frac{1}{n}\sum_i E(Xi) = \frac{1}{n}\sum_i \mu = \frac{1}{n}n\mu = \mu.$$

وهذا يعني أن متوسط مجتمع المتوسطات هو نفسه متوسط المجتمع الأصلي.

<u>**مثال1:**</u> ليكن لدينا المجتمع المتكون من المفردات التالية:1، 3، 5.

- أحسب متوسط المجتمع μ ؟.
- ما هي القيمة المتوقعة لمتوسط عينة مسحوبة من المجتمع في حالة السحب بالإرجاع وفي حالة السحببدون إرجاع مكونة من مفردتين $(n=2)$ ؟.
- قارن بين $\mu_{\bar{x}}$ و μ ؟

<u>الحل:</u>

- حساب متوسط المجتمع: μ = (1 + 3 + 5) /3 = 3
- إيجادالقيمة المتوقعة لمتوسط عينة مسحوبة من المجتمع في حالة السحب بالإرجاع وفي حالة السحببدون إرجاع:

أ- <u>حالة السحب بالإرجاع:</u>

من أجل تحديد ذلك يجب حساب جميع الحالات الممكنة للمتوسط $\overline{x_i}$ حسب كل عينة.

العينات الممكنة ذات الحجم $n=2$ من مجتمع حجمه 3 عددها: $N^n = 3^2 = 9$.

العينات الممكنة		
(1, 1)	(3, 1)	(5, 1)
(1, 3)	(3, 3)	(5, 3)
(1, 5)	(3, 5)	(5, 5)

المتوسطات الممكنة للعينة (معاينة بالإرجاع) $\overline{x_i}$		
1	2	3
2	3	4
3	4	5

القيمة المتوقعة لمتوسط العينة $E(\overline{X})$ هي متوسط قيمها وهي: $E(\overline{X}) = \mu_{\bar{x}} = \sum_i \overline{x_i} / 9 = 3$

ب- <u>حالة السحب بدون إرجاع (في الأغلب الترتيب غير مهم):</u>

العينات الممكنة ذات الحجم $n=2$ من مجتمع حجمه 3 عددها: $C_N^n = C_3^2 = \frac{(3)!}{(3-2)!(2)!} = 3$

العينات الممكنة بدون إرجاع					المتوسطات الممكنة للعينة أو توزيع المعاينة للمتوسطات (معاينة بدون إرجاع) $\overline{x_i}$			
(1, 3)					2			
(1, 5)	(3, 5)				3	4		

القيمة المتوقعة لمتوسط العينة $E(\overline{X})$ هي متوسط قيمها وهي: $E(\overline{X}) = \mu_{\bar{x}} = \sum_i \overline{x_i} \Big/ 3 = 3$

- إذا ما قارننا بين $\mu_{\bar{x}}$ و μ نجد: $E(\overline{X}) = \mu_{\bar{x}} = \mu = 3$

3-2-2- تباين توزيع المعاينة للمتوسطات:

إذا كانت X متغيرة عشوائية تمثل مجتمع ما حجمه N و $\overline{x_i}$ متغيرة عشوائية تمثل متوسطعينة حجمها n مسحوبة من ذات المجتمع،فإن تباين توزيع المعاينة للمتوسطات يكتب كما يلي:

- حالة السحب بالإرجاع: $\delta_{\bar{x}}^2 = \frac{\delta^2}{n} \Rightarrow \delta_{\bar{x}} = \frac{\delta}{\sqrt{n}}$.

- حالة السحب بدون إرجاع: $\delta_{\bar{x}}^2 = \frac{\delta^2}{n}\left(\frac{N-n}{N-1}\right) \Rightarrow \delta_{\bar{x}} = \frac{\delta}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$.

- تسمى النسبة $\frac{N-n}{N-1}$ معامل الإرجاع.

وهذا يعني أن الانحراف المعياري هوالانحراف المعياري للمجتمع الأصلي مضروبا في عامل معين، وهذه المعلومات صحيحة مهما اختلف التوزيع الاحتمالي لمجتمع المتوسطات عن التوزيع الاحتمالي للمجتمع الأصلي.

وإذا نظرنا إلى الانحراف المعياري لمجتمع المتوسطات نلاحظ أن مجتمع المتوسطات أكثر تجانسا من المجتمع الأصلي، أي أن مفرداته متجانسة وغير متشتتة اذا قورنت بمفردات

المجتمع الأصلي وهذا من أهم الأسباب التي تجعلنا نعتمد على مجتمع المتوسطات في الاستدلال حول معالم المجتمع الأصلي.

مثال2:في نفس المثال السابق (المثال1).

- أحسب تباين المجتمع؟
- أحسب التباين لتوزيع المعاينة للمتوسطات $\delta_{\bar{x}}^2$ في حالة السحب بالإرجاع وفي حالة السحببدون إرجاع؟.
- قارن بين تباين المجتمع وتباين متوسطات العينات الممكنة (توزيع المعاينة للمتوسطات) في كل حالة؟.

الحل:

- حساب تباين المجتمع: $\delta^2 = \left[\dfrac{\sum_i (X_i - \mu)^2}{N}\right] = 2.66$
- حساب التباين لتوزيع المعاينة للمتوسطات $\delta_{\bar{x}}^2$:

أ- **حالة السحب بالإرجاع:**

$\bar{x}_i$ السحب بالإرجاع		
1	2	3
2	3	4
3	4	5

تباين المتوسطات الممكنة للعينة $\delta_{\bar{x}}^2$:

$$\delta_{\bar{x}}^2 = \frac{\sum_i \left(\bar{x}_i - \mu_{\bar{x}}\right)^2}{9} = 1.33$$

- إذا ما قارننا بين $\delta_{\bar{x}}^2$ و δ^2 نجد أن: $\delta_{\bar{x}}^2 = 1.33 = \dfrac{\delta^2}{n} = \dfrac{2.66}{2} = 1.33$

ب- **حالة السحب بدون إرجاع:**

المتوسطات الممكنة للعينة $\bar{x}_i$ السحب بدون إرجاع			
2			
3	4		

تباين المتوسطات الممكنة للعينة $\delta_{\bar{x}}^2$:

$$\delta_{\bar{x}}^2 = \frac{\sum_i \left(\bar{x}_i - \mu_{\bar{x}}\right)^2}{3} = 0.66$$

– إذا ما قارننا بين $\delta_{\bar{x}}^2$ و δ^2 نجد أن: $\delta_{\bar{x}}^2 = 0.66 = \frac{\delta^2}{n}\left(\frac{N-n}{N-1}\right) = \frac{2.66}{2}\left(\frac{3-2}{3-1}\right) = 0.66$

3-2-3- طبيعة توزيع المتوسط:

حتى الآن لم نتعرض لتوزيع مجتمع المتوسطات الحسابية، كل ما تعرضنا له الوسط الحسابي والانحراف المعياري لهذا التوزيع ولكن شكل ونوع التوزيع الاحتمالي نفسه يعتمد على توزيع المجتمع الأصلي المسحوب منه العينات. وفيما يلي النظرية الإحصائية التي تعطي التوزيع الاحتمالي لمجتمع المتوسطات الحسابية نذكرها بدون إثبات وذلك لأننا سوف نستخدمها كثيرا في الاستدلال حول معالم المجتمع الأصلي.

نظرية1: إذا كان المجتمع موزع طبيعيا بمتوسط μ وتباين δ^2 فإن متوسط العينة المسحوبة منه يتبع أيضا التوزيع الطبيعيبمتوسط μ وتباين $\frac{\delta^2}{n}$ ، ونكتب:

$$\bar{x} \rightarrow N\left(\mu, \frac{\delta^2}{n}\right)$$

– **نظرية النهاية المركزية2:** إذا كان المجتمع الذي تسحب منه العينة ذو متوسط μ وتباين δ^2 لكن ليس بالضرورة طبيعيا فإن المتغيرة المعيارية لـ $\bar{x}$ أي $z = \frac{\bar{x}-\mu}{\delta/\sqrt{n}}$ تؤولإلى التوزيع الطبيعي المعياري عندما يكون n كبيرا $(n \geq 30)$ ونكتب:

$$z \rightarrow N\,(0\,.\,1)$$

– في حالة المجتمع محدود والمعاينة بدون إرجاع نستبدل العبارة $\delta/\sqrt{n}$ بالعبارة:

$$\delta_{\bar{x}} = \frac{\delta}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

عمليا تستخدمهذه الصيغة المعدلة بمعامل الإرجاع للانحراف المعياري عندما $\frac{n}{N} \geq 0.05$

<u>مثال3</u>:مجتمع حجمه 1200 بمتوسط $\mu = 40$ و $\delta = 15$. نستخرج كل العينات الممكنة.

– المطلوب: أحسب المتوسط والانحراف المعياري لتوزيع المعاينة للمتوسطات في حالة:

– حجم العينة $n = 49$، – حجم العينة $n = 81$.

الحل:

$$1 - n = 49 : \frac{n}{N} = \frac{49}{1200} = 0.04 < 0.05 \Rightarrow \delta_{\bar{x}} = \frac{\delta}{\sqrt{n}} = \frac{15}{\sqrt{49}} \approx 2.14$$

$$E\left(\overline{X}\right) = \mu = 20$$

$$2 - n = 81 : \frac{n}{N} = \frac{81}{1200} = 0.0675 > 0.05 \Rightarrow \delta_{\bar{x}} = \frac{\delta}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

$$= \frac{15}{\sqrt{81}}\sqrt{\frac{1200-81}{1200-1}} \approx 1.61$$

$$E\left(\overline{X}\right) = \mu = 20$$

3-3- توزيع المعاينة للنسبة:

عندما تكلمنا عن توزيعات المعاينة ذكرنا أنها توزيعات احتمالية للمقاييس الإحصائية التي نحسبها من العينات كما ذكرنا أن الوسط الحسابي $\overline{X}$ والانحراف المعياري للعينة، ونسبة المفردات في العينة التي لها صفة معينة هي بعض هذه المقاييس، ولقد حصلنا فيما سبق على التوزيع الاحتمالي لمجتمع المتوسطات الحسابية والآن نتعرف على التوزيع الاحتمالي

لمجتمع النسب وسوف نتبع نفس الأسلوب الذي اتبعناه في تقديم التوزيع الاحتمالي لمجتمع المتوسطات الحسابية.

نظرية3: لتكن X متغيرة عشوائية تمثل مجتمع ما غير محدود وموزع طبيعيا حيث p نسبة المفردات في المجتمع ذات صفة معينة، ولتكن p' متغيرة عشوائية تمثل نسبة المفردات ذات الصفة المذكورة في العينة المسحوبة من ذات المجتمع،نحصل على توزيع للإحصائية p' حيث معالمه $E(p')$ و $\delta^2_{p'}$، هذه المعالم تساوي:

$$\delta^2_{p'} = \frac{pq}{n} \quad \text{و} \quad E(p') = \mu_{p'} = p$$

عند $n \geq 30 : \quad p' \rightarrow N\left(p, \delta^2_{p'}\right)$

عندما يكون المجتمع محدودا والمعاينة بدون إرجاع نضرب في معامل الإرجاع عند حساب الانحراف المعياري.

<u>مثال4</u>.لاحظت إدارة الجامعة أنه في مجتمع الطلبة هناك نسبة $p = 0.4$ تحصلوا على الشهادة من بين الطلبة، فقامت الجامعة بسحب عينة حجمها $n = 100$. تريد الإدارة معرفة توزيع للإحصائية p' والتي تمثل نسبة الطلبة المحصلين على الشهادة في العينة.

<u>الحل</u>: نحصل على توزيع للإحصائية p' حيث معالمه $E(p')$ و $\delta^2_{p'}$، هذه المعالم تساوي:

$$\delta^2_{p'} = \frac{pq}{n} \quad \text{و} \quad E(p') = \mu_{p'} = p$$

ومنه بالتعويض نجد:

$$E(p') = \mu_{p'} = p = 0.4$$
$$\delta^2_{p'} = \frac{pq}{n} = \frac{(0.4)(0.6)}{100} = 0.0024$$

ومنه فان توزيع المعاينةللإحصائية p' يكتب كما يلي:

$$p' \rightarrow N\left(p, \delta^2_{p'}\right)$$
$$p' \rightarrow N\left((0.4); (0.0024)\right)$$

3-4-توزيع المعاينة للفروق والمجاميع

3-4-1- المتوسط والتباين

ليكن لدينا مجتمعين نسحب من كل منهما عينة عشوائية، نحسب في كل عينة محسوبة من المجتمع الأول الإحصائية S_1 ونحسب نفس الإحصائية (المتوسط مثلا أو التباين ...) في كل عينة من المجتمع الثاني ونسميها S_2. إن الفرق $S_1 - S_2$ يشكل بدوره متغيرة عشوائية لها المتوسط والتباين التاليين:

$$\sigma^2_{S1-S2} = \sigma^2_{S1} + \sigma^2_{S2} \qquad\qquad \mu_{S-S2} = \mu_{S1} - \mu_{S2}$$

<u>مثال5</u>. إذا كانت الإحصائية هيالمتوسط، وكان لدينا مجتمعين كبيرين من المفردات، وكان متوسط الأول μ_1وتباينه δ_1^2 ومتوسط الثاني μ_2 وتباينه δ_2^2 فإذا سحبنا عينة من الأول حجمها n_1 مفردة وسحبنا عينة من المجتمع الثاني حجمها n_2فإن:

$$\mu_{\bar{x}_1-\bar{x}_2} = \mu_{\bar{x}_1} - \mu_{\bar{x}_2} = \mu_1 - \mu_2$$

$$\delta^2_{\bar{x}_1-\bar{x}_2} = \delta^2_{\bar{x}_1} + \delta^2_{\bar{x}_2} = \frac{\delta_1^2}{n_1} + \frac{\delta_2^2}{n_2}$$

<u>مثال6</u>. إذا كانت الإحصائية هيالنسبة فإن:

$$\mu_{p_1-p_2} = \mu_{p_1} - \mu_{p_2} = p_1 - p_2$$

$$\delta^2_{p_1-p_2} = \delta^2_{p_1} + \delta^2_{p_2} = \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}$$

- إذا كان الاهتمام هو على مجموع الإحصائيتين بدلا من الفرق بينهما فإن:

$$\sigma^2_{S1+S2} = \sigma^2_{S1} + \sigma^2_{S2} \qquad\qquad \mu_{S1+S2} = \mu_{S1} + \mu_{S2}$$

3-4-2- طبيعة توزيع المعاينة للفرق بين متوسطين

نظرية4:في حالة$n_1 \geq 30$ وn_2 ، يقترب توزيـع المتغيرة المعيارية للفرق بين متوسطين من التوزيع الطبيعي المعياري. ونكتب:

$$\mu_{\bar{x}_1 - \bar{x}_2} \rightarrow N\,(0.1)$$

مثال7: ليكن المجتمع V_1 : 2، 6، 10.

والمجتمع V_2: 1، 3. تحقق من أن:

$\boldsymbol{\mu_{V1-V2} = \mu_{V1} - \mu_{V2}}$; $\boldsymbol{\sigma^2_{V1-V2} = \sigma^2_{V1} + \sigma^2_{V2}}$

الحل:

V_1			$V_1 - V_2$	
10	**6**	**2**		
9	5	1	**1**	V_2
7	3	-1	**3**	

$\mu_{V1} = (2 + 6 + 10)/3 = 6$;

$\mu_{V2} = (1 + 3)/2 = 2 \Rightarrow$

$\mu_{V1} - \mu_{V2} = 6 - 2 = 4$

$\mu_{V1-V2} = (1 + 5 + 9 - 1 + 3 + 7)/6 = 4$

$\sigma^2_{V1} = (2^2 + 6^2 + 10^2)/3 - 6^2 = 10{,}67$;

$\sigma^2_{V2} = (1^2 + 3^2)/2 - 2^2 = 1 \Rightarrow \sigma^2_{V1} + \sigma^2_{V2} = 11.67$

$\sigma^2_{V1-V2} = (1^2 + 5^2 + 9^2 + 1^2 + 3^2 + 7^2) / 6 - 4^2 = 11{,}67$

مثال8: اذا كان $X_1 \to N(30;;25)$ وسحبت من مجتمعه عينة حجمها 30 مشاهدة و $X_2 \to N(20;;16)$ وسحبت من مجتمعه عينة حجمها 35 مشاهدة، أوجد توزيع الفرق بين متوسطي العينتين؟.

الحل:

بالتطبيق المباشر للعلاقتين التاليتين:

$$\mu_{\bar{x}_1-\bar{x}_2} = \mu_{\bar{x}_1} - \mu_{\bar{x}_2} = \mu_1 - \mu_2$$

$$\delta^2_{\bar{x}_1-\bar{x}_2} = \delta^2_{\bar{x}_1} + \delta^2_{\bar{x}_2} = \frac{\delta_1^2}{n_1} + \frac{\delta_2^2}{n_2}$$

نجد:

$$\mu_{\bar{x}_1-\bar{x}_2} = \mu_{\bar{x}_1} - \mu_{\bar{x}_2} = \mu_1 - \mu_2 = 30 - 20 = 10$$

$$\delta^2_{\bar{x}_1-\bar{x}_2} = \delta^2_{\bar{x}_1} + \delta^2_{\bar{x}_2} = \frac{\delta_1^2}{n_1} + \frac{\delta_2^2}{n_2} = \frac{25}{30} + \frac{16}{35} = 1,29$$

3-5- توزيع المعاينة للتباين وتوزيع المعاينة لنسبة تبايني عينتين

3-5-1-توزيع المعاينة للتباين: إذا كانت متغيرة عشوائية تمثل مجتمع ما وS^2 متغيرة عشوائية تمثل تباين عينة مسحوبة من ذات المجتمع، فإن القيمة المتوقعة لتباين العينة تكتب كما يلي:

- في حالة السحب بالإرجاع: $E(s^2) = \mu_{s^2} = \delta^2\left(\frac{n-1}{n}\right)$

- في حالة السحب بدون إرجاع: $E(s^2) = \mu_{s^2} = \delta^2\left(\frac{n-1}{n}\right)\left(\frac{N}{N-1}\right)$

عندما يكون N كبير جدا (N/ (N-1 تؤول إلى 1)

مثال9: من خلال المثال1.

- أحسب القيمة المتوقعة لتباين العينة المسحوبة من المجتمع من خلال متوسط تباينات العينات الممكنة وذلك في حالة السحب بالإرجاع و في حالة السحب بدون إرجاع؟

– ما ذا تلاحظ؟.

الحل:

– حساب القيمة المتوقعة لتباين العينة المسحوبة من المجتمع من خلال متوسط تباينات العينات الممكنة:

أ– **حالة السحب بالإرجاع:**

التباينات الممكنة S^2_i		
0	1	4
1	0	1
4	1	0

$$\sum_i S_i^2 / 9 = 12/9 = 1.33 \Rightarrow E[S^2] = 1.33$$

– نلاحظ: $E(s^2) = \mu_{s^2} = 1.33 = \delta^2\left(\frac{n-1}{n}\right) = 2.66\left(\frac{2-1}{2}\right) = 1.33$

ب– **حالة السحب بدون إرجاع:**

التباينات الممكنة S^2_i			
1			
4	1		

$$\sum_i S_i^2 / 3 = 6/3 = 2 \Rightarrow E[S^2] = 2$$

– نلاحظ: $E(s^2) = \mu_{s^2} = 2 = \delta^2\left(\frac{n-1}{n}\right)\left(\frac{N}{N-1}\right) = 2.66\left(\frac{2-1}{2}\right)\left(\frac{3}{3-1}\right) = 2$

ملاحظة: لدينا: $E\left(S^2\frac{n}{n-1}\right) = \sigma^2$ ونقول عن $S^2\frac{n}{n-1}$ أنه مقدر "غير منحرف" لـσ^2 ويرمز له ب $\hat{S}^2$حيث:

$$\hat{S}^2 = S^2 \frac{n}{n-1}$$

- إذا أخذنا عينات عشوائية حجمها n من مجتمع طبيعي، فإن :

$$\frac{nS^2}{\sigma^2} = \frac{(n-1)\hat{S}^2}{\sigma^2} \sim \chi^2_{n-1}$$

<u>مثال10</u> : ليكن مجتمع طبيعي حجمه 100 نسحب منه عينة حجمها $n = 16$. ما هو احتمال أن يكون تباين العينة S^2 أقل من أو يساوي 10 علما أن تباين المجتمع 80.

$$P(S^2 \leq 10) = P(\chi^2_{15} \leq 2) = P(S^2 \frac{16}{80} \leq X \sim N(\mu, \sigma)) \Rightarrow \frac{nS^2}{\sigma^2} \sim \chi^2_{n-1}$$

من الجدول $p\left(\chi^2_{15} \leq 2\right) < 0.005$

3-5-2- توزيع المعاينة لنسبة تباينين:

يعتبر هذا التوزيع من التوزيعات الهامة التي تبحث في تجانس المجتمعات ونلجأ لحساب النسب بين التباينات وليس الفرق بينهما لسهولة دراسة النسب وتفسيرها.

لدينا: $X = \dfrac{X_1 / v_1}{X_2 / v_2} \rightarrow F_{v_1 ; v_2}$ في حالة المتغيرتان العشوائيتان مستقلتان و $X_1 \rightarrow \chi^2_{v_1}$ و $X_2 \rightarrow \chi^2_{v_2}$ ومنه نستنتج ما يلي:

- ليكن لدينا مجتمعان طبيعيان تبايناهما σ^2_1 , σ^2_2، نسحب من المجتمعين عينتين عشوائيتين حجمهما علىالتوالي n_1 , n_2 :

$$F = \frac{\left[\dfrac{S_1^2 n_1}{n_1 - 1}\right] \dfrac{1}{\sigma_1^2}}{\left[\dfrac{S_2^2 n_2}{n_2 - 1}\right] \dfrac{1}{\sigma_2^2}} = \frac{\hat{S}_1^2 / \sigma_1^2}{\hat{S}_2^2 / \sigma_2^2} \rightarrow F_{n_1 - 1;\ n_2 - 1}$$

– وفي حالة تساوي تبايني المجتمعين فان:

$$F = \frac{\widehat{S}_1^2}{\widehat{S}_2^2} \to F_{n_1-1, n_2-1}$$

مثال11. عينتين حجمهما 8 و10 مسحوبتين من مجتمعين طبيعيين تبايناهما على التوالي 20 و36. ما احتمال أن يكون تباين الأولى أكبر من ضعف تباين الثانية؟

الحل:

$$P(S_1^2 > 2S_2^2) = P(\frac{S_1^2}{S_2^2} > 2) = P\left(\frac{S_1^2\left(\frac{n_1}{n_1-1}\right)\frac{1}{\sigma_1^2}}{S_2^2\left(\frac{n_2}{n_2-1}\right)\frac{1}{\sigma_2^2}} > 2\frac{\left(\frac{n_1}{n_1-1}\right)\frac{1}{\sigma_1^2}}{\left(\frac{n_2}{n_2-1}\right)\frac{1}{\sigma_2^2}} \right) = P\left(\frac{S_1^2\left(\frac{n_1}{n_1-1}\right)\frac{1}{\sigma_1^2}}{S_2^2\left(\frac{n_2}{n_2-1}\right)\frac{1}{\sigma_2^2}} > 2\frac{\left(\frac{8}{7}\right)\frac{1}{20}}{\left(\frac{10}{9}\right)\frac{1}{36}} \right) =$$

$$= p\left(F_{7;9} > 3.7\right) \approx 0.036$$

مثال12: سحب عينة حجمها 13 من مجتمع طبيعي تباينه 9، وسحبت عينة أخرى حجمها 21 من مجتمع طبيعي تباينه 25 مستقل عن المجتمع الأول. أوجد احتمال النسبة بين تبايني العينتين المعدل أقل من 0.8؟.

الحل:

لدينا: $$F = \frac{\widehat{S}_1^2 / \delta_1^2}{\widehat{S}_2^2 / \delta_2^2} \to F_{n_1-1, n_2-1} \Rightarrow F = \frac{\widehat{S}_1^2 / 9}{\widehat{S}_2^2 / 25} \to F_{12,20}$$

اذن: $$p\left(\frac{\widehat{S}_1^2}{\widehat{S}_2^2} < 0.8\right) = p\left(\frac{\widehat{S}_1^2 / 9}{\widehat{S}_2^2 / 25} < 0.8\left(\frac{25}{9}\right)\right)$$

$$= p\left(F < 2.22\right) = 1 - p\left(F > 2.22\right) = 0.95$$

3-6- تمارين محلولة:

التمرين1:

- إذا كان المتغير العشوائي X له التوزيع الطبيعي $N(3;;4)$..

1- ما هو احتمال أن تقع X بين القيمتين3 و5 ؟.

2- ما هو احتمال أن يكون X أكبر من 1 ؟

حل التمرين1:

لدينا: $Z = \frac{X-\mu}{\delta}$ والدالة Φ هي دالة التوزيع الطبيعي المعياري (قيمها تستخرج من الجدول مباشرة).

1- احتمال أن تقع X بين القيمتين3 و5:

$$p(3 < X < 5) = p\left(\frac{3-3}{2} < Z < \frac{5-3}{2}\right)$$
$$= p(0 < Z < 1) = \Phi(1) - \Phi(0)$$
$$= 0.8413 - 0.5 = 0.3413$$

2- احتمال أن يكون X أكبر من 1:

$$p(X > 1) = p\left(Z > \frac{1-3}{2}\right)$$
$$= p(Z > -1) = 1 - p(Z \leq -1)$$
$$= 1 - \Phi(-1) = 1 - 1 + \Phi(1)$$
$$= 0.8413$$

التمرين2:

- مجتمع إحصائي يخضع للتوزيع الطبيعي، يتكون من المفردات التالية: $(4;;3;;2)$، نسحب بإرجاع عينة من هذا المجتمع ذات حجم $n = 2$.

1- أحسب الوسط والتباين للمجتمع؟

2- أحسب عدد العينات الممكنة؟

3- أحسب معالم العينة $\left(\left(\overline{X}\right)\mu_{\overline{X}},,\delta^2_{\overline{X}}\right)$ بطريقة الجداول؟

حل التمرين2:

1- حساب الوسط والتباين للمجتمع:

$$\mu = \frac{\sum X_i}{N} = \frac{2+3+4}{3} = 3$$

$$\delta^2 = \frac{\sum (X_i - \mu)^2}{N} \approx 0.67$$

2- حساب عدد العينات الممكنة:

العينات الممكنة ذات الحجم $n = 2$ من مجتمع حجمه 3 عددها: $N^n = 3^2 = 9$.

3- أحسب معالم العينة $\left(\left(\overline{X}\right)\mu_{\overline{X}},,\delta^2_{\overline{X}}\right)$ بطريقة الجداول:

المتوسطات الممكنة للعينة (معاينة بالإرجاع) $\overline{x_i}$		
3	2,5	2
3,5	3	2,5
4	3,5	3

العينات الممكنة		
(2,4)	(2,3)	(2, 2)
(3,4)	(3, 3)	(3,2)
(4,4)	(4,3)	(4,2)

- القيمة المتوقعة لمتوسط العينة $E\left(\overline{X}\right)$ هي متوسط قيمها وهي: $E\left(\overline{X}\right) = \mu_{\bar{x}} = \sum_i \overline{x_i} \big/ 9 = 3$
- تباين المتوسطات الممكنة للعينة $\delta^2_{\bar{x}}$:

$$\delta^2_{\bar{x}} = \frac{\sum_i \left(\overline{x_i} - \mu_{\bar{x}}\right)^2}{9} = 0.33$$

التمرين3:

1- إذا كان ضغط الدم لمجموعة من الأفراد يتبع التوزيع الطبيعي بمتوسط يساوي 90 وانحراف معياري يساوي 9، سحبنا عينة حجمها 36 فردا من هذه المجموعة.

ما هو احتمال أن يكون متوسط ضغط الدم في العينة أكبر من 86 ؟.

2- إذا علم أن نسبة البيض التالف الذي ينتجه أحد مراكز إنتاج الدواجن هي 0.03 ولها توزيع طبيعي، اشترى شخص 400 بيضة من إنتاج هذا المركز.

- ما هو احتمال أن يجد من بينها 20 بيضة على الأقل تالفة؟

حل التمرين3:

1- ايجاد احتمال أن يكون متوسط ضغط الدم في العينة أكبر من 86:

لدينا: $\mu = 90;;n = 36;;\delta = 9$

$$\mu_{\bar{x}} = \mu = 90$$
$$\delta_{\bar{x}} = \frac{\delta}{\sqrt{n}} = \frac{9}{\sqrt{36}} = 1.5$$

$$p\left(\bar{x} > 86\right) = p\left(\frac{\bar{x} - \mu_{\bar{x}}}{\delta_{\bar{x}}} > \frac{86 - 90}{1.5}\right)$$
$$= p\left(z > -2.67\right) = 1 - p\left(z \leq -2.67\right)$$
$$= 1 - \Phi\left(-2.67\right) = \Phi\left(2.67\right) = 0.9962$$

2- ايجاد احتمال أن يجد 20 بيضة على الأقل تالفة:

لدينا: $p = 0.03;;n = 400$

$$\mu_{p'} = p = 0.03$$

$$\delta^2_{p'} = \frac{pq}{n} = \frac{(0.03)(0.97)}{400}$$

$$\delta_{p'} = \sqrt{\frac{(0.03)(0.97)}{400}} = 0.0085$$

$$p(x \geq 20) = p\left(p' \geq \frac{20}{400}\right) = p(p' \geq 0.05)$$

$$= p\left(z \geq \frac{0.05 - 0.03}{0.0085}\right) = p(z \geq 2.35)$$

$$= 1 - p(z < 2.35) = 1 - 0.9906 = 0.0094$$

التمرين4: اذا كان لدينا البيانات التالية:

المجتمع	متوسط المجتمع	تباين المجتمع
1	40	18
2	35	25

– تم سحب عينتين مستقلتين من المجتمعين حيث $n_1 = 12$ و $n_2 = 10$. مع العلم أن المجتمعين طبيعيين.

– أحسب التوزيع العيني للفرق بين متوسطي العينتين $(\mu_{\bar{x}_1 - \bar{x}_2}; \delta^2_{\bar{x}_1 - \bar{x}_2})$

حل التمرين4:

– حساب التوزيع العيني للفرق بين متوسطي العينتين $(\mu_{\bar{x}_1 - \bar{x}_2}; \delta^2_{\bar{x}_1 - \bar{x}_2})$:

$$\mu_{\bar{x}_1 - \bar{x}_2} = \mu_{\bar{x}_1} - \mu_{\bar{x}_2} = \mu_1 - \mu_2 = 40 - 35 = 5$$

$$\delta^2_{\bar{x}_1 - \bar{x}_2} = \delta^2_{\bar{x}_1} + \delta^2_{\bar{x}_2} = \frac{\delta_1^2}{n_1} + \frac{\delta_2^2}{n_2} = \frac{18}{12} + \frac{25}{10} = 4$$

$$\bar{x}_1 - \bar{x}_2 \to N(5;;4)$$

3-7- تمارين غير محلولة:

التمرين1:

إذا كان لدينا مجتمع يتكون من المفردات التالية (0، 2، 4، 6) نسحب عينة من هذا المجتمع ذات الحجم n=2.

أوجد متوسط وتباين المجتمع؟

أحسب عدد العينات الممكنة ومعالم العينة$\left(\boldsymbol{\delta}_{\bar{x}}^{2}; \boldsymbol{\mu}_{\bar{x}}\right)$ باستعمال طريقة الجداول)، وقارن بين معالم المجتمع ومعالم العينة في الحالات التالية:

في حالة السحب بالإرجاع؟

في حالة السحب بدون إرجاع؟

التمرين2:

تدرس شركة طيران إمكانية السماح بحمولة يدوية للزبون مجانية، وقد وجد أن الوزن المتوسط للحمولة بالكيلوغرام هو $\mu = 5$ وانحراف معياري $\delta = 0.5$، إذا أخذت عينة من 100 راكب.

ما هو المتوسط المتوقع للحمولة اليدوية $\mu_{\bar{x}}$ في العينة والانحراف المعياري $\delta_{\bar{x}}$ ؟.

أحسب احتمال أن يكون الوزن الإجمالي للأمتعة:

أ- محصور بين 500 و 515كلغ؟ ب- أقل من 515كلغ؟.

التمرين3:

1- أوجد توزيع المعاينة للمتوسط الحسابي لعينة من الحجم 3 مسحوبة من المجتمع التالي: 4، 6، 8، 5، 7، و ذلك في الحالات التالية:

- السحب بالإرجاع؟

- السحب بدون إرجاع؟

2- إذا كان لدينا متغير عشوائي x يتبع التوزيع الطبيعي بمتوسط $\mu=80$ وتباين $\delta^2=49$.

- أوجد توزيع المعاينة لمتوسط العينة من الحجم 25 مسحوبة من المجتمع؟

- أحسب $p(\bar{x}>78)$ ؟

3- إذا كان لدينا متغيران عشوائيان مستقلان X_1 و X_2 حيث X_1 يتبع التوزيع الطبيعي بمتوسط يساوي 12 و تباين يساوي 40، و X_2 يتبع التوزيع الطبيعي بمتوسط يساوي8 وتباين يساوي 30.

- ما هو التوزيع العيني للفرق بين متوسطي عينتين الأولى من المجتمع(1) حجمها 20 والثانية من المجتمع(2) حجمها 15؟

- ما هو احتمال أن لا يزيد الفرق بين المتوسطين عن 4؟

4- نظرية التقدير

المقصود بالتقدير هو تقدير معالم المجتمع الإحصائي (أو التوزيع الاحتمالي) والتي غالباً ما تكون مجهولة ويكون المطلوب هو الحصول على تقديرات لها -عادة- من بيانات العينة فقد يكون المطلوب تقدير متوسط دخل الدولة، أو تقدير متوسط عمر الناخب... وغيرها.

وهناك نوعان (أو أسلوبان) للتقدير يسمى الأول تقدير النقطة (أو القيمة الواحدة)، ويسمى الثاني التقدير بمجال (أو فترة التقدير).

ففي حالة تقدير النقطة نحصل على قيمة واحدة فقط، يأخذها الثابت الإحصائي المقدر بدلالة التابع الإحصائي المقابل والمحسوب من تلك العينة المسحوبة من المجتمع الإحصائي، والمراد تقدير أحد ثوابته الإحصائي أي تقدير الثابت بقيمة واحدة مثلا : يكون أفضل تقدير للوسط الحسابي للمجتمع الإحصائي هو الوسط الحسابي المحسوب من عينة عشوائية سحبت من ذلك المجتمع على اعتبار أن هناك احتمالا كبيرا ليكون الوسط الحسابي للعينة قريبا جدا من الوسط الحسابي للمجتمع غير المعلوم، وما ينطبق على الوسط الحسابي ينطبق على الثوابت الإحصائي الأخرى

أما في التقدير بمجال أو فترة التقدير فنحصل على مدى Range أو فترة تتحدد بحدين (حد أدنى وحد أعلى) - نحصل عليهما من العينة. ونلاحظ هنا أن التقدير بمجال (أو

تقدير الفترة) تحتوي على أكثر من قيمة بل قد يكون عدد القيم غير محدود أو لا نهائياً في كثير من الحالات. فمثلاً : إذا قدرنا أن الوسط الحسابي لأعمار الناخبين يتراوح بين:20 سنة كحد أدنى و 60 سنة كحد أعلى نكون قد حصلنا على تقدير مجال للوسط الحسابي لأعمار الناخبين في المجتمع – ونلاحظ أن هذا المجال (20،60) يحتوي على عدد لا نهائي من الأعمار، بمعنى أن العدد لا يقتصر فقط على الأعداد الصحيحة والتي تشمل السنوات، ولكنها تشمل أيضا كسور السنوات، والأيام والشهور، والساعات... الخ وسوف نرى كيف نحدد التقدير بمجال في بعض الحالات.

وتتميز التقديرات بمجال بالإضافة إلى أنها تحتوي على عدد كبير جداً من القيم، بأنه يمكن حساب احتمال أن يكون التقدير صحيحاً، وبالتالي فإنه يمكن معرفة مدى دقة التقديرات. لذا فإن التقدير بمجاليسمى أيضاً " مجال الثقة " لأن هذه الفترات تعتمد في تكوينها الإحصائي على درجات أو مستويات ثقة معينة مثل 95% أو 99 % وغيرها، بمعنى أن احتمال أن تكون فترة التقدير صحيحة هو 0.95 أو 0.99 وهكذا...

فإذا كان متوسط أعمار الناخبين يتراوح ما بين 20 و60 سنة، ودرجة الثقة هي 95 % فإن هذا معناه أنه لو تكررت التجربة مائة مرة، فإن التقدير سيكون محصوراً بين هذين الرقمين في 95 من الحالات (أي احتمال أن يكون صحيحاً هو 95%).

4–1– مفاهيم أساسية

4–1–1– بعض خصائص المقدر

لتقدير معلمة من معالم مجتمع محل دراسة، نحتاج إلى اختيار الإحصائية المناسبة في العينة لتقدير هذه المعلمة. غالبا ما تكون المعلمة المناظرة في العينة هي أحسن مقدر، كأن نقدر متوسط المجتمع μ من خلال متوسط العينة $\mu_{\bar{x}}$. تسمى الإحصائية المستخدمة في التقدير المقدر.

– **المقدر غير المتحيز**

نقول عن إحصائية ما بأنها مقدر غير متحيز sans biais لمعلمة المجتمع إذا كان متوسطها أو توقعها الرياضي مساويا لمعلمة المجتمع.

مثال1: نقول عن متوسط العينة $\bar{x}$ أنه مقدر غير متحيز لمتوسط المجتمع μ لأن $E(\overline{X}) = \mu$.
في المقابل نسمي الإحصائية S^2 في معاينة بالإرجاع أنها مقدر متحيز لـσ^2 لأن $E(S^2) = \delta^2 \frac{(n-1)}{n} \neq \delta^2$.

– **الكفاءة:**

تتعلق كفاءة (efficacité) مقدر ما بمقدار التباين لتوزيع المعاينة للإحصائية، فإذا كان لمقدرين (إحصائيتين) نفس المتوسط نقول عن المقدر ذو توزيع المعاينة الأقل تباينا أنه الأكثر كفاءة.

من البديهي أن استخدام مقدرات فعالة وغير متحيزة هو الأفضل، إلا أنه قد يلجأ لمقدرات أخرى لسهولة الحصول عليها.

– **التقارب convergeance**

نقول عن مقدر أنه متقارب إذا كان يؤول إلى قيمة المعلمة المقدرة عندما يؤول حجم العينة إلى ما لا نهاية.

مثال2: يعتبر متوسط العينة مقدرا متقاربا لمتوسط المجتمع لأن:

$$E(\overline{X}) = \mu \text{ ، } V(\bar{x}) = \delta_{\bar{x}}^2 = \frac{\delta^2}{n} \xrightarrow[n\to\infty]{} 0$$

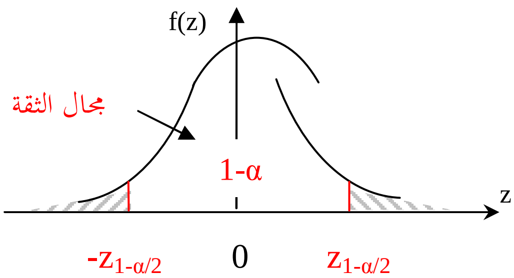
4-1-2 درجة التأكد

لكي يكون التقدير علميا ينبغي تقييم احتمال أن تكون المعلمة تنتمي فعلا إلى المجال المحدد، لذلك نلحق بالمجال ما يسمى بدرجة أو مستوى الثقة،ويرمز له ب $(1-\alpha)$. الاحتمال المعاكس يسمى احتمال الخطأ ويرمز له ب α، ويسمى أيضا ""مستوى المعنوية".

مثال3: دخل الأسرة في المنطقة(A) ينتمي إلى المجال [16000، 20000]بمستوى معنوية 5 % أي بمستوى ثقة 95 % . وتسمى الحدود 16000 و 20000بحدود الثقة.

4-1-3- تعيين حدود مجال الثقة

تحدد حدود الثقة من خلال معاملات الثقة التي بدورها تحدد من خلال مستوى المعنوية (مستوى الثقة). ففي حالة استخدام التوزيع الطبيعي للتقدير تكون القيمتين ± 1.96 معاملات الثقة من أجل مستوى ثقة 95% بينما القيمتين ±2.58 تمثلان معاملات الثقة من أجل مستوى ثقة 99 % .



مجال الثقة للتوزيع الطبيعي

مثال4:ليكن μ_s وσ_s متوسط وانحراف معياري توزيع المعاينة لإحصائية ما s حيث

$\mu_s = \mu$. إذا كان توزيع المعاينة ل s توزيعا طبيعيا (كما هو الحال بالنسبة لأغلب الإحصائيات عندما$(n \geq 30)$) فإننا نقدر مثلا وبالنظر إلى توزيع s أن:

القيمتين $\mu_s \pm 1.96\sigma_s$ تمثلان **حدود الثقة** ب 95%، و $\mu_s \pm 2.58\sigma_s$ حدود الثقة ب 99%.

في حالة التوزيع الطبيعي يرمز لحدود الثقة ب Z_c أو $Z_{1-\alpha/2}$ (أنظر الرسم).

4-2- التقدير بمجال

4-2-1- مجال الثقة للمتوسط

يقدر متوسط المجتمع μ من خلال الإحصائية $\bar{x}$.

4-2-1-1- تقدير μ باستخدام التوزيع الطبيعي

إذا كان المجتمع المسحوب منه العينة ذا توزيعاً طبيعياً وتباينه معروفاً أو كانت العينة كبيرة (أي حجمها ثلاثون مفردة أو أكثر) فإن توزيع المعاينة للوسط الحسابي يكون توزيعاً طبيعياً، وطالما نحن نتكلم عن تقدير متوسط المجتمع فإن أول ما نفكر فيه هو الوسط الحسابي للعينة. ومجال الثقة للوسط الحسابي للمجتمع يأخذ الشكل التالي :

تقدير متوسط المجتمع = الوسط الحسابي للعينة ± الخطأ المعياري للوسط وبالرموزفإن:

$$IC_{\mu} = \left[\bar{x} \pm 1\delta_{\bar{x}}\right]$$

حيث $\bar{x}$ هو الوسط الحسابي للعينة، $\delta_{\bar{x}}$ هو الخطأ المعياري للوسط، + تشير للجمع فنحصل على الحد الأعلى لفترة التقدير، – تشير للطرح فنحصل على الحد الأدنى، ولكن احتمال أن يكون هذا الكلام صحيحا هو % 68.26 فقط، أي أن درجة الثقة هنا لا تتعدى % 68.26 فإذا أضفنا وطرحنا ضعف الخطأ المعياري يرتفع الاحتمال إلى 95.44% أي ترتفع درجة الثقة إلى % 95.44 وفي هذه الحالة يأخذ مجال الثقة الشكل التالي :

$$IC_{\mu} = \left[\bar{x} \pm 2\delta_{\bar{x}}\right]$$

وإذا أضفنا وطرحنا ثلاثة أمثال الخطأ المعياري يصبح الاحتمال % 99.72 أي ترتفع درجة الثقة إلى 99.72% وتأخذ فترة الثقة الشكل التالي : $IC_{\mu} = \left[\bar{x} \pm 3\,\delta_{\bar{x}}\right]$

أي أنه بزيادة درجة الثقة يزيد طول الفترة. ومما سبق نلاحظ ما يلي:

– أن هناك علاقة وثيقة بين درجة الثقة والرقم أو " المعامل المضروب في الخطأ المعياري فهو إما 1 أو 2 أو 3 على حسب درجة الثقة 68.26% أو 95.44% أو % 99.72 ولذلك فإن هذا المعامل هو الذي يسمى " معامل الثقة ". فبناء على درجة الثقة المطلوبة يتحدد معامل الثقة.

– أن درجات ومعاملات الثقة التي ذكرناها تخص التوزيع الطبيعي، وأن المعاملات 1 أو 2 أو 3 ما هي إلا الدرجة المعيارية (Z) والتي نحصل عليها من جدول التوزيع الطبيعي المعياري وذلك بقسمة درجة الثقة (أو الاحتمال) على 2 (حيث أن المساحة موزعة بالتساوي على يمين ويسار الوسط) ثم بالكشف في المساحات تحت المنحنى الطبيعي المعياري عن خارج القسمة (أو أقرب رقم له) فنحصل على Z المقابلة وهذا يرجع إلى أن توزيع المعاينة للوسط هو التوزيع الطبيعي.

– يمكن الحصول على مجال الثقة بأي درجة ثقة أخرى كما يلي:

لدينا لما: $X \to N(\mu,\delta^2)$ فان $\bar{x} \to N\left(\mu,\frac{\delta^2}{n}\right)$ وكذلك $\frac{\bar{x}-\mu}{\frac{\delta}{\sqrt{n}}} \to N(0,1)$ ، ومنه نعرف مجال الثقة لـ$(1-\alpha)$ أو نقول مجال الثقة عند مستوى معنوية α بأنه المجال الذي يحقق ما يلي:

$$p\left(-Z_{\frac{\alpha}{2}} < \frac{\bar{x}-\mu}{\frac{\delta}{\sqrt{n}}} < Z_{\frac{\alpha}{2}}\right) = 1-\alpha$$

وبتبسيط العبارة نحصل على ما يلي:

$$\bar{x} - Z_{\frac{\alpha}{2}} \frac{\delta}{\sqrt{n}} < \mu < \bar{x} + Z_{\frac{\alpha}{2}} \frac{\delta}{\sqrt{n}}$$

ملاحظة: نلاحظ أن المجال مركزه $\bar{x}$ وحيث $\bar{x}$ متغير عشوائي سيتغير مركز المجال بتغير العينة وبالتالي سيتغير مجال الثقة، ولكن كل المجالات تشترك في أن وسط العينة μ يقع داخل كل مجال بنسبة $(1-\alpha)$.

خلاصة: أن مجال الثقة للوسط الحسابي للمجتمع في حالة δ معلوميأخذ الشكل التالي:

$$IC_{\mu} = \left[\bar{x} \pm z_c \delta_{\bar{x}}\right]$$

حيث: $\delta_{\bar{x}} = \frac{\delta}{\sqrt{n}}$

ومنه يصبح مجال الثقة كما يلي:

$$IC_{\mu} = \left[\bar{x} \pm z_c \frac{\delta}{\sqrt{n}}\right]$$

ويمكن تلخيص خطوات تقدير الوسط الحسابي للمجتمع فيما يلي:

–حساب الوسط الحسابي للعينة $\bar{x}$.

– حساب الخطأ المعياري للوسط $\delta_{\bar{x}}$ والذي يساوي $\frac{\delta}{\sqrt{n}}$

– ضرب الخطأ المعياري للوسط في معامل الثقة المناسب (أو الدرجة المعيارية) حسب درجة الثقة المطلوبة أي أحسب $Z\frac{\sigma}{\sqrt{n}}$.

– طرح حاصل الضرب السابق من الوسط الحسابي للعينة فنحصل على الحد الأدنى لفترة التقدير، واجمع حاصل الضرب مرة أخرى على الوسط الحسابي للعينة فتحصل على الحد الأعلى لفترة التقدير.

– إذا كان الانحراف المعياري للمجتمع δ غير معروف (مجهول)– وهو غالبا ما يحدث في الواقع فيمكن استخدام الانحراف المعياري للعينة S أو الانحراف المعياري المعدل للعينة $\hat{S}$ بدلا منه طالما كان حجم العينة كبيرا بدرجة كافية $(n \geq 30)$ ويصبح مجال الثقة للوسط الحسابي للمجتمع كما يلي[3] :

$$IC_{\mu} = \left[\bar{x} \pm z_c \frac{\hat{S}}{\sqrt{n}} \right] \quad \text{أو} \quad IC_{\mu} = \left[\bar{x} \pm z_c \frac{S}{\sqrt{n-1}} \right]$$

ملاحظة: جميع الصيغ السابقة تستعمل في حالة المعاينة بالإرجاع، أما في حالة المعاينة بدون إرجاع تصبح الصيغة كالآتي:

$$IC_{\mu} = \left[\bar{x} \pm z_c \frac{\delta}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \right]$$

وفي حالة ما يكون الانحراف المعياري للمجتمع δ مجهولا، نعوض δ في الصيغة السابقة بالمقدر S أو $\hat{S}$.

– يمكن كتابة أشهر وأهم درجات ومعاملات الثقة z_c (للتوزيع الطبيعي) في الجدول التالي (مع ملاحظة أن 95%، 99% هي أشهرها على الإطلاق)

0.5	0.8	0.90	0.95	0.98	0.99	مستوى الثقة $1-\alpha$
0.5	0.2	0.10	0.05	0.02	0.01	مستوى المعنوية α
0.75	0.9	0.95	0.975	0.99	0.995	$1-\alpha/2$
0.674	1.282	1.645	1.96	2.326	2.58	$Z_{1-\alpha/2}$

مثال5: نقدر أن μ يوجد داخل المجال $\left[\bar{x} \pm 1.96\delta_{\bar{x}}\right]$ بمستوى ثقة 95% (0.95) أي بمستوى معنوية 5 % (0.05)....

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

مثال6: لو أردنا معرفة متوسط الدخل اليومي لمجموعة من الناخبين في دولة ما، فإن ذلك يبدو أمرا صعباً من الناحية العملية نظراً لكبر حجم مجتمع الناخبين، إضافة إلى طول الوقت والتكاليف. لذا فإن الأسلوب العلمي المتبع في حالة كهذه هو اختيار عينة عشوائية نستطيع من خلال معرفة نتائجها تقدير متوسط دخول الناخبين في هذه الدولة.

فلو سحبت عينة عشوائية من مجموع مجتمع الناخبين في دولة ما حجمها 101 ناخب فإذا كان الوسط الحسابي والانحراف المعياري للدخل اليومي للناخبين بالعينة هما على الترتيب 90 دولار و 25 دولار، أوجد مجال الثقة للوسط الحسابي للدخل اليومي لمجموع الناخبين في هذه الدولة بمستوى ثقة 95% ؟

الحل :

لدينا مجال الثقة للوسط الحسابي للمجتمع هو:

$$IC_{\mu} = \left[\bar{x} \pm z_c \frac{\delta}{\sqrt{n}} \right]$$

بما أن δ الانحراف المعياري للمجتمع مجهولة، نستعمل في هذه الحالة الانحراف المعياري للعينة، ومنه يصبح مجال الثقة يكتب على الشكل التالي:

$$IC_{\mu} = \left[\bar{x} \pm z_c \frac{S}{\sqrt{n-1}} \right]$$

لدينا حجم العينة $n = 101$ والوسط الحسابي للعينة $\bar{x} = 90$، والانحراف المعياري للعينة $S = 25$.

وبما أن مستوى الثقة هي % 95 فإن : $z_c = 1.96$ حسب ما هو موضح في الجدول السابق. وبالتالي فإن فترة تقدير الوسط الحسابي للدخل اليومي لمجتمع الناخبين بمستوى ثقة % 95 هي :

$$IC_{\mu} = \left[\bar{x} \pm z_c \frac{S}{\sqrt{n-1}} \right] = \left[90 \pm 1.96 \frac{25}{\sqrt{101-1}} \right]$$
$$= [90 \pm 1.96 (2.5)] = [90 \pm 4.9] = [(85.1), (94.9)]$$

– أي أن الوسط الحسابي للدخل اليومي لمجتمع الناخبين يتراوح بين 85.1 دولاراً كحد أدنى، 94.9 كحد أعلى، وذلك بمستوى ثقة % 95.

4–2–1–2– تقدير متوسط المجتمعμ 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 عليهتوزيع 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.

4–2–1–3– 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).

<u>خلاصة</u>: إنمجال الثقة للوسط الحسابي للمجتمع في حالة δ مجهولة وحجم العينة أقل من 30يأخذ الشكل التالي:

$$IC_{\mu} = \left[\bar{x} \pm t_{\frac{\alpha}{2},n-1} \frac{\hat{S}}{\sqrt{n}} \right] \quad \textbf{أو} \quad IC_{\mu} = \left[\bar{x} \pm t_{\frac{\alpha}{2},n-1} \frac{S}{\sqrt{n-1}} \right]$$

ولعل من أهم الملاحظات على المعادلة الإحصائية السابقة احتواؤها على مفهومين مهمين هما :

- مستوى المعنوية أو الدلالة والذي رمزنا له بالرمز اللاتيني ألفا α والذي يعني أنه المكمل لدرجة الثقة أي نسبة الخطأ. وبالتالي فإذا كانت درجة الثقة % 99 أي احتمال أن يكون التقدير صحيحاً بنسبة % 99، فإن مستوى المعنوية، والذي يعني هنا درجة احتمال الخطأ يساوي 1%. وعند الكشف في جدول(t)، ولأنه توزيع متماثل، فإنه يتم قسمة مستوى المعنوية على 2.

- درجات الحرية، وهو ما سبق شرحه ويساوي في هذه الحالة n - 1. حيث n هو حجم العينة وطرحنا منه 1 لأنه تم تقدير الانحراف المعياري للمجتمع المجهول باستخدام الانحراف المعياري للعينة S أو الانحراف المعياري المعدل للعينة $\hat{S}$.

مثال7: إذا كانت دخول مجموعة من الأفراد في دولة ما تتبع التوزيع الطبيعي، وسحبت منهم عينة عشوائية حجمها 10 أفراد بوسط حسابي72 دولاراً وانحراف معياري بلغ6.4 دولار، أوجد مجال الثقة للوسط الحسابي للدخل اليومي لجميع الأفراد بمستوى ثقة 95%؟.

الحل :

نلاحظ: أن العينة صغيرة (حجمها $n=10$ أفراد فقط) وأن المجتمع طبيعي وانحرافه المعياري غير معروف (مجهول) لذلك نستخدم مجال الثقة التالي:

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

وحيث أن n = 10 فإن درجات الحرية لها هي :

n -1 = 10 - 1 = 9

ولدينا مستوى الثقة المطلوبة هي $1-\alpha = 0.95$ فإن مستوى المعنوية $\alpha = 0.05$ وبالتالي فإن نصف مستوى المعنوية هو : $\frac{\alpha}{2} = \frac{0.05}{2} = 0.025$

أي يتم الكشف في جدول ستيودنت عند درجات حرية تساوي 9 تحت احتمال (نصف مستوى المعنوية) 0.025 أي أن : $t_{0.025,9} = 2.262$

وبالتعويض في مجال الثقة للوسط نحصل على ما يلي:

$$IC_{\mu} = \left[\overline{x} \pm t_{\frac{\alpha}{2},n-1} \frac{S}{\sqrt{n-1}} \right] = \left[72 \pm 2.262 \frac{6.4}{\sqrt{10-1}} \right]$$
$$= \left[72 \pm 2.262 \frac{6.4}{3} \right] = [72 \pm 4.82]$$
$$= [(67.18), (76.82)]$$

أي أن الوسط الحسابي للدخول اليومية يتراوح بين67.18دولاراً كحد أدنى، و 76.82 دولاراً كحد أعلى وذلكبمستوى ثقة 95%.

4-2-1-4- تحديد حجم العينة لتقدير الوسط الحسابي للمجتمع :

يعتبر تحديد حجم العينة المناسب منالمشاكل المهمة والشائعة التي تواجه الباحثين في مختلف المجالات، وبالذات عند دراسة الظواهر الاقتصادية، ويختلف تحديد حجم العينة باختلاف الهدف من التقدير.

فإذا كان المطلوب هو تقدير الوسط الحسابي للمجتمع، فإن فترة تقدير الوسط هي كما سبق وأن أوضحنا :

$$IC_{\mu} = \left[\overline{x} \pm z \frac{\delta}{\sqrt{n}} \right]$$

ومنه نجد أن حجم العينة يأخذ الشكل التالي:

$$n = \frac{z^{2} \delta^{2}}{e^{2}}$$

حيث: e هو أقصى خطأ مسموح به في تقدير الوسط، وهو عادة ما يحدده الباحث، وتتوقف على أهمية الموضوع أو الظاهرة السياسية المراد دراستها، ومدى الدقة المطلوبة في التقدير، ويسمى اختصاراً "الخطأ في تقدير الوسط".

مثال8:

إذا كانت دخول الأفراد اليومية في إحدى دول العالم النامية تتبع التوزيع الطبيعي بانحراف معياري $\delta = 15$ دولاراً، فما هو حجم العينة المناسب لتقدير متوسط دخول الأفراد في هذه الدولة بحيث لا يتعدى الخطأ في تقدير متوسط الدخل اليومي 5 دولارات، وذلك بمستوى ثقة 99 % ؟

<u>الحل :</u>

في هذا المثال نجد أن :

درجة الثقة 99 % أي أن : Z = 2.58

أقصى خطأ مسموح به هو 5 دولارات، أي أن : e = 5

والانحراف المعياري للمجتمع : $\delta = 15$

وبالتعويض بهذه القيم في المعادلة التي تحدد حجم العينة وهي:

$$n = \frac{z^2 \delta^2}{e^2} = \frac{(2.58)^2 (15)^2}{5} = 59.9 \approx 60$$

فإن حجم العينة مقرباً لأقرب عدد صحيح هو: 60فردا.

أي أنه يجب على الباحث أن يأخذ عينة لا يقل حجمها عن 60 فرداً حتى يكون لديه تقديراً دقيقاً عن متوسط دخول الأفراد في هذه الدولة بحيث لا يتعدى الخطأ في تقديره لمتوسط الدخل عن خمس دولارات، وذلك بدرجة ثقة 99 %.

4-2-2- مجال الثقة للنسبة وحجم العينة المناسب.

4-2-2-1- مجال الثقة للنسبة.

إن تقدير النسبة في المجتمع تعتبر من الحالات المهمة لقياس الظواهر الاقتصادية، وبالذات التحليلية منها كتحليل اتجاهات النمو الاقتصادي، وقياس نسبة مواليد العام، ونسبة الدول التي لم توفي بالتزاماتها في المنظمات الدولية أو الإقليمية... وغيرها ونظراً لأنه من الصعوبة بمكان في كثير من الأحيان حساب هذه النسبة مباشرة من المجتمع، فإننا غالبا ما نلجأ لتقدير هذه النسبة من عينة عشوائية مسحوبة من هذا المجتمع.

فلو افترضنا أن نسبة المؤيدين للسياسة الاقتصادية التي تنتهجها دولة ما هي p وأن العينة العشوائية كبيرة بدرجة كافية$(n \geq 30)$و المعاينة بالإرجاع وأن نسبة مؤيدي هذه السياسة في العينة هي p' فإن مجال الثقة للنسبة في المجتمع يكتب كما يلي:

$$IC_p = \left[p^{/} \pm z_c \delta_{p^{/}} \right]$$

ولدينا: $\delta_{p^{/}} = \sqrt{\frac{pq}{n}}$، ومنه يصبح مجال الثقة كما يلي: $IC_p = \left[p^{/} \pm z_c \sqrt{\frac{pq}{n}} \right]$

وبما أن p مجهولة ونريد إيجاد مجال الثقة لها ومنه لحساب $\delta_{p^{/}}$ نستبدل p بدلالة $p^{/}$ النسبة في العينة وبذلك يصبح مجال الثقة يكتب كما يلي:

$$IC_p = \left[p^{/} \pm z_c \sqrt{\frac{p^{/} q^{/}}{n}} \right]$$

أما في حالة كون المجتمع محدودا ذا حجم N والمعاينة بدون إرجاع، فإننا نضرب في معامل الإرجاع، ومنه يصبح يكتب مجال الثقة للنسبة في المجتمع من الشكل التالي:

$$IC_p = \left[p^{/} \pm z_c \sqrt{\frac{p^{/} q^{/}}{n}} \sqrt{\frac{N - n}{N - 1}} \right]$$

مثال9: عينة عشوائية حجمها 144 ناخباً سحبت من إحدى المدن فوجد أن عدد المؤيدين في العينة لمرشح معين هو 60 ناخباً، أوجد مجال الثقة لنسبة المؤيدين لهذا المرشح في المدينة كلها بمستوى ثقة % 95؟.

الحل :

– نحسب أولاً نسبة المؤيدين للمرشح في العينة $p^{/}$ التي نحصل عليها بقسمة عدد المؤيدين له على العدد الكلي للعينة (حجم العينة) أي أن:

$$p^{/} = \frac{60}{144} = 0.42$$

ولدينا مستوى الثقة المطلوبة يساوي % 95، ومنه معامل الثقة المناسب هو: $z_c = 1.96$ ومجال الثقة لنسبة المؤيدين لهذا المرشح في المدينة تأخذ الشكل التالي :

$$IC_p = \left[p' \pm z_c \sqrt{\frac{p'q'}{n}} \right]$$

وبالتعويض بحجم العينة والانحراف المعياري للعينة والنسبة في العينة نجد مجال الثقة كما يلي:

$$IC_p = \left[p' \pm z_c \sqrt{\frac{p'q'}{n}} \right] = \left[0.42 \pm 1.96 \sqrt{\frac{(0.42)(0.58)}{144}} \right]$$
$$= [0.42 \pm 0.08] = [(0.34),(0.50)]$$

أي أن نسبة المؤيدين للمرشح في المدينة تتراوح بين 0.34 , 0.50 وذلك بدرجة ثقة% 95 بمعنى آخر أن نسبة مؤيدي هذا المرشح في هذه المدينة لا تتجاوز % 50 كحد أعلى، وبالتالي ففرصته في الفوز كمرشح قد لا تكون كبيرة وذلك بدرجة ثقة % 95 بمعنى أن هذا الحكم لا تتجاوز نسبة الخطأ فيه % 5.

4-2-2-2-تحديد حجم العينة لتقدير النسبة في المجتمع :

وبالطريقة نفسها يمكن تحديد حجم العينة اللازمة للحصول على درجة ثقة معينة عند تقدير النسبة في المجتمع بافتراض أن أقصى خطأ في التقدير مسموح به هو e تبعاً حيث يكتب حجم العينة كما يلي:

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

أي أن حجم العينة المناسب في هذه الحالة يساوي حاصل ضرب مربع z في النسبة، ثم في النسبة المكملة مقسوماً على مربع الخطأ المسموح به.

مثال10:

يدعي أحد مراكز استطلاعات الرأي العام أنه عند دراسته لاتجاهات آراء الناخبين لاثنين من المتنافسين على أحد مقاعد السلطة التشريعية بأن نتائج دراسته هي من الدقة بحيث لا يتعدى نسبة الخطأ في التقدير 2%، فما هو حجم العينة المناسب التي نستطيع من خلالها الحكم على مدى صحة إدعاء هذا المركز بافتراض أن نسبة المؤيدين للمرشح هي 50 % وذلك بدرجة ثقة 95 %؟.

الحل :

بما أن درجة الثقة 95 % فإن: $z = 1.96$ بافتراض أن نسبة المؤيدين للمرشح هي: $p = 0.5$

وبالتالي فإن النسبة المكملة $q = 1 - p$ هي : $q = 1 - p = 1 - 0.5 = 0.5$

ولدينا أقصى خطأ مسموح به هو: $e = 0.02$

ومنه فان حجم العينة اللازم هو:

$$n = \frac{z^2 . p(1 - p)}{e^2} = \frac{(1.96)^2 (0.5)(0.5)}{(0.02)^2}$$
$$= \frac{0.9604}{0.0004} = 2401$$

أي أن حجم العينة المناسب الذي يعطي درجة الدقة المطلوبة هو 2401 ناخب. بمعنى آخر فإن على هذا المركز أن يستطلع حجم عينة لا يقل عددها عن هذا العدد.

4-3-مجال الثقة للتباين ونسبة تباينين.

4-3-1- مجال الثقة للتباين.

لتقدير التباين والانحراف المعياري لمجتمع بمجال ثقة نستعمل الخاصية :

$$\cdot \frac{nS^2}{\sigma^2} = \frac{(n-1)\hat{S}^2}{\sigma^2} \sim \chi^2{}_{n-1}$$

$$\frac{nS^2}{\chi_{\left(\frac{\alpha}{2},n-1\right)}} = \frac{(n-1)\hat{S}^2}{\chi_{\left(\frac{\alpha}{2},n-1\right)}} < \delta^2 < \frac{nS^2}{\chi_{\left(1-\frac{\alpha}{2},n-1\right)}} = \frac{(n-1)\hat{S}^2}{\chi_{\left(1-\frac{\alpha}{2},n-1\right)}}$$

مثال11: مجال الثقة ب95 % يحدد كما يلي:

$$\chi^2{}_{0.025} \le \frac{nS^2}{\sigma^2} = \frac{(n-1)\hat{S}^2}{\sigma^2} \le \chi^2{}_{0.975}$$

ومنه نستنتج مجال الثقة ل σ كما يلي:

$$\frac{\sqrt{n}\,S}{\chi_{0.975}} \le \sigma \le \frac{\sqrt{n}\,S}{\chi_{0.025}} \qquad ou \qquad \frac{\sqrt{n-1}\,\hat{S}}{\chi_{0.975}} \le \sigma \le \frac{\sqrt{n-1}\,\hat{S}}{\chi_{0.025}}$$

مثال12: أوجد مجال الثقة لتباين مجتمع سحبت منه عينة حجمها 6 وتباينها المعدل $\hat{S}^2 = 11.8$ بمستوى معنوية $\alpha = 0.05$؟.

الحل:

لدينا من جدول توزيع كاي تربيع χ^2 أن:

$$\chi^2_{(0.025,5)} = 12.833;;;\chi^2_{(0.975,5)} = 0.831$$

فيكون مجال الثقة كما يلي:

$$\frac{(n-1)\hat{S}^2}{\chi_{\left(\frac{\alpha}{2},n-1\right)}} < \delta^2 < \frac{(n-1)\hat{S}^2}{\chi_{\left(1-\frac{\alpha}{2},n-1\right)}}$$

$$\frac{(5)(11.8)}{12.833} < \delta^2 < \frac{(5)(11.8)}{0.831}$$

$$4.6 < \delta^2 < 71$$

$$IC_{\delta^2} = [(4.6);(71)]$$

4 -3-2- مجالات الثقة لنسبة تباينين

رأينا سابقاأنه إذا كان لدينا مجتمعان طبيعيان تبايناهما σ^2_1 , σ^2_2 وسحبنا منهما عينتين عشوائيتين حجمهما على التوالي n_1 , n_2 فإن :

$$F = \frac{\left[\frac{S_1^2 n_1}{n_1 - 1}\right]\frac{1}{\sigma_1^2}}{\left[\frac{S_2^2 n_2}{n_2 - 1}\right]\frac{1}{\sigma_2^2}} = \frac{\hat{S}_1^2 / \sigma_1^2}{\hat{S}_2^2 / \sigma_2^2} \rightarrow F_{n_1-1;\ n_2-1}$$

ومنه مجال الثقة يكتب كما يلي:

$$\frac{\hat{S}_1^2}{\hat{S}_2^2}\frac{1}{F_{\left(\frac{\alpha}{2}, n_1-1, n_2-1\right)}} \leq \frac{\delta_1^2}{\delta_2^2} \leq \frac{\hat{S}_1^2}{\hat{S}_2^2}\frac{1}{F_{\left(1-\frac{\alpha}{2}, n_1-1, n_2-1\right)}}$$

إذا يمكن تكوين تقدير بمجال لـFعند مستوى ثقة 0.98 كما يلي:

$$F_{0.01} \leq \frac{\hat{S}_1^2 / \sigma_1^2}{\hat{S}_2^2 / \sigma_2^2} \leq F_{0.99}$$

و من ثم يمكن تقدير النسبة بين تبايني المجتمعين كما يلي:

$$\frac{1}{F_{0.99}}\frac{\hat{S}_1^2}{\hat{S}_2^2} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{1}{F_{0.01}}\frac{\hat{S}_1^2}{\hat{S}_2^2}$$

- وتكمن أهمية إيجاد مجال الثقة في بحث وجود تجانس بين المجتمعين، فكلما كانت النسبة قريبة من 1 كان المجتمعان أكثر تجانسا أي لهما نفس التباين تقريبا.

4-4- مجالات الثقة للفروق والمجاميع

إذا كانت s_1 وs_2 إحصائيتا معاينة لها توزيع يقترب من التوزيع الطبيعي، والعينتان مستقلتان، تكتب حدود الثقة للفروق بين المعالم التي تمثلها الإحصائيتين كما يلي:

$$S_1 - S_2 \ \pm \ z_c \cdot \sigma_{S_1 - S_2} = S_1 - S_2 \ \pm \ z_c \cdot \sqrt{\sigma^2{}_{S_1} + \sigma^2{}_{S_2}}$$

في حالة المجموع :

$$S_1 + S_2 \ \pm \ z_c \cdot \sigma_{S_1 - S_2} = S_1 + S_2 \ \pm \ z_c \cdot \sqrt{\sigma^2{}_{S_1} + \sigma^2{}_{S_2}}$$

مثال13: إذا كانت الإحصائيتان هما متوسطا عينتين مستقلتين، مسحوبتين من مجتمعينغير محدودين وتباينهما δ_1^2 و δ_2^2 معلومتين، نحدد مجال الثقة للفرق (و للمجموع) بين متوسطي المجتمعين $\mu_1 - \mu_2$ كما يلي :

$$\bar{x}_1 - \bar{x}_2 \pm z_c \delta_{\bar{x}_1 - \bar{x}_2} = \bar{x}_1 - \bar{x}_2 \pm z_c \sqrt{\frac{\delta_1^2}{n_1} + \frac{\delta_2^2}{n_2}}$$

- أما في حالة تباينا المجتمعين متساويين ومجهولين $\left(\delta_1^2 = \delta_2^2\right)$، فنحسب التباين المشترك S_p^2 أولا حيث يساوي:

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

ومنه يصبح مجال الثقة للفرق (و للمجموع) بين متوسطي المجتمعين $\mu_1 - \mu_2$ كما يلي:

$$\bar{x}_1 - \bar{x}_2 \pm t_{n_1 + n_2 - 2; \frac{\alpha}{2}} \left(S_p\right) \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

مثال14: إذا كانت الإحصائيتان هما نسبتان في عينتين مستقلتين، مسحوبتان من مجتمعين غير محدودين :

$$p'_1 - p'_2 \ \pm \ z_c \cdot \sigma_{p'_1 - p'_2} = p'_1 - p'_2 \ \pm \ z_c \cdot \sqrt{\frac{pq_1}{n_1} + \frac{pq_2}{n_2}}$$

4-5- طريقة المعقولية العظمى في التقدير(طريقة الاحتمال الأكبر):

نريد تقدير معلمة α للمجتمع، ولدينا عينة غير نفادية (المتغيرات التي تمثل قيم المحصل عليها في العينة مستقلة) لها نفس التوزيع للمجتمع، إن احتمال تحقق عينة بذاتها مرتبط بقيمة المعلمة المجهولة، هناك قيمة لـ α تعظم احتمال الحصول على العينة المحصل عليها، ونفترض أن تلك القيمة هي الصحيحة بما أن العينة حصلت بالفعل، تتمثل طريقة المعقولية العظمى في البحث عن هذه القيمة، أي البحث عن α التي تعظم $L(\underline{x},\alpha)$.

حيث $L(\underline{x},\alpha)$ ترمز لمعقولية العينة لتوزيع الاحتمال للشعاع العشوائي $(x_1, x_2, x_3,............x_n)$.

ملاحظة:

- نستطيع كتابة العبارة $L(\underline{x},\alpha)$ على الشكل التالي: $L(x_1, x_2, x_3,............x_n,\alpha)$.
- مقدر المعقولية العظمى ليس بالضرورة غير متحيز وكذلك ليس بالضرورة وحيد.

إن طريقة المعقولية العظمى تهدف إلى اختيار مقدر لـ α وهو $\widehat{\alpha}$ وهي القيمة الأكثر معقولية، إن التقدير المتحصل عليه هو القيمة الأكبر احتمالا من أجل القيم الملاحظة للعينة. والتقدير بطريقة المعقولية العظمى يعطى بتعظيم دالة المعقولية حيث:

$$L(x_1, x_2, x_3,............x_n, \alpha) = L(\underline{x}, \alpha) = \prod_{i=1}^{n} f(x_i, \alpha)$$

حيث: $f(x,\alpha)$ تمثل توزيع المجتمع.

- ولكي يكون $\widehat{\alpha}$ مقدر بطريقة المعقولية العظمى يجب أن يكون حل للمعادلة الآتية:

$$\begin{cases} \dfrac{d\ln L(x_1, x_2, x_3,..........x_n,\alpha)}{d\alpha} = 0 \\ \dfrac{d^2 \ln L(x_1, x_2, x_3,..........x_n,\alpha)}{d\alpha^2} < 0 \end{cases} \Leftrightarrow \begin{cases} \dfrac{dL(x_1, x_2, x_3,..........x_n,\alpha)}{d\alpha} = 0 \\ \dfrac{d^2 L(x_1, x_2, x_3,..........x_n,\alpha)}{d\alpha^2} < 0 \end{cases}$$

مثال15:ليكن x متغير عشوائي يتبع التوزيع ذات دالة الكثافة التالية:

$$f(x,\alpha) = \alpha e^{-\alpha x}$$

– أعطي مقدر $\hat{\alpha}$ لـ α باستعمال طريقة المعقولية العظمى؟

الحل:

لدينا لكي يكون $\hat{\alpha}$ مقدر بطريقة المعقولية العظمى يجب أن يكون حل للمعادلة الآتية:

$$\begin{cases} \dfrac{d \ln L(x_1, x_2, x_3, \ldots\ldots\ldots \ldots x_n, \alpha)}{d\alpha} = 0 \\ \dfrac{d^2 \ln L(x_1, x_2, x_3, \ldots\ldots\ldots \ldots x_n, \alpha)}{d\alpha^2} < 0 \end{cases}$$

لدينا دالة المعقولية العظمى تكتب كما يلي:

$$L(x_1, x_2, x_3, \ldots\ldots\ldots \ldots x_n, \alpha) = L(\underline{x}, \alpha) = \prod_{i=1}^{n} f(x_i, \alpha) = \prod_{i=1}^{n} \alpha e^{-\alpha x_i}$$

$$= \alpha^n e^{-\alpha \sum_{i=1}^{n} x_i}$$

$$\ln L(\underline{x}, \alpha) = n \ln \alpha - \alpha \sum_{1}^{n} x_i$$

$$\Rightarrow \frac{d \ln L(\underline{x}, \alpha)}{d\alpha} = \frac{n}{\alpha} - \sum_{1}^{n} x_i$$

$$\frac{d \ln L(\underline{x}, \alpha)}{d\alpha} = 0 \Rightarrow \frac{n}{\hat{\alpha}} - \sum_{1}^{n} x_i = 0$$

ولدينا:

$$\Rightarrow \frac{n}{\hat{\alpha}} = \sum_{1}^{n} x_i \Rightarrow \hat{\alpha} = \frac{n}{\sum_{1}^{n} x_i} = \frac{1}{\overline{x}}$$

$$\frac{d^2 \ln L(\underline{x}, \alpha)}{d\alpha^2} = \frac{-n}{\alpha^2} < 0$$

وبما أن الشرطين محققين فإننا نستنتج أن $\hat{\alpha} = \frac{1}{\overline{x}}$ هو مقدر لـ α.

4–6– تمارين محلولة:

التمرين1: ضمن خطتها لإصلاح حركة المرور في المدينة قامت ولاية جيجل بإجراء مسح ميداني لتحديد حجم الحركة عبر تقاطع رئيسي في المدينة خلال فترات الصباح، باختيار يوم السبت لثمانية أسابيع متوالية تم عد المركبات التي تمر عبر التقاطع بين الساعة السابعة والساعة التاسعة صباحا ووجد أن متوسط عدد المركبات للعينة يساوي 1500 والانحراف المعياري للعينة يساوي 300.

المطلوب: أوجد مجال الثقة لمتوسط عدد المركبات في مجتمع الحركة عبر التقاطع باعتبار أن توزيع المجتمع طبيعي؟ (مع العلم أن مستوى الثقة يساوي 99٪)...

حل التمرين1:

-إيجاد مجال الثقة لمتوسط عدد المركبات في مجتمع الحركة عبر التقاطع باعتبار أن توزيع المجتمع طبيعي:

لدينا عموما:

$$IC_{\mu} = \left[\bar{x} \pm z_c \delta_{\bar{x}}\right]$$

$$IC_{\mu} = \left[\bar{x} \pm z_c \frac{\delta}{\sqrt{n}}\right]$$

– بما أن حجم العينة صغيرة $n < 30$ وعدم معرفة الانحراف المعياري للمجتمع نستخدم التوزيع ستيودنت t لإيجاد مجال الثقة، ومنه يصبح مجال الثقة كما يلي:

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

– ولدينا: $\bar{x} = 1500;; t_{0.005,7} = 3.499;; S = 300;; n = 8$ بالتعويض نجد:

$$IC_{\mu} = \left[\overline{x} \pm t_{\frac{\alpha}{2},n-1} \frac{S}{\sqrt{n-1}}\right] = \left[1500 - (3.499)\left(\frac{300}{\sqrt{7}}\right); 1500 + (3.499)\left(\frac{300}{\sqrt{7}}\right)\right]$$
$$IC_{\mu} = [(1103.25);(1896.75)]$$

<u>التمرين02</u>: إذا كان لدينا متغيران عشوائيان x_1 وx_2يتبعان التوزيع الطبيعي بتباين متساوي ومتوسطات مختلفة، تم سحب عينتان عشوائيتان مستقلتان من المجتمعين محل الدراسة وكانت لدينا البيانات التالية:

	المجتمع01	المجتمع02
حجم العينة	$n_1 = 15$	$n_2 = 10$
متوسط العينة	$\overline{x_1} = 40$	$\overline{x_2} = 30$
تباين العينة	$s_1^2 = 18$	$s_2^2 = 16$

- أوجد مجال الثقة للفرق بين متوسطي المجتمعين $\mu_1 - \mu_2$ مع العلم أن مستوى الثقة هو 95%؟.

<u>حل التمرين2</u>:

- لدينا مجال الثقة للفرقبين متوسطي المجتمعين $\mu_1 - \mu_2$، وفي حالة تباينا المجتمعين متساويين ومجهولين $(\delta_1^2 = \delta_2^2)$،يكتب كما يلي:

$$IC_{\mu_1-\mu_2} = \left[\overline{x}_1 - \overline{x}_2 \pm t_{n_1+n_2-2;\frac{\alpha}{2}} (S_p)\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right]$$

- إذن سوف نحسب التباين المشترك S_p^2 أولا حيث يساوي:

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} = \frac{(14)(18)+(9)(16)}{23} \approx 17.22$$
$$t_{23;0.025} = 2.069$$

ومنه مجال الثقة هو:

$$IC_{\mu_1-\mu_2} = \left[\overline{x}_1 - \overline{x}_2 \pm t_{n_1+n_2-2;\frac{\alpha}{2}} (S_p) \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right]$$
$$= \left[(40 - 30) \pm (2.069)(4.15) \sqrt{\frac{1}{15} + \frac{1}{10}} \right]$$
$$= [(10 - 3.5);(10 + 3.5)] = [(6.5);(13.5)]$$

<u>التمرين3</u>:

1- لتقدير نسبة عملاء بنك معين في منطقة ما، أخذت عينة عشوائية مكونة من 200 شخص فكان عدد عملاء البنك 120شخصا.

- المطلوب: أوجد مجال الثقة لنسبة عملاء البنك في المنطقة؟ (مع العلم أن $1-\alpha = 95\%$ $(z = 1.96)$).

2- إذا علمنا أن عدد سكان تلك المنطقة يقدر ب 20000نسمة.

المطلوب: أوجد مجال تقدير لعدد عملاء البنك في تلك المنطقة؟(يرمز لعدد عملاء البنك بالرمز A)

<u>حل التمرين3</u>:

1- إيجاد مجال الثقة لنسبة عملاء البنك:

لدينا: $p' = \frac{120}{200} = 0.6$

ولدينا مجال الثقة للنسبة في المجتمع يكتب كما يلي:

$$IC_p = \left[p' \pm z_c \delta_{p'} \right]$$

ولدينا: $\delta_{p'} = \sqrt{\frac{pq}{n}}$، ومنه يصبح مجال الثقة كما يلي: $IC_p = \left[p' \pm z_c \sqrt{\frac{pq}{n}} \right]$

وبما أن p مجهولة ونريد إيجاد مجال الثقة لها ومنه لحساب $\delta_{p'}$ نستبدل p بدلالة p' النسبة في العينة وبذلك يصبح مجال الثقة يكتب كما يلي:

$$IC_p = \left[p' \pm z_c \sqrt{\frac{p'q'}{n}} \right] = \left[(0.6) \pm (1.96)\sqrt{\frac{(0.6)(0.4)}{200}} \right]$$
$$= [(0.532);(0.668)]$$

2–إيجاد مجال تقدير لعدد عملاء البنك يرمز للمجال بـ IC_A:

$$IC_A = [(0.532)(20000);(0.668)(20000)]$$
$$IC_A = [(10640);(13360)]$$

4–7– تمارين غير محلولة:

<u>التمرين1</u>: من بين المؤسسات العاملة في قطاع معين وعددها 194مؤسسة، تم سحب عينة (عينة عشوائية بسيطة بدون إرجاع) من 40مؤسسة فوجد أن عدد العمال في المؤسسات كما يلي:

91	168	171	53	114	37	126	12	71	95
33	43	158	137	2	115	99	190	32	140
81	147	68	78	11	86	127	64	57	194
131	141	93	25	105	26	79	23	69	101

1– قدر متوسط عدد العمال للمؤسسة العاملة في القطاع؟

2– تقرر منح إعفاء ضريبي لمؤسسات القطاع التي تشغل 135عاملا أو أكثر. قدر نسبة المؤسسات المستفيدة من الإجراء؟.

3– أحسب $(\delta_{p'})$خطأ المعاينة للمقدر المستخدم في السؤال 2 ؟

4- أعط مجال ثقة لنسبة مؤسسات القطاع المستفيدة من الإعفاء؟(مستوى معنوية= 5 **%**).

5- كم يجب أن يكون حجم العينة إذا أردنا أن يكون تباين المقدر لنسبة المؤسسات المستفيدة $\left(\delta_{p'}^{2}\right)=0.01$؟.

<u>التمرين2</u>:

نسحب من مجتمع الأعمار في بلد معين عينة عشوائية بالإرجاع حجمها 410، متوسط القيم المحصلة 38.1سنة.

1- أوجد مجال الثقة لمتوسط الأعمار في البلد μ بمستوى ثقة 0.95، إذا علمت أن تباين المجتمع هو $\delta^{2}=36$؟.

2- كيف يكون المجال إذا كان الانحراف المعياري للمجتمع مجهولا، علما أن الانحراف المعياري للعينة $S=6.2$؟.

<u>التمرين3</u>:

1) لتقدير التفرقة بين الأجور بين الإناث والذكور وجد لعينتين من 80 ذكر و60 أنثى أن متوسطي الدخول يساويان 20600دج بانحراف معياري 3000 دج للذكور و19700 دج بانحراف معياري 2500 دج للإناث.

- أحسب مجال الثقة للفرق بين الدخول بافتراض التوزيع طبيعي ومستوى الثقة هو 95% ؟.

2) إذا استبدلنا في السؤال الأول الأوساط الحسابية بالنسب في العينتين مثلا أن نسبة الذكور ذوي الدخول أكثر من 10000دج تساوي 40% وللإناث تساوي 30%، بحيث لم يتم تغيير حجم العينتين.

– أحسب مجال الثقة للفرق في النسب، التوزيع طبيعي ومستوى الثقة هو 95% ؟.

<u>التمرين4</u>:لاحظ أستاذ بخبرته أن وسط درجات الطلاب في مادة الإحصاء 75 علامة وبانحراف معياري 9 علامات، إذا رغب الأستاذ في تطوير أسلوب تدريس مادة الإحصاء ومن تم تقدير الوسط الحسابي للعلامات وفق الأسلوب الجديد بحيث يكون متأكد بنسبة 95% أن الخطأ في التقدير الناتج يساوي 3 علامات.

– كم طالبا يحتاج الأستاذ لإخضاعهم للتجربة؟.

<u>5–اختبارات الفروض وتطبيقاتها</u>

المقصود بالفروض هنا الفروض الإحصائية بمعنى الفروض التي تتعلق بالمجتمع الإحصائي المسحوبة منه العينة، أو توزيع هذا المجتمع أو معالمه كالوسط الحسابي أو النسبة في المجتمع.

والفرض ما هو إلا تخمين أو استنتاج ذكي مبني على حيثيات معقولة أو منطقية ولكنه ليس مبنياً على حسابات دقيقة خاصة بالمجتمع لأننا نفترض أنه لا يمكن دراسة المجتمع بالكامل عن طريق الحصر الشامل بل نحاول استنتاج أو الاستدلال على مقاييس المجتمع باستخدام بيانات ونتائج العينة.

فمثلاً: قد يفترض الباحث أن متوسط الدخل الشهري للفرد هو 20000 دينار جزائري (بناءً على ما يراه من مستوى المعيشة في البلد وأوضاعه الاقتصادية)، ويحتاج إلى اختبار علمي (إحصائي) لمعرفة مدى صحة هذا الفرض أو قد يفترض باحث آخر أن نسبة الناخبين في إحدى الدوائر الذين يؤيدون مرشحاً معيناً لا تقل عن % 30 وهكذا... والمطلوب هو اختبار مدى صحة هذه الفروض. أي أن يصل الباحث إلى قرار إما بقبول الفرض أو عدم قبوله (أي رفضه) وذلك باحتمال معين. وقبل تناول كيفية إجراء الاختبارات الإحصائية نستعرض أولاً بعض المفاهيم والتعريفات الأساسية اللازمة لهذا الموضوع حتى تكون الصورة أكثر وضوحاً..

5-1- مفاهيم أساسية:

5-1-1- الفرض العدم (أو الصفري)

فرضية العدم هي "الفرضية الأساسي المراد اختبارها". ويرمز لها عادة بالرمز : H_0. وهي فرضية حول معلمة المجتمع التي نجري اختبارعليها باستخدام بيانات من عينة والتي تشير أن الفرق بين معلمة المجتمع والإحصائي من العينة ناتج عن الصدفة ولا فرق حقيقي بينهما. وهي الفرضية التي ننطلق منها ونرفضها عندما تتوفر دلائل على عدم صحتها، فمثلاً إذا كان الفرض العدم المراد اختباره هو أن متوسط دخل الفرد هو 20000 دينار جزائري شهرياً فإن هذا الفرض يكتب بالرموز كما يلي :

$$H_0 : \mu = 20000$$

ويقرأ بالشكل التالي :

الفرض العدم هو : أن متوسط دخل الفرد هو20000 دينار جزائري شهرياً.

وليس شرطاً أن يصاغ الفرض العدم بالرموز، فقد يتم التعبير عنه بدون رموز. فقد يريد الباحث أن يختبر ما إذا كانت هناك علاقة بين المؤهل العلمي ودرجة الوعي السياسي. فقد يصيغ الباحث الفرض العدم بالشكل التالي (على سبيل المثال) :

الأمية والاستعداد للانحراف مستقلان (أي لا توجد علاقة بينهما، أو أن العلاقة بينهما منعدمة).

5-1-2- الفرض البديل:

في اختبارات الفروض يتحتم وضع فرض آخر غير الفرض العدم المراد اختباره يسمى الفرض البديل. وهذا الفرض " هو الذي سيقبل في حالة رفض الفرض العدم " أي لابد من تحديد فرض آخر بديل في الوقت الذي نحدد فيه الفرض العدم، وبالتالي فإن الفرض البديل يعرف كما يلي: "الفرض البديل هو الفرض الآخر الذي سيقبل في حالة رفض الفرض العدم" ويرمز له عادة بالرمز : H_1

والفرض البديل له أهمية كبيرة وبالذات في قياس الظواهر الاقتصادية والاجتماعية، فهو الذي يحدد نوع الاختبار المستخدم لذلك فهو يأخذ أحد الأشكال الثلاثة هي:

أ- أن يأخذ شكل " لا يساوي ". وفي هذه الحالة نستخدم ما يسمى: اختبار ثنائي الاتجاه (اختبار الطرفين).

مثال1: إذا كان الفرض العدم هو أن متوسط الدخل الشهري لفئة معينة في المجتمعهو 20000 دينار جزائري.

$$H_0 : \mu = 20000$$

فإن الفرض البديل في هذه الحالة يأخذ الشكل التالي :

$$H_1 : \mu \neq 20000$$

بمعنى أن متوسط دخل هذه الفئة من المجتمع " لا يساوي " 20000 دينار جزائري.شهرياً.

ب- أو أن يأخذ شكل " أكبر من ". وفي هذه الحالة نستخدم ما يسمى الاختبار أحادي الاتجاه من اليمين (اختبار الطرف الأيمن).

مثال2: قد يكون الفرض البديل كما يلي :

$$H_1 : \mu > 20000$$

أي أن متوسط الدخل لهذه الفئة من المجتمع أكبر من 20000 دينار جزائري.شهرياً.

ج- وأخيراً قد يأخذ الفرض البديل شكل " أقل من ".وفي هذه الحالة نستخدم ما يسمى الاختبار أحادي الاتجاه من اليسار(اختبار الطرف الأيسر).

مثال3: قد يكون الفرض البديل هو :

$$H_1 : \mu < 20000$$

أي أن متوسط الدخل لهذه الفئة من المجتمع أقل من 20000 دينار جزائري.شهرياً.

و**الخلاصة:** هي لابد للباحث من تحديد الفرض البديل الذي لا يخرج عن أحد الأشكال الثلاثة السابقة، وهذا التحديد مهم جداً قبل الدخول في تفاصيل الاختبار الإحصائي وذلك لأنه هو الذي يحدد نوع الاختبار المستخدم كما سوف نرى.

5-1-3- الخطأ في اتخاذ القرار :

ففي حالة قبول الباحث لفرضه العدم، فلا مجال للبحث في الفرض البديل، أما في حالة حدوث العكس بمعنى رفض الفرض العدم فإنه يتحتم في هذه الحالة قبول الفرض البديل، على أنه من الجدير بالذكر أن الباحث هنا عرضة للوقوع في الخطأ عند اتخاذ قراره بقبول الفرض العدم أو رفضه، فقد يرفض فرضاً هو في الواقع صحيح، وقد يقبل فرضا هو في الواقع غير صحيح. لذلك فقد تم تصنيف هذه الأخطاء إلى نوعين هما :

5-1-3-1- الخطأ من النوع الأول:

الخطأ من النوع الأول هو "رفض الفرض العدم بينما هو صحيح ". أي أنه على الرغم من أن الفرض العدم في الواقع صحيح وكان من الواجب قبوله فقد تم أخذ قرار خاطئ برفضه. وباختصار شديد فإن الخطأ من النوع الأول هو : " رفض فرض صحيح". واحتماله α، ويكتب : $p(RH_0 / H_0) = \alpha$

5-1-3-2- الخطأ من النوع الثاني:

وفي المقابل فإن الخطأ من النوع الثاني يعني " قبول الفرض العدم بينما هو خاطئأي أنه على الرغم من أن الفرض العدم خاطئ وكان من الواجب رفضه فقد تم أخذ قرار خاطئ بقبوله وباختصار شديد فإن الخطأ من النوع الثاني هو " قبول فرض خاطئ ".

واحتماله 1-α ويكتب: $p(\overline{R}H_0 / H_1) = 1 - \alpha$

جدول: القرارات الخاطئة والصائبة في عملية اختبار الفرضيات

نتيجة الاختبار	الوضع الحقيقي	
	الادعاء غير صحيح	الادعاء صحيح

	H_oصحيحة	رفض H_o وإقرار H_1
عدم رفض H_o	قرار صائب	قرار خاطئ خطاء من النوع الثاني
رفض H_o وقبول H_1	قرار خاطئ خطاء من النوع الأول	قرار صائب

وقد يتساءل البعض عند مدى إمكانية تصغير الخطأين معاً ولكن لسوء الحظ لا يمكن تصغيرهما معاً إلى أدنى حد ممكن، ويبدو أن الطريقة الوحيد ة المتاحة لذلك هي زيادة (أو تكبير) حجم العينة، الأمر الذي قد لا يكون ممكنا في كل الحالات. لذلك فإن الذي يحدث عادة هو تثبيت أحدهما كأن يكون نسبة أو احتمال حدوث الخطأ من النوع الأول ومحاولة تصغير الآخر، و يقيس احتمال رفض الفرضية الصفريةقوة الاختبارفيما يقيس احتمال قبولهافعالية الاختبار. ويتوقف كلا الاحتمالين على القيمة الحقيقية ل µ .

والفكرة الأساسية في اختبار الفرض هي تقسيم المساحة تحت المنحنى إلى منطقتين: أحداهما تسمى " منطقة القبول " أي منطقة قبول الفرض العدم. والأخرى تسمى " منطقة الرفض"، أي منطقة رفض الفرض العدم والتي تسمى أحيانا " بالمنطقة الحرجة، والنقطة الجديرة بالملاحظة هنا هي أن منطقة القبول تمثل درجة الثقة، بينما تمثل منطقة الرفض مستوى المعنوية.

5-1-3-3-خطوات الاختبار الإحصائي:

- تحديد الفرضيات (الصفرية والبديلة)
- تحديد قاعدة القرار
- حساب القيمة الجدولية للمتغيرة
- حساب القيمة الفعلية للمتغيرة.
- اتخاذ القرار.

تتحدد كيفية إتمام كل خطوة حسب طبيعة الاختبار (ثنائي أو أحادي الاتجاه)، حسب طبيعة المجتمع و طبيعة و حجم العينة، ... و تسخدم في ذلك نظريات توزيع المعاينة.

5-2- اختبار المتوسط

يتناول هذا الاختبار متوسط المجتمع (μ)، مثل متوسط الدخل، متوسط وزن منتج معين، ويؤكد اختبار المتوسط فرضية مساواته لقيمة ما μ_0. وللقيام بالاختبار نستخرج عينة عشوائية نحسب فيها المتوسط$\bar{x}$ ثم نستخدم التوزيع الاحتمالي ل$\bar{x}$ لقياس قرب أو بعد هذه القيمة من μ_0.

5-2-1- اختبار ثنائي الاتجاه للمتوسط.

نحتاج إلى الخطوات التالية: تحديد الفرضيات، تحديد قاعدة القرار، حساب القيمة الجدولية للمتغيرة، حساب القيمة الفعلية للمتغيرة، اتخاذ القرار.

تحديد الفرضيات (الصفرية والبديلة):

$$\leftrightarrow H_1 : \mu \neq \mu_0 H_0 : \mu = \mu_0$$

الفرضية H_0 الفرضية الصفرية أو فرضية العدم، ويؤدي الاختبار إما إلى رفضها ونكتب RH_0وفي هذه الحالة نقبل الفرضية البديلة أو المعاكسة أو عدم رفضها ونكتب $\overline{R}H_0$.

μ_0هي القيمة الافتراضية لـμعادة ما تكون μ_0 محددة بناءا على بيانات عينة عشوائية بسيطة، وفي هذه الحالة يمكن استخدام الخاصية $\bar{x} \rightarrow N\left(\mu, \frac{\delta^2}{n}\right)$، لإجراء الاختبار.

حيث أنه تحت H_0فإن **:** $\bar{x} \rightarrow N\left(\mu_0, \frac{\delta^2}{n}\right)$

وبصفة عامة نكتب: $p\left(-z_{1-\frac{\alpha}{2}} \leq \frac{\bar{x}-\mu_0}{\delta_{\bar{x}}} \leq z_{1-\frac{\alpha}{2}}\right) = 1-\alpha$

حيث:

- $\frac{\bar{x}-\mu_0}{\delta_{\bar{x}}}$: (متغيرة القرار) هي المتغيرة المعيارية ل $\bar{x}$ ونرمز لها ب z_c.
- $\delta_{\bar{x}}$ تحدد كما يلي: $\delta_{\bar{x}}=\frac{\delta}{\sqrt{n}}$ في حالة المعاينة بالإرجاع و $\delta_{\bar{x}}=\frac{\delta}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$ في الحالة المعاينة بدون إرجاع.
- $1-\frac{\alpha}{2}$: المساحة على يسار z .
- n : حجم العينة.

يمكن إذا كان $\bar{x}$ خارج المجال 1-α، أن نرفض الفرضية الصفرية التي حدد على أساسها هذا المجال ونقبل بالتالي الفرضية البديلة.

تسمى هذه (الخطة) قاعدة القرار.

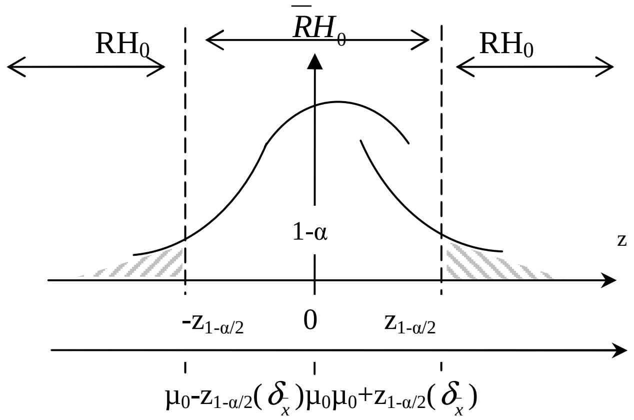
تحديد قاعدة القرار

تكتب قاعدة القرار، وهي قاعدة اختبار ثنائي الاتجاه، كما يلي:

$$\begin{cases} RH_0 \ si\ z_c=\dfrac{\bar{X}-\mu_0}{\sigma_{\bar{X}}}\notin\left[-z_{1-\alpha/2};\ z_{1-\alpha/2}\right]. \\ \overline{R}H_0 \ \text{sinon.} \end{cases}$$

أو

$$\begin{cases} RH_0 \ si \left|\dfrac{\bar{X}-\mu_0}{\sigma_{\bar{X}}}\right| > z_{1-\alpha/2}\ . \\ \overline{R}H_0 \ \text{sinon.} \end{cases}$$



الشكل:منطقتي القبول و الرفض في حالة قاعد القرار الثنائية

حساب z الجدولية:

ويرمز لها ب z_tحيث، وهي المشار إليها في قاعدة القرار.

حساب z الفعلية:

ويرمز لها ب z_c وهي المتغير المعيارية ل $\bar{x}$(أنظر قاعدة القرار) : $z_c = \dfrac{\overline{X} - \mu_0}{\delta_{\bar{x}}}$

القرار:

نقرر قبول أو رفض H_0 حسب قاعدة القرار.

5-2-2- اختبار أحادي الاتجاه للمتوسط.

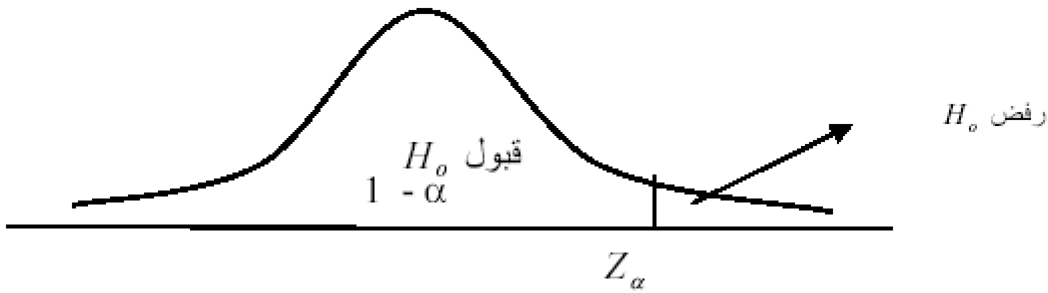
يتميز الاختبار الثنائي عن الأحادي في الفرضية البديلة التي هي عدم مساواة في الاختبار الثنائي وأكبر تماما أو أصغر تماما (حسب الحالة) في الاختبار الأحادي، وهذا يترتب عليه تغيير في قاعدة القرار.

5-2-2-1- اختبار أحادي الاتجاه من اليمين.

- **الفرضيات :** $H_0 : \mu = \mu_0 \quad H_1 : \mu > \mu_0 \leftrightarrow$

- **قاعدة القرار:**

$$\begin{cases} RH_0 \ si \dfrac{\overline{X}-\mu_0}{\sigma_{\overline{X}}} > z_{1-\alpha} \ . \\ \overline{R}H_0 \ \text{sinon.} \end{cases}$$



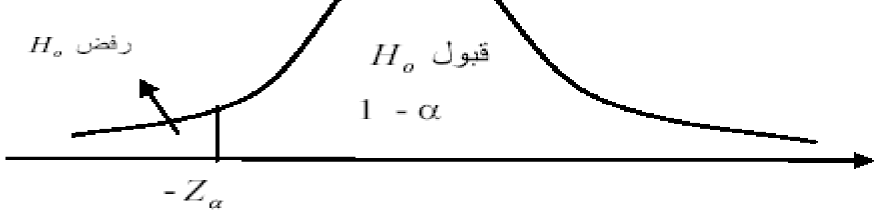
الشكل:منطقتي القبول و الرفض في حالة قاعد القرار اختبار أحادي الاتجاه من اليمين

5-2-2-2- اختبار أحادي الاتجاه من اليسار

- **الفرضيات :** $H_0 : \mu = \mu_0 \quad H_1 : \mu < \mu_0 \leftrightarrow$

- **قاعدة القرار:**

$$\begin{cases} RH_0 \ si \dfrac{\overline{X}-\mu_0}{\sigma_{\overline{X}}} < -z_{1-\alpha} \ . \\ \overline{R}H_0 \ \text{sinon.} \end{cases}$$



الشكل:منطقتي القبول و الرفض في حالة قاعد القرار اختبار أحادي الاتجاه من اليسار.

5-3- استخدام S كمقدر لـσوالتوزيع t في اختبار المتوسط

5-3-1- استخدام S كمقدر لـσ في اختبار المتوسط.

فيما سبق افترضنا أن σ معلوم وحجم العينة $n \geq 30$، في الواقع غالبا ما يكون الانحراف المعياري مجهولا ونحتاج بالتالي إلى استخدام الانحراف المعياري للعينة (S) عند حساب $\delta_{\bar{x}}$، حيث نعوض في عبارة متغيرة القرار أو قاعدة القرار:

$$\delta_{\bar{x}} = \frac{\hat{S}}{\sqrt{n}} \quad \text{أو} \quad \delta_{\bar{x}} = \frac{S}{\sqrt{n-1}} \quad \text{بـ}\ \delta_{\bar{x}} = \frac{\delta}{\sqrt{n}}$$

5-3-2- استخدام التوزيع t في اختبار المتوسط.

في حالة n< 30 وσ (الانحراف المعياري للمجتمع) مجهولا، لا يمكن استخدام التوزيع الطبيعي، ولكن لدينا :

$$\frac{\bar{X}-\mu_0}{S/\sqrt{n-1}} \sim t_{n-1} : (\mu = \mu_0)\ H_0 \ \text{تحت و} \qquad \frac{\bar{X}-\mu}{S/\sqrt{n-1}} \sim t_{n-1}$$

يمكن إذا استخدام التوزيع ستيودنت (بشرط أن يكون توزيع المجتمع طبيعيا) .

و تتغير قاعدة القرار تبعا لهذا التغيير فتكتب في حالة الاختبار الثنائي كما يلي:

$$\begin{cases} RH_0 \ si \left| \dfrac{\bar{X}-\mu_0}{S/\sqrt{n-1}} \right| > t_{n-1;\ 1-\alpha/2} \ . \\ \bar{R}H_0 \ \text{sinon.} \end{cases}$$

في حالة اختبار من اليمين:

$$\begin{cases} RH_0 \ si \dfrac{\bar{X}-\mu_0}{S/\sqrt{n-1}} > t_{n-1,\ 1-\alpha} \\ \bar{R}H_0 \ \text{sinon.} \end{cases}$$

في حالة اختبار من اليسار:

$$\begin{cases} RH_0 \ si \dfrac{\bar{X}-\mu_0}{S/\sqrt{n-1}} < -t_{n-1;\ 1-\alpha} \\ \bar{R}H_0 \ \text{sinon.} \end{cases}$$

مثال4: عينة عشوائية حجمها 49 شخصاً اختيرت من أفراد دولة ما، فإذا كان الوسط الحسابي لدخول الأفراد الأسبوعية في العينة هو 75 دولاراً. كيف يمكن اختبار الفرضية بأن متوسط الدخل الأسبوعي لمواطني هذه الدولة يساوي 72 دولاراً مقابل الفرض البديل أنه لا يساوي 72 وذلك بمستوى معنوية % 5 إذا علمت أن الانحراف المعياري للمجتمع يساوي 14 دولاراً.

الحل :

– الفرض الصفرية: هو أن متوسط المجتمع يساوي 72.

– الفرض البديل : هو أن المتوسط لا يساوي 72.

وبالرموز : $\begin{cases} H_0 : \mu = 72 \\ H_1 : \mu \neq 72 \end{cases}$

– بما أن العينة كبيرة $n \geq 30$ و $\delta = 14$ و $\overline{X} = 75$ والاختبار هو اختبار ثنائي الاتجاه، فان متغيرة القرار تحسب كما يلي:

$$z_c = \frac{\overline{X} - \mu_0}{\delta_{\bar{x}}} = \frac{75 - 72}{\frac{14}{\sqrt{49}}} = \frac{3}{2} = 1.5$$

أي أن قيمة الإحصائية تساوي 1.5

– حدود منطقتي القبول والرفض: نحصل عليها من التوزيع الطبيعي المعياري حيث مستوى المعنوية 5%. وبالكشف في جدول التوزيع الطبيعي المعياري عن Z التي تقابل مستوى المعنوية 5% نجد أنها تساوي 1.96

– المقارنة والقرار : وبمقارنة قيمة الإحصائية المحسوبةوالتي تساوي $(z_c = 1.5)$ 1.5 بـ z الفعلية $(z_t = 1.96)$ نلاحظ أن $z_c < z_t$ ، ومنه فإن القرار هو :

قبول الفرض الصفرية بأن متوسط دخول الأفراد الأسبوعية في هذه الدولة يساوي 72 دولاراً وذلك بمستوى معنوية % 5.

5-4- اختبار النسبة

يتعلق هذا الاختبار بنسبة مفردات المجتمع التي تتصف بخاصية ما (p)، حيث يؤكد الاختبار أو ينفي صحة فرضية معينة بخصوص قيمة p . يرمز للقيمة الافتراضية ب p_0

وتكتب الفرضية كما يلي: $H_0 : p = p_0$

للقيام بالاختبار نستخدم خصائص p’ النسبة في العينة:

$$E(p') = \mu_{p'} = p \qquad ; \qquad \sigma^2_{p'} = \frac{pq}{n}$$

عند $n \geq 30$: $p' \rightarrow N\left(p, \delta^2_{p'}\right)$

استنادا إلى هذه الخصائص وتحت H0 :

$$p' \rightarrow N\left(p_0, \frac{p_0 q_0}{n}\right)$$

و من ثم يمكن تحديد قاعدة القرار بحسب طبيعة الاختباركما يلي:

في حالة الاختبار الثنائي:

$$\begin{cases} RH_0 \ si \left|\dfrac{p' - p_0}{\sqrt{p_0 q_0 / n}}\right| > z_{1-\alpha/2} \ . \\ \overline{R}H_0 \ \text{sinon.} \end{cases}$$

في حالة اختبار من اليمين:

$$\begin{cases} RH_0 \ si \ \dfrac{p' - p_0}{\sqrt{p_0 q_0 / n}} > z_{1-\alpha} \\ \overline{R}H_0 \ \text{sinon.} \end{cases}$$

في حالة اختبار من اليسار:

$$\begin{cases} RH_0 \ si \ \dfrac{p' - p_0}{\sqrt{p_0 q_0 / n}} < -z_{1-\alpha} \\ \overline{R}H_0 \ \text{sinon.} \end{cases}$$

مثال5: تقدر الدوائر الرسمية نسبة المتخرجين الجامعيين الذين يحصلون على عمل في السنة الأولى التي تلي تخرجهم ب 70 % . وجدت دراسة أجريت على عينة من 900 طالب أن نسبة الحصول على عمل 67 %. كيف يمكن اختبار ما إذا كانت النسبة الرسمية صحيحة أم مبالغ فيها، بمستوى معنوية 5 %.

$$H_0 : p = 0.70 \leftrightarrow H_1 : p < 0.70$$

$$\frac{p' - p_0}{\sqrt{p_0 q_0 / n}} = \frac{0.67 - 0.7}{\sqrt{0.7(0.3)/900}} \cong -196.34 < -z_{1-0.05} = -1.64$$

ومنه نرفض الفرضية H_0.

5-5- اختبار الفرق بين وسطين حسابيين :

قد يرغب الباحث في إجراء اختبار عما إذا كان متوسط الدخل في إحدى الدول يساوي متوسط الدخل في دولة أخرى، أو إجراء اختبار عما إذا كان متوسط عمر الناخب في إحدى المناطق يساوي متوسط عمر الناخب في منطقة أخرى... وهكذا بمعنى آخر قد يرغب الباحث في إجراء اختبار عما إذا كان متوسط المجتمع الأول يساوي متوسط المجتمع الثاني.. في مثل هذه الحالات يسمى الاختبار اختبار الفرق بين وسطين حسابيين، وتكتب الفرضياتكما يلي:

- الفرضية الصفرية : أن متوسط المجتمع الأول يساوي متوسط المجتمع الثاني (أي لا يوجد فرق بين متوسطي المجتمعين).

وبالرموز : $H_0 : \mu_1 = \mu_2$

- الفرضية البديلة : أن المتوسطين غير متساويين وبالرموز :

$H_1 : \mu_1 \neq \mu_2$

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

لتحديد متغيرة القرار نعتمد في الاختبار على متغيرة القرار T أو T’ بحسب الحالة، حيث نميز بين حالة كون تباينا المجتمعين معلومين وحالة كون تباينا المجتمعين مجهولين.

5-5-1- تباينا المجتمعين معلومين

- المجتمعين طبيعيين:

$$T=\frac{\bar{X}_1-\bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\sim N(0.1)$$

حيث : يرمز بـ n1 إلى حجم العينة الأولى.

يرمز بـ n2 إلى حجم العينة الثانية.

يرمز بـ $\bar{X}1$ إلى الوسط الحسابي للعينة الأولى.

يرمز بـ $\bar{X}2$ إلى الوسط الحسابي للعينة الثانية.

يرمز بـ σ_1^2 إلى تباين المجتمع الأول.

يرمز σ_2^2 إلى تباين المجتمع الثاني.

– **مجتمعين ما $(n_1 , n_2 \geq 30)$:**

$$T=\frac{\bar{X}_1-\bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\approx N(0,1)$$

مثال6: إذا كانت لدينا نتائج عينتين عشوائيتين مستقلتين مسحوبتين من منطقتين لمقارنة متوسط عمر الناخب فيهما حيث: $\bar{X}_1=40, \bar{X}_2=34, \delta_1^2=84, \delta_2^2=27, n_1=120, n_2=90$

– اختبر الفرضية أن متوسط عمر الناخب في المنطقة الأولى يساوي متوسط عمر الناخب في المنطقة الثانية بمستوى معنوية 5% مقابل الفرض البديل أنهما غير متساويين؟

الحل:

– الفرض الصفرية أن المتوسطين متساويان.

– الفرضية البديلة أن المتوسطين غير متساويين.

ومنه: $\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases}$

-الإحصائية: تأخذ الشكل التالي:

$$T = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\delta_1^2}{n_1} + \frac{\delta_2^2}{n_2}}}$$

وبالتعويض عن :

$$\overline{X}_1 = 40, \overline{X}_2 = 34, \delta_1^2 = 84, \delta_2^2 = 27, n_1 = 120, n_2 = 90$$

نحصل على:

$$T = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\delta_1^2}{n_1} + \frac{\delta_2^2}{n_2}}} = \frac{40 - 34}{\sqrt{\frac{84}{120} + \frac{27}{90}}}$$
$$= \frac{6}{\sqrt{0.7 + 0.3}} = 6$$

أي أن قيمة $T = z_c$ الإحصائية تساوي 6

– حدود منطقتي القبول والرفض التي نحصل عليها من جدول التوزيع الطبيعيZ لأن العينات كبيرة، والاختبار هو اختبار الطرفين (لأن الفرض البديل لا يساوي)ومستوى المعنوية المطلوب هو 5 %.

أي أن منطقة القبول تبدأ من 1.96- إلى 1.96+ ومنطقة الرفض هي القيم التي أصغر من 1.96 - والتي أكبر من 1.96 +. أي سوف نقارن بالقيمة $z_t = 1.96$.

– المقارنة والقرار: بما أن قيمة الإحصائية T والتي تساوي 6 تقع في منطقة الرفض أي $T = z_c > 1.96$ فإن القرار هو رفض الفرض الصفرية وقبول الفرضية البديلة بمستوى معنوية 5% أي أننا نرفض الفرض القائل بأن متوسط عمر الناخب في المنطقة الأولى يساوي متوسط عمر الناخب في المنطقة الثانية وذلك بمستوى معنوية 5 %.

5-5-2- تباينا المجتمعين مجهولين

- المجتمعان طبيعيان:

إذا كانت العينات صغيرة (مجموع العينتين أقل من 30 مفردة أو حتى 31 مفردة) فإن الإحصائية في هذه الحالات بافتراض أن المجتمعين طبيعيان، وأن تباين المجتمع الأول يساوي تباين المجتمع الثاني ولكنه مجهول (بمعنى أن $\sigma_1^2 = \sigma_2^2$ ولكن قيمة هذا التباين غير معروفة) وأن العينتين مستقلتان فإن إحصائية الاختبار تأخذ الشكل التالي والتي لها توزيع t بدرجات حرية تساوي $n_1 + n_2 - 2$:

$$T' = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \rightarrow t_{n_1+n_2-2}$$

حيث :
$$S^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n1 + n_2 - 2}$$

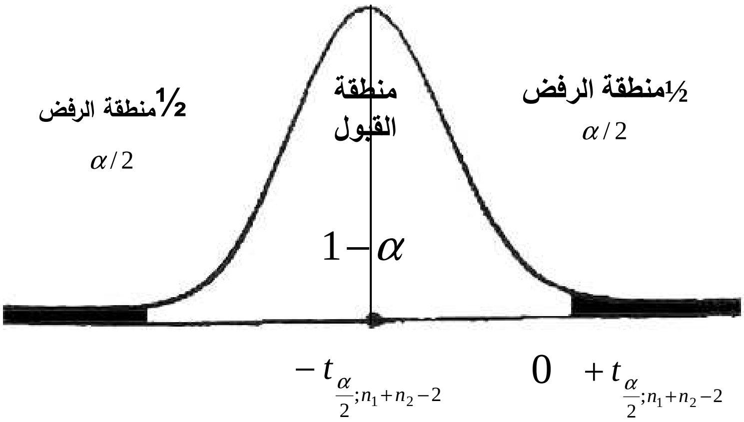
أي يتم حساب S^2 أولاً قبل التعويض في الإحصائية وتكون خطوات الاختبار هي:

هي :

- الفرضية الصفرية $H_0 : \mu_1 = \mu_2$
- الفرضية البديلة $H_1 : \mu_1 \neq \mu_2$

- الإحصائية هي المكتوبة أعلاه (وهي في هذه الحالة t وليست Z)
- حدود منطقتي القبول والرفض ونحصل عليها في هذه الحالة من جدول t عند درجات حرية تساوي $n_1 + n_2 - 2$ وعند مستوى معنوية يساوي $\frac{\alpha}{2}$ كما في الشكل التالي:

− +



– المقارنة والقرار : كما سبق.

–مجتمعين ما $(n_1 , n_2 \geq 30)$: أن تباين المجتمعين غير متساويين:

$$T = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{S_1^2}{n_1 - 1} + \frac{S_2^2}{n_2 - 1}}} \approx N(0,1)$$

مثال7: البيانات التالية تمثل نتائج عينتين عشوائيتين مستقلتين مسحوبتين من مدينتين عن أعمار الناخبين بهما (بافتراض أن تباينهما هو نفسه):

$$n1 = 10, n_2 = 10, \overline{X}1 = 28\,\overline{X}2 = 26, S_1^2 = 50, S_2^2 = 30$$

اختبر الفرضية الصفرية: $Ho: \mu_1 = \mu_2$ مقابل الفرضية البديلة $HI: \mu_1 \neq \mu_2$

وذلك بمستوى معنوية % 5 بافتراض أن الأعمار في المدينتين لهما توزيع طبيعي؟.

الحل :

– الفرضية الصفرية: $Ho: \mu_1 = \mu_2$

أي متوسط أعمار الناخبين في المدينتين متساوٍ

– الفرضية البديلة : $HI: \mu_1 \neq \mu_2$

أي متوسط أعمار الناخبين في المدينتين غير متساوٍ

– الإحصائية لاحظ (أن العينات صغيرة، وأن تباين المجتمعين هو نفسه، وأن المجتمعين طبيعيان). فإن الإحصائية المناسبة في هذه الحالة هي $T^{/}$:

$$T^{/} = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

حيث

$$S^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n1 + n_2 - 2}$$

نحسب أولاً S^2 كما يلي :

$$S^2 = \frac{(10 - 1)\times 50 + (10 - 1)\times 30}{10 + 10 - 2}$$

$$= \frac{9\times 50 + 9\times 30}{18}$$

$$= \frac{450 + 270}{18}$$

$$= \frac{720}{18}$$

$$S^2 = 40$$

وبالتعويض في الإحصائية عن :

$$\overline{X}_1 = 28, \overline{X}_2 = 26, S^2 = 40, n1 = 10, n_2 = 10$$

نحصل على :

$$T^{/} = \frac{28-26}{\sqrt{40\left(\frac{1}{10}+\frac{1}{10}\right)}}$$

$$= \frac{2}{\sqrt{8}} = \frac{2}{2.828} = 0.7$$

أي أن قيمة الإحصائية تساوي 0.7

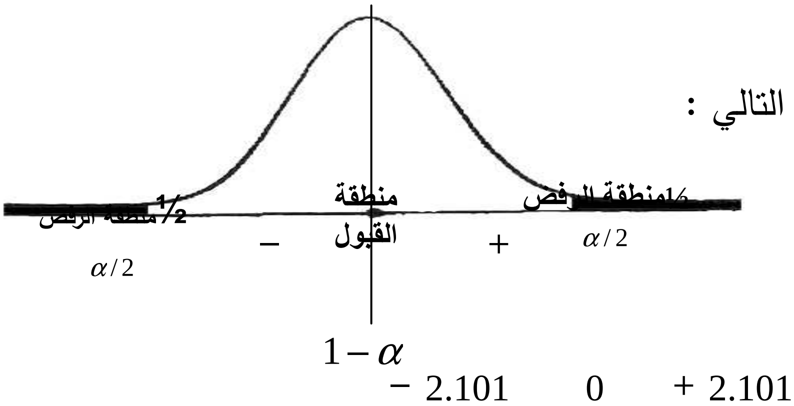
– حدود منطقتي القبول والرفض:

ونحصل عليها في هذه الحالة من جدول t عند درجات حرية تساوي $n_1 + n_2 - 2$

أي تساوي $10 + 10 - 2$ والتي تساوي 18 وذلك عند مستوى معنوية يساوي

$\alpha = 0.05$ أي أن نصف مستوى المعنوية $\frac{0.05}{2} = 0.025$ أي أن $t_{0.025,18} = 2.101$

كما في الشكل التالي :



أي أن منطقة القبول تبدأ من 2.101- وحتى 2.101 +

– المقارنة والقرار :

حيث أن قيمة الإحصائية تساوي 0.7أقل من قيمة ستيودنت الجدولية والتي تساوي 2.101 ، وهي تقع في منطقة القبول وبالتالي فإن القرار هو قبول الفرضية الصفرية بأن متوسط أعمار الناخبين في المدينة الأولى يساوي متوسط أعمار الناخبين في المدينة الثانية وذلك بمستوى معنوي 5.% (حل المثال السابق بافتراض أن تباين المجتمعين غير متساويين).

5–6– اختبار الفرق بين نسبتين :

ربما نرغب في اختبار ما إذا كانت نسبة المؤيدين لمرشح ما في الانتخابات التشريعية تساوي نسبة المؤيدين لمرشح آخر في الانتخابات نفسها، في مثل هذه الحالات فإن المطلوب هو اختبار ما إذا كانت النسبة في المجتمع الأول تساوي النسبة في المجتمع الثاني، ويسمى الاختبار : اختبار الفرق بين نسبتين وتكون خطوات هذا الاختبار ما يلي :

– الفرضية الصفرية: أن النسبة في المجتمعين متساوية أي: $H_0 : p_1 = p_2$

– الفرضية البديلة: أن النسبتين في المجتمعين غير متساوية: $H_0 : p_1 \neq p_2$

(ويمكن اختيار شكل آخر للفرض البديل مثل: أكبر من أو أقل إذا دعت الحاجة لذلك).

– الإحصائية (متغيرة القرار): بافتراض أن العينتين كبيرتان بدرجة كافية تكون الإحصائية كما يلي:

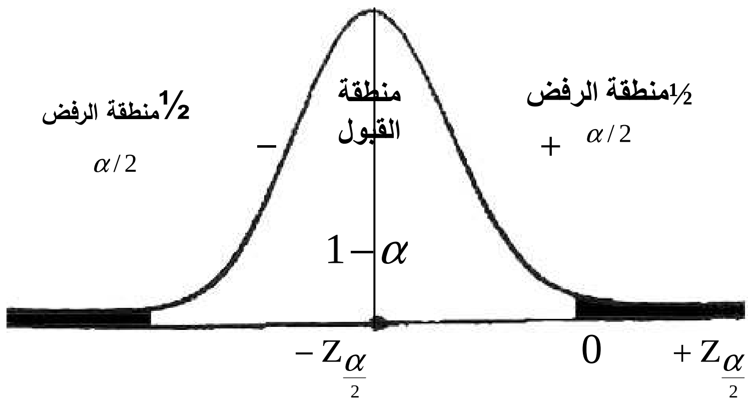
$$Z_{p_1^{/}-p_2^{/}} = \frac{p_1^{/} - p_2^{/}}{\sqrt{p^{/} q^{/} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \rightarrow N(0,1)$$

حيث: $p^{/} = \frac{n_1 p_1^{/} + n_2 p_2^{/}}{n_1 + n_2}$

$$q^{/} = 1 - p^{/}$$

أي يتم أولاً حساب $p^{/}$ (والتي تمثل متوسط مرجح من نسبتي العينتين) قبل التعويض في الإحصائية والتي لها توزيع طبيعي معياري.

– حدود منطقتي القبول والرفض ونحصل عليها من جدول التوزيع الطبيعي، والاختبار هنا هو اختبار الطرفين (لأن الفرض البديل لا يساوي) وتحدد المنطقتين بناءً على مستوى المعنوية المطلوب، وذلك كما في الشكل التالي:



– المقارنة والقرار: كما سبق

مثال8: لاختبار ما إذا كانت نسبة المؤيدين لبرنامج اقتصادي معين في المدينة A يساوي نسبة المؤيدين لهذا البرنامج في المدينة Bتم اختيار عينتين عشوائيتين مستقلتين من المدينتين حيث : حجم العينة الأولى يساوي حجم العينة الثانية يساوي 100 وكانت نسبة

المؤيدين للبرنامج في عينة المدينة A هي: $p_1^/ = 0.70$ونسبة المؤيدين للبرنامج في عينة المدينة B هي: $p_2^/ = 0.50$.

– اختبر الفرضية الصفرية أن النسبة في المدينتين متساوية مقابل الفرضية البديلة أنها غير متساوية وذلك بمستوى معنوية 1 %؟.

<u>الحل</u>:

– الفرضية الصفرية:النسبة في المدينة A تساوي النسبة في المدينة B .

– الفرضية البديلة : النسبة في المدينتين غير متساوية.

ومنه تكتب الفرضيتين كما يلي: $\begin{cases} H_0 : p_1 = p_2 \\ H_1 : p_1 \neq p_2 \end{cases}$

– متغيرة القرار:

$$Z_{p_1^/ - p_2^/} = \frac{p_1^/ - p_2^/}{\sqrt{p^/ q^/ \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

حيث: $p^/ = \frac{n_1 p_1^/ + n_2 p_2^/}{n_1 + n_2}$

$q^/ = 1 - p^/$

وبالتعويض عن: $p_1^/ = 0.7; p_2^/ = 0.5; n_1 = 100; n_2 = 100$ نجد:

$$p^/ = \frac{(100)(0.7) + (100)(0.5)}{100 + 100}$$

$$= \frac{70 + 50}{200} = \frac{120}{200}$$

$$p^/ = 0.60$$

$$q^/ = 1 - 0.6 = 0.40$$

وبالتعويض في متغيرة القرار نجد:

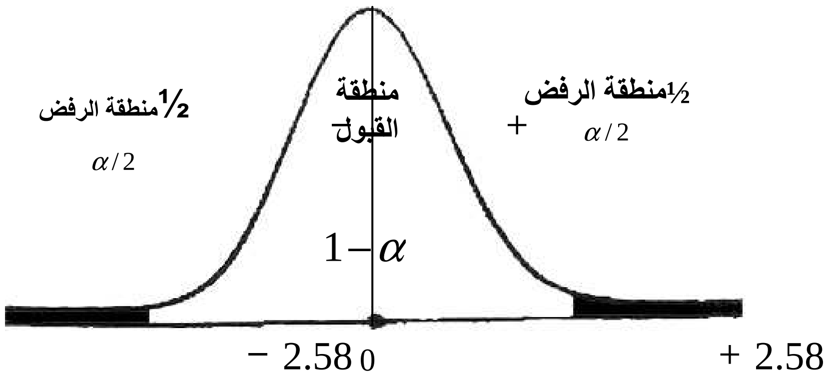
$$Z_{p_1^/ - p_2^/} = \frac{p_1^/ - p_2^/}{\sqrt{p^/ q^/ \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{0.70 - 0.50}{\sqrt{(0.6)(0.4)\left(\frac{1}{100} + \frac{1}{100} \right)}}$$

$$= \frac{0.20}{0.069} = 2.899$$

أي أن قيمة الإحصائية تساوي 2.899

– حدود منطقتي القبول والرفض

نحصل عليها من التوزيع الطبيعي، واختبار الطرفين بمستوى معنوية 1% كما في الشكل التالي :



أي أن منطقة القبول تبدأ من 2.58- وحتى 2.58 +أي أن $z_t = 2.58$

– المقارنة والقرار :

بما أن قيمة الإحصائية تساوي 2.899 فهي تقع في منطقة الرفض أي $z_c > z_t$ وبالتالي فإن القرار هو :رفض الفرضية الصفرية وقبول الفرضية البديلة أي رفض الفرض القائل بأن نسبة المؤيدين للبرنامج الاقتصادي في المدينة Aتساوي نسبة المؤيدين له في المدينة Bوذلك بمستوى معنوية 1%، وقبول الفرض البديل بأن النسبتين غير متساويتين.

5-7- اختبار التباين وتساوي تبايني مجتمعين.

5-7-1-اختبار التباين.

لاختبار صدقية فرضية بخصوص قيمة تباين مجتمع ما،

$$H_0 : \sigma^2 = \sigma_0^2 \leftrightarrow H_1 : \sigma^2 = \sigma_0^2$$

نستعمل المقدر غير المنحاز $\hat{S}^2 = \frac{\sum_i (X_i - \overline{X})^2}{n-1}$. حيث في حالة العينة الكبيرة ($n \geq 50$ في أحسن الأحوال)، وتحت H_0 فإن $\mu_4 = E(X-\mu)^4$

$$T = \frac{\hat{S}^2 - \sigma_0^2}{\sqrt{(\mu_4 - \sigma_0^4)/n}} \approx N(0.1).$$

حيث μ_4 هو العزم المركزي من الدرجة الرابعة. وبهذا الشكل تكتب قاعدة القرار للاختبار الثنائي كما يلي:

$$\begin{cases} RH_0 \ si \left| \dfrac{\hat{S}^2 - \sigma_0^2}{\sqrt{(\mu_4 - \sigma_0^4)/n}} \right| > z_{1-\alpha/2} \\ \overline{R}H_0 \ \text{sinon.} \end{cases}$$

وفي حالة μ_4 مجهول يمكن استخدام كمقدر:

$$\bar{x}_4 = E\left(x_i - \bar{x}\right)^4$$

وإذا كان المجتمع طبيعيا، حيث $\mu_4 = 3\sigma^4$ ، فإن متغيرة القرار يمكن أن تكتب كما يلي:

$$T = \frac{\hat{S}^2 - \sigma_0^2}{\sigma_0^2 \sqrt{2/n}} \approx N(0.1).$$

5-7-2- اختبار تساوي تبايني مجتمعين

الغرض من الاختبار هو تأكيد أو نفي تساوي تباينا مجتمعين من خلال عينتين عشوائيتين بسيطتين مستقلتين.

تكتب الفرضيات (في حالة الاختبار الثنائي)كما يلي:

$$H_0 : \sigma^2_1 = \sigma^2_2 \ \leftrightarrow \ H_1 : \sigma^2_1 \neq \sigma^2_2$$

نعتمد في الاختبار على متغيرة القرار T أو 'T بحسب الحالة، حيث نميز بين حالة كون المجتمعين طبيعيين أم غير ذلك.

5-7-2-1- مجتمعين طبيعيين.

-الحالة العامة:

$$T' = \frac{\hat{S}_1^2}{\hat{S}_2^2} \sim F_{n_1-1;\ n_2-1}$$

- في حالة $n_1, n_2 \geq 30$

$$T=\frac{1}{2}\ln\left(\frac{\hat{S}_1^2}{\hat{S}_2^2}\right)/\sqrt{\frac{1}{2}\left(\frac{1}{n_1-1}+\frac{1}{n_2-1}\right)}\approx N(0;1)$$

5-7-2-2- مجتمعين ما $(n_1, n_2 \geq 50)$

- $\mu_4^{(1)};\mu_4^{(2)}$ معروفين :

$$T=(\hat{S}_1^2-\hat{S}_2^2)/\sqrt{\frac{\mu_4^{(1)}-\hat{S}_1^4}{n_1}-\frac{\mu_4^{(2)}-\hat{S}_2^4}{n_2}}\approx N(0;1)$$

- في حالة $\mu_4^{(1)};\mu_4^{(2)}$ غير معروفين : نعوض μ_4 بـ $\bar{x}_4$.

مثال9: نسحب من مجتمعين طبيعيين عينتين حجم الأولى 18 وحجم الثانية 21. وجدنا النتائج التالية: $S_1^2=9;S_2^2=8;$

- كيف يمكن إجراء اختبارتأكيد أو نفي تساوي تباينا مجتمعين بمستوى معنوية 5% ؟

الحل:

$$H_0:\sigma^2_1=\sigma^2_2 \leftrightarrow H_1:\sigma^2_1\neq \sigma^2_2$$

- متغيرة القرار في هذه الحالة هي:

$$T'=\frac{\hat{S}_1^2}{\hat{S}_2^2}\sim F_{n_1-1;\,n_2-1}$$

لدينا:

$$\hat{S}_1^2=S_1^2\frac{n_1}{(n_1-1)}=9\,\frac{18}{17}\approx 9.53$$

$$\hat{S}_2^{/}=S_2^2\frac{n_2}{n_2-1}=8\,\frac{21}{20}\approx 8.4$$

$$T^{/}=\frac{\hat{S}_1^2}{\hat{S}_2^{/}}=\frac{9.53}{8.4}\approx 1.135$$

$$F_{0.05;17;20}\approx 2.17$$

$$T^{/}=1.135<F_{0.05;17;20}\approx 2.17$$

$$\Rightarrow \overline{R}H_0$$

5-8- اختبارالتجانس والتعديل.

5-8-1- اختبار التجانس

لنعد إلى اختبار النسبة، ونفترض أن لدينا عددا k من الخصائص المتنافية، نسبة تحقق كل منها في المجتمع p_i حيث $\sum p_i = 1$. نريد اختبار فرضية تساوي هذه النسب:

$$H_0 : p_i = p_{i0} \quad (i = 1, 2, \ldots k) \leftrightarrow H_1 : p_i \neq p_{i0}$$

(الفرضية البديلة هي أن إحدى النسب النظرية p_{i0}على الأقل غير مساوية للقيمة الحقيقية.)

متغيرة القرار: لإنجاز الاختبار نستخرج عينة نحسب فيها عدد مرات تحقق الخصائص (n_i) . إذا تحققت الشروط

$n \geq 30$ ، $p_{i0} \geq 1$ وعلى الأقل في 80 % من الحالات $np_{i0} \geq 5$ نبرهن أن :

$$T = \sum_i \frac{(n_i - np_{i0})^2}{np_{i0}} \approx \chi^2_{k-1}$$

5-8-2- اختبار التعديل

تستخدم هذه الطريقة أيضا لاختبار تعديل توزيع معين بتوزيع آخر، وفي هذه الحالة نقارن بين تكرارات العينة (التكرارات الحقيقية) n_i وتكرارات افتراضية n_{i0}، حيث تصاغ الفرضيات كما يلي :

على الأقل إحدى التكرارات النظرية n_{i0} غير مساوية للتكرار الحقيقي

$$H_0: n_i = n_{i0} (i = 1, 2, \ldots k) \leftrightarrow H_1: n_i \neq n_{i0}$$

$$T = \sum_i \frac{(n_i - n_{i0})^2}{n_{i0}} \approx \chi^2_{k-m-1}$$

m عدد من معالم من المجتمع المقدرة انطلاقا من بيانات العينة لتحديد التكرارات النظرية.

مثال10: يتقدم إلى انتخابات معينة 3 مرشحين: أ، ب وج. نريد اختبار فرضية بمستوى معنوية 5 % حول شعبيتهم كما يلي:

$$H_0 : p_1 = 0.4, p_2 = 0.35, p_3 = 0,25$$

أجري استجواب ل 400 ناخب فكان توزع فئات المساندين على التوالي: 170، 135، 95 .

لدينا $n = 400 \geq 30$، الأعداد الافتراضية $np_{i0} = 160, 140, 100 \geq 1$ ، وأكثر من 80 % من $np_{i0} \geq 5$.

$$T = \sum_i \frac{(n_i - np_{i0})^2}{np_{i0}} = \frac{(170-160)^2}{160} + \frac{(135-140)^2}{140} + \frac{(95-100)^2}{100} = 1.05$$

$$\chi^2_{2;0.95} = 5.99 > 1.05 \Rightarrow \overline{R}H_0$$

5-9- تمارين محلولة:

التمرين1:

اشترت شركة خمسة منتجات من مورد يقول أن متوسط عمر المنتج هو 1050ساعة، ولما قامت الشركة باستعمال عينة المنتجات عاشت عدد 964، 1082، 1136، 825، 863ساعة، ولم تكن هذه النتائج مرضية بالنسبة للشركة، المطلوب إرسال شكوى للمورد تحلل النتائج وتقدم التوصيات.

- هل يمكن القول بصحة زعم المورد بمستوى معنوية 0.05؟

حل التمرين1:

- نلاحظ بأن الاختبار هنا هو اختبار المتوسط، وبما أن النتائج لم تكن مرضية بالنسبة للشركة فهذا يعني أن الاختبار هو اختبار أحادي الاتجاه من اليسار ومنه الفرضيات تكتب من الشكل التالي:

$$\begin{cases} H_0 : \mu = 1050 \\ H_1 : \mu < 1050 \end{cases}$$

ومنه قاعدة القرار تكتب من الشكل التالي:

$$\begin{cases} RH_0 \ si \dfrac{\overline{X} - \mu_0}{\sigma_{\overline{X}}} < -z_{1-\alpha} \ . \\ \overline{R}H_0 \ \text{sinon.} \end{cases}$$

بما أن عدد المشاهدات للعينة هو 5، ولدينا δ الانحراف المعياري للمجتمع مجهول، إذن نحتاج إلى الانحراف المعياري للعينة S عند حساب $\delta_{\bar{x}}$. ونستخدم كذلك توزيع ستيودنت t ومنه تصبح قاعدة القرار تكتب كما يلي:

$$\begin{cases} RH_0 \ si \dfrac{\overline{X}-\mu_0}{S/\sqrt{n-1}} < -t_{n-1;\,1-\alpha} \\ \overline{R}H_0 \ \text{sinon.} \end{cases}$$

– لدينا: $n = 5;;\ \overline{X} = \dfrac{\sum x_i}{n} = 974\ ;;;\ S \approx 134.66$

ومنه بالتعويض في متغيرة القرار t_c نجد:

$$t_c = \frac{974 - 1050}{134.66 \Big/ \sqrt{4}} \approx -1.13$$

$$-t_t = -t_{4,1-\alpha} = -2.132$$

– القرار: بما أن $t_c = -1.13 > -t_t = -2.132$ فإننا نقبل الفرضية الصفرية H_0، بمعنى أننا يمكن أن نجزم بصحة ادعاء المورد، وبالتالي لا يمكن رفع شكوى حول عدم صحة ادعائه.

التمرين2:

لدينا اختبار لقياس قدرة الطفل على تعلم كلمات جديدة، والمطلوب معرفة ما إذا كان متوسط درجات هذا الاختبار لأطفال إحدى المجتمعات يختلف عنه لأطفال مجتمع أخر، افترض أن درجات أطفال هاتين المجتمعين تتبع التوزيع الطبيعي بحيث لهما نفس التباين، أعطى الاختبار لمجموعتين عشوائيتين مستقلتين من الأطفال بعمر 4 سنوات البيانات التالية:

المجموعة الأولى	المجموعة الثانية
$n_1 = 10$	$n_2 = 8$
$\overline{X}_1 = 95$	$\overline{X}_2 = 97$
$S_1^2 = 40$	$S_2^2 = 36$

– اختبر الفرض القائل بتساوي قدرات أطفال المجموعتين على تعلم كلمات جديدة، بمستوى معنوية $\alpha = 0.05$ ؟.

حل التمرين2:

لدينا الاختبار هو اختبار الفرق بين متوسطين ومنه فان الفرضيتين تكتب من الشكل التالي:

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases}$$

– وبما أن العينات صغيرة فإن الإحصائية في هذه الحالات بافتراض أن المجتمعين طبيعيان، وأن تباين المجتمع الأول يساوي تباين المجتمع الثاني ولكنه مجهول (بمعنى أن $\sigma_1^2 = \sigma_2^2$ ولكن قيمة هذا التباين غير معروفة) وأن العينتين مستقلتان فإن إحصائية الاختبار تأخذ الشكل التالي والتي لها توزيع t بدرجات حرية تساوي $n_1 + n_2 - 2$:

$$T^{/} = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \rightarrow t_{n_1 + n_2 - 2}$$

والقرار يكتب من الشكل التالي:

$$\begin{cases} RH_0 ..Si ...T^{/} > t_{n_1 + n_2 - 2} \\ \overline{R}H_0 ..Sinon \end{cases}$$

– بالتعويض بالقيم في متغيرة القرار نجد:

$$T^{/} = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$= \frac{95 - 97}{\sqrt{\frac{(9)(40) + (7)(36)}{16}\left(\frac{1}{10} + \frac{1}{8}\right)}} \approx -0.68$$

$$t_{n_1 + n_2 - 2} = t_{16 ; 0.025} = 2.12$$

– القرار: بما أن $T^{/} = -0.68 < t_t = 2.12$ فإننا نقبل فرضية العدم ونستنتج أن الفرض القائل بتساوي قدرات أطفال المجموعتين على تعلم كلمات جديدة صحيح.

التمرين3:

سحبنا من مجتمعين طبيعيين عينتين حجم الأولى $n_1=51$ وحجم الثانية $n_2=67$ ومستقلتين، وجدنا النتائج التالية:

$$S_1^2 = 13 \;;\; S_2^2 = 11 \;;\; \overline{X}_1 = 101 \;;\; \overline{X}_2 = 96 .$$

– كيف يمكن إجراء اختبار تساوي تبايني المجتمعين بمستوى معنوية 0.05؟.

<u>حل التمرين3:</u>

الغرض من الاختبار هو تأكيد أو نفي تساوي تباينا مجتمعين من خلال عينتين عشوائيتين بسيطتين مستقلتين.

تكتب الفرضياتفي حالة الاختبار الثنائيكما يلي:

$$H_0 : \sigma^2_1 = \sigma^2_2 \leftrightarrow H_1 : \sigma^2_1 \neq \sigma^2_2$$

بما أن المجتمعين طبيعيين وحجم العينتين كبير يعني أكبر من 30، فإننا نعتمد في الاختبار على متغيرة القرار T والتي تساوي:

$$T = \frac{1}{2}\ln\left(\frac{\hat{S}_1^2}{\hat{S}_2^2}\right) / \sqrt{\frac{1}{2}\left(\frac{1}{n_1-1}+\frac{1}{n_2-1}\right)} \approx N(0;1)$$

$$\hat{S}_1^2 = S_1^2 \frac{n_1}{n_1-1} = 13\frac{51}{50} = 13.26$$

$$\hat{S}_2^2 = S_2^2 \frac{n_2}{n_2-1} = 11\frac{67}{66} \approx 11.17$$

$$T = \frac{1}{2}\ln\left(\frac{13.26}{11.17}\right) / \sqrt{\frac{1}{2}\left[\left(\frac{1}{50}\right)+\left(\frac{1}{66}\right)\right]} \approx 0.65$$

$$\alpha = 0.05 \Rightarrow z_t = 1.96$$

– القرار: بما أن $T = 0.65 < Z_t = 1.96$ فإننا نقبل الفرضية الصفرية، أي نقبل بتساوي التباينين.

5–10– تمارين غير محلولة:

<u>التمرين1</u>: إذا كان من المعروف أن جسم الإنسان البالغ يحتاج يوميا في المتوسط 800 مليغرام من الكالسيوم لكي يقوم بوظائفه خير قام.ويعتقد احد علماء التغذية أن الأفراد ذوي الدخل المنخفض لا يستطيعون تحقيق هذا المتوسط ، ولاختبار ذلك تم اختيار عينة من 50 شخصا بالغا من بين ذوي الدخل المنخفض فكان متوسط ما يتناوله من كالسيوم يوميا هو

755.3مليغرام والانحراف المعياري هو 239.3مليغرام.فهل تدل هذه النتائج على أن متوسط ما يتناوله الأشخاص البالغون من ذوي الدخل المنخفض من كالسيوم يقل عن 800مليغرام ؟ استخدمي مستوى معنوية 0.05

التمرين2: البيانات التالية تمثل نتائج تجربة أجريت على عشرين شخصاً لاختبار مدى فعالية نظام خاص من الغذاء لتخفيف الوزن، حيث تم قياس أوزانهم قبل البدء في تطبيق هذا النظام، وبعد إتباع هذا النظام الخاص لمدة ثلاثة شهور.

قبل	96	110	90	94	107	93	89	120	103	92
بعد	90	96	85	87	104	85	76	103	95	84
قبل	86	94	86	110	105	123	95	90	111	123
بعد	78	84	80	102	95	109	89	83	102	107

المطلوب: هل تستطيع أن تستنتج أن نظام الغذاء كان فعالاً في تخفيف الوزن مستخدماً مستوى دلالة $\alpha = 0.05$؟

التمرين3: أجريت التجربة التالية لمقارنة نوعين من الأعلاف، خصص لكل علف 10 أبقار وكان معدل الزيادة في أوزان الأبقار كما يلي:

علف(1)	31	34	29	26	32	35	38	34	30	32
علف(2)	28	26	24	28	29	30	29	26	31	29

- إذا كان المجتمعين يتبعان التوزيع الطبيعي، ولهما نفس التباين.

هل هناك فرق معنوي بين نوعي العلف، وذلك بمستوى معنوية $(\alpha = 0,05)$؟.

المراجع باللغة العربية:

1- عدنانعباس حميدان، مطانيوس مخول، فريد جاعوني، عمار ناصر آغا، الإحصاء التطبيقي، منشورات جامعة، دمشق2006.

2- محمدأمانى موسى، التحليل الإحصائي للبيانات، مركز تطوير الدراسات العلياالقاهرة، مصر 2007.

3- جلال الصياد، عبد الحميد محمد ربيع، مبادئ الطرق الإحصائية، الطبعة الأولى، دار الناشر تهامة، جدة- الجزائر العربية السعودية1983.

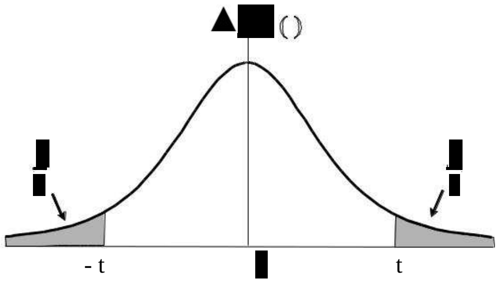
المراجع باللغة الأجنبية:

1-Amany Mousa: Cairo (2005), Statistical Data Analysis, Center for Advancement of Postgraduate Studies and Research, Faculty of Engineering, Cairo University.
2-Freund, J (2001) Modern Elementary Statistics 10th Edition, Printice Hall.
3-Hamburg, Morris (1989) Statistical Analysis for Decision Making, Harcoort Brace Jovanwich, USA.
4-Keller, G and Waracck, B (2001): Statistics for Management and Economics 6th Edition Duxbury.
5-Mills, Frederick (1955), Statistical Methods, Holt, Rinehart and Winston, New York, USA.
6-Neeter, J, Waserman, Whitmare, (1993): Applied Statistics. 4th Edition, Louise Richardson.
7-Wonnacott, Thomas and Wonnacott, Ronald (1990) Introductory Statistics for Business and Economics, John wiley & soon, New York, USA.

الملاحق:

Table de la loi de Student

Valeurs de T ayant la probabilit´e P d’ˆetre d´epass´ees en valeur absolue



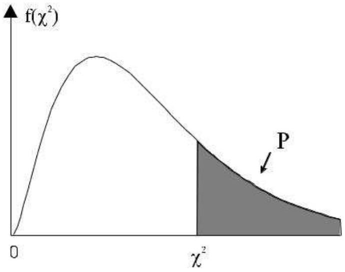
v	P =0,90	0,80	0,70	0,60	0,50	0,40	0,30	0,20	0,10	0,05	0,02	0,01
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604
5	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707
7	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499
8	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355
9	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106
12	0,128	0,260	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861
20	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845
21	0,127	0,257	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831
22	0,127	0,256	0,390	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819
23	0,127	0,256	0,390	0,532	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807
24	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,492	2,797
25	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787
26	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779
27	0,127	0,256	0,389	0,531	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771
28	0,127	0,256	0,389	0,530	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750
∞	0,126	0,253	0,385	0,524	0,674	0,842	1,036	1,282	1,645	1,96	2,326	2,576

Nota. v est le nombre de degr´es de libert´.
Le quantile d’ordre $1 - \frac{\alpha}{2}$ se lit dans la colonne P = α.

Le quantile d’ordre 1 − α se lit dans la colonne P = 2α.

Table de la loi du Khi-deux

Valeurs de χ^2 ayant la probabilit´e P d’ˆetre d´epass´ees



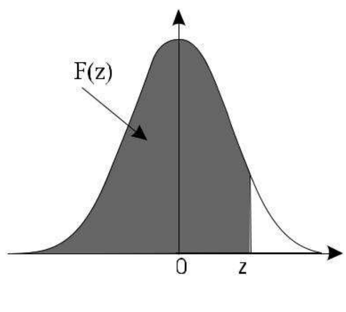
v	P = 0,995	0,99	0,975	0,95	0,90	0,10	0,05	0,025	0,01	0,005
1	0,00004	0,0002	0,001	0,0039	0,0158	2,706	3,841	5,024	6,635	7,879
2	0,010	0,020	0,051	0,103	0,211	4,605	5,991	7,378	9,210	10,597
3	0,072	0,115	0,216	0,352	0,584	6,251	7,815	9,348	11,345	12,838
4	0,207	0,297	0,484	0,711	1,064	7,779	9,488	11,143	13,277	14,860
5	0,412	0,554	0,831	1,145	1,610	9,236	11,070	12,833	15,086	16,750
6	0,676	0,872	1,237	1,635	2,204	10,645	12,592	14,449	16,812	18,548
7	0,989	1,239	1,690	2,167	2,833	12,017	14,067	16,013	18,475	20,278
8	1,344	1,646	2,180	2,733	3,490	13,362	15,507	17,535	20,090	21,955
9	1,735	2,088	2,700	3,325	4,168	14,684	16,919	19,023	21,666	23,589
10	2,156	2,558	3,247	3,940	4,865	15,987	18,307	20,483	23,209	25,188
11	2,603	3,053	3,816	4,575	5,578	17,275	19,675	21,920	24,725	26,757
12	3,074	3,571	4,404	5,226	6,304	18,549	21,026	23,337	26,217	28,300
13	3,565	4,107	5,009	5,892	7,042	19,812	22,362	24,736	27,688	29,819
14	4,075	4,660	5,629	6,571	7,790	21,064	23,685	26,119	29,141	31,319
15	4,601	5,229	6,262	7,261	8,547	22,307	24,996	27,488	30,578	32,801
16	5,142	5,812	6,908	7,962	9,312	23,542	26,296	28,845	32,000	34,267
17	5,697	6,408	7,564	8,672	10,085	24,769	27,587	30,191	33,409	35,718
18	6,265	7,015	8,231	9,39	10,865	25,989	28,869	31,526	34,805	37,156
19	6,844	7,633	8,907	10,117	11,651	27,204	30,144	32,852	36,191	38,582
20	7,434	8,260	9,591	10,851	12,443	28,412	31,410	34,170	37,566	39,997
21	8,034	8,897	10,283	11,591	13,240	29,615	32,671	35,479	38,932	41,401
22	8,643	9,542	10,982	12,338	14,041	30,813	33,924	36,781	40,289	42,796
23	9,260	10,196	11,689	13,091	14,848	32,007	35,172	38,076	41,638	44,181
24	9,886	10,856	12,401	13,848	15,659	33,196	36,415	39,364	42,980	45,559
25	10,520	11,524	13,120	14,611	16,473	34,382	37,652	40,646	44,314	46,928
26	11,160	12,198	13,844	15,379	17,292	35,563	38,885	41,923	45,642	48,290
27	11,808	12,879	14,573	16,151	18,114	36,741	40,113	43,195	46,963	49,645
28	12,461	13,565	15,308	16,928	18,939	37,916	41,337	44,461	48,278	50,993
29	13,121	14,256	16,047	17,708	19,768	39,087	42,557	45,722	49,588	52,336
30	13,787	14,953	16,791	18,493	20,599	40,256	43,773	46,979	50,892	53,672

Nota. v est le nombre de degrés de liberté.

Pour v > 30, on peut admettre que la quantité $\sqrt{2\chi^2} - \sqrt{2v-1}$ suit la loi normale centrée réduite.

Fonction de répartition de la loi normale centrée réduite

(probabilité F(z) de trouver une valeur inférieure à z)



z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986

Table pour les grandes valeurs de z

z	3,0	3,1	3,2	3,3	3,4	3,5	3,6	3,7	3,8	3,9
F(z)	0,998650	0,999032	0,999313	0,999517	0,999663	0,999767	0,999841	0,999892	0,999928	0,999952

z	4,0	4,1	4,2	4,3	4,4	4,5	4,6	4,7	4,8	4,9
F(z)	0,999968	0,999979	0,999987	0,999991	0,999995	0,999997	0,999998	0,999999	0,999999	1,000000

Nota. La table donne F(z) pour z positif. Pour z n´egatif, il faut prendre le compl´ement a` l'unit´e de la valeur lue dans la table. Exemple : F(-1,37) = 1 - F(1,37) =1 - 0,9147 = 0,0853.