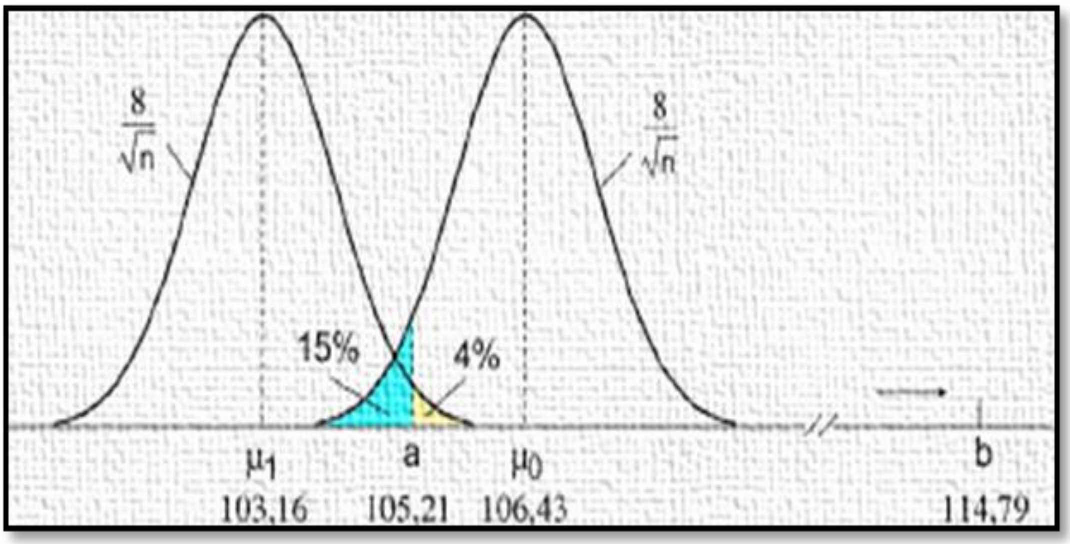
جامعة محمد الصديق بن يحيى-جيجل
كلية العلوم الاقتصادية والتجارية وعلوم التسيير

مطبوعة تحت عنوان:

دروس في الإحصاء (03)



إعداد الدكتور: بودغدغ أحمد

السنة الجامعية 2017/2016

فهرس المحتويات

مقدمة

الإحصاء علم يهتم بالمعلومات والبيانات – ويهدف إلى تجميعها وتبويبها وتنظميها وتحليلها واستخلاص النتائج منها بل وتعميم نتائجها – واستخدامها في اتخاذ القرارات، وأدى التقدم المذهل في تكنولوجيا المعلومات واستخدام الحاسبات الآلية إلى مساعدة الدارسين والباحثين ومتخذي القرارات في الوصول إلى درجات عالية ومستويات متقدمة من التحليل ووصف الواقع ومتابعته ثم إلى التنبؤ بالمستقبل .

ولم تعد 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 أسلوب الإقناع والتقدير والقياس.

ولقد أصبح الاتجاه العام في مثل هذه البحوث والدراسات هو استخدام طرق القياس الكمية ووسائل الإقناع الإحصائية وذلك لتحديد الخصائص وإبراز الاتجاهات العامة في الظواهر الاجتماعية والإدارية، وتحليل العلاقات المتشابكة والمتبادلة بين الظواهر على أساس موضوع غير متميز.

وعلم الإحصاء 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 المتغيرات المحددة 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.

فبحوث الرأي العام على سبيل المثال تقوم على مقابلة ودراسة عينة صغيرة من أفراد المجتمع ولكن نتائجها تستخدم في الاستدلال على اتجاهات الرأي العام في المجتمع ككل، وبذلك يمكن القول بان الإحصاء يشير إلي طرق تنظيم وتلخيص البيانات وإلى الأساليب التي تستخدم في تحليل وتفسير النتائج واستخلاصاتها والتي يمكن 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 الوصفي والإحصاء (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- مفهوم الاستدلال: ويقصد بوظيفة الاستدلال اشتقاق النتائج من دراسة وفحص المقدمات والبيانات المتوافرة عن ظاهرة معينة. ولهذا يطلق على العملية الإحصائية التي تستخدم الاستدلالي على أساس المنطق الاستدلالي المبني على نظرية الاحتمالات الرياضية فمن عينة محددة من أعمال أحد المصانع وباستخدام الأسلوب الإحصاء 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 الأدوات الإحصائية المستخدمة وحجم العينة محل الدارسة والإجراءات الإحصائية اتخذت عند اختيارها.

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

إذا كانت الظاهرة موضوع الدراسة والتحليل ممثلة للمجتمع الذي تنتمي إليه فانه يمكن الحصول على نتائج معنوية عن المجتمع بتحليل بيانات هذه الظاهرة وهو ما يعرف بالاستدلال ويعتمد هذا الأسلوب في 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 تأتى أهمية 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 الإحصائي أساسا لتطور المعرفة العلمية باعتباره البرهان لهذه القوانين. ووظيفة الاستقراء تحقق مطلبين أساسيين في البحث: الأول تقدير خواص المجتمع والثاني اختبارات الفروض حول هذه الخواص. 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 أية خاصية تأخذ قيما باختلاف المتغيرات (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 المثال عندما يراد معرفة مدى ممارسة طلبة الجامعة للأنشطة الرياضية يكونوا إما (ممارسين لها) أو (غير ممارسين لها) ..

1-2-3- <u>المتغيرات العشوائية:</u> 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:

- **<u>المتغير العشوائي المتقطع(المنفصل):</u>**

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.

- **<u>المتغير العشوائي المستمر:</u>**

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- <u>مفهوم العينات والمجتمع الإحصائي:</u>

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

يعرف المجتمع بأنه عبارة 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 أو طلاب 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 أعداد الطلاب أو 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.

<u>**1-3-2 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 استخدام العينة (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- <u>أنواع العينات:</u>

<u>– العينات غير الاحتمالية</u>: 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.

<u>– العينات الاحتمالية:</u>

– 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- مفهوم العينة النفادية والعينة غير النفادية Echantillon exhaustif et non exhaustif:

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- مفهوم إحصائية المعاينة Statistique de l'échantillonnage:

لتقدير معالم المجتمع (متوسط المجتمع μ، تباين المجتمع σ^2 النسبة p ...) ننطلق من بيانات العينة، حيث نحتاج إلى حساب معالم مثل متوسط العينة $\bar{x}$ ، تباين العينة S^2، النسبة في العينة p'. 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- نظرية توزيع المعاينة

يعتبر أسلوب المعاينة من أهم الأساليب الإحصائية 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 إلا إذا تم اختيار المفردات بطريقة نضمن بها أن تكون العينة ممثلة للمجتمع تمثيلا صادقا.

2-1- تعريف توزيع المعاينة وعدد العينات المسحوبة:

2-1-1- 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.

2-1-2- 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.

2-2- توزيع المعاينة 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 $\overline{X}$ كأي مجتمع أخر له توزيع احتمالي يتمتع بجميع صفات وخواص التوزيعات الاحتمالية وبالطبع له متوسط وانحراف معياري.

2-2-1- 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 $E(\overline{X})$ تكتب كما يلي:

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

<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 السحب بدون إرجاع مكونة من مفردتين $(n=2)$؟.
- قارن بين $\mu_{\bar{x}}$ و μ؟

<u>الحل:</u>

- حساب متوسط المجتمع: $\mu = (1 + 3 + 5) / 3 = 3$
- إيجاد القيمة المتوقعة لمتوسط عينة مسحوبة من المجتمع في حالة السحب بالإرجاع وفي حالة السحب بدون 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: $N^n = 3^2 = 9$.

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

العينات الممكنة		
(1, 1)	(3, 1)	(5, 1)
(1, 3)	(3, 3)	(5, 3)
(1, 5)	(3, 5)	(5, 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: $C_N^n = C_3^2 = \frac{(3)!}{(3-2)!(2)!} = 3$

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

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

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

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

2-2-2- تباين توزيع المعاينة للمتوسطات:

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

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

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

وهذا يعني أن الانحراف المعياري هو الانحراف 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.

<u>**مثال2:** </u>في نفس المثال السابق (المثال1).

- أحسب تباين المجتمع؟
- أحسب التباين لتوزيع المعاينة للمتوسطات $\delta_{\bar{x}}^2$ في حالة السحب بالإرجاع وفي حالة السحب بدون 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:**

السحب بالإرجاع $\overline{x_i}$		
1	2	3
2	3	4
3	4	5

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

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

– 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 $\overline{x_i}$ السحب بدون إرجاع			
2			
3	4		

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

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

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

2-2-3- 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 الطبيعي بمتوسط 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 $z = \frac{\bar{x}-\mu}{\delta/\sqrt{n}}$ تؤول إلى التوزيع الطبيعي المعياري عندما يكون n كبيرا $(n \geq 30)$ ونكتب:

$$z \longrightarrow N(0.1)$$

- 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 تستخدم هذه 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:

$$1-n=49:\frac{n}{N}=\frac{49}{1200}=0.04<0.05\Rightarrow \delta_{\bar{x}}=\frac{\delta}{\sqrt{n}}=\frac{15}{\sqrt{49}}\approx 2.14$$
$$E(\overline{X})=\mu=20$$
$$2-n=81:\frac{n}{N}=\frac{81}{1200}=0.0675>0.05\Rightarrow \delta_{\bar{x}}=\frac{\delta}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$
$$=\frac{15}{\sqrt{81}}\sqrt{\frac{1200-81}{1200-1}}\approx 1.61$$
$$E(\overline{X})=\mu=20$$

2-4- توزيع المعاينة للفروق والمجاميع

2-4-1- 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:

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

<u>مثال5</u>. إذا كانت الإحصائية هي المتوسط، وكان لدينا مجتمعين كبيرين من المفردات، وكان متوسط الأول μ_1 وتباينه 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_2 فإن:

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

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

<u>مثال6</u>. إذا كانت الإحصائية هي النسبة فإن:

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

$$\delta^2_{p_1-p_2} = \delta^2_{p_1} + \delta^2_{p_2} = \frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}$$

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

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

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

نظرية4: في حالة $n_1 \geq 30$ و n_2 ، يقترب توزيع المتغيرة المعيارية للفرق بين متوسطين من التوزيع الطبيعي المعياري. ونكتب:

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

<u>مثال7</u>: ليكن المجتمع V_1 : 2، 6، 10.

والمجتمع V_2 : 1، 3. تحقق من أن:

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

الحل:

V_1				
10	**6**	**2**	$V_1 - V_2$	
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 $X_1 \rightarrow N(30;;25)$ وسحبت من مجتمعه عينة حجمها 30 مشاهدة و $X_2 \rightarrow N(20;;16)$ 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:

$$\mu_{\bar{x}_1-\bar{x}_2} = \mu_{\bar{x}_1} - \mu_{\bar{x}_2} = \mu_1 - \mu_2 = 30 - 20 = 10$$
$$\delta^2_{\bar{x}_1-\bar{x}_2} = \delta^2_{\bar{x}_1} + \delta^2_{\bar{x}_2} = \frac{\delta_1^2}{n_1} + \frac{\delta_2^2}{n_2} = \frac{25}{30} + \frac{16}{35} = 1,29$$

2-5- توزيع المعاينة للتباين وتوزيع المعاينة لنسبة تبايني عينتين

2-5-1- 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" لـ σ^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 :

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

2-6- تمارين محلولة:

التمرين1:

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

1- ما هو احتمال أن تقع X بين القيمتين 3 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 القيمتين 3 و5:

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

2- احتمال أن يكون X أكبر من 1:

$$
\begin{aligned}
&p(X > 1) = p\left(Z > \frac{1-3}{2}\right) \\
&= p(Z > -1) = 1 - p(Z \le -1) \\
&= 1 - \Phi(-1) = 1 - 1 + \Phi(1) \\
&= 0.8413
\end{aligned}
$$

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؟

<u>حل التمرين3:</u>

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

لدينا: $\mu = 90;;n = 36;;\delta = 9$

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

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

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 المجتمعيين طبيعيين.

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

حل التمرين4:

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

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

$$\delta^2_{\bar{x}_1 - \bar{x}_2} = \delta^2_{\bar{x}_1} + \delta^2_{\bar{x}_2} = \frac{\delta_1^2}{n_1} + \frac{\delta_2^2}{n_2} = \frac{18}{12} + \frac{25}{10} = 4$$
$$\bar{x}_1 - \bar{x}_2 \to N(5;;4)$$

2-7- 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$\left(\delta_{\bar{x}}^2; \mu_{\bar{x}}\right)$(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 يساوي 8 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؟

3- 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 بمجال يسمى 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%).

3-1- مفاهيم أساسية

3-1-1- 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 لـ σ^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:

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

3-1-2 درجة التأكد

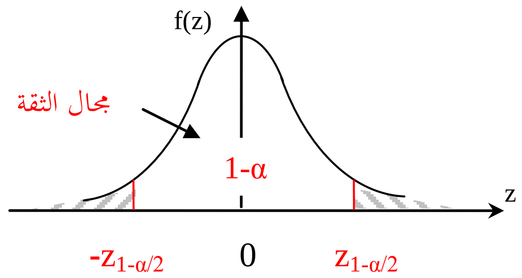
لكي يكون التقدير علميا ينبغي تقييم احتمال أن تكون المعلمة تنتمي فعلا إلى المجال المحدد، لذلك نلحق بالمجال ما يسمى بدرجة أو مستوى الثقة، ويرمز له ب$(1-\alpha)$. الاحتمال المعاكس يسمى احتمال الخطأ ويرمز له ب α، ويسمى أيضا ""مستوى المعنوية".

مثال3: دخل الأسرة في المنطقة(A) ينتمي إلى المجال [16000، 20000] بمستوى معنوية 5 % أي بمستوى ثقة 95 % . وتسمي الحدود 16000 و20000 بحدود الثقة.

3-1-3- 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 و σ_s متوسط وانحراف معياري توزيع المعاينة لإحصائية ما s حيث $\mu = \mu_s$. 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).

3-2- التقدير بمجال

3-2-1- مجال الثقة للمتوسط

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

3-2-1-1- 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 وبالرموز فإن:

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

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 معلوم يأخذ 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[1] :

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

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 $\hat{S}$.

---

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

– يمكن كتابة أشهر وأهم درجات ومعاملات الثقةz_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 μيوجد داخل المجال $\left[\bar{x} \pm 1.96\delta_{\bar{x}}\right]$ بمستوى ثقة 95% (0.95) أي بمستوى معنوية 5 % (0.05)....

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.

3-2-1-2- تقدير متوسط المجتمع μ باستخدام التوزيع t :

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

ولعل الاختلاف الأساسي بين توزيع t والتوزيع الطبيعي هو أن الانحراف المعياري للعينة هو المستخدم في الأول بدلا من الانحراف المعياري للمجتمع في الثاني، وفيما عدا ذلك

فالتوزيعان متماثلان وكلما زادت قيمة n كلما اقترب توزيع t من توزيع z ويعتمد توزيع t على ما يعرف بدرجات الحرية.

3-2-1-3- 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>: إن مجال الثقة للوسط الحسابي للمجتمع في حالة δ مجهولة وحجم العينة أقل من 30 يأخذ الشكل التالي:

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

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 حسابي 72 دولاراً وانحراف معياري بلغ 6.4 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:

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

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:

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

أي أن الوسط الحسابي للدخول اليومية يتراوح بين 67.18 دولاراً كحد أدنى، و 76.82 دولاراً كحد أعلى وذلك بمستوى ثقة 95%.

3-2-1-4- تحديد حجم العينة لتقدير الوسط الحسابي للمجتمع :

يعتبر تحديد حجم العينة المناسب من المشاكل 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 :

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

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 % ؟

الحل :

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: 60 فردا.

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 %.

3-2-2- مجال الثقة للنسبة وحجم العينة المناسب.

3-2-2-1- 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.

3-2-2-2- 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.

3-5- 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 $\hat{\alpha}$ 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 $\hat{\alpha}$ مقدر بطريقة المعقولية العظمى يجب أن يكون حل للمعادلة الآتية:

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

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؟

<u>الحل:</u>

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:

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

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

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

3-6- تمارين محلولة:

<u>التمرين1:</u> 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.

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

<u>حل التمرين1</u>:

- 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:

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

التمرين02: إذا كان لدينا متغيران عشوائيان x_1 و x_2 يتبعان 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	$\bar{x}_1 = 40$	$\bar{x}_2 = 30$
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%؟.

حل التمرين2:

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

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

- 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:

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

التمرين3:

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)

حل التمرين3:

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:

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

3-7- تمارين غير محلولة:

التمرين1: 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$(\delta^2_{p'})=0.01$؟.

التمرين2:

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 $S = 6.2$؟.

التمرين3:

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 % ؟.

التمرين4:

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؟.

4- اختبارات الفروض وتطبيقاتها

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..

4-1- مفاهيم أساسية:

4-1-1- الفرض العدم (أو الصفري)

فرضية العدم هي "الفرضية الأساسي المراد اختبارها". ويرمز لها عادة بالرمز : H_0. وهي فرضية حول معلمة المجتمع التي نجري اختبار عليها 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 هو 20000 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).

4-1-2- 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 المجتمع هو 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**<u>الخلاصة:</u>** 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.

4-1-3- 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 :

4-1-3-1- 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 : $p(RH_0 / H_0) = \alpha$

4-1-3-2- الخطأ من النوع الثاني:

وفي المقابل فإن الخطأ من النوع الثاني يعني " قبول الفرض العدم بينما هو خاطئ أي 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 α -1 ويكتب : $p\left(\overline{R}H_0 / H_1\right)=1-\alpha$

جدول: القرارات الخاطئة والصائبة في عملية اختبار الفرضيات

نتيجة الاختبار	الوضع الحقيقي	
	الادعاء غير صحيح H_oصحيحة	الادعاء صحيح 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 الصفرية قوة الاختبار فيما يقيس احتمال قبولها فعالية 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.

4-1-3-3- 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.

4-2- اختبار المتوسط

يتناول هذا الاختبار متوسط المجتمع (μ)، مثل متوسط الدخل، متوسط وزن منتج معين، ويؤكد اختبار المتوسط فرضية مساواته لقيمة ما μ_0. و للقيام 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.

4-2-1- 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):

$$H_0 : \mu = \mu_0 \leftrightarrow H_1 : \mu \neq \mu_0$$

الفرضية H_0 الفرضية الصفرية أو فرضية العدم، ويؤدي الاختبار إما إلى رفضها ونكتب RH_0 وفي هذه الحالة نقبل الفرضية البديلة أو المعاكسة أو عدم رفضها ونكتب $\overline{R}H_0$.

μ_0 هي القيمة الافتراضية لـ μ عادة ما تكون μ_0 محددة بناءا على بيانات عينة عشوائية بسيطة، وفي هذه الحالة يمكن استخدام الخاصية $\bar{x} \to N\left(\mu, \frac{\delta^2}{n}\right)$، لإجراء الاختبار.

حيث أنه تحت H_0 فإن : $\bar{x} \to N\left(\mu_0, \frac{\delta^2}{n}\right)$

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 α-1، 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:

$$\begin{cases} RH_0 \ si \left|\frac{\bar{X}-\mu_0}{\sigma_{\bar{X}}}\right| > z_{1-\alpha/2} \ . \\ \overline{R}H_0 \ \text{sinon.} \end{cases} \quad \text{أو} \quad \begin{cases} RH_0 \ si \ z_c = \frac{\bar{X}-\mu_0}{\sigma_{\bar{X}}} \notin \left[-z_{1-\alpha/2}; \ z_{1-\alpha/2}\right] . \\ \overline{R}H_0 \ \text{sinon.} \end{cases}$$

حساب z الجدولية:

ويرمز لها ب z_t حيث، 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.

4-2-2- اختبار أحادي الاتجاه للمتوسط.

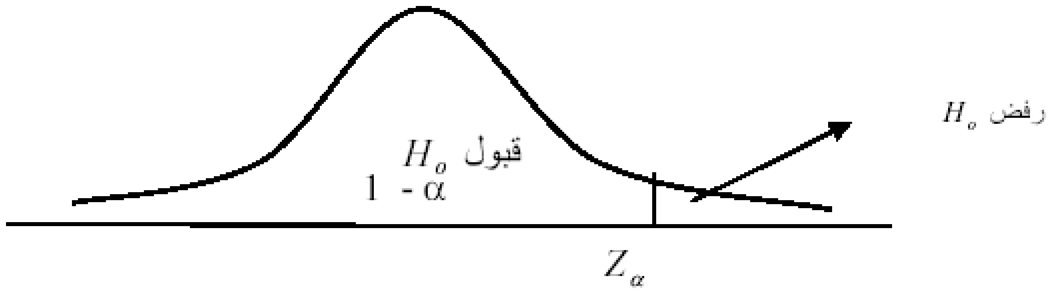
يتميز الاختبار الثنائي عن الأحادي في الفرضية البديلة التي هي عدم مساواة في الاختبار الثنائي وأكبر تماما أو أصغر تماما (حسب الحالة) في الاختبار الأحادي، وهذا يترتب عليه تغيير في قاعدة القرار.

4-2-2-1- اختبار أحادي الاتجاه من اليمين.

- **الفرضيات :** $H_0 : \mu = \mu_0 \leftrightarrow H_1 : \mu > \mu_0$

- **قاعدة القرار:**

$$\begin{cases} RH_0 \ si \dfrac{\overline{X} - \mu_0}{\sigma_{\overline{X}}} > z_{1-\alpha} \ . \\ \overline{R}H_0 \ \text{sinon.} \end{cases}$$



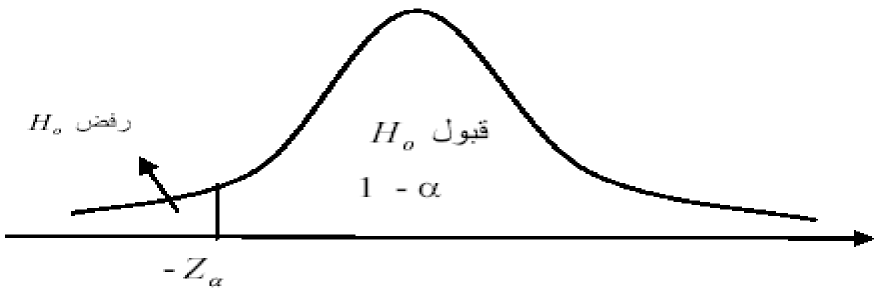
الشكل: منطقتي القبول و الرفض في حالة قاعد القرار اختبار أحادي الاتجاه من اليمين

4-2-2-2- اختبار أحادي الاتجاه من اليسار

- **الفرضيات :** $H_0 : \mu = \mu_0 \leftrightarrow H_1 : \mu < \mu_0$

- **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.

4-3- استخدام S كمقدر لـ σ والتوزيع t في اختبار المتوسط

4-3-1- استخدام S كمقدر لـ σ 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:

$$\delta_{\bar{x}} = \frac{\hat{S}}{\sqrt{n}} \text{ أو } \delta_{\bar{x}} = \frac{S}{\sqrt{n-1}} \text{ بـ } \delta_{\bar{x}} = \frac{\delta}{\sqrt{n}}$$

4-3-2- استخدام التوزيع t في اختبار المتوسط.

في حالة n< 30 و σ (الانحراف المعياري للمجتمع) مجهولا، لا يمكن استخدام التوزيع الطبيعي، ولكن لدينا :

$$\frac{\bar{X}-\mu}{S/\sqrt{n-1}} \sim t_{n-1} \quad \text{و تحت } H_0 \ (\mu = \mu_0) : \quad \frac{\bar{X}-\mu_0}{S/\sqrt{n-1}} \sim t_{n-1}$$

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 المحسوبة والتي تساوي 1.5 $(z_c = 1.5)$ 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.

4-4- 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 $n \geq 30$: $p' \to N\left(p, \delta^2_{p'}\right)$

استنادا إلى هذه الخصائص وتحت H0 :

$$p' \to N\left(p_0, \frac{p_0 q_0}{n}\right)$$

و من ثم يمكن تحديد قاعدة القرار بحسب طبيعة الاختبار كما 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 .

4-5- 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 الفرضيات كما 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.

4-5-1- تباينا المجتمعين معلومين

- **المجتمعين طبيعيين:**

$$T = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0.1)$$

حيث : يرمز بـ n1 إلى حجم العينة الأولى.

يرمز بـ n2 إلى حجم العينة الثانية.

يرمز بـ $\overline{X}1$ إلى الوسط الحسابي للعينة الأولى.

يرمز بـ $\overline{X}2$ إلى الوسط الحسابي للعينة الثانية.

يرمز بـ σ_1^2 إلى تباين المجتمع الأول.

يرمز σ_2^2 إلى تباين المجتمع الثاني.

مثال6: إذا كانت لدينا نتائج عينتين عشوائيتين مستقلتين مسحوبتين من منطقتين لمقارنة متوسط عمر الناخب فيهما حيث: $\overline{X}_1 = 40, \overline{X}_2 = 34, \delta_1^2 = 84, \delta_2^2 = 27, n_1 = 120, n_2 = 90$

– 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 الطبيعي Z 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 %.

4–5–2– 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 $n_1 + n_2 - 2$:

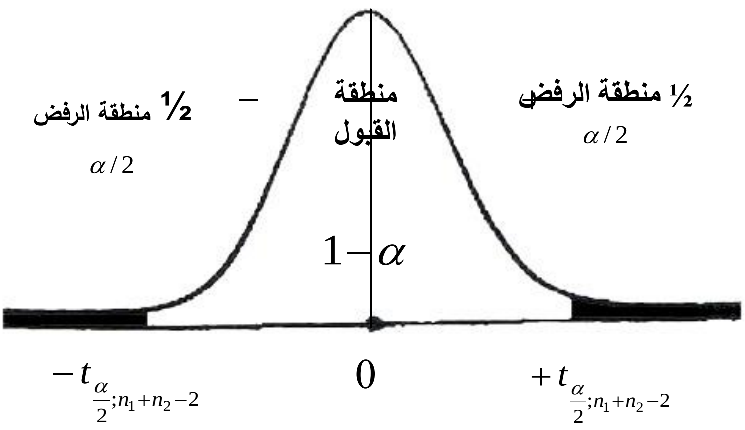
$$T' = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \rightarrow t_{n_1 + n_2 - 2}$$

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** $(\mathbf{n_1 , n_2 \geq 30})$: 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؟.

<u>الحل</u> :

– 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

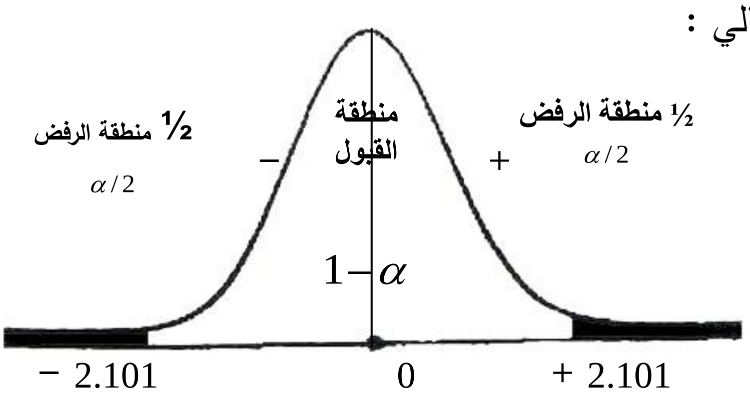
– حدود منطقتي القبول والرفض:

ونحصل عليها في هذه الحالة من جدول t عند درجات حرية تساوي $n_1 + n_2 - 2$

أي تساوي 2 - 10 + 10 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 0.7 أقل 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).

4-6- 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:

$$Z_{p_1' - p_2'} = \frac{p_1' - p_2'}{\sqrt{p' q' \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \rightarrow N(0,1)$$

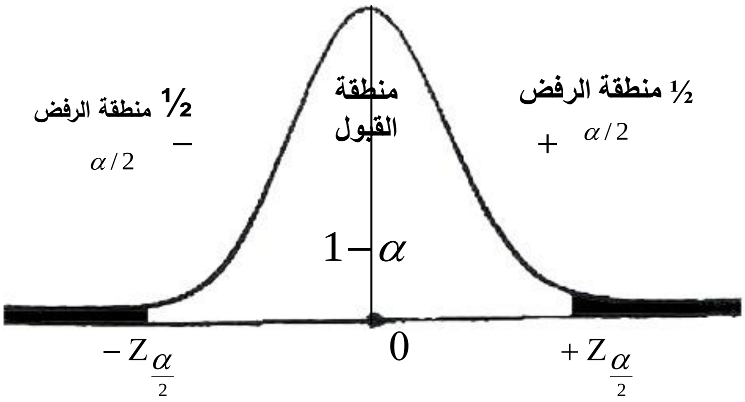
حيث:

$$p' = \frac{n_1 p_1' + n_2 p_2'}{n_1 + n_2}$$

$$q' = 1 - p'$$

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 B تم 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:

$$Z_{p_1'-p_2'} = \frac{p_1' - p_2'}{\sqrt{p'q'\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$$

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:

$$p^{/} = \frac{(100)(0.7)+(100)(0.5)}{100+100}$$
$$= \frac{70+50}{200} = \frac{120}{200}$$
$$p^{/} = 0.60$$
$$q^{/} = 1-0.6 = 0.40$$

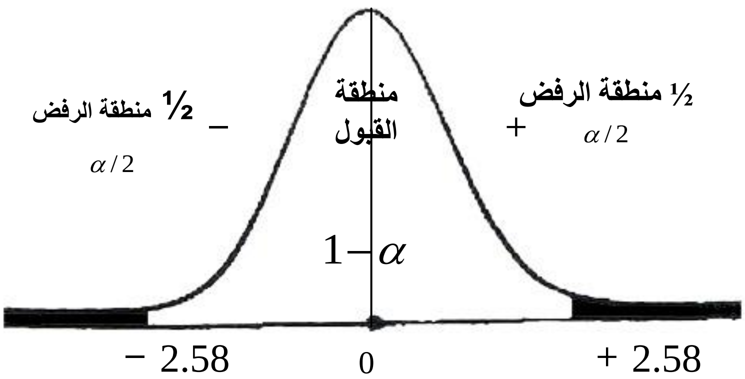
وبالتعويض في متغيرة القرار نجد:

$$Z_{p_1^{/}-p_2^{/}} = \frac{p_1^{/} - p_2^{/}}{\sqrt{p^{/}q^{/}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{0.70-0.50}{\sqrt{(0.6)(0.4)\left(\frac{1}{100}+\frac{1}{100}\right)}}$$
$$= \frac{0.20}{0.069} = 2.899$$

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 A تساوي نسبة المؤيدين له في المدينة B وذلك بمستوى معنوية 1%، وقبول الفرض البديل بأن النسبتين غير متساويتين.

4-8- 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: $n=5;;\overline{X}=\frac{\sum x_i}{n}=974;;;S\approx 134.66$

ومنه بالتعويض في متغيرة القرار t_c نجد:

$$t_c=\frac{974-1050}{134.66/\sqrt{4}}\approx -1.13$$

$$-t_t=-t_{4,1-\alpha}=-2.132$$

- 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

ولكن قيمة هذا التباين غير معروفة) وأن العينتين مستقلتان فإن إحصائية الاختبار $\sigma_1^2 = \sigma_2^2$ تأخذ الشكل التالي والتي لها توزيع t بدرجات حرية تساوي $n_1 + n_2 - 2$:

$$T' = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \to t_{n_1+n_2-2}$$

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.

<u>التمرين3:</u>

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 الفرضيات في حالة الاختبار الثنائي كما يلي:

$H_0 : \sigma^2_1 = \sigma^2_2 \leftrightarrow H_1 : \sigma^2_1 \neq \sigma^2_2$

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.

4-10- 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 755.3 مليغرام والانحراف المعياري هو 239.3 مليغرام. فهل تدل هذه النتائج على أن متوسط ما يتناوله الأشخاص البالغون من ذوي الدخل المنخفض من كالسيوم يقل عن 800 مليغرام ؟ استخدمي مستوى معنوية 0.05

<u>التمرين2:</u> 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 $\alpha = 0.05$؟

<u>التمرين3:</u> 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- عدنان عباس حميدان، مطانيوس مخول، فريد جاعوني، عمار ناصر آغا، الإحصاء التطبيقي، منشورات جامعة، دمشق 2006.

2- محمد أمانى موسى، التحليل الإحصائي للبيانات، مركز تطوير الدراسات العليا القاهرة، مصر 2007.

3- جلال الصياد، عبد الحميد محمد ربيع، مبادئ الطرق الإحصائية، الطبعة الأولى، دار الناشر تهامة، جدة- المملكة العربية السعودية 1983.

4- بوعبد الله صالح، محاضرات الإحصاء الرياضي، كلية العلوم الاقتصادية، جامعة المسيلة، 2005-2006.

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.