وهناك طريقتان هامتان بين توزيع المعاينة للوسط والمجتمع الأصلي.

نظرية (1)

إذا أخذنا عينات متكررة حجمها n من مجتمع ما فإن:

$$\mu_{\bar{x}} = \mu$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \quad \text{or} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

حيث تستخدم المعادلة الأخيرة للمجتمعات المحدودة ذات الحجم N عندما تكون:

$$N \geq 0.05N$$

نظرية (2)

مع تزايد حجم العينة (أي عندما $n \to \infty$) فإن توزيع المعاينة للمتوسط يقترب من التوزيع الطبيعي، بصرف النظر عن شكل المجتمع الأصلي. ويعتبر التقريب جيدا عندما تكون $n \geq 30$. (وهذه هي فكرة النظرية النهائية المركزية التي تم شرحها).

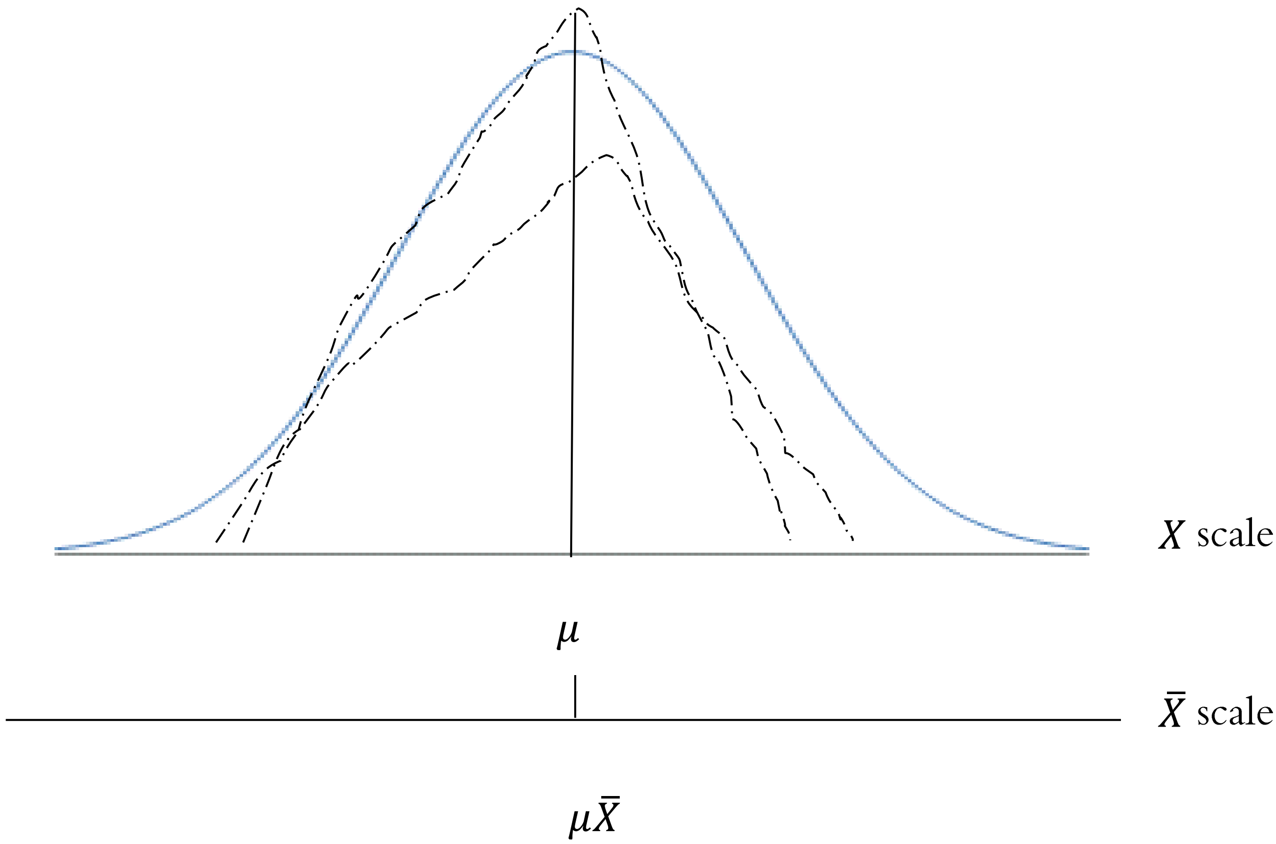
ويمكن إيجاد أ احتمال أن يكون الوسط $\bar{x}$ لعينة عشوائية داخل فترة معينة، لحساب قيم z للفترة، حيث أن:

$$Z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

تم الكشف عن القيم في جداول التوزيع الطبيعي القياسي (الملاحق).

مثال(3-1)

في شكل (5-1) يكون متوسط المعاينة للمتوسط $\mu_{\bar{x}}$، يساوي متوسط المجتمع μ، بصرف النظر عن حجم العينة n، ولكنن كلما كانت n، كلما صغر الخطأ المعياري للمتوسط $\sigma_{\bar{x}}$، فإذا كان توزيع المجتمع هو التوزيع الطبيعي، فإن توزيع المجتمع هو التوزيع الطبيعي، فإن توزيع المعاينة للوسط هو أيضا التوزيع الطبيعي، حتى للعينات الصغيرة، وطبقا لنظرية النهاية المركزية، حتى إذا كان المجتمع غير طبيعي التوزيع، فإن توزيع المعاينة للمتوسط يكون طبيعيا تقريبا عندما تكون $n \geq 30$.



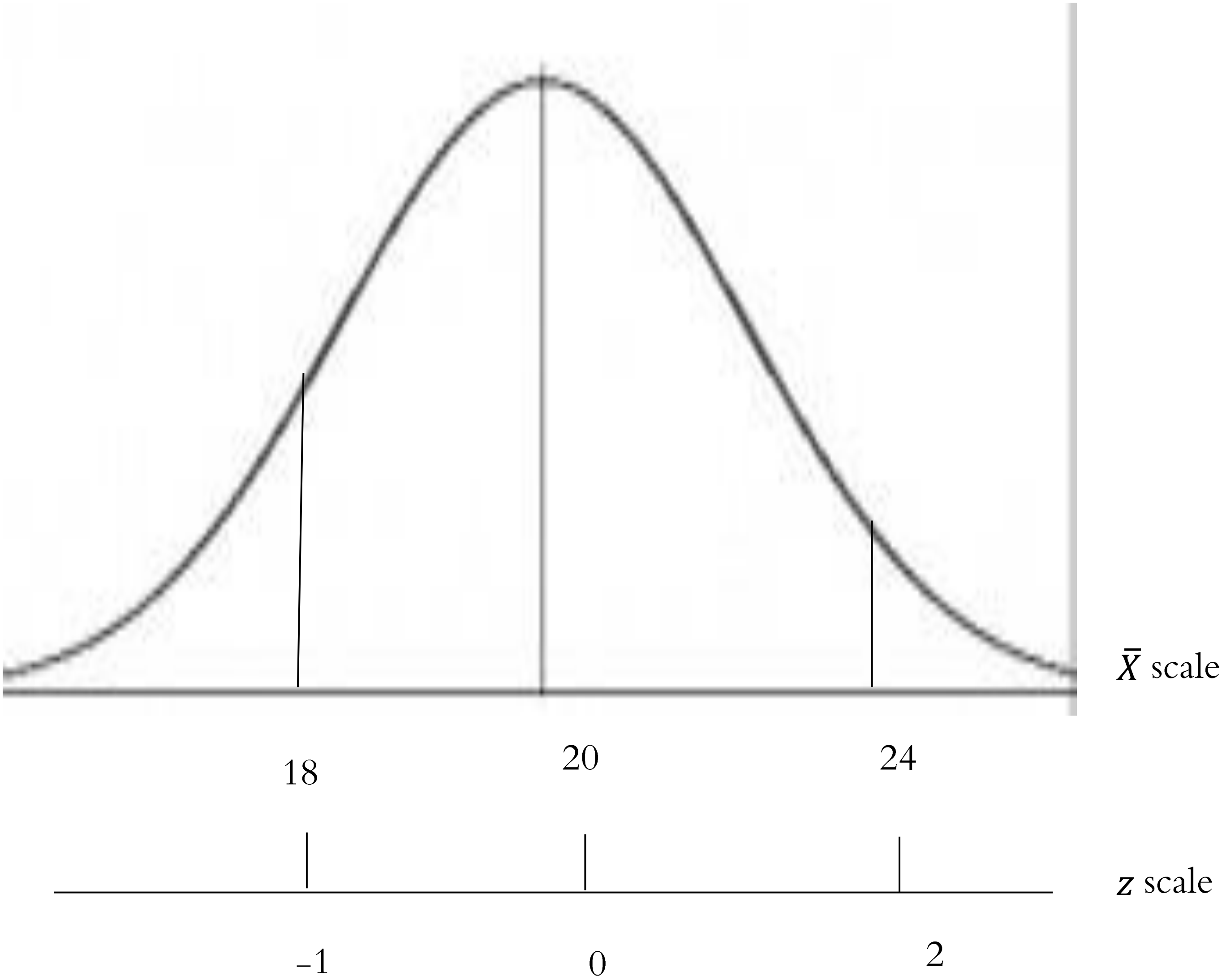
الشكل (1)

مثال (3-2)

افترض . المجتمع يتكون من 900 عنصر بوسط حسابي 20 وانحراف معياري 12 وحدة. الوسط الحسابي والانحراف المعياري لتوزيع المعاينة لوسط عينة حجمها36 هما.

$\mu_{\bar{x}}= \mu$=20 units

$\sigma_{\bar{x}}=\frac{\sigma}{\sqrt{n}}=\frac{12}{\sqrt{36}}=\frac{16}{6}=2$



الشكل (2)

لو كانت n تساوي 64 بدلا من 36 (بحيث) $n>0.05\ N$، فإن:

$$\sigma_{\bar{x}}=\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-1}{N-1}}=\frac{12}{\sqrt{64}}\sqrt{\frac{900-64}{900-1}}=\frac{12}{8}\sqrt{\frac{9836}{9899}}=(1.5\text{x}0.96)=1.44$$

وهذه تكون في حالة استخدام معامل التصحيح للمجتمعات المحدودة.

مثال (3-3)

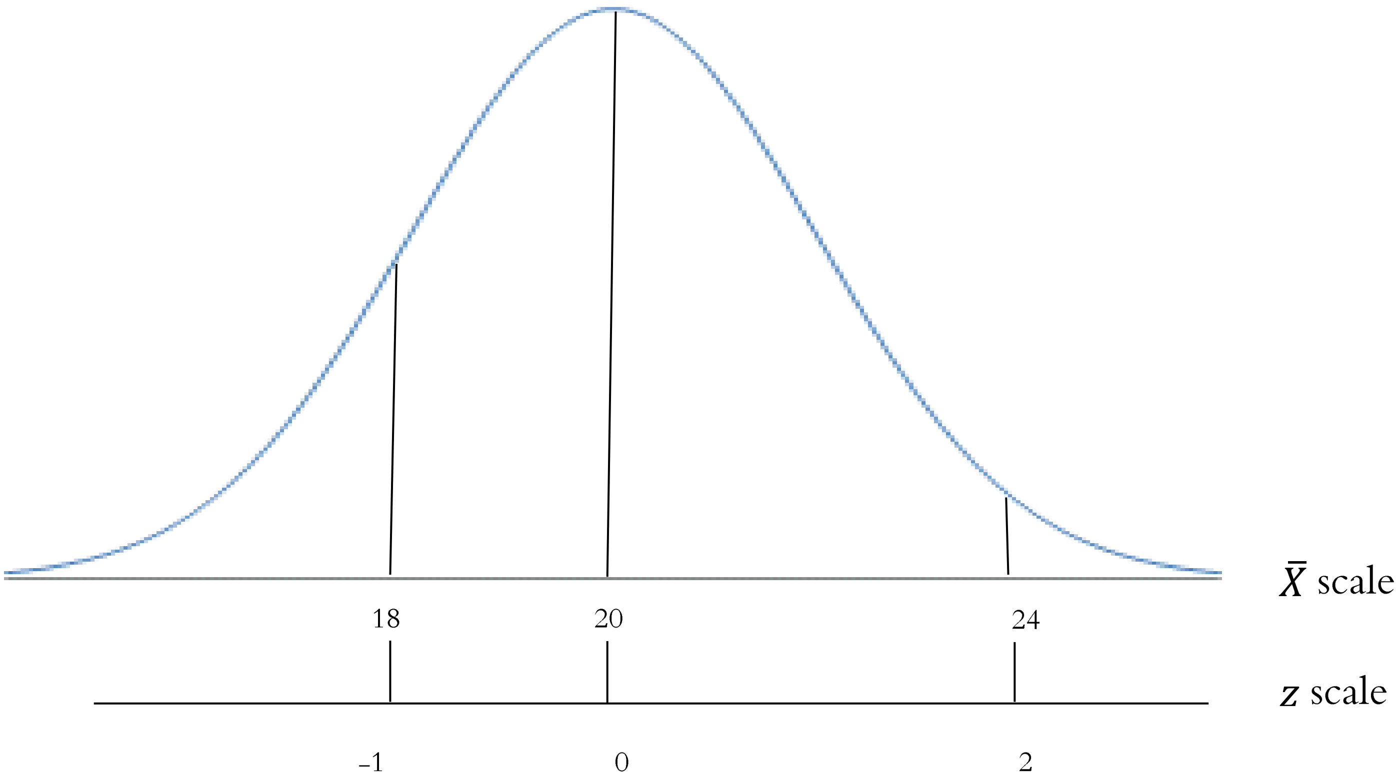
يمكن حساب احتمال أن يقع وسط عينة عشوائية $\bar{x}$ حجمها 36 مأخوذة من مجتمع طبيعي وسطه 0.20 بين 18 و 24كما يأتي:

$Z_1=\frac{\bar{x}_1-\mu_{\bar{x}}}{\sigma_{\bar{x}}}=\frac{18-20}{2}=-1$ or $Z_2=\frac{\bar{x}_2-\mu_{\bar{x}}}{\sigma_{\bar{x}}}=\frac{24-20}{2}=2$

وبالبحث مقابل $Z_1 Z_2$ في جدول التوزيع الطبيعي القياسي(الملاحق)، نحصل على:

$P(18<\bar{x}<24)=0.3413+0.4772=0.8185$,or 81.85%.

كما في الشكل (3) أدناه:



الشكل (3)

3-1-1- فترة تقدير الوسط الحسابي للمجتمع في حالة العينات الكبيرة (فترة الثقة للوسط)

إذا كان المجتمع المسحوب منه العينة ذا توزيع طبيعي وتباينه معروفاً أو كانت العينة كبيرة (أي حجمها ثلاثون مفردة أو أكثر) فإن توزيع المعاينة للوسط الحسابي يكون توزيعا طبيعيا، ولطالما نحن نتكلم عن تقدير متوسط المجتمع فإن أول ما نفكر فيه هو الوسط الحسابي للعينة. وفترة التقدير (أو الثقة) للوسط الحسابي للمجتمع تأخذ الشكل التالي:

تقدير متوسط المجتمع= الوسط الحسابي للعينة$\pm$ الخطأ المعياري للوسط فإن:

$$\hat{\mu} = \bar{x} \pm 1\sigma_{\bar{x}}$$

حيث $\hat{\mu}$ تقدير الوسط الحسابي للمجتمع $\bar{x}$ هو الوسط الحسابي للعينة:

$\sigma_{\bar{x}}$ هو الخطأ المعياري للوسط، ولكن احتمال أن يكون هذا الكلام صحيحا هو 68.2% فقط، أي أن درجة الثقة هنا لا تتعدى 68.2%، فإذا أضفنا وطرحنا ضعف الخطأ المعياري يرتفع الاحتمال إلى 95.44% أي ترتفع درجة الثقة إلى 95.44% وفي هذه الحالة تأخذ الحالة تأخذ فترة الثقة الصيغة الآتية:

$$\hat{\mu} = \bar{x} \pm 2\sigma_{\bar{x}}$$

فإذا أضفنا وطرحنا ثلاثة أمثال الخطأ المعياري يصبح الاحتمال 99.72 % أي ترتفع درجة الثقة إلى 99.72%

وتأخذ فترة الثقة الصيغة الآتية:

$$\hat{\mu} = \bar{x} \pm 3\sigma_{\bar{x}}$$

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 95.44%، ولذلك فإن هذا المعامل هو الذي يسمى معامل الثقة. فبناء على درجة الثقة المطلوبة يتحدد معامل الثقة.

2.إن درجات وعاملات الثقة التي ذكرناها تخص التوزيع الطبيعي، وإن المعاملات 1 أو 2 أو 3 ما هي إلا الدرجة المعيارية (z) والتي نحصل عليها من جدول التوزيع الطبيعي المعياري (القياسي) وذلك بقسمة درجة الثقة (أو الاحتمال) على 2 (حيث إن 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.

3.يمكن الحصول على فترات تقدير بأي درجة ثقة أخرى (غير الثلاث التي ذكرناها) وذلك بقسمة درجة الثقة المطلوبة-كما ذكرناها- على 2 ثم الكشف في المساحات حتى نحصل على Z المناسبة.

4.والخلاصة أن فترة التقدير (أو الثقة) للوسط الحسابي للمجتمع تأخذ الصيغة الآتية:

$$\hat{\mu} = \bar{x} \pm Z\sigma_{\bar{x}}$$

حيث تحدد Z درجة الثقة المطلوبة. وحيث إن الخطأ المعياري للوسط $\sigma_{\bar{x}}$ يأخذ الشكل الآتي:

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

فإن فترة تقدير الوسط تأخذ الصيغة النهائية الآتية:

$$\hat{\mu}=\bar{x} \pm Z\frac{\sigma}{\sqrt{n}}$$

ويمكن تلخيص خطوات تقدير الوسط الحسابي للمجتمع فيما يلي:

أ. أحسب الخطأ المعياري للوسط $\sigma_{\bar{x}}$ والذي يساوي $\frac{\sigma}{\sqrt{n}}$.

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

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

ث. يمك كتابة أهم درجات ومعاملات الثقة (للتوزيع الطبيعي) في الجدول التالي (مع ملاحظة 95%، 99% هي أكثر استخداما).

درجة الثقة	معامل الثقة Z
%68.26	1
%90	1.65
%95	1.96
%95.44	2
%99	2.58
%99.72	3

جـ. إذا كان الانحراف المعياري σ غير معروف، وهو غالبا في الواقع، فيمكن استخدام الانحراف المعياري للعينة S بدلا منه طالما كان حجم العينة كبيرا بدرجة كافية وتصبح فترة تقدير الوسط الحسابي للمجتمع كما يلي:

$$\hat{\mu}=\bar{x} \pm Z\frac{s}{\sqrt{n}}$$

ولإيضاح هذه النقطة بشيء من التفاصيل نأخذ المثال التالي:

مثال (3–4)

لو لأردنا معرفة متوسط الدخل اليومي لمجموعة من العاملين في بلد ما، فإن ذلك يبدو أمرا صعبا من الناحية العملية نظرا لكبر حجم مجتمع العاملين إضافة إلى طول الوقت والتكاليف. لذا فإن الأسلوب العلمي المتبع في حالة كهذه هو اختيار عينة عشوائية نستطيع من خلالها معرفة نتائجها تقدير متوسط دخول العاملين إلى البلد.

فلو سحبت عينة عشوائية م مجموع مجتمع العاملين في بلد ما حجمها 100 عامل فإذا كان الوسط الحسابي والإنحراف المعياري للدخل اليومي للعاملين بالعينة هما على الترتيب 90 دولارا و25 دولارا، فأوجد فترة تقدير للوسط الحسابي للدخل اليومي لمجموع العاملين في هذا البلد بدرجة ثقة 95%؟

الحل:

بما أن فترة تقدير الوسط الحسابي للمجتمع هي:

$$\hat{\mu}=\bar{x} \pm Z\frac{s}{\sqrt{n}}$$

والمعلومات المعطاة هي:

حجم العينة $n = 100$

الوسط الحسابي للعينة $\bar{x}=90$

والإنحراف المعياري للعينة $S=25$

وبحث أن درجة الثقة هي 95% فإن معامل الثقة هو: $Z=1.96$ حس ما هو موضح في جدول التوزيع الطبيعي القياسي(الملاحق). وبالتالي فإن فترة تقدير الوسط الحسابي للدخل اليومي لمجتمع العاملين بدرجة ثقة 95% هي:

$$\hat{\mu}=90 \pm 1.96\frac{25}{\sqrt{100}}$$

$$=90 \pm 1.96(2.5)$$

$$=90 \pm 4.9$$

$$\hat{\mu}=\begin{Bmatrix} 85.1 \\ 94.9 \end{Bmatrix}$$

أي أن الوسط الحسابي للدخل اليومي لمجتمع العاملين يتراوح بين دولار الحد الأدنى، كحد أعلى، وذلك بدرجة ثقة 95%.

ملاحظة: يمكن حساب مستوى الثقة α في الحالة أعلاه من خلال العلاقة التالية:

$1-\alpha=0.95$ وعليه مستوى الثقة $\alpha=1-0.95=0.05$

ويمكن حساب قيمة z من خلال حساب القيمة:

$$Z_{1-\frac{\alpha}{2}}= Z_{1-\frac{0.05}{2}}= Z_{1-0.025}= Z_{0.0975}=1.96$$

وهذه الأخيرة تحسب من جدول التوزيع الطبيعي القياسي بسهولة.

إن التقدير بفترة يشير إلى مدى من القيم مقرونا باحتمال أن يضم المدى معلمة المجتمع غير المعروفة، ويسمى هذا الاحتمال مستوى الثقة. ومعلومية الانحراف المعياري للمجتمع أو تقديره، وإذا علم أن توزيع المجتمع طبيعي وعلم أن العينة العشوائية تساوي أو تزيد عن 30 يمكننا إيجاد فترة الثقة 95% لوسط المجتمع μ غير المعروف كالآتي:

$$P(\bar{x}-1.96\sigma_{\bar{x}}<\mu < \bar{x}+1.96\sigma_{\bar{x}})=0.95$$

مثال (3–5)

أخذت عينة عشوائية حجمها 144 بوسط مقداره 100 وانحراف معياري مقداره 60 من مجتمع حجمه 1000، أوجد فترة الثقة 95% لوسط المجتمع غير المعلوم؟

الحل:

$$\hat{\mu}=\bar{x} \pm 1.96\sigma_{\bar{x}} \qquad \text{since } n> 30$$

$$=\bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \qquad \text{since } n> 0.05N$$

$$100 \pm 1.96\frac{60}{\sqrt{144}}\sqrt{\frac{1000-144}{1000-1}} \qquad \text{using S an estimate of } \sigma$$

$$=100\pm1.96(5)(0.93)$$

$$=100\pm9.11$$

أي أن μ تقع بين 90.89، 109.11 بدرجة ثقة 95%. وكثيرا ما تستخدم أيضا درجات الثقة 90، 99%، وهي مناظرة لقيمة z=1.64، z=2.58 على الترتيب من جدول التوزيع الطبيعي القياسي (المعياري).

3-2 فترة تقدير النسبة للمجتمع (أو فترة الثقة للنسبة)

إن تقدير النسبة في المجتمع تعتبر من الحالات المهمة لقياس الظواهر السياسية، والعسكرية والاجتماعية والاقتصادية، وبالذات الوصفية منها كقياس اتجاهات الرأي العام، وقياس نسبة الولادات أو الوفيات، وغيرها ونظرا 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 هي p وأن العينة العشوائية كبيرة بدرجة كافية وأن نسبة مؤيدي هذه السياسة في العينة $\hat{p}$ فإن خطوات تقدير النسبة في المجتمع تكون كما يلي:

1.أحسب النسبة في العينة $\hat{p}$.

2. أحسب الخطأ المعياري للنسبة والتي تساوي في هذه الحالة

$$\boldsymbol{\sigma}_{\hat{p}}=\sqrt{\frac{\hat{p}(1-\hat{p})}{\boldsymbol{n}}}$$

3.اضرب الخطأ المعياري للنسبة في معامل الثقة المناسب Z (أحسب درجة الثقة المطلوبة) والتي نحصل عليها من جدول التوزيع الطبيعي القياسي (أو من الجدول الذي يحوي أهم درجات وعاملات الثقة والذي ذكرناها سابقا). أي تحسب:

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

4. للحصول على الحد الأدنى لتقدير النسبة نطرح حاصل الضرب(السابق) من نسبة $\hat{p}$، وللحصول على الحد الأعلى نجمع حاصل الضرب مع النسبة في العينة فنحصل على فترة تقدير النسبة. وبالتالي فإن فترة تقدير النسبة تكون في صيغتها النهائية كالآتي:

$$p=\hat{p} \pm z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

ولتوضيح هذه الخطوات بشيء من التفاصيل نورد المثال الآتي:

مثال(3-6)

عينة عشوائية حجمها 144 ناخبا سحبت من إحدى المدن فوجد أن عدد المؤيدين في العينة لمرشح معين هو 60 ناخبا، أنشئ فترة تقدير نسبة المؤيدين لهذا المرشح في المدينة كلها بدرجة ثقة 95%.

الحل:

نحسب أولا نسبة المؤيدين للمرشح في العينة $\hat{p}$ التي نحصل عليها بقسمة عدد المؤيدين له على العدد الكلي للعينة (حجم العينة) أي أن:

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

وبحيث إن درجة الثقة المطلوبة هي 95% فإن معامل الثقة المناسب هو Z=196 وفترة تقدير نسبة المؤيدين لهذا المرشح في المدينة تأخذ الشكل الآتي:

$$p=\hat{p}\pm z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

وبالتعويض عن حجم العينة$\mathbf{n}=144$ والنسبة في العينة $\hat{\mathbf{p}}=0.42$، $1-\hat{\mathbf{p}}=1-0.42=0.58$ ومعامل الثقة $\mathbf{Z}=1.96$

$$p=0.42\pm 1.96\sqrt{\frac{0.42x0.58}{144}}$$

$$=0.42\pm 0.08$$

$$p=\begin{Bmatrix}0.34\\0.50\end{Bmatrix}$$

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

3-3 تحديد حجم العينة للتقدير الوسط الحسابي للمجتمع

يعتبر تحديد حجم العينة المناسب من المشاكل المهمة والشائعة التي تواجه الباحثين في مختلف المجالات، وبالذات عند دراسة الظواهر السياسية، 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:

$\mu = \bar{x} \pm Z\frac{\sigma}{\sqrt{n}}$

ونلاحظ بأننا عندما نقدر متوسط المجتمع فإننا نرتكب خطأ مقدارا e وأن:

$e = |\bar{x} - \mu|$ وهو يمثل مقدار الخطأ المطلق، وبما أن $e = |\bar{x} - \mu| < Z\frac{\sigma}{\sqrt{n}}$

نستنتج أن $e = Z\frac{\sigma}{\sqrt{n}}$، لكي يكون الخطأ أقل ما يمكن وبتربيع الطرفين نحصل على: $e^2 = (Z\frac{\sigma}{\sqrt{n}})^2$ ، ومنها نجد أن حجم العينة يأخذ الصيغة التالية:

$$n = \frac{z^2 . \sigma^2}{(e)^2}$$

-Z: هو معامل الثقة (أو الدرجة المعيارية) المقابل لدرجة الثقة المطلوبة، ونحصل عليها من جدول التوزيع الطبيعي القياسي.

-(σ^2) هو تباين المجتمع (أو هو مربع الانحراف المعياري).

-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 اختصارا في تقدير الوسط.

ولتوضيح كيفية تحديد حجم العينة المناسب عند تقدير الوسط الحسابي للمجتمع، نأخذ المثال التالي:

مثال(3-7)

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

الحل:

في هذا المثال نجد أن:

درجة الثقة 99% أي أن: Z=2.85

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

والانحراف المعياري: $\sigma = 15$.

وبالتعويض بهذه القيم في المعادلة التي تحدد حجم العينة هي:

$$n=\frac{z^2.\sigma^2}{(e)^2}=59.9=60.$$

فإن حجم العينة مقربا لأقرب عدد صحيح هو:

$$n=\frac{(2.58)^2.(15)^2}{5^2}\approx 60$$ فرداً

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

3-4 تحديد حجم العينة لتقدير النسبة في المجتمع:

إن الخطأ المطلق الأعظم المرتكب في تقدير p بمستوى ثقة % (1-α) هو:

$$e=\left|z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right| = \varepsilon$$

وعليه وبالاستفادة من تفريغ هذه العلاقة يمكن تحديد حجم العينة اللازمة للحصول على درجة ثقة معينة عند تقدير النسبة في المجتمع وبافتراض أن أقصى خطأ في التقدير مسموح به هو e تبعا للمعادلة التالية:

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

حيث:

-z هو معامل الثقة (أو الدرجة المعيارية) المقابل لدرجة الثقة ونحصل عليه من جدول التوزيع الطبيعي القياسي.

-p هي النسبة في المجتمع (أو تقدير لها).

-1-p: هي النسبة المكملة بمعنى إذا كانت نسبة المؤيدين 60% فإن نسبة غير المؤيدين 40%.

-e أقصى خطأ في التقدير مسموح به.أو الخطأ في تقدير النسبة.

أي أن حجم العينة المناسب في هذه الحالة يساوي حاصل ضرب مربع z في النسبة، ثم في النسبة المكملة مقسوما على مربع الخطأ كما في المثال الآتي:

مثال (3-8)

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 50%.

الحل:

بما أن درجة الثقة 95% فإن $z=1.96$ بافتراض أن نسبة المؤيدين للمرشح هي $p.=0.50$.

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

وحيث إن أقصى خطأ مسموح به هو: $e=0.02$.

فإن حجم العينة للازم هو:

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

وبالتعويض نحصل على:

$$n=\frac{(1.96)^2(0.5)(0.5)}{(0.02)^2}$$

$$n=\frac{0.9604}{0.0004}$$

$$n=2401$$

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

3-5 فترات الثقة للمتوسط باستخدام توزيع t

تناولنا فيما سبق التقدير الإحصائي للوسط الحسابي للمجتمع في الحالات التي يكون فيها الانحراف المعياري للمجتمع معلوما، و (أو) أن العينة كبيرة بدرجة كافية. ولكن إذا كانت العينة صغيرة بمعنى أن حجمها أقل من (30) مفردة، والانحراف المعياري للمجتمع الطبيعي غير معلوم، فإن التوزيع الإحصائي المتبع في مثل هذه الحالات هو ما يطلق عليه توزيع

t فعند تقدير متوسط عمر الناخب في مدينة ما عن طريق سحب عينة صغيرة (حجمها أقل من 30 ناخبا) فإن التوزيع الطبيعي يكون في مثل هذه الحالات غير مناسب لصغر حجم العينة أولا، ثم عدم معرفة الانحراف المعياري لعمر الناخب ثانيا.

لذا فإن الأسلوب الإحصائي المتبع في حالات كهذه هو استخدام "توزيعt" 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 التوزيعان متماثلان وكلما زادت قيمة n كلما اقترب توزيع t من توزيع z ويعتمد توزيع tعلى ما يعرف بدرجات الحرية.

3-5-1 درجات الحرية:

تعرف درجات الحرية بأنها عدد المشاهدات المستقلة في العينة والتي تساوي حجم العينة مطروحا منه عدد القيود أو معالم المجتمع التي يتم تقديرها من بيانات العينة. وكمثال مبسط لشرح فكرة درجات الحرية نفترض أن لدينا 3 قيم واشترطنا أن مجموع القيم يساوي 10 فإن لدى الباحث في هذه الحالة حرية في اختبار الرقم الأول (وليكن12) والثاني (وليكن3) لذلك فإن قيمة الثالثة لا بد وأن تكون (5) وبالتالي نستطيع القول بأن درجة الحرية المتاحة لدى الباحث هي (2) أي 2=1-3، أي أن درجات الحرية في هذه الحالة هي n-1، حيث n تساوي حجم العينة (والتي يساوي في المثال السابق3) والرقم(1) والذي طرحناه يعني الشرط ويحتم أن مجموع القيم=10 وبصفة عامة إذا كان عدد القيود k فإن درجات الحرية n-k.

3-5-2 شروط استخدام توزيع t

يمكن تحديد الشروط الثلاثة أدناه لاستخدام توزيع t:

1-أن يكون المجتمع المسحوبة منه العينة له توزيع طبيعي.

2-والانحراف المعياري للمجتمعσ غير معروف أو مجهول.

3-والعينة صغيرة (حجمها أقل من 30 مفردة).

3-6 تقدير الوسط الحسابي للمجتمع في حالة العينات الصغيرة

تأخذ فترة تقدير الوسط الحسابي للمجتمع في حالة العينات الصغيرة للشكل التالي:

$$\widehat{M}= \bar{x} \pm (t_{(\frac{a}{2},n-1)}) \frac{s}{\sqrt{n}}$$

ولعل أهم الملاحظات على المعادلة الإحصائية السابقة احتوائها على مفهومين هما:

1. مستوى المعنوية أو الدلالة والذي رمزنا له بالرمز α والذي يعني أنه المكمل لدرجة الثقة أي نسبة الخطأ. وبالتالي فإذا كانت درجة الثقة 99% أي احتمال أن يكون التقدير صحيحا 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 على2.

2. درجات الحرية، وهو ما سبق شرحه في هذه الحالية، ويساوي في هذه الحالة n-1. حيث n هو حجم العينة وطرحنا منه 1 لأن تقدير الانحراف المعياري للمجتمع المجهول باستخدام الانحراف المعياري للعينة S.

مثال (3-9)

إذا كانت دخول مجموعة من الأفراد في دولة ما تتبع التوزيع الطبيعي، وسحبت منهم عينة عشوائية حجمها 10 بوسط حسابي $\bar{x}$=72 دولارا، وانحراف معياري S=6.4 دولارا، أنشئ فترة تقدير للوسط الحسابي للدخل اليومي لجميع الأفراد بدرجة ثقة 95%.

الحل:

إن العينة صغيرة (حجمها 10 أفراد فقط) وأن المجتمع طبيعي وانحرافه المعياري غير معروف لذلك نستخدم فترة تقدير الوسط للعينات الصغيرة التي تعتمد على توزيع t.

وحيث إن n=10 فإن درجات الحرية لها هي:

n-1=10-1=9

وحيث إن درجة المعنوية المطلوبة هي0.95= α -1 فإن مستوى المعنوية $\alpha = 0.05$ وبالتالي فإن مستوى المعنوية هو: $0.025=\frac{0.05}{2}=\frac{\alpha}{2}$

أي تم الكشف في جدول توزيع t (الملاحق) عند درجات حرية تساوي 9 تحت احتمال (نصف مستوى المعنوية) 0.025 أي أن:

$t_{0.025.9}$=2.262

وبالتعويض عن فترة تقدير الوسط نحصل على المعادلة التالية:

$\hat{\mu}=\bar{x} \pm (t_{(0.025.9)} \frac{s}{\sqrt{n}}$

$\hat{\mu}=72 \pm 2.262 \frac{6.4}{\sqrt{10}}$

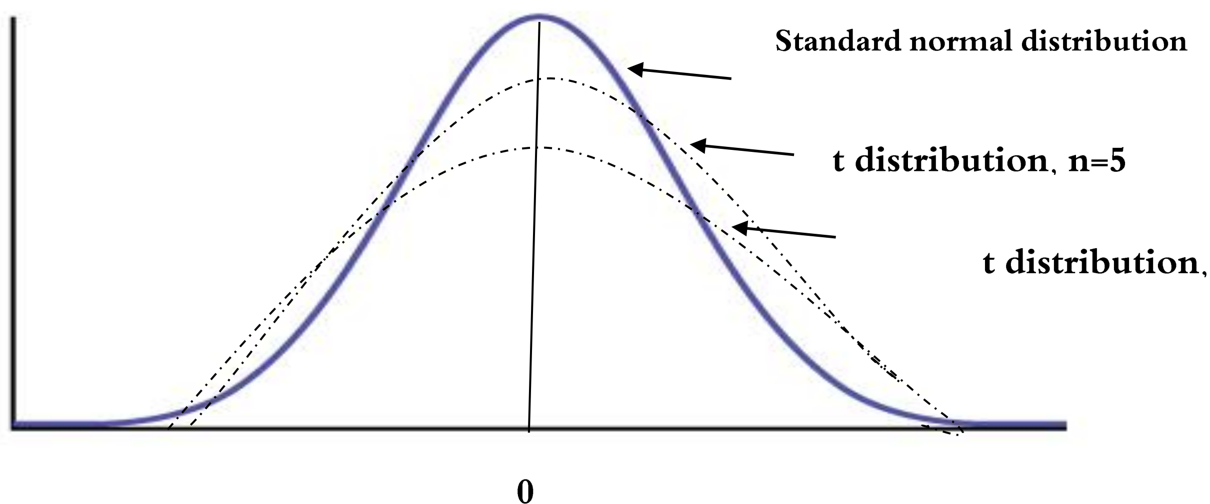
$\hat{\mu}=72 \pm \frac{14.78}{3.16}$

$\hat{\mu}=72 \pm 4.6$

$\hat{\mu}=\begin{Bmatrix} 67.4 \\ 76.6 \end{Bmatrix}$

أي أن الوسط الحسابي للدخول اليومية يتراوح بين 67.4 دولارا كحد أدنى، 76.6 كحد أعلى وذلك بدرجة ثقة 95%.

ونستنتج من ذلك بأنه عندما يكون التوزيع طبيعيا، ولكن σ غير معلومة و n>30، فإنا لا نستطيع استخدام التوزيع الطبيعي لتحديد فترات الثقة لمتوسط المجتمع غير المعلوم، ولكن يمكننا استخدام توزيع t. هذا التوزيع متماثل حول متوسط الصفر ولكنه منبسط عن التوزيع الطبيعي القياسي، ولهذا فإن جزاء أكبر من مساحته تقع عند الأطراف. وبينما يوجد توزيع طبيعي قياسي واحد، فإن هناك توزيعا t مختلفا لكل حجم للعينة n. ولكن مع تزايد n فإن توزيع t يقترب من التوزيع الطبيعي القياسي، شكل (3-5) إلى أن تكون n≥30، وعندئذ يتساويان تقريبا.



يعطى جدول توزيع t قيم t، التي على يمينها تمثل المساحة تحت المنحنى 10%، 5%، 2.5%، 1%، 0.5%. من المساحة الكلية تحت المحنى لدرجات الحرية المختلفة. وفترة الثقة 95% لوسط المجتمع غير المعلوم عند استخدام توزيع t هي:

$$P(\bar{x} - t_{(\frac{\alpha}{2},n-1)}\frac{s}{\sqrt{n}} < \mu < (\bar{x} + t_{(\frac{\alpha}{2},n-1)}\frac{s}{\sqrt{n}}) = 0.95$$

حيث تشير t إلى قيمة t التي تقع عندها 2.5% م المساحة الكلية للمنحنى عند كل طرف (عند درجات الحرية المستخدمة) وتستخدم $s/\sqrt{n}$ بدلا من $\sigma_{\bar{x}} = \sigma/\sqrt{n}$.

مثال (3–10)

سحبت عينة عشوائية $n=10$ بطارية فلاش بمتوسط $\bar{x}=5$ ساعة، وبالانحراف معياري للعينة $s=1$ ساعة، من خط إنتاج من المعروف أنه ينتج بطاريات عمرها موزع طبقا للتوزيع الطبيعي. أوجد فترة الثقة 95% للمتوسط غير المعلوم لعمر البطاريات في المجتمع كله؟

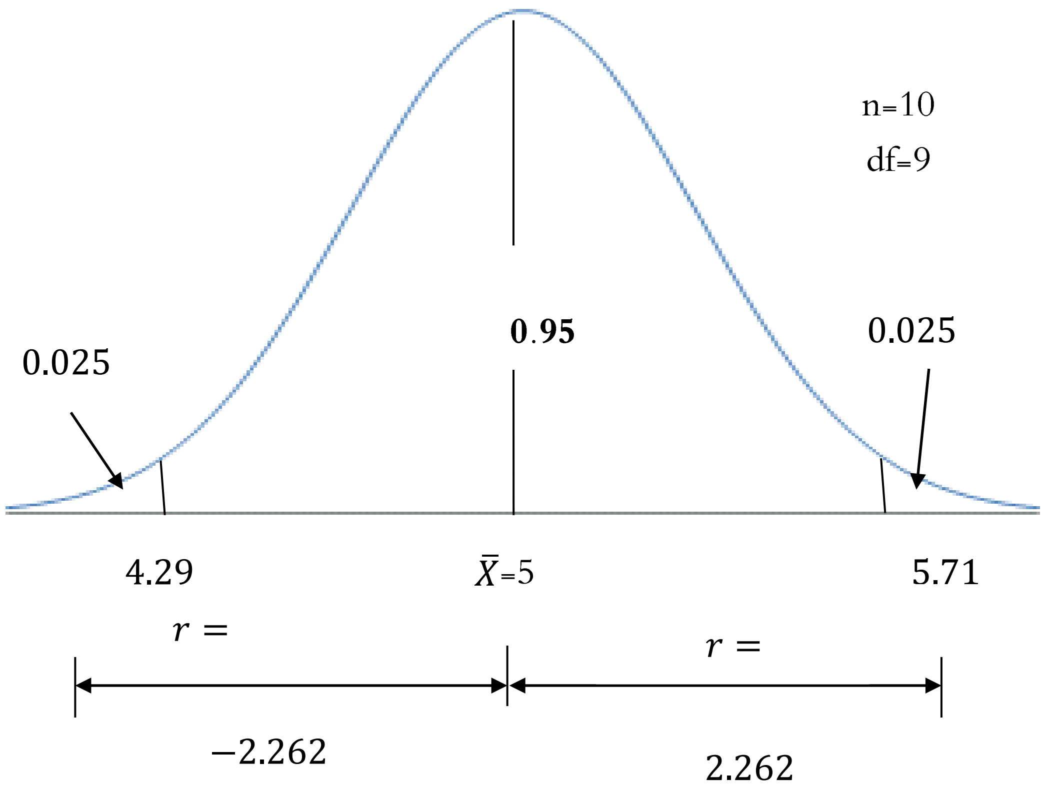
الحل:

نوجد أولا قيمة $\pm t=0.025$ والتي يكون معها 2.5% من المساحة عند الأطراف لدرجة حرية $n-1$ ونحصل على القيمة من جدول توزيع t بالتحرك تحت عمود 0.025 حتى درجات حرية 9 والقيمة التي نحصل عليها هي 2.262 فيكون:

$$\mu = \bar{x} \pm (2.262\frac{s}{\sqrt{n}} = 5 \pm 2.62\frac{1}{\sqrt{10}} \cong 5 \pm 2.262(0.316) \cong 5 \pm 0.71$$

وتقع μ بين 4.29، 5.71 بدرجة ثقة 95% أنظر الشكل (5).

وعندما تكون $n<30$ والتوزيع غير طبيعي، فيجب استخدام نظرية تشبتشيف.



الشكل(5)

4–1 مجال الثقة للفرق بين متوسطي متغيرين عشوائيين

بنفس الأسلوب السابق الذي تناولنا به حساب فترات الثقة للمتوسطات فإننا يمكن أن نحسب مجال الثقة للفرق بين متوسطين في الحالات التالية:

4–1–1 مجال الثقة للفرق بين متوسطي متغيرين عشوائيين

عندما يكون حجما المجتمعين كبيرين أي أن : $\mathbf{n_1}$ ،$\mathbf{30 \geq n_2}$، وإذا كان:

-$\mathbf{X_1}$ متغير عشوائي له توزيع طبيعي بوسط $\boldsymbol{\mu_1}$ وتباين $\boldsymbol{\sigma_1}^2$ معلوم.

-$\mathbf{X_2}$ متغير عشوائي له توزيع طبيعي بوسط $\boldsymbol{\mu_2}$ وتباين $\boldsymbol{\sigma_1}^2$ معلوم.

وأن $\bar{\mathbf{X}}_1$،$\bar{\mathbf{X}}_2$هما الوسطان الحسابيان لعينتين عشوائيتين مستقلتين مسحوبتين من التوزيع فإن للمتغير$\bar{\mathbf{X}}_2 - \bar{\mathbf{X}}_1$توزيعا طبيعيا بمتوسط يمثل متوسط الفرق للعينتين وهو$\boldsymbol{\mu_{\bar{X}1-\bar{X}2}} = \boldsymbol{\mu_1} - \boldsymbol{\mu_2}$، وكذلك لهما التباين:

$\sigma^2_{\bar{X}1-\bar{X}2}=\frac{\sigma_1 2}{n_1}+\frac{\sigma_1 2}{n_2}$ ومنه نحصل على:

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

وهي تمثل القيمة المحورية التي تتوزع توزيعا طبيعيا معياريا ($Z\approx N(0,1)$)

وعليه بنفس الطريقة السابقة نستنتج أن فترة الثقة للفرق بين المتوسطين وصيغتها النهائية هي:

$$\left[(\bar{X}_1-\bar{X}_2)\pm Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1{}^2}{n_1}+\frac{\sigma_2{}^2}{n_2}}\right]$$

أما في حالة كون التباينين مجهولين n_1 ، $n_2\geq 30$ فإننا نستبدل $\boldsymbol{\sigma_1}^2$ و $\boldsymbol{\sigma_1}^2$ بشكل $\mathbf{s_1}^2$ و $\mathbf{s_2}^2$ فنحصل على فترة مشابهة للصيغة أعلاه وكما يأتي:

$$\left[(\bar{X}_1-\bar{X}_2)\pm Z_{1-\frac{\alpha}{2}}\sqrt{\frac{s_1{}^2}{n_1}+\frac{s_2{}^2}{n_2}}\right]$$

مثال (3-11)

مجتمعان من المصابيح أخذت منهما عينتان $\mathbf{n_1}$=150 و $\mathbf{n_2}$=200 وانحرافهما المعياري s_1=120 و s_2=80،

أوجد مجال الثقة 97% للفرق بين متوسطي أعمار مجتمعي المصابيح؟

الحل:

متروك للطالب (نستخدم علاقة التباينين معلومين والتوزيع طبيعي معياري)

4-1-2 مجال الثقة للفرق بين متوسطين في حالة التباين مجهول لكلي المجتمعين

عندما يكون حجما المجتمعين كبيرين أي أن : $\mathbf{n_1}$، $\mathbf{30\geq n_2}$، وإذا كان:

- $\mathbf{X_1}$ متغير عشوائي له توزيع طبيعي بوسط $\boldsymbol{\mu_1}$ وتباين $\boldsymbol{\sigma_1}^2$ معلوم.

- $\mathbf{X_2}$ متغير عشوائي له توزيع طبيعي بوسط $\boldsymbol{\mu_2}$ وتباين $\boldsymbol{\sigma_1}^2$ معلوم.

وبفرض أن التباينين متساويين أي أن $\boldsymbol{\sigma_1}^2=\boldsymbol{\sigma_2}^2=\boldsymbol{\sigma}^2$.

وأن $\bar{\mathbf{X}}_1$،$\bar{\mathbf{X}}_2$ هما الوسطان الحسابيان لعينتين عشوائيتين مستقلتين مسحوبتين من التوزيع فإن للمتغير $\bar{\mathbf{X}}_2 - \bar{\mathbf{X}}_1$ توزيعا طبيعيا بمتوسط يمثل متوسط الفرق للعينتين وهو $\boldsymbol{\mu}_{\bar{X}1-\bar{X}2}=\boldsymbol{\mu_1}-\boldsymbol{\mu_2}$، وكذلك لهما:

- **التباين المجمع (pooled Variance)**

$S_p^2=\frac{(n_1-1)S_1^2+((n_2-1)S_2^2}{n_1+n_2-2}$ وهو تقدير غير متحيز للتباين $\boldsymbol{\sigma^2}$ (**لماذا**)؟

ومن المعلوم أن:

$$\frac{(n_1-1)S_1^2}{\sigma^2} \approx \chi^2\,(n_1-1)$$

و كذلك :

$$\frac{(n_2-1)S_2^2}{\sigma^2} \approx \chi^2\,(n_2-1)$$

ومن خاصية التجميع للتوزيع نحصل على:

$$y=\left[\frac{(n_1-1)S_1^2}{\sigma^2}+\frac{(n_2-1)S_2^2}{\sigma^2}\right] \approx \chi^2\,(n_1+n_2-2)$$

وبما أن توزيع **t** هو حاصل قسمة توزيع طبيعي معياري على الجذر التربيعي لمتغير مستقل مقسوما على درجات حريته أي أن:

$$t=\left[\frac{Z}{\sqrt{y/(n_1+n_2-2)}}=\frac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\right]$$

وعليه فإن **t** له توزيع **t** بدرجات حرية (n_1+n_2-2) ومنه نحصل على فترة الثقة التالية:

$$\left[(\overline{X}_1-\overline{X}_2)\pm t_{1-\frac{\alpha}{2}}S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right]$$

ملاحظة: إذا كان حجم العينتين صغيرا والتباين مجهولين وغير متساويين فإن إنشاء فترة ثقة يكون على درجة عالية من التعقيد مما دعا العلمين فشر- بهترين (fisher- Behren) لحل هذه المشكلة من خلال اقتراح درجات الحرية المعدلة وهذا سوف لن نتطرق له في دراستنا الحالية.

مثال (3-12)

عينتان منن السجائر إحداهما تحتوي 10 والأخرى تحتوي 8 سجائر ومتوسط كمية النيكوتين في الأولى 1.3 وانحرافها المعياري 1.3 بينما متوسط كمية النيكوتين في الثانية 2.7 وانحرافها المعياري 0.7 علما أن لكلي الكميتين التوزيع الطبيعي والتباين يكون معلوما ومتساويا، أوجد مجال ثقة 95% لمتوسط الفرق بين الكميتين؟

الحل:

$\mathbf{1}-\alpha=0.95 \Rightarrow \alpha=1=0.95=0.5 \Rightarrow 1-\frac{\alpha}{2}=1-\frac{0.05}{2}$

$=1-0.025=0.975$

وأن: $16=2-8+10=2-\;\; n_2 + n_1$ومن جدول توزيع t نحسب:

$t_{1-\frac{\alpha}{2}}(n_1+n_2-2)=t_{0.975}(16)=2.120$

ثم نحسب قيمة التباين:

$S_p^2=\frac{(10-1)(3.1)^2+(8-1)(0.7)^2}{10+8-2}$ ، وبعد حساب الناتج وإيجاد الجذر التربيعي له نحصل على $S_p=0.596$،

ونعوض في صيغة فترة الثقة:

$$\left[(\bar{X}_1-\bar{X}_2) \pm t_{1-\frac{\alpha}{2}} S_p \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right]=(3.1-2.7) \pm (2.12)(0.596)\sqrt{\left(\frac{1}{10}+\frac{1}{8}\right)}$$

3-1-4 فترة الثقة للتباين σ^2

إذا كان $\bar{X}$ و S^2 هما متوسط وتباين عينة عشوائية حجمها n على التوالي، تتوزع بتغير عشوائي كاي-سكوير χ^2 توزيعا طبيعيا أي أن:

$\chi^2=\frac{(n-1)s^2}{\sigma^2}$ بدرجة حرية n-1 فإن فترة الثقة يمكن اشتقاقها بنفس الأسلوب السابق وتكون صيغتها النهائية كالآتي:

$$\left[\frac{(n-1)s^2}{\chi_{\frac{\sigma^2}{2}}(n-1)}, \frac{(n-1)s^2}{\chi_{1-\frac{\sigma^2}{2}}(n-1)}\right]$$

مثال(3-13)

سحبت عينة من 20 شخصا وقسيت أوزانهم فوجد أن الانحراف المعياري له 9 كلغ. فإذا كان التوزيع الطبيعي المعياري فأوجد 95% مجال ثقة لتباين الأوزان في المجتمع؟

الحل: يترك تمري للطالب (درجة الحرية تكون n-1=19) ونستخدم توزيع كاي-سكوبر

4-1-4 فترة الثقة للنسبة بين تباينين

إذا كان s_1 و s_2 عينتين عشوائيتين مستقلتين حجمهما n_1وn_2 مسحوبتان من مجتمعين طبيعيين فإن:

$f=\frac{S_1{}^2}{\sigma_1{}^2}\div\frac{S_2{}^2}{\sigma_2{}^2}=\frac{S_1{}^2}{\sigma_1{}^2}\div\frac{S_2{}^2}{\sigma_2{}^2}=\frac{S_1{}^2\sigma_2{}^2}{S_2{}^2\sigma_1{}^2}$تتوزع توزيع F بدرجة حرية (n_1-1) و (n_2-1)(للبسط والمقام فإن فترة الثقة بين هذين التباينين تشتق بنفس الأسلوب السابق وهو:

$$\left[F_{\frac{\alpha}{2}}(n_1-1,n_2-1)\frac{S_1{}^2}{S_2{}^2},F_{\frac{\alpha}{2}}(n_1-1,n_2-1)\frac{S_1{}^2}{S_2{}^2}\right]$$

مثال (3-14)

عينتان مستقلتان مسحوبتان من مجتمع طبيعي حجمها 10 و 8 على التوالي، وتبلينهما 0.25و 0.499، أوجد فترة ثقة للنسبة بين التباينين عند مستوى ثقة 90%؟

الحل: تمرين

5-1-4 مجال الثقة للفرق بين نسبتي مجتمعينp_1-p_2

إن مقارنة النسب في المجتمعات بالنسبة لصفة معينة مثل نسبة إناث كلية التربية مع نسبة إناث كلية العلوم يتطلب تقدير الفرق بين هاتين النسبتين.

نأخذ عينتين عشوائيتين لمتغيرين مستقلينX_1و X_2أحجامهما n_1 وn_2، وكان y_1 وy_2يمثل عدد مرات النجاح لكليهما، فإن مقدري p_1 و p_2 هما:

$p_1=\frac{y_1}{n_1}$, $p_2=\frac{y_2}{n_2}\Rightarrow\hat{p}_1-\hat{p}_2$

وهذا يكون مقدارا غير متحيز لـ $p_1 - p_2$ ، لأن:

$$E(\hat{p}_1-\hat{p}_2)=E(\hat{p}_1)-E(\hat{p}_2)= p_1 - p_2 \qquad \hat{p}_1-\hat{p}_2$$

وبما أن $(n_1, n_2 \geq 30)$ فيكون للمتغير العشوائي $\hat{p}_1-\hat{p}_2$ التوزيع الطبيعي، أي أن:

$$M\hat{p}_1-\hat{p}_2= p_1 - p_2$$

والتباين يكون:

$$\sigma^2_{\hat{p}_1-\hat{p}_2}{}^2 = V(\hat{p}_1+\hat{p}_2)=v(\hat{p}_1)+v(\hat{p}_2)=\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}$$

$$Z=\frac{(\hat{p}_1-\hat{p}_2)-p_1-p_2}{\sqrt{\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}}}\approx N(0,1)$$

$$p\left[-Z_{1-\frac{\alpha}{2}} < Z < Z_{1-\frac{\alpha}{2}}\right]=1-\alpha$$

نبدل p_1 بـ $\hat{p}_1$ *ونبدل* p_2 بـ $\hat{p}_2$ فنحصل على فترة الثقة التالية:

$$\left[(\hat{p}_1-\hat{p}_2)-Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}+\frac{\hat{p}_2\hat{q}_2}{n_2}},(\hat{p}_1-\hat{p}_2)+Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}+\frac{\hat{p}_2\hat{q}_2}{n_2}}\right]$$

مثال (3-15)

لمقارنة نسبة المدخنين في مدينتي صنعاء وعدن أخذت عينة من سكان صنعاء حجمها1000 شخص، ووجد من بينهم 700 مدخن، وأخذت عينة من سكان حجمها 600 شخص، ووجد من بينهم 4388 مدخنا فأوجد **%**92 مجال ثقة بين نسبتي المدخنين في صنعاء وعدن؟

الحل:

$$\hat{p}_1 = \frac{700}{1000}=0.7 \ , \ \hat{p}_2=\frac{438}{600}=0.73$$

$$(\hat{p}_1-\hat{p}_2)=0.70-0.73=-0.03$$

من جدول التوزيع الطبيعي المعياري يكون:

$$1-\alpha=1-0.92=0.08=\alpha \Rightarrow \frac{\alpha}{2}=\frac{0.08}{2}=0.04$$

$$Z_{1-\frac{\alpha}{2}} = Z_{0.96} = 1.75$$

$$\left[(\hat{p}_1-\hat{p}_2)-Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}+\frac{\hat{p}_2\hat{q}_2}{n_2}},(\hat{p}_1-\hat{p}_2)+Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}+\frac{\hat{p}_2\hat{q}_2}{n_2}}\right]$$

$$\left[\begin{array}{l}-0.03-1.75\sqrt{\frac{(0.7)(0.30)}{100}+\frac{(0.73)(0.27)}{600}},\\ -0.03+1.75\sqrt{\frac{(0.7)(0.30)}{100}+\frac{(0.73)(0.27)}{600}}\end{array}\right]$$

(-0.075 ,0.011)

نلاحظ أن طرفي المجال بإشارتين مختلفتين مما دل على أنه لا يوجد فرق جوهري في نسبة المدخنين في مدينتي صنعاء وعدن بمستوى ثقة 92%.

تمارين مختارة

1.يرغب مدير في تقدير متوسط عدد الدقائق التي يستغرقها العمال لإنجاز عملية صناعية معينة في حدود ± 3 دقيقة وبدرجة ثقة 90%. ويعلم المدير من خبرته الماضية أن الانحراف المعياري σ هو 15 دقيقة. أوجد الحد الأدنى لحجم العينة المطلوب (n>30)؟

2.وجدت إدارة التعليم لإحدى الولايات، أن في عينة من 100 شخص مختارين عشوائيا، من بين الملتحقين بالجامعات 40% منهم، قد حصلوا على درجات جامعية. أوجد فترة الثقة 99% لنسبة الحاصلين على درجات جامعية من بين جميع الملتحقين بالجامعة؟

3.ما المقصود بالمعاينة العشوائية؟ وما هي أهميتها؟

4.ماذا يقصد بتوزيع المعاينة للوسط؟ وكيف يتم الحصول عليه؟

5.مجتمع مكون من 12 عنصرا بوسط مقداره 100 وانحراف معياري مقداره 60. أوجد الوسط والخطأ المعياري لتوزيع المعاينة للمتوسط عندما يكون حجم العينة100؟

6.أوجد احتمال أن يكون وسط عينة عشوائية من 25 عنصرا، مأخوذة من مجتمع طبيعي بمتوسط 90 وانحراف معياري 60 أكبر من 100؟

7.عينة عشوائية من 64 مفردة وسطها 50 وانحرافها المعياري 20 أخذت من مجتمع عدد مفرداته 800، أوجد تقديرا بفترة لوسط المجتمع نكون معه واثقين 95%، أن تتضمن وسط المجتمع؟

8.تم تخدير 100 شخص من المرضى كبار السن وتبين أن 36 منهم حصلت على مضاعفات جراء التخدير:

أ.أوجد 95% مجال ثقة لنسبة الذين يعانون مضاعفات جراء التخدير؟

ب.عين الخطأ المطلق الأعظم المرتكب عندما نفترض أن $\hat{p}$= p *بنسبة* 98%؟

ج.ما هو حجم العينة التي ينبغي دراستها لكي لا تتجاوز الخطأ في تقدير pالمقدار ε=0.04 وبثقة 99%؟

مقدمة:

أشرنا في الفصل السابق إلى أن هناك أسلوبين أساسيين للاستدلال الإحصائي حول معلمة من معالم المجتمع $\boldsymbol{\theta}$. يعتمد الأسلوب الأول على تقدير قيمة (التقدير النقطي) أو مدى هذه المعلمة(التقدير بفترة). أما الأسلوب الثاني للاستدلال الإحصائي حول معلمة المجتمع فيقوم على اختبار الفرضيات. تبعا لهذا الأسلوب نفترض قيمة معينة لمعلمة المجتمع ثم نعتمد على عينة مسحوبة عشوائيا من هذا المجتمع في اتخاذ قرار قبول أو رفض هذه القيمة، وفي ذلك لا يختلف الأسلوب الأول عن الثاني. وكما سنرى فيما بعد أن الأسلوبين مرتبطين ارتباطا وثيقا حيث يمكن أن يستخدم الأسلوب الأول في اتخاذ أو رفض فرض معين حول معلمة من معالم المجتمع. ومن الطبيعي. أن نتوقع أننا سواءا استخدمنا أسلوب فترات الثقة أو اختبار الفرضيات سوف نتخذ نفس القرار حول قبول أو رفض فرضية معينة.

تعريف: الفرضية الإحصائية هي جملة حول معلمة أو أكثر من معالم المجتمع قد تكون صحيحة أو خاطئة وبالتالي يحتاج في قبولها أو رفضها إلى قرار. وتجدر الإشارة إلى أن اختبار الفرضية يدعم الفرضية ولا يرقى إلى البرهان.

إن الهدف من الاختبار الإحصائي هو اختبار فرضية حول معلمة أو أكثر من معالم المجتمع الإحصائي. وبالتالي فإن الاختبار الإحصائي يتكون من أربعة عناصر.

إن تخصيص أي من هذه العناصر الأربع يعرف اختبار خاص وتغيير واحد أو أكثر من هذه العناصر ينشئ اختبار جديد. إن الفرضية البديلة هي تلك الفرضية التي يدعمها الباحث، أما الفرضية الصفرية(فرضية العدم) فهي التي تناقض الفرضية البديلة أي بمعنى إذا كانت فرضية العدم خاطئة فإن الفرضية البديلة صحيحة.

إن قرار أو قبول الفرضية الصفرية يعتمد على معلومات من العينة المسحوبة من المجتمع قيد الدراسة. إن قيم العينة تستخدم لحساب عدد معين يقابل نقطة أو خط والذي يوظف لاتخاذ القرار. إن صانع القرار يسمى إحصاء الاختبار. إن مجموعة قيم صانع القرار الكلية يقسم منطقة اتخاذ القرار إلى منطقتين، واحدة تحوي القيم التي تدعم الفرضية البديلة وتسمى منطقة الرفض(rejection region). إن قيمة حالة الاختبار (إحصاء الاختبار) إذا وقعت في منطقة الرفض فإن فرضية العدم سوف ترفض وبالتالي فإن الفرضية البديلة يتم قبولها. أما إذا وقعت قيمة دالة الإختبار(القيمة المحسوبة) في منطقة القبول. فإن الفرضية سوف يتم قبولها.

إن اتخاذ القرار ينطوي على نوعين من الأخطاء في أي من اختبارات الفروض الإحصائية وهذا يقودنا إلى التعريف التالي:

تعريف:

1) الخطأ من النوع الأول(error of type I) في اختبار الفروض هو الخطأ الذي يحدث عند رفض فرضية العدم عندما تكون صحيحة. والاحتمال الذي يحدث نتيجة الخطأ سوف نرمز له بالرمزα.

2) الخطأ من النوع الثاني (error of type II) في اختبار الفروض وهو الخطأ الذي يحدث عند قبول (عدم رفض) فرضية العدم عندما تكون خاطئة، والاحتمال يحدث نتيجة الخطأ الذي سوف نرمز له بالرمز β.

الجدول التالي يوضح هذين النوعين من الأخطاء

القرار	صحيح	خاطئ
رفضH_0	خطأ من النوع الأول	قرار صحيح
قبولH_0	قرار صحيح	خطأ من النوع الثاني

إن الملائم أن اختبار الفروض يقاس بالاحتمالات الناتجة عن الخطأ الأول والثاني والتي نرمز لها بالرموز αو β هو الاحتمال الناتج عن رفضH_0 عندما تكون صحيحة فإن منطقة الرفض تزداد بازديادα وفي نفس الوقت ينقص β عند ثبات حجم العينة. وعند إنقاص منطقة الرفض فإن قيمة β تزداد. أما إذا زدنا حجم العينة فإن معلومات أكثر ستوفر لاتخاذ القرار على أساسه وعندئذ ستنقص قيمتي α وβ إن قيمة β وهو الاحتمال الناتج عن الخطأ من النوع الثاني يتغير بالاعتماد أساسا على القيم الحقيقية لمعلمة المجتمع.

تعريف: قوة إحصاء الإختبار (power of a statistical test) يعطى بالعلاقة (رفضH_0 عندما تكون خاطئة) $p=1-\beta$

حيث تقيس قوة إحصاء الاختبار مقدرة إحصاء الإختبار في صنع القرار.

بوجه عام، سنستخدم الرمزθ لنرمز إلى واحدة من المعالم الأربعة الخاصة بالمجتمع الإحصائي وهي μ أو($\mu_1-\mu_2$) أو p

(p_1-p_2) والرمز($\hat{\theta}$) لترمز إلى التقدير النقطي غير المتحيز المقابل ($\hat{\mu}=\bar{x}$) و($\hat{p}=\frac{x}{n}$) وهكذا.

تعريف: (القيمة الحرجة critical value) هي قيمة لإحصاء الاختبار تقسم منطقة الاختبار إلى منطقة قبول وأخرى منطقة الرفض.

بالنسبة للفرض الذي ترغب في اختياره والذي رمزنا له بالرمز (H_0) أي افتراض عدم وجود فرق بين معلمة المجتمع والقيمة الافتراضية ويقابل ذلك فرضا آخر يرمز بالرمز H_a. ومما يذكر أن تحديد H_a يحدد تلقائيا H_0 عموما لا تخرج صيغة الفرض المطلوب اختباره عن واحد من الصيغ الثلاثة التالية:

الصيغة الأولى: $H_0 : \theta = \theta_0$

$H_a\ \theta \neq \theta_0$

ففي هذه الصيغة واعتمادا على المعلمات التي حصلنا عليها من العينة سوف نقبل H_0 أي أن معلمة المجتمع (θ) تساوي القيمة الإفتراضية(θ_0) أو سوف نرفض H_0 بمعنى أن معلومات العينة لم تمدنا بما يمكننا من عدم رفض H_0 وبالتالي نرفضه، بعدم رفضنا H_0 فنحن في الحقيقة نرفض H_a ويطلق على هذه الصيغة اختبار فرضية ذات ذيلين **tow-tails test**)

الصيغة الثانية: $H_0 : \theta = \theta_0$

$H_a\ \theta < \theta_0$

في هذه الصيغة عدم رفضنا H_0 **يعني أن** θ تساوي أو تزيد عن θ_0 وبالتالي نرفض H_a والعكس إذا رفضنا H_0 فهذا يعني أن θ أقل من θ_0 وعليه نكون قد قبلنا ضمنا H_a ويطلق على هذه الصيغة اختبار فرضية ذات ذيل أيسر left-tail test)

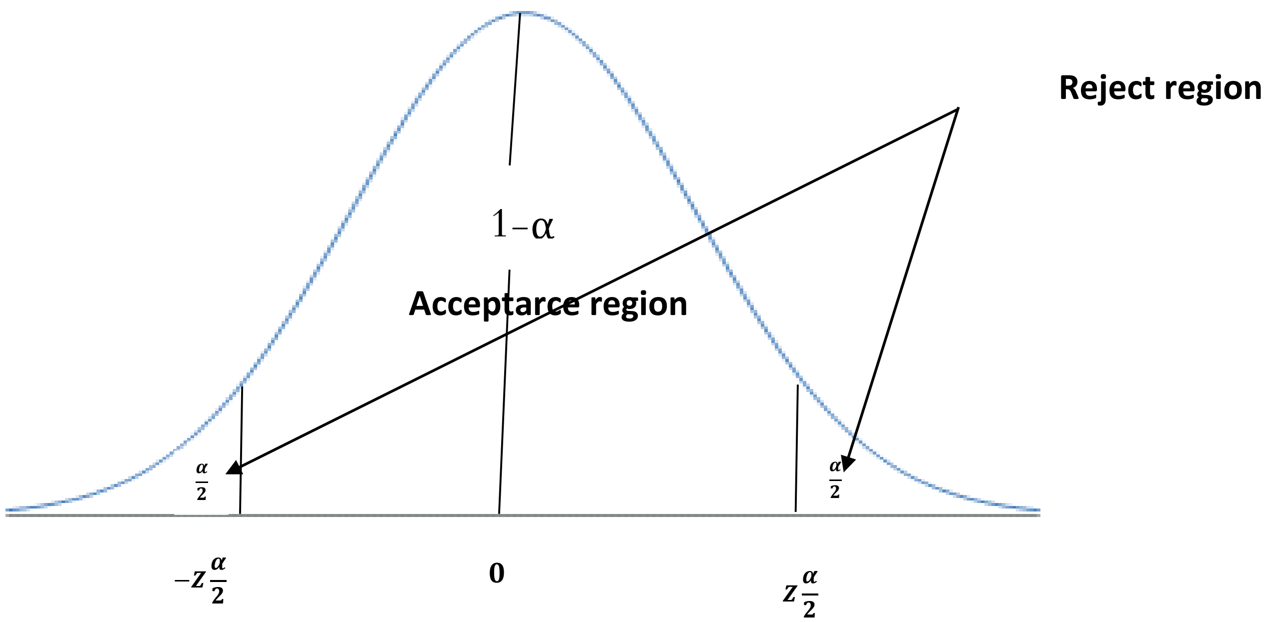
الصيغة الثالثة: $H_0 : \theta = \theta_0$

$H_a : \theta > \theta_0$

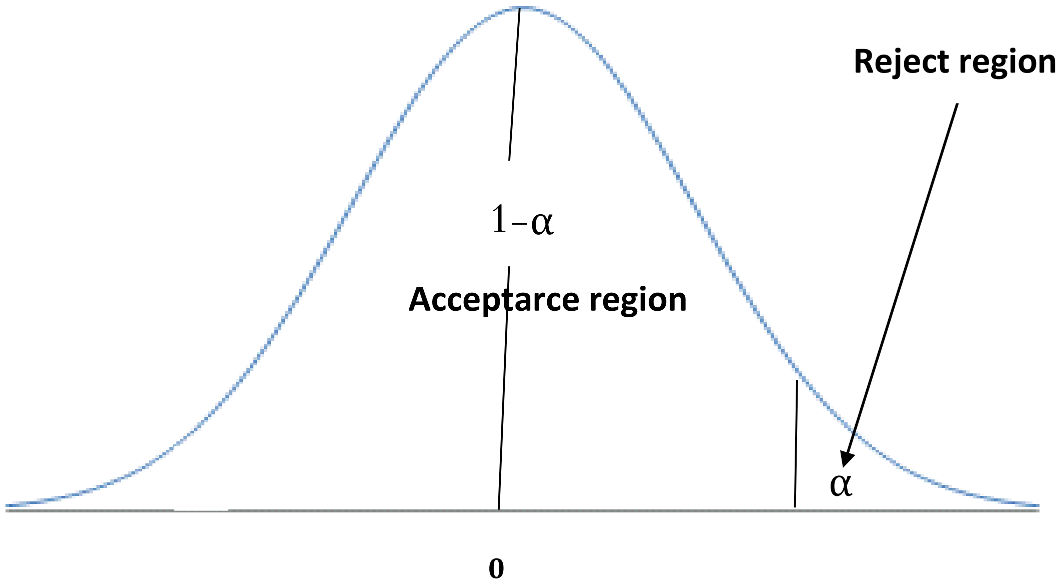
وفي هذه الصيغة فإن عدم رفضنا H_0 يعني أن ثبت لدينا من معلومات العينة أن معلمة المجتمع θ أقل من θ_0 *وعدم* رفضنا H_0 يعني ضمنيا رفضنا H_a والعكس إذا رفضنا H_0 يعني لم يثبت لدينا أن θ أقل من θ_0 وبالتالي نكون ضمنا قد قبلنا H_a ويطلق على هذه الصيغة اختبار فرضية ذات ذيل أيمن **(right-tail test)**

ولكن ما نود الإشارة إليه هنا هو أن ما يحدد نوعية طرف الاختبار هو الفرض البديل حيث يظهر فرض العدم دائما تساوي معلمة المجتمع مع القيمة الافتراضية ويحوي ضمنا مكمل الفرض البديل. فإذا كان:

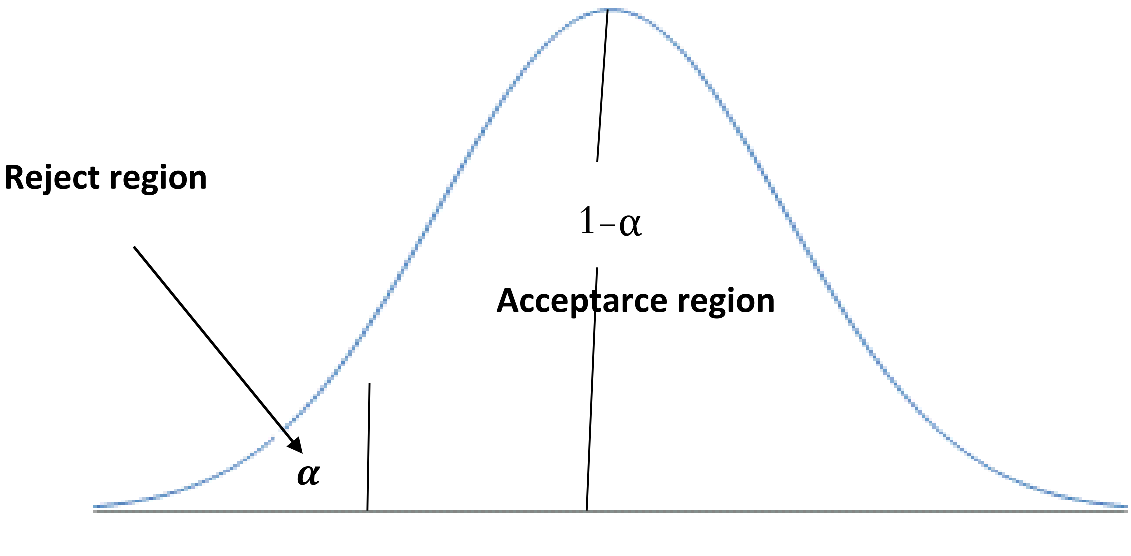
الفرض البديل $\theta \neq \theta_0$ فذلك يعني أن الغير مرغوب فيه هو الطرف الذي تزيد فيه θ عن θ_0 وبالتالي فإن منطقة الرفض هما الطرفين (الطرف الأعلى والطرف الأسفل) (الاتجاه الموجب لمتغير اتخاذ القرار أو الاتجاه السالب لمتغير اتخاذ القرار) وبقية المساحة تمثل منطقة عدم رفض(قبول)، والشكل التالي يمثل ذلك.



الفرض البديل ($\theta > \theta_0$) فذلك يعني أن الغير مرغوب فيه هو الطرف الذي تزيد فيه θ عن θ_0، إذن طرف الرفض هو الطرف الأيمن وبقية المساحة تمثل منطقة عدم رفض (قبول) والشكل التالي يوضح هذه الحالة.



الفرض البديل $(\theta < \theta_0)$ فذلك يعني أن الغير مرغوب فيه هو الطرف الذي تقل θ عن θ_0 ويكون *الطرف* هو *ذلك الطرف الذي تقل فيه* θ عنθ_0 وهو الطرف الأيسر وبقية المساحة منطقة عدم رفض(قبول) والشكل التالي يوضح الحالة



ويخلص قولنا إلى أن معلومات العينة هي التي تساعد في اتخاذ القرار حول رفض أو عدم رفض H_0 ويجب ملاحظة أن القرار دائما حول H_0 ويأتي القرار حول هذا القرار ضمنيا.

خطوات اختبار الفرضيات الإحصائية:

يتم إجراء اختبار الفرضية الإحصائية على ست خطوات:

الخطوة الأولى: صياغة الفرضيات H_0 و H_a.

الخطوة الثانية: اختيار مقياس العينة الذي سيتخذ القرار على أساسه.

الخطوة الثالثة: اختيار مستوى احتمال حدوث خطأ من النوع الأول α والتي يطلق عليه مستوى الدلالة الإحصائية.

الخطوة الرابعة: اختبار مقياس اتخاذ القرار الذي على أساسه يتم رفض H_a **أو** H_0.

الخطوة الخامسة: اختبار مفردات العينة وحساب قيمة مقياس العينة.

الخطوة السادسة: اتخاذ القرار برفض أو عدم رفض H_0 اعتمادا على مقياس اتخاذ القرار وقيمة مقياس العينة.

(2-11) اختبار الفرضيات الإحصائية حول متوسط المجتمع (μ)

رأينا أن الوسط الحسابي المحسوب من العينة($\bar{x}$) يمثل تقدير **نقطة جيدة لمتوسط المجتمع(μ)** وبالتالي يؤخذ كمقياس عيني مناسب. عند اختبار الفرضيات حول متوسط المجتمع μ هناك حالتين أساسيتين:

الحالة الأولى: التباين للمجتمع(σ^2) معلوم.

الحالة الثانية: تباين المجتمع (σ^2) مجهول.

في كل حالة من هاتين الحالتين سوف ندرس الأسلوب المناسب لاختبار الفرضيات والذي سيعتمد على كون المجتمع قيد الدراسة يتبع التوزيع الطبيعي أو لا يتبعه.

الحالة الأولى: تباين المجتمع معلوم

من دراستنا للتوزيع العيني ونظرية النهاية المركزية عرفنا أن التوزيع العيني للوسط ($\bar{X}$) يتوزع تبعا للتوزيع الطبيعي بوسط حسابي(μ) وانحراف معياري $\frac{\sigma}{\sqrt{n}}$.

أولا-مجتمع قيد الدراسة يتبع التوزيع الطبيعي:

إذا كان لدينا مجتمع طبيعي بمتوسط (μ) مجهول وتباين (σ^2) معلوم فإن مقياس اتخاذ القرار هو المتغير العشوائي (Z) حيث $Z:N(0,1)$ ويعطى بالعلاقة التالية:

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

وذلك بغض النظر عن حجم العينة

ثانيا-مجتمع قيد الدراسة يتبع التوزيع الطبيعي:

إذا كان لدينا مجتمع بمتوسط (μ) مجهول لكن تباينه (σ^2) معلوم لاختبار فرضية متوسط المجتمع (μ) فإن مقياس اتخاذ القرار هو Z حيث $Z:N(0,1)$ شريطة أن يكون حجم العينة ($n\geq 30$) وبالتالي فإن دالة الاختبار (مقياس اتخاذ القرار تعطى بالعلاقة التالية:

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

وذلك بغض النظر عن حجم العينة.

ثالثا-مجتمع قيد الدراسة لا يتبع التوزيع الطبيعي:

إذا كان لدينا مجتمع بمتوسط (μ) مجهول لكن تباينه (σ^2) معلوم لاختبار فرضية متوسط المجتمع (μ) فإن مقياس اتخاذ القرار هو Z حيث $Z:N(0,1)$ شريطة أن يكون حجم العينة ($n\geq 30$) وبالتالي فإن دالة الاختبار (مقياس اتخاذ القرار تعطى بالعلاقة التالية:

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

الحالة الثانية: تباين المجتمع قيد الدراسة معلوم

أولا: قيد الدراسة يتوزع تبعا للتوزيع الطبيعي

إذا كان لدينا مجتمع معتدل (طبيعي) مجهول الوسط والتباين ولاختبار حول وسط هذا المجتمع فيتم الاعتماد على العينة في تقدير التباين حيث تعتبر(μ)غير متحيزة للمعلمة(σ^2). وبناءا عليه فإن مقياس اتخاذ القرار هو المتغير العشوائي بدرجات حرية (n-1) ويعطى بالعلاقة التالية:

$$t=\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$$ المحسوبة

ثانيا: توزيع المجتمع غير معلوم

إذا كان لديننا مجتمع مجهول الوسط والتباين فيتم استخدام S^2 كتقدير للمعلمة (σ^2) وتكون دالة الاختبار هي:

$$t=\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$$ المحسوبة

حيث t **تقرب من توزيع (t) بدرجات حرية** (n-1) إذا كان حجم العينة ($n\geq 30$) .

مثال(1): يدعي أحد موردي آلات التعبئة أن الوسط الحسابي له قد بلغ 0.495. وللتأكد من صحة إدعاء المورد تم سحب عينة حجمها حيث كان متوسط العبوة (0.495 كقسم) فإذا علمت أن الأوزان النسبية للتوزيع كان بانحراف معياري (24) غرام، المطلوب مساعدة الشركة في اتخاذ قرار حول شراء هذه الآلات وذلك عند مستوى دلالة إحصائية ($\alpha=$ 0.05)

الحل: (1) صياغة الفرضيات: لاحظ بأن المصنع يرغب أن يكون متوسط وزن الكيس (0.5 كلغ) بدون زيادة أو نقصان وهذا يعني أن:

H_0:μ=0.5

H_0:$\mu \neq$0.5

(1) (اختبار فرضية ذات ذيلين)

(2) مقياس العينة هو الوسط $\bar{x}$

(3) مستوى الدلالة ($\alpha = 0.05$)

(4) قاعدة اتخاذ القرار

المجتمع يتبع التوزيع الطبيعي بانحراف معياري معلوم (σ=0.024)

إذن القاعدة تعتمد على المتغير z الذي يتبع التوزيع الطبيعي المعياري.

$$Z_{\text{المحسوبة}}=\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}=\frac{0.495-0.5000}{0.024/\sqrt{36}}=-1.25$$

(5) حيث أن $(\alpha = 0.05)$ والفرضية بذيلين فإن z الحرجة يتم حسابها كالتالي:

$Z=z_1$ الجدولية =z الحرجة

أوجد z_1 بحيث $p(0<Z<z_1)=\frac{1-\alpha}{2}=0.4750$

ومنها $z=1.96$ الحرجة

(06) القرار: بما أن z المحسوبة ليست أكبر من z الجدولية أو z *المحسوبة* ليست أقل من z الجدولية الأمر الذي يجعلنا لا نرفض فرضية العدم ونقول بأن متوسط وزن الكيس المعبأ بواسطة هذه الآلات لا تختلف عن (0.5) كلغ عند مستوى دلالة $(\alpha = 0.05)$ وبناءا على ما توفر لنا منن العينة فإننا نوصي بشراء الآلات ونؤيد إدعاء مورد الآلات.

مثال(02): يدعي أحد منتجي المصابيح الكهربائية أن متوسط عمر المصباح الذي ينتجه لا يقل عن (3000) ساعة. للتأكد من صحة إدعاء المنتج سحبت عينة حجمها (81) مصباح وكان عمر المصابيح من العينة (2980) ساعة. فإذا كان معلوما لهذا النوع من الصناعة أن الانحراف المعياري (250) ساعة. المطلوب عند مستوى دلالة (0.01) مناقشة هذا الادعاء.

الحل:

(1): الادعاء أن متوسط عمر المصباح لا يقل عن (3000) ساعة.

$H_0=\mu=3000$

مقابل $H_a=\mu < 3000$ (الاختبار فرضية ذات ذيل أسفل)

(2) مقياس العينة هو الوسط $(\bar{x})$

(3) مستوى الدلالة $\alpha = 0.01$

(4) σ معلومة=250

توزيع المجتمع غير معلوم وأن n=82>30

إذن إحصاء الاختبار (دالة اتخاذ القرار) هو Z

$$Z_{\text{المحسوبة}}=\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$$

(05) حيث أن n=84، x=2980، σ=250، μ_0=3000

المحسوبة $Z=\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}=\frac{2980-30000}{250/\sqrt{81}}=-0.72$

الجدولية. = $-Z_{1-\alpha}=-Z_{0.99}=-2.33$

(06) القرار: حيث أن Z المحسوبة ليست أقل من Z الجدولية فإنا لا نرفض فرض العدم H_0 ونقول بصحة ادعاء المنتج بأن متوسط المصابيح المنتجة لا يقل عن (3000) ساعة.

مثال(3): وجد في دراسة سابقة أن معدل قيم الفواتير في أحد المستشفيات تساوي (70.2) وأن توزيعها يقترب من التوزيع الطبيعي.أردننا اختبار فرضية أن قيم الفواتير قد تغيرت، فدرست (452) فاتورة أخذت عشوائيا فوجد أن $\bar{x}=73.7$ وأن S=1.12 باستعمال مستوى دلالة $\alpha=0.05$ اختبر الفرضية $H_0=\mu=70.2$ مقابل الفرضية البديلة المناسبة.

الحل: (1) صياغة الفرضيات: قيم الفواتير تغيرت

(2) مقياس العينة (x).

(3) مقياس الدلالة $H_0=\mu=70.2$

مقابل $H_1=\mu\neq 70.2$

$$\alpha=0.05$$

(04) σ مجهولة، توزيع المجتمع التوزيع المعتدل.

لإحصاء الاختبار هو المتغير العشوائي (t) حيث

$$t_{\text{المحسوبة}}=\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$$

(5) حيث أن: n=25 , $\bar{x}$=73.7 , S =11.2 , μ_0=70.2

المحسوبة $t=\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}=\frac{73.7-70.2}{11.2/\sqrt{25}}=1.5625$

الجدولية$t= t\left[1-\frac{\propto}{1};n-1\right]= t[0.975;24]=2.064$

(6) القرار: حيث أن t المحسوبة أكبر من t الجدولية فإننا لا نرفض H_0 ونقول بأن قيمة الفواتير قد تغيرت.

مثال(4): يبلغ معدل طول لاعب السلة في المنتخب الوطني (190) سم وفي السنوات الأخيرة بدأ الإقبال على هذه اللعبة وبالتالي فإن المنتخب وضع شروطا أشد بخصوص الطول. اختر الفرضية التي تقول إن معدل اللاعب في المنتخب قد ازداد علما بأن عينة عشوائية حجمها (81) من أفراد المنتخب الوطني أعطت معدل الطول(198) سم بتباين مقداره (36). استعمل مستوى دلالة ($\alpha = 0.01$).

الحل:

(1) صياغة الفرضيات: حيث الادعاء بأن الطول قد ازداد فالمطلوب اختبار $H_0:\mu=190$ مقابل $H_0:\mu>190$

(2) مقياس العينة هو الوسط ($\bar{x}$)

(3) مستوى الدلالة $\alpha = 0.01$

(4) σ مجهولة وتوزيع المجتمع غير معلوم وبما أن n=81>30 *فإن* إحصاء الاختبار هو المتغير العشوائي هو (t) ويقترب من توزيع (t) بدرجات حرية (n-1).

(5)حيث أن حيث أن: n=81 , $\bar{x}$= 198 , S =6 , μ_0=190 وبالتالي فإن

t المحسوبة $=\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}=\frac{198-190}{6/\sqrt{81}}=12$

t الجدولية $=t[1-\alpha\,;\ n-1]=t[0.99;80]=2.66$

(06) القرار: حيث أن (t) المحسوبة أكبر من (t) الجدولية فإننا نرفض H_0 ونقول بأن معدل الطول لم يزداد.

(3-11) اختبار الفرضيات الإحصائية حول نسبة المجتمع(p):

من دراستا السابقة بأن الخطأ المعياري للنسبة ($\bar{p}$) والذي رمزنا له بالرمز ($\sigma_{\bar{p}}$) يحسب بالعلاقة : $\sigma_p=\sqrt{\frac{p(1-p)}{n}}$

ومن هذه العلاقة نلاحظ أنها تعتمد على p (نسبة المجتمع)، وحيث أن p تكون مجهولة ونهدف باختبار الفرضيات حولها أن نتعرف عليها أو نقترب من قيمتها، لهذا سوف لا نقابل حالات معلومية أو عدم معلومية (σ_p) كما حدث في حالة اختبار الفرضيات حول متوسط المجتمع (μ)

كما ذكرنا فإن التوزيع الطبيعي العيني للنسبة (p) يتبع توزيع إما ذات الحدين أو للتوزيع الهندسي ومن ناحية أخرى إذا كان حجم العينة كبيرا ($n \geq 100$) فسوف تساعدنا نظرية النهاية المركزية في أن يقترب توزيع المعاينة للنسبة من التوزيع الطبيعي.

وقد يقول البعض: إذا كانت (σ_p) مجهولة وسيتم تقديرها والأجدر أن يقترب توزيع المعاينة للنسبة (p) من توزيع (t) مشابه تماما للتوزيع الطبيعي المعياري. لذلك سوف نستخدم التوزيع الطبيعي بصفة دائمة في اختبار الفرضيات حول النسبة (p) في حالة ($n \geq 100$) .

سوف نستخدم نسبة العينة ($\bar{p} = \frac{m}{n}$) حيث (m) نسبة من يملكون خاصية معينة في العينة، (n) حجم العينة الكلي كتقدير نعطي لنسبة المجتمع (p) ويكون مقياس اتخاذ القرار كالآتي:

$$Z=\frac{\bar{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} : N(0.1)$$

حيث(p_0)**:** القيمة الافتراضية التي يقاس عنده المجتمع (p)

ملحوظة: سنرمز لـــــــ Z *المحسوبة بـــــــ* **(**Z_{cal}**)**

مثال (5) بإهمال العمر حوالي (20%) من الأمريكان البالغون يشتركون ببرامج تخفيف السمنة على الأقل مرتين أسبوعيا. علاوة على ذلك إن برامج تخفيف السمنة تتغير مع تقدم العمر، ولأسباب عارضة فإن هذه البرامج تتوقف عند عمر معين في مسح محلي أخذت عينة عشوائية حجمها (n=100) دون اعتبار للعمر فوجد من بينهم (27) شخص مشتركون ببرامج تخفيف السمنة على الأقل مرتين أسبوعيا . هل تشير هذه البيانات إلى أن معدل الاشتراك المحلي يختلف. عند مستوى دلالة (α =0.05).

الحل: (1) صياغة الفرضيات $H_0: p=0.20$ مقابل $H_0: p \neq 0.20$

(2) مقياس العينة ($\bar{p}$) حيث أن $\bar{p}=\frac{27}{100}=0.27$.

$$Z_{cal}=\frac{\bar{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} : N(0.1)$$

(5) حيث أن $n = 100$ ، $p_0=0.20$ ، $\bar{p}=0.27$.

$$Z_{cal} = \frac{0.27-0.20}{\sqrt{\frac{(0.20)(0.80)}{100}}}$$

الجدولية $Z = Z_{1-\frac{\alpha}{2}} = Z_{0.975} = 1.96$

(6) القرار: حيث **أن** Z المحسوبة ليست أكبر من Z الجدولية فإننا لا نرفض H_0 ونقول بأن البيانات لا تختلف.

مثال(6): قررت إدارة البرامج الثقافية إلغاء برنامج من سيربح المليون إذا كان هناك ما يؤكد أن أقل من (**37%**) من مشاهدي التلفزيون لا يشاهدون البرنامج. لهذا تم اختيار عينة عشوائية من (5000) مشاهد بلغ عدد مشاهدي البرنامج (1800) مشاهد.

اختبر فرضية أن هذه الإدارة ستلغي البرنامج أم لا عند مستوى دلالة (α=0.05)؟

الحل: 1) صياغة الفرضيات $H_0: p=0.37$ مقابل $H_a: p < 0.37$

(2) مقياس العينة ($\bar{p}$) حيث أن $360 = \frac{1800}{5000} = \bar{p}$.

(3) مستوى الدلالة : $\alpha=0.05$

(4) مقياس اتخاذ القرار هو المتغير العشوائي (Z) حيث:

$$Z_{cal} = \frac{\bar{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} : N(0.1)$$

(5)حيث أن $n = 5000$ ، $p_0=0.20$ ، $\bar{p}=0.27$.

$$Z_{cal} = \frac{036-0.37}{\sqrt{\frac{(0.35)(065)}{5000}}}$$

الجدولية $Z = Z_{1-0.05} = -Z_{0.95} = -1.645$

(7) القرار: حيث أن Z المحسوبة ليست أقل من Z الجدولية فإن إدارة البرنامج لن تلغي برنامج من سيربح المليون.

مثال (7): قبل تسويق إنتاجها المتطور على نطاق واسع قررت الشركة أن تتأكد من أن أكثر من (**30%**) من المستهلكين سوف يشترون هذا الإنتاج المتطور. في بعض مراكز البيع والتي يعرض فيها إنتاجها المتطور تمت ملاحظة (1000) مشتري فوجد أن (320) منهم قد اشتروا الإنتاج المتطور. عند مستوى معنوية (**2.5%**) هل ننصح الشركة بالبدء في تسويق إنتاجها المتطور على نطاق واسع أم لا؟

الحل: 1) صياغة الفرضيات H_0:p=0.30 مقابل H_a:p >0.30

2) مقياس العينة ($\bar{p}$) حيث أن 0.32 =$\frac{320}{1000}$=$\bar{p}$.

3) مستوى الدلالة : α=2.5%

4) مقياس اتخاذ القرار هو المتغير العشوائي (Z) حيث:

$$Z_{cal} = \frac{\bar{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} : N(0.1)$$

5)حيث أن1000= n ، p_0=0.30 ، $\bar{p}$=0.32.

$$Z_{cal} = \frac{0.32-0.30}{\sqrt{\frac{(0.30)(0.30)}{1000}}}$$

الجدولية $Z=Z_{0.}957=1.96$

06) القرار: حيث أن Z المحسوبة أكبر من Z الجدولية فإننا نقبل فرض العدم ونقول بأنه تأكد لدينا من بيانات العينة وبمستوى دلالة (2.5%) أن (30%) من المستهلكين سيشترون من المنتج، وبالتالي ننصح الشركة بعدم البدء بالتسويق على نطاق واسع.

4-11) اختبار الفرضيات الإحصائية للفرق بين نسبتين (p_1-p_2):

من المعلوم بأن التوزيع العيني للنسبة يتبع التوزيع الطبيعي إذا كان كلا من (np) ، (n(1-p)) يزيد عن (0.5) وبصدد اختبار الفرضيات حول نسبة المجتمع سوف تكون أحجام العينات كبيرة حتى تضمن افتراض أن التوزيع العيني للنسبة يتبع التوزيع الطبيعي.

الآن.... نفترض أن لدينا مجتمعين (1) و(2) ونسب خاصية معينة في هذين المجتمعين هما (p_1) و (p_2) على الترتيب. سحبت عينة من المجتمع (1) حجمها (n_1) وكانت نسبة الخاصية في العينة هي($\bar{p}_1$)، وسحبت عينة من المجتمع (n_2) [مستقلة عن العينة الأولى] حجمها ونسبة الخاصية في العينة هي($\bar{p}_2$)وحيث أن أحجام العينات كبيرة فإنه يمكن القول بان التوزيع العيني لكل من ($\bar{p}_1$) و ($\bar{p}_2$) يتبع التوزيع (Z) عندئذ:

(1)　$\mu_{\bar{p}_1}$=p_1, $\sigma_{\bar{p}_1}=\sqrt{\frac{p_1(1-p_1)}{n_1}}$

(2)　$\mu_{\bar{p}_2}$=p_2, $\sigma_{\bar{p}_2}=\sqrt{\frac{p_2(1-p_2)}{n_2}}$

على غرار دراسة التوزيع العيني لأي مقياس إحصائي يمكن دراسة التوزيع العيني للفرق بين النسبتين ($\bar{p}_1$) و($\bar{p}_2$)

أي ($\bar{p}_1 - \bar{p}_2$) ، ويمكن إثبات أن متوسط التوزيع العيني للفرق ($\bar{p}_1 - \bar{p}_2$) والخطأ المعياري للفرق هما:

(a)　$\mu_{\bar{p}_1-\bar{p}_2}$=$p_1-p_2$

(b)　$\sigma_{\bar{p}_1-\bar{p}_2}=\sqrt{\frac{p_{1(1-p_1)}}{n_1}+\frac{p_{2(1-p_2)}}{n_2}}$

حيث أن التوزيع العيني للفرق ($\bar{p}_1 - \bar{p}_2$) يتبع التوزيع الطبيعي فيمكن استخدام المغير العشوائي (Z) كمقياس لاتخاذ القرار عند استخدام ($\bar{p}_1 - \bar{p}_2$) لاختبار فرضية حول الفرق بين نسبتي مجتمعين (p_1-p_2).

حيث أن (p_1) و (p_2) مجهولتين فحساب الخطأ المعياري($\sigma_{\bar{p}_1-\bar{p}_2}$) سيعتمد على بدائل ($p_1$) و ($p_2$) .

بداية نفترض دائما أن ($p_1=p_2$) حيث يثبت لدينا العكس. ونتيجة لذلك فإن:

قيمة p_1 = قيمة p_1= قيمة p (النسبة المشتركة) ويكون الخطأ المعياري:

$$\sigma_{\bar{p}_1-\bar{p}_2}=\sqrt{\frac{p_{1(1-p_1)}}{n_1}}+\sqrt{\frac{p_{2(1-p_2)}}{n_2}}=\sqrt{\bar{p}(1-\bar{p})\left[\frac{1}{n_1}+\frac{1}{n_2}\right]}$$

وحيث أن قيمة(p) مجهولة لدينا فيستخدم تقديرا لها من العينات ($\bar{p}$) ويكون الخطأ المعيار من العينة:

$$s_{\bar{p}_1-\bar{p}_2}=\sqrt{\frac{p_{1(1-p_1)}}{n_1}}+\sqrt{\frac{p_{2(1-p_2)}}{n_2}}$$:حيث: $\bar{p}=\frac{x_1+x_2}{n_1+n_2}$

حيث (x_1): عدد حالات النجاح في العينة (1) والتي حجمها (n_1).

(x_2): عدد حالات النجاح في العينة (2) والتي حجمها (n_2).

الآن: يمكن اتخاذ كتابة مقياس (Z) على النحو التالي:

$$Z=\frac{\bar{p}_1-\bar{p}_2}{\bar{p}(1-\bar{p})\left[\frac{1}{n_1}+\frac{1}{n_2}\right]}:N(0.1)$$

مثال(8): في دراسة للتعرف على الفرق بين نسبتي الشفاء في مجتمعين كل منهما يستخدم دواء معين. أخذت عينتين حجم كل منهما (100) مريض فوجد أن عدد الذين شفوا في العينة الأولى (78) مريضا وكان في العينة الثانية (82) مريض. هل توافق على أن تأثير الدواء في العينة الثانية أكبر منه في العينة الأولى. استخدم مستوى دلالة (5%)

الحل: (1) صياغة الفرضيات: $H_0 = p_1 = p_2$ مقابل $H_a : p_1 < p_2$ (اختبار فرضية ذات ذيل أسفل)

(2)مقياس العينة: $(\bar{p}_1 - \bar{p}_2)$:

$$\bar{p}_1 = \frac{78}{100} = 0.78$$

$$\bar{p}_2 = \frac{82}{100} = 0.82$$

(3) مستوى الدلالة: $\alpha = 0.05$.

(4) مقياس اتخاذ القرار هو المتغير العشوائي (Z) حيث

$$Z_{cal} = \frac{\bar{p}_1 - \bar{p}_2}{\bar{p}(1-\bar{p})\left[\frac{1}{n_1} + \frac{1}{n_2}\right]} : N(0.1)$$

حيث:

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{78+82}{100+100} = 0.80$$

(5) حيث أن:

$n_1 = n_2$، $\bar{p}_1 = 0.78$ ، $\bar{p}_2 = 0.82$، $\bar{p} = 0.80$ وبالتالي فإن:

$$Z_{cal} = \frac{0.78 - 0.82}{(0.80)(0.20)\left[\frac{1}{100} + \frac{1}{100}\right]} = 0.707.$$

$$Z_{\text{الجدولية}} = -z_{0.95} = -1.645$$

(6) القرار: حيث أن القيمة المطلقة (Z) المحسوبة ليست أكبر من القيمة لـــــ (Z) الجدولية فإننا نقبل بفرض العدم ونقول بان الدواء له نفس التأثير في الشفاء في المجتمعين.

مثال(9): يبدو بأن النساء اللواتي تجاوز أعمارهن (50) عاما واللواتي لم يلدن من قبل أكثر مقاومة لنوبات القلب وذلك نتيجة بحث قام به أحد الباحثين بعد مقابلة مرضى القلب. وتشير دراسة للسجلات الطبية للنساء اللواتي كن ضحية لنوبات القلب وجد أن (12) من بين (51) ضحية لنوبات القلب كن من العاقر وبمجموعة متماثلة مكونة من (47) امرأة درست على المصابات بنوبة القلب اثنين فقط كن من العاقر. هل تشير هذه البيانات عند مستوى دلالة (α=0.05) إلى فروق ذات دلالة إحصائية بين النساء العاقر المصابات بنوبات القلب وبين النساء اللواتي لسن عاقر بنفس الفئة العمرية.

الحل: (1) صياغة الفرضيات:

$H_0=p_1=p_2$ مقابل $H_a:p_1\neq p_2$

(2) مقياس العينة: $(\bar{p}_1 - \bar{p}_2)$:

$$\bar{p}_1=\frac{12}{51}=0.23529$$

$$\bar{p}_2=\frac{2}{47}=0.04255$$

(3) مستوى الدلالة: α=0.05.

(4) مقياس اتخاذ القرار هو المتغير العشوائي (Z) حيث:

$$Z=\frac{\bar{p}_1-\bar{p}_2}{\bar{p}(1-\bar{p})\left[\frac{1}{n_1}+\frac{1}{n_2}\right]}:N(0.1)$$

(5) حيث أن $n_1=51$ ، $n_2=47$، $\bar{p}_1=0.23596$ ، $\bar{p}_2=0.04255$، وبالتالي فإن:

وبالتالي:

$$\bar{p}=\frac{x_1+x_2}{n_1+n_2}=\frac{12+2}{15+47}=0.80$$

$$Z_{cal}=\frac{\bar{p}_1-\bar{p}_2}{\bar{p}(1-\bar{p})\left[\frac{1}{n_1}+\frac{1}{n_2}\right]}:N(0.1)$$

$$Z_{cal}=\frac{0.23529-0.04255}{(0.14286)(0.85714)\left[\frac{1}{51}+\frac{1}{47}\right]}=\frac{0.19274}{0.07076}=0.707.$$

$$Z_{\text{الجدولية}}=z_{0.95}=1.645$$

(6) القرار: حيث أن القيمة المطلقة (Z) المحسوبة أكبر من القيمة لـــــ (Z) الجدولية فإننا نرفض بفرض العدم ونقبل الفرض لبديل ونقول بأن هناك فروق ذات دلالة إحصائية.

(5-11) اختبار الفرضيات للفرق بين وسطي مجتمعين ($\mu_1 - \mu_2$):

يتطلب التعرض للحالات المختلفة التي يكون عليها حجم العينات وتوزيع المجتمعات ومعلومية التباين سنتعرض في هذا الموضوع إلى الحالتين التاليتين:

الحالة الأولى: σ_1^2، σ_2^2 معلومتان:

بافتراض أن لدينا مجتمعين يتوزعان تبعا للتوزيع الطبيعي، المجتمع الأول بمتوسط (μ_1) وتباين (σ_1^2) والمجتمع الثاني (μ_2) **وتباين** (σ_2^2) . سحبت جميع أزواج العينات المستقلة التي حجم العينة الأولى(N_1) وحجم العينة الثانية (N_2). من هذه العينات يمكن إثبات أن التوزيع العيني للفرق بين وسطي أزواج العينات يتوزع تبعا للتوزيع الطبيعي حيث:

$$\text{الوسط}=\mu_{\bar{x}-\bar{y}}=(\mu_1 - \mu_2)$$

$$\sigma_{\bar{x}-\bar{y}}=\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$

ويكون مقياس اتخاذ القرار هو المتغير العشوائي (Z) حيث:

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

وتجدر الإشارة إلى أنه إذا كان المجتمعين لا يتوزعا تبعا للتوزيع الطبيعي فيجب أن يكون حجم كلا العينتين أكبر من (25) مفردة فالاختبار أيضا يتم عن طريق (Z).

الحالة الثانية: σ_1^2، σ_2^2 مجهولتين ولكن σ_1^2= σ_2^2:

بافتراض أن لدينا مجتمعين يتوزعان تبعا للتوزيع الطبيعي الأول(μ_1) **وتباين** (σ_1^2) وكلا منهما مجهولين والوسط والتباين للمجتمع الثاني هما (μ_1)، (σ_2^2) على الترتيب وكذلك مجهولين. وإذا كان معلوم مسبقا **أن** σ_1^2= σ_2^2. وسحبنا جميع أزواج العينات المستقلة التي حجم الأولى منها (n_1) **والثانية** (n_2) . من هذه العينات يمكن إثبات أن التوزيع العيني للفرق ($\bar{x}-\bar{y}$) يتوزع تبعا للتوزيع الطبيعي والخطأ المعياري للتوزيع العيني هو:

$$\sigma_{\bar{x}-\bar{y}}=\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$

$$=\sqrt{\sigma^2\left[\frac{1}{n_1}+\frac{1}{n_2}\right]}=\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

وحيث أن المجتمعين لهما نفس التباين فإن التباينين المحسوبين من العينتين s_1^2، s_2^2 يعتبرا تقديرين لمعلمة واحدة لذلك فمن المنطقي أن نستخدمهما ونحسب الوسط الحسابي المرجح ليكون تقدير مشترك للتباين (σ^2).

$$s_c^2=\frac{(n_1-1)s_1^2+(n_1-1)s_2^2}{n_1+n_2-2}$$

ويكون الخطأ المعياري من العينات:

$$s_{\bar{x}-\bar{y}}=s_c\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

ويكون مقياس اتخاذ القرار هو المتغير العشوائي (t) بدرجات حرية $(v=n_1+n_2-2)$

$$t=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{s_c\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

ومن الجدير بالذكر بأنه إذا كان المجتمعين لا يتوزعا تبعا للتوزيع الطبيعي فيجب أن لا يصل حجم كل عينة من العينتين المسحوبتين عن (25) مفردة وفي هذه الحالة سوف نستخدم مقياس (t) أيضا.

قبل أن نورد بعض الأمثلة سنستعرض الصيغ المختلفة لاختبار الفرضيات.

$H_0: \mu_1=\mu_2$ مقابل(1) $H_a: \mu_1 \neq \mu_2$ (اختبار فرضية ذات ذيلين).

(2) $H_a: \mu_1 < \mu_2$ (اختبار فرضية ذات ذيل أسفل).

(3) $H_a: \mu_1 > \mu_2$ (اختبار فرضية ذات ذيل أعلى)

مثال (9): عينات مستقلتان حجم العينة الأولى (n_1=50) وحجم العينة الثانية (n_2=45) سحبت من مجتمعين يتوزعا تبعا للتوزيع الطبيعي حيث أن التباين للمجتمع الأول (σ_1^2=400) **والتباين للمجتمع الثاني (σ_2^2=900).**

المطلوب: اختبار فرضية($\mu_1=\mu_2$)مقابل0< ($\mu_1=\mu_2$) علما أن الوسط الحسابي للعينة الأولى (350) والوسط الحسابي للعينة الثانية (340) وذلك عند مستوى دلالة (α=0.05).

الحل: (1) صياغة الفرضيات:

$H_0: \mu_1 = \mu_2$ مقابل $H_a (\mu_1 - \mu_2) > 0$

(2)مقياس العينة$(\bar{X} - \bar{Y})$ حيث: $\bar{Y}$=340، $\bar{X}$=350

(3)مستوى الدلالة (α=0.05).

(4) مقياس اتخاذ القرار هو المتغير العشوائي (Z) حيث:

$$Z_{\text{المسحوبة}} = \frac{(\bar{x}-\bar{y})}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$$

(5)حيث أن: حيث أن σ_2^2 =900 ،σ_1^2 =400 ،σ_2^2 n_2=45، n_1=50، $\bar{y}$=340 ، $\bar{x}$. =350

$$Z_{\text{المسحوبة}} = \frac{(350-340)}{\sqrt{\frac{400}{50}+\frac{900}{45}}} = \frac{10}{\sqrt{28}} = \frac{10}{\sqrt{8+20}} = 1.889$$

$Z_{\text{الجدولية}} = z_{0.95} = 1.645$

(6) القرار: حيث أن Z المحسوبة **أكبر من** Z الجدولية فإننا نرفض فرض العدم بأن($\mu_1 = \mu_2$) **ونقبل البديل** ($\mu_1 > \mu_2$) .

مثال(10) في المثال السابق إذا لم يكن معلوما تبايني المجتمعي وكان معلوما فقط أنهما متساويان وتم حساب s_1^2، s_2^2 **من العينتان وكانا على الترتيب** 600، 1350 فاختبر الفرضية القائلة($\mu_1 = \mu_2$) **مقابل الفرضية(** $\mu_1 \neq \mu_2$**)**

الحل: المجتمعين يتبعا التوزيع الطبيعي ومجهولي التباين ومعلوما فقط $\sigma_1^2 = \sigma_2^2$

(1) صياغة الفرضيات:

$H_0: \mu_1 \neq \mu_2$ مقابل $H_{a:} (\mu_1 = \mu_2)$

(2)مقياس العينة$(\bar{X} - \bar{Y})$ حيث: $\bar{Y}$=340، $\bar{X}$=350

(3)مستوى الدلالة (α=0.05).

(4) مقياس اتخاذ القرار هو المتغير العشوائي (Z) حيث:

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

وكذلك:

$$s_c^2=\frac{(n_1-1)s_1^2+(n_1-1)s_2^2}{n_1+n_2-2}$$

5)حيث أن:

$n_1=50$ ، $n_2=45$،

$\bar{x}.=350$ ، $\bar{y}=340$

$s_2^2.=1350$ ، $s_1^2=600$

وبالتالي فإن:

$$s_c^2=\frac{(49)(600)+(44)(1350)}{50+45-2}=954.838$$

$$s_c^2=30.9$$

$$t_{\text{المحسوبة}}=\frac{(350-340)}{(30.9)\sqrt{\frac{1}{50}+\frac{1}{45}}}=\frac{(10)}{(30.9)(0.207)}=1.563$$

$$t=t[1-(9/2)\,;\,n_1+n_2-2]$$

$=t[0.975;\,93]=1.98$

6)القرار:حيث أن t المحسوبة أكبر من t الجدولية فإننا نقبل فرضية فرض العدم ونقول بأن معدلي المجتمع متساويين.

مثال(11): شركة زراعية ترغب بالمفاضلة بين نوعين من الأسمدة يدعي المورد أن النوع الأول يؤدي إلى زيادة متوسط الإنتاجية عن النوع الثاني بمناقشة ادعاء المورد سحبت عينتين كل منهما تتكون من 40 كيس. عند استخدام النوعين كان معدل الإنتاجية المحصول الذي استخدم السماد الأول (18.2 كلغ) والذي استخدم السماد الثاني (18.8 كلغ) حسب التباين في العينتين$s_1^2=0.64$، $s_2^2=1.44$ كلغ تربيع. اختبر عند مستوى دلالة $\alpha=0.05$. بأن معدل الإنتاجية للمحصول الأول أقل من معدل الإنتاجية للمحصول الثاني.

الحل:

(1) صياغة الفرضيات: $H_0: \mu_1 = \mu_2$ مقابل $H_a: (\mu_1 < \mu_2)$

(2) مقياس العينة $(\bar{X} - \bar{Y})$ حيث: $\bar{Y}$=18.8، $\bar{X}$=12.2

(3) مستوى الدلالة .α=0.0

(4) مقياس اتخاذ القرار هو المتغير العشوائي (t) بدرجات حرية $(v=n_1 + n_2 - 2)$ بحيث

$$t=\frac{(\bar{X}-\bar{Y})}{s_c\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} : t[1-\alpha; n_1+n_2-2]$$

(5) حيث أن:

$n_1=n_2=40$

s_2^2 =1.44 , s_1^2=0.64

وبالتالي فإن:

$$S_c=\frac{(n_1-1)s_1^2+(n_1-1)s_2^2}{n_1+n_2-2}$$

$$S_c=\frac{(39)(0.64)+(39)(1.44)}{540+400-2}=1.04=1.02$$

$$t_{\text{المحسوبة}}=\frac{(18.2-18.8)}{1.02\sqrt{\frac{1}{40}+\frac{1}{40}}}=\frac{-0.6}{(1.02))0.224)} \quad =-2.63$$

$$t_{\text{الجدولية}}=-t[0.95; 78]=-1.685$$

(6) القرار: حيث أن $|t_{\text{المحسوبة}}|$ المحسوبة أكبر من $|t_{\text{الجدولية}}|$ الجدولية فإننا نرفض فرض العدم ونقبل الفرضية البديلة ونقول بأن معدل الإنتاجية الذي استخدم السماد النوع الأول أقل من معدل الإنتاجية للمحصول الذي استخدم فيه السماد من النوع الثاني.

(6-11) اختبار الفرضيات للفرق بين ووسطي مجتمعين $(\mu_1 - \mu_2)$ في حالة العينات غير المستقلة:

في بعض الأحيان تهدف الدراسة إلى المقارنة بين وسط العينة قبل إجراء تجربة معينة بوسط نفس العينة بعد إجراء التجربة يمك اختبار الفرق بين معدلي المجتمعين بواسطة وسطي العينتين. ونلاحظ هنا أن بيانات العينة قبل إجراء التجربة ستكون مسحوبة من مجتمع مختلف تماما عنها بعد إجراء التجربة فكأنهما عينتان مختلفتان محسوبتان من مجتمعين مختلفين.

فإذا رمز لمفردات العينة الأولى (العينة قبل التجربة) بالرموز

$$X_1, X_2, \ldots\ldots X_n$$

ولمفردات العينة الثانية (العينة بعد التجربة)

$$Y_1, Y_2, \ldots\ldots Y_n$$

فأزواج المفردات التي تتم المقارنة بينهما.. أو بمعنى آخر مفروق الأزواج التي ستستخدم لاختبار الفرق بين μ_1 و μ_2 هي:

$$d_1 = x_1 - y_1, d_2 = x_2 - y_2, \ldots\ldots, d_n = x_n - y_n$$

وعندئذ فإن:

1-الوسط الحسابي للفروق:

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

2-التباين للفروق:

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

وإذا افترضنا أن مجموع الفروق $d_1, d_2, \ldots\ldots d_n$ تكون عينة عشوائية من مجتمع طبيعي بمعدل ($\mu_1 - \mu_2) = \mu_d$

وبالتالي فإن مقياس اتخاذ القرار هو المتغير العشوائي (t) حيث

$$t = \frac{\mu_d - \bar{d}}{s_{d/\sqrt{n}}}$$

ويخضع التوزيع(t) بدرجات حرية (n-1)

مثال(12): عشر أشخاص يرغبون في تخفيض أوزانهم.اشتركوا في نظام غذائي للمساعدة في تخفيض الوزن. كانت أوزانهم قبل وبعد الاشتراك كالآتي:

رقم الشخص	1	2	3	4	5	6	7	8	9	10
قبل الإشتراك	62	53	39	37	92	53	68	67	47	82
بعد الاشتراك	60	50	40	35	90	55	70	65	45	80

هل أدى الاشتراك في النظام الغذائي إلى تخفيض الوزن؟

استخدم مستوى الدلالة .(α=0.05).

الحل: (1) صياغة الفرضيات: $H_0: \mu_d = 0$ مقابل $H_a: (\mu_1 < 0)$ (اختبار فرضية ذات ذيل أسفل).

(2) مقياس العينة ($\bar{d}$) حيث

$$\bar{d}=\frac{\sum d_i}{n}=\frac{10}{10}=1$$

(3)مستوى الدلالة: (α=0.05)

(4)مقياس اتخاذ القرار هو المتغير العشوائي(t) حيث:

$$t=\frac{\mu_d-\bar{d}}{s_{d/\sqrt{n}}};\ t[1-\alpha;n-1]$$

(5)من البيانات:

$$n=10, \bar{d}=1, s_d=\frac{\sum(d_i-\bar{d})^2}{n-1}=1.8856$$

المحسوبة $t=\frac{1}{1.8856/\sqrt{10}}=1.667$

الجدولية $t=-t[1-\alpha;n-1]$

$=-t[0.95;9]=-1.833$

(6)القرار: حيث أن t المحسوبة أكبر من tالجدولية فإننا نقبل H_0 ونقول بأن النظام الغذائي لم يؤدي إلى تخفيض متوسط الوزن.

تمارين مختارة

س1: أخذت عينة من 81 طالبا من إحدى المدارس فوجد أن متوسط أطوالهم 16 سم بانحراف معياري 6 سم.اختبر الفرض القائل

$H_0{:}\mu=165$ مقابل $H_a{:}\mu \neq 165$ وذلك عند مستوى دلالة $\alpha=0.05$.

س2: إذا كان أحد مصانع الأغذية ينتج نوعا معينا من الألبان حيث متوسط وزن العلبة حوالي 500 غم وذلك بانحراف معياري 36 غم حيث أن أوزان العبوات تتبع التوزيع الطبيعي.ولمراقبة التزام المصنع بمعايير الجودة تم اختيار عينة من 20 عبوة فوجد أن متوسط الأوزان يساوي 4800 غم. هل ترى أن هناك عيبا في الإنتاج أدى إلى انخفاض متوسط الأوزان للعبوات. استعمل مستوى دلالة $\alpha=0.05$.

س3: إذا كان متوسط ربح شركة البوتاس العربية 500 دينار في العام الماضي. تم اختيار عينة من 20 مساهم عن توقعاتهم عن متوسط الربح للعام الحالي فوجد أنه 480 دينارا بانحراف معياري 20. هل توافق المساهمين عن توقعاتهم بانخفاض ربح السهم هذا العام وذلك بمستوى دلالة ($\alpha=0.01$).

س4: أخذت عينة مكونة من 150 طالبا من جامعة خاصة فوجد أن متوسط ذكائهم يساوي 115 بتباين ممقداره360، كذلك تم اختيار عينة من الجامعة الأردنية فوجد أن متوسط ذكائهم يساوي 125 بتباين 225. اختبر الفرض القائل بأن متوسط الذكاء لطلبة الجامعة الأردنية أعلى من الجامعة الخاصة وذلك بمستوى دلالة ($\alpha=0.01$).

س5: قامت إحدى شركات الكمبيوتر باستيراد شحنة من أجهزة الكمبيوتر وقد تعهدت الشركة المصدرة للأجهزة بأن نسبة الأجهزة المعيبة تزيد عن 6**%** تم اختيار عينة عشوائية من 200 جهاز فوجد من بينها 30 جهاز معيب، عند مستوى دلالة $\alpha=0.05$.

س6: مجموعتان A,B تتكون كل مجموعة من 200 شخص مصابين بمرض معين. أراد باحث اختبار مصل ضد هذا المرض قام بإعطاء هذا المصل للمجموعة A، بينما استمر بإعطاء المجموعة B العلاج المعتاد. وبعد فترة من إعطاء هذا المصل وجد أن 150 شخص شفي من المجموعة A بينما شفي 120 شخص من المجموعة B**.** اختر الفرض القائل بأن المصل قد ساعد في العلاج أكثر من العلاج المعتاد. استعمل مستوى دلالة $\alpha=0.05$**.**

س7: في دراسة لمتوسط أعمار المتقدمين لشغل وظيفة معينة. أخذت عينة مكونة من 60 شخص فوجد أن متوسط أعمارهم 24 سنة بانحراف معياري 4 سنوات. هل يمكن القول أن متوسط أعمار المتقدمين يساوي 27 سنة وذلك عند مستوى دلالة $\alpha=0.25$**.**

س8: في دراسة على نسبة غياب العاملين في قطاعين من قطاعات إحدى المؤسسات وجد أن نسبة الغياب في القطاع A هي 30**%** في حين أن نسبة الغياب في القطاع B هي 20**%**. أخذت عينة عشوائية من كل قطاع حجمها 1000 عامل فوجد أن عدد الغائبين في القطاع A يساوي 350 عامل بينما عدد الغائبين في القطاع B يساوي 300. هل يمكن القول أن نسبة الغياب الحقيقية في القطاعين مختلفة كما أشارت البيانات وذاك عند مستولى دلالة $\alpha=0.25$**.**

س9: في دراسة للمقارنة بين مجموعتين من الدارسين لتحديد مدى قدراتهم للتحصيل في الرياضيات. أخذت عينة حجمها 200 دارس منكل مجموعة فكانت النتائج على النحو التالي:

	العينة B	العينة A
متوسط القدرة	120	110
تباين العينة	64	81

هل يمكن القول أن متوسط التحصيل في المجموعتين مختلف وذلك عند مستوى دلالة $\alpha=0.25$**.**

س10: أردنا إجراء دراسة عن مدى فاعلية برنامجين لتدريب العاملين. أخذت عينة حجمها 100 عامل من المتدربين في البرنامج الأول فوجد أن متوسط درجاتهم 20 بانحراف معياري 5 درجات. كذلك أخذت عينة حجمها 80 عامل من المتدربين في البرنامج الثاني فوجد أن متوسط درجاتهم 15 درجة بانحراف معياري 3 درجات. هل يمكن القول بأن البرنامجين متكافئين عند مستوى دلالة $\alpha=0.25$**.**

س11: عينة مكونة من 44 طالب تم اختيارهم بشكل عشوائي من مدرسة ومن ثم تم تقسيمهم إلى 22 مجموعة، كل مجموعة لها نفس معدل الذكاء. وقد تم اختيار عشوائيا عضو واحد من كل زوج لتلقي تدريب خاص. ثم أخضعوا جميعهم لاختبار الذكاء .

ملخص نتائج اختبار كما في الجدول:

رقم المجموعة	مدرب	غير مدرب	رقم المجموعة	مدرب	غير مدرب
1	90	95	**12**	83	85
2	90	95	**13**	83	87
3	73	76	**14**	83	85
4	90	92	**15**	82	85
5	90	91	**16**	65	68
6	53	53	**17**	79	81
7	68	67	**18**	83	84
8	90	88	**19**	60	71
9	78	75	**20**	47	46
10	89	85	**21**	77	75
11	95	9083	**22**	83	80

هل هذه النتائج تقدم دليلا على أن التدريب الخاص ساعد في أداء الطلبة؟

استخدم مستوى دلالة α=0.05 باعتبار أن الاختلافات تخضع تقريبا لتوزيع طبيعي.

س12: يدعي الرئيس التنفيذي لشركة الطاقة الكهربائية لتزويد الطاقة أن 80 في المائة من الزبائن الذين عددهم مليون راضون جدا عن الخدمة التي يحصلون عليها لاختبار هذا الادعاء تم اختيار عينة مكونة من 100 عميل وأجريت عليهم الدراسة وبينت أن 73 في المائة يقولون أنهم راضون جدا. وبناءا على النتائج هل يمكن أن نرفض فرضية الرئيس التنفيذي بأن 80% من الزبائن راضون جدا؟

س13: لنفترض في المثال السابق بشكل مختلف قليلا. لنفترض أن يدعي الرئيس التنفيذي لشركة الطاقة الكهربائية لتزويد الطاقة لا يقل عن 80 في المئة من عملاء الشركة1,000,000 راضون جدا. لاختبار هذا الادعاء تم اختيار عينة مكونة من

100 عميل وأجريت عليهم الدراسة وبينت أن 73 في المئة يقولون أنهم راضون جدا. وبناءا على هذه النتائج هل يمكن أن نرفض فرضية الرئيس التنفيذي بأن 80% من الزبائن راضون جدا؟ استخدام مستوى معنوية $\alpha=0.05$.

قائمة المراجع

1- عزام صبري (2010)، الاحصاء الرياضي، دار الصفاء للنشر والتوزيع، الطبعة الاولى، عمان، الاردن.

2- عاروري فتحي (2009)، المعاينة الاحصائية طرقها استخداماتها، الأكاديميون للنشر والتوزيع، الطيعة الاولى، عمان، الاردن.

3- موساوي عبد النور، بركان يوسف (2010)، الاحصاء 2، دار العلوم للنشر والتوزيع، عنابة، الجزائر.

4- ثائر فيصل شاهر (2013) ، اختبار الفرضيات الإحصائية ، دار و مكتبة الحامد للنشر و التوزيع، الطيعة الاولى، عمان، الأردن.

5- عبد الحفيظ محمد فوزي مصطفى (2000) ، نظرية التقدير، مجموعة النيل العربية، الطبعة الاولى، القاهرة، مصر.

6- البلداوي عبد الحميد عبد المجيد (2009) ، أساليب الاحصاء للعلوم الاقتصادية و ادارة الاعمال مع استخدام برنامج SPSS، دار وائل للنشر و التوزيع ، الطبعة الأولى ، عمان ، الأردن.

7- معتوق أمحمد (2007) ، الإحصاء الرياضي و النماذج الإحصائية، ديوان المطبوعات الجامعية، الجزائر.

8- إبراهيم أبو عقيل (2012)، مبادئ في الإحصاء، دار أسامة للنشر و التوزيع ، الطبعة الأولى ، عمان ، الأردن.

9- علاء محمد القاضي، بكر عمر حمدان (2011) ، الرياضيات و الإحصاء، دار الاعصار العلمي للنشر و التوزيع، الطبعة الاولى، عمان ، الاردن.

10- حسن ياسين طعمة (2015)، الاختبارات الإحصائية أسس و تطبيقات، دار صفاء للنشر و التوزيع، الطبعة الثانية، عمان ، الأردن.

11- جبار عبد مضحي (2015) ، مقدمة في الإحصاء الرياضي، دار الميسرة للنشر و التوزيع و الطباعة ، الطبعة الأولى، عمان ، الأردن.

12- محمد حسين محمد رشيد القادري، منى عطا الله الشويلات (2012)، مبادئ الإحصاء و الاحتمالات و معالجتها باستخدام برنامج SPSS ، دار صفاء للنشر و التوزيع ، الطبعة الأولى ، عمان ، الأردن.

13- زروخي صباح (2018)، محاضرات في مادة الإحصاء الاستدلالي (إحصاء 3)، كلية العلوم الاقتصادية التجارية و علوم التسير ، جامعة محمد بوضياف المسيلة.

14- بوعبد الله صالح (2006)، محاضرات الإحصاء الرياضي، كلية العلوم الاقتصادية التجارية و علوم التسير.

15- Vincent Rivoirard (2022), Statistique Mathématique, Département MIDO, Université Paris Dauphine –PSL

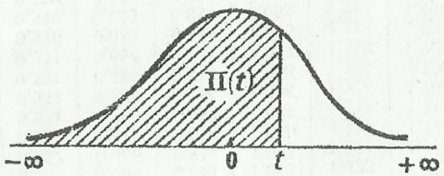
16- Jean-Pierre Lecoutre (2016), Cours et exercices corrigés, 6e édition, Dunod.

17- Thérèse Phan & Jean-Pierre Rowenczyk (2012), Exercices et problèmes de statistique et probabilités, 2e édition, Dunod.
18- Jean-Pierre Lecoutre (2019), Statistique et Probabilités, 7e édition, Dunod.
19- Chamoun Chamoun (2010), Elément de statistiques et de probabilités, office des publications universitaires 1ére Edition, Alger
20- Chibat Ahmed (2000), Notions sur le calcul des probabilités, Fascicule 2, imprimerie Top Colors, Constantine.

III. EXTRAITS DE LA TABLE DE LA FONCTION INTÉGRALE DE LA LOI NORMALE CENTRÉE, RÉDUITE $\mathcal{N}(0, 1)$

$$\Pi(t) = \Pr\{ T < t \}.$$

I



t	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,500 0	0,504 0	0,508 0	0,512 0	0,516 0	0,519 9	0,523 9	0,527 9	0,531 9	0,535 9
0,1	0,539 8	0,543 8	0,547 8	0,551 7	0,555 7	0,559 6	0,563 6	0,567 5	0,571 4	0,575 3
0,2	0,579 3	0,583 2	0,587 1	0,591 0	0,594 8	0,598 7	0,602 6	0,606 4	0,610 3	0,614 1
0,3	0,617 9	0,621 7	0,625 5	0,629 3	0,633 1	0,636 8	0,640 6	0,644 3	0,648 0	0,651 7
0,4	0,655 4	0,659 1	0,662 8	0,666 4	0,670 0	0,673 6	0,677 2	0,680 8	0,684 4	0,687 9
0,5	0,691 5	0,695 0	0,698 5	0,701 9	0,705 4	0,708 8	0,712 3	0,715 7	0,719 0	0,722 4
0,6	0,725 7	0,729 0	0,732 4	0,735 7	0,738 9	0,742 2	0,745 4	0,748 6	0,751 7	0,754 9
0,7	0,758 0	0,761 1	0,764 2	0,767 3	0,770 4	0,773 4	0,776 4	0,779 4	0,782 3	0,785 2
0,8	0,788 1	0,791 0	0,793 9	0,796 7	0,799 5	0,802 3	0,805 1	0,807 8	0,810 6	0,813 3
0,9	0,815 9	0,818 6	0,821 2	0,823 8	0,825 4	0,828 9	0,831 5	0,834 0	0,836 5	0,838 9
1,0	0,841 3	0,843 8	0,846 1	0,848 5	0,850 8	0,853 1	0,855 4	0,857 7	0,859 9	0,862 1
1,1	0,864 3	0,866 5	0,868 6	0,870 8	0,872 9	0,874 9	0,877 0	0,879 0	0,881 0	0,883 0
1,2	0,884 9	0,886 9	0,888 8	0,890 7	0,892 5	0,894 4	0,896 2	0,898 0	0,899 7	0,901 5
1,3	0,903 2	0,904 9	0,906 6	0,908 2	0,909 9	0,911 5	0,913 1	0,914 7	0,916 2	0,917 7
1,4	0,919 2	0,920 7	0,922 2	0,923 6	0,925 1	0,926 5	0,927 9	0,929 2	0,930 6	0,931 9
1,5	0,933 2	0,934 5	0,935 7	0,937 0	0,938 2	0,939 4	0,940 6	0,941 8	0,942 9	0,944 1
1,6	0,945 2	0,946 3	0,947 4	0,948 4	0,949 5	0,950 5	0,951 5	0,952 5	0,953 5	0,954 5
1,7	0,955 4	0,956 4	0,957 3	0,958 2	0,959 1	0,959 9	0,960 8	0,961 6	0,962 5	0,963 3
1,8	0,964 1	0,964 9	0,965 6	0,966 4	0,967 1	0,967 8	0,968 6	0,969 3	0,969 9	0,970 6
1,9	0,971 3	0,971 9	0,972 6	0,973 2	0,973 8	0,974 4	0,975 0	0,975 6	0,976 1	0,976 7
2,0	0,977 2	0,977 9	0,978 3	0,978 8	0,979 3	0,979 8	0,980 3	0,980 8	0,981 2	0,981 7
2,1	0,982 1	0,982 6	0,983 0	0,983 4	0,983 8	0,984 2	0,984 6	0,985 0	0,985 4	0,985 7
2,2	0,986 1	0,986 4	0,986 8	0,987 1	0,987 5	0,987 8	0,988 1	0,988 4	0,988 7	0,989 0
2,3	0,989 3	0,989 6	0,989 8	0,990 1	0,990 4	0,990 6	0,990 9	0,991 1	0,991 3	0,991 6
2,4	0,991 8	0,992 0	0,992 2	0,992 5	0,992 7	0,992 9	0,993 1	0,993 2	0,993 4	0,993 6
2,5	0,993 8	0,994 0	0,994 1	0,994 3	0,994 5	0,994 6	0,994 8	0,994 6	0,995 4	0,995 2
2,6	0,995 3	0,995 5	0,995 6	0,995 7	0,995 9	0,996 0	0,996 1	0,996 2	0,996 3	0,996 4
2,7	0,996 5	0,996 6	0,996 7	0,996 8	0,996 9	0,997 0	0,997 1	0,997 2	0,997 3	0,997 4
2,8	0,997 4	0,997 5	0,997 6	0,997 7	0,997 7	0,997 8	0,997 9	0,997 9	0,998 0	0,998 1
2,9	0,998 1	0,998 2	0,998 2	0,998 3	0,998 4	0,998 4	0,998 5	0,998 5	0,998 6	0,998 6

TABLE POUR LES GRANDES VALEURS DE t

t	3,0	3,1	3,2	3,3	3,4	3,5	3,6	3,8	4,0	4,5
$\Pi(t)$	0,998 65	0,999 04	0,999 31	0,999 52	0,999 66	0,999 76	0,999 841	0,999 928	0,999 968	0,999 997

Nota. — La table donne les valeurs de $\Pi(t)$ pour t positif. Lorsque t est négatif il faut prendre le complément à l'unité de la valeur lue dans la table.

Exemple : pour $t = 1,37$ $\quad \Pi(t = 1,37) = 0,914\,7$
pour $t = -1,37$ $\quad \Pi(t = -1,37) = 1 - 0,9147 = 0,085\,3$

IV. TABLE DE LA LOI NORMALE, CENTRÉE, RÉDUITE $\mathcal{N}(0, 1)$ (DITE TABLE DE L'ÉCART RÉDUIT)

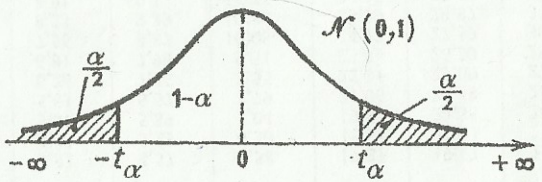
La table donne la probabilité α pour que l'écart réduit t égale ou dépasse, en valeur absolue, une valeur donnée t_α, c'est-à-dire la probabilité extérieure à l'intervalle $(-t_\alpha, +t_\alpha)$.

$$\alpha = \Pr\{ t \notin (-t_\alpha, +t_\alpha) \},$$

ou

II

$$1 - \alpha = \Pr\{ -t_\alpha < t < +t_\alpha \}.$$



$$\alpha = 2[1 - \Pi(t_\alpha)]$$

α	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,00	∞	2,576	2,326	2,170	2,064	1,960	1,881	1,812	1,751	1,695
0,10	1,645	1,598	1,555	1,514	1,476	1,440	1,405	1,372	1,341	1,311
0,20	1,282	1,254	1,227	1,200	1,175	1,150	1,126	1,103	1,080	1,058
0,30	1,036	1,015	0,994	0,974	0,954	0,935	0,915	0,896	0,878	0,860
0,40	0,842	0,824	0,806	0,789	0,772	0,755	0,739	0,722	0,706	0,690
0,50	0,674	0,659	0,643	0,628	0,613	0,598	0,583	0,568	0,553	0,539
0,60	0,524	0,510	0,496	0,482	0,468	0,454	0,440	0,426	0,412	0,399
0,70	0,385	0,372	0,358	0,345	0,332	0,319	0,305	0,292	0,279	0,266
0,80	0,253	0,240	0,228	0,215	0,202	0,189	0,176	0,164	0,151	0,138
0,90	0,126	0,113	0,100	0,088	0,075	0,063	0,050	0,038	0,025	0,013

La probabilité α s'obtient par addition des nombres inscrits en marge.

Ex : Pour $t_\alpha = 1,96$ la probabilité est $\alpha = 0,00 + 0,05 = 0,05$.

TABLE POUR LES PETITES VALEURS DE α.

α	0,001	0,000 1	0,000 01	0,000 001	0,000 000 1	0,000 000 01	0,000 000 001
t_α	3,290 53	3,890 59	4,417 17	4,891 64	5,326 72	5,730 73	6,109 41

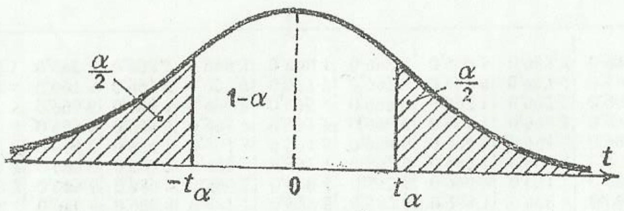
VII. TABLE DE DISTRIBUTION DE t

(Loi de Student-Fisher)

La table donne, en fonction du nombre de degrés de liberté v, la probabilité α pour que t égale ou dépasse, en valeur absolue, une valeur donnée t_α

$$1 - \alpha = \Pr\{ -t_\alpha < t < +t_\alpha \}.$$

III



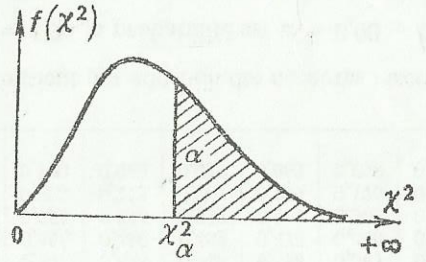
v \ α	0,90	0,80	0,70	0,60	0,50	0,40	0,30	0,20	0,10	0,05	0,02	0,01
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841
4	0,134	0,271	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	0,259	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	2,957
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
> 30	0,125 66	0,253 35	0,385 32	0,524 40	0,674 49	0,841 62	1,036 43	1,281 55	1,644 85	1,959 96	2,326 34	2,575 82

VIII. TABLE DE DISTRIBUTION DE χ^2

(Loi de K. Pearson)

La table donne la probabilité α, en fonction du nombre de degrés de liberté v, pour que χ^2 égale ou dépasse une valeur donnée χ^2_α.

$$\alpha = \Pr\{ \chi^2 \geqslant \chi^2_\alpha \}.$$



v \ α	0,990	0,975	0,950	0,900	0,100	0,050	0,025	0,010
1	0,031 6	0,039 8	0,023 9	0,015 8	2,71	3,84	5,02	6,63
2	0,02	0,05	0,10	0,21	4,60	5,99	7,38	9,21
3	0,12	0,22	0,35	0,58	6,25	7,81	9,35	11,24
4	0,30	0,48	0,71	1,06	7,78	9,49	11,1	13,28
5	0,55	0,83	1,15	1,61	9,24	11,07	12,8	15,09
6	0,87	1,24	1,64	2,20	10,64	12,59	14,0	16,81
7	1,24	1,69	2,17	2,83	12,02	14,07	16,0	18,47
8	1,65	2,18	2,73	3,49	13,36	15,51	17,5	20,09
9	2,09	2,70	3,33	4,17	14,68	16,92	19,0	21,66
10	2,56	3,25	3,94	4,86	15,99	18,31	20,5	23,21
11	3,05	3,82	4,57	5,58	17,27	19,67	21,9	24,72
12	3,57	4,40	5,23	6,30	18,55	21,03	23,3	26,22
13	4,11	5,01	5,89	7,04	19,81	22,36	24,7	27,69
14	4,66	5,63	6,57	7,79	21,06	23,68	26,1	29,14
15	5,23	6,26	7,26	8,55	22,31	25,00	27,5	30,58
16	5,81	6,81	7,96	9,31	23,54	26,30	28,8	32,00
17	6,41	7,56	8,87	10,08	24,77	27,59	30,2	33,41
18	7,01	8,23	9,39	10,86	25,99	28,87	31,3	34,80
19	7,63	8,91	10,1	11,65	27,20	30,14	32,9	36,19
20	8,26	9,59	10,9	12,44	28,41	31,41	34,2	37,57
21	8,90	10,3	11,6	13,24	29,61	32,67	35,5	38.93
22	9,54	11,0	12,3	14,04	30,81	33,92	36,8	40,29
23	10,2	11,7	13,1	14,85	32,01	35,17	38,1	41,64
24	10,9	12,4	13,8	15,66	33,20	36,41	39,4	42,98
25	11,5	13,1	14,6	16,47	34,38	37,65	40,6	44,31
26	12,2	13,8	15,4	17,29	35,56	38,88	41,9	45,64
27	12,9	14,6	16,2	18,11	36,74	40,11	43,2	46,96
28	13,6	15,3	16,9	18,94	37,92	41,34	44,5	48,28
29	14,3	16,0	17,7	19,77	39,09	42,56	45,7	49,59
30	15,0	16,8	18,5	20,60	40,26	43,77	47,0	50,89

Quand v est supérieur à 30, on utilise la table de la loi normale (table de l'écart réduit) avec

$$t = \sqrt{2\chi^2} - \sqrt{2v - 1}.$$