الفصل الحادي عشر

الاستدلال الإحصائي

(١١-١) مقدمة :

عند استخدام المصادر الميدانية للبيانات فإن الباحث يحاول أن يستخدم أسلوب الحصر الشامل والذي يقوم بدراسة جميع مفردات المجتمع الإحصائي محل الدراسة ، ولكن قد لا يتمكن الباحث من استخدام أسلوب الحصر الشامل لعدة عوامل أهمها عوامل الوقت والجهد والتكلفة ، بالإضافة إلى طبيعة المجتمع والذي قد يكون غير محدود ، وطبيعة الدراسة ذاتها والتي قد تؤثر على وحدات المعاينة (راجع الفصل الأول) ، لذا فإن الباحث يستخدم أسلوب العينات والذي يعتمد على دراسة جزء أو عينة من المجتمع محل الدراسة (وسوف نفترض أن العينة عشوائية) فالباحث إذن يهدف إلى دراسة المجتمع كله ، ولكنه لا يتمكن من ذلك ، والوسيلة أو الأداة المتاحة أمامه هي عينة عشوائية مسحوبة من هذا المجتمع بحيث تمثل المجتمع المدروس بشكل جيد كما سبق إيضاحه في الفصل الأول . لذلك فإن المقاييس الإحصائية للمجتمع - والتي تسمى معلمات - تكون مجهولة مثل متوسط المجتمع الإحصائي (μ) وانحرافه المعياري (σ) . بينما تكون المقاييس الإحصائية للعينة - والتي تسمى إحصائيات أو توابع إحصائية - معلومة حيث يستطيع الباحث حسابها من بيانات العينة مثل متوسط العينة ($\overline{X}$) والانحراف المعياري للعينة (S) .

والسؤال هو كيف نستنتج أو نستدل على مقاييس المجتمع والتي تكون مجهولة من مقاييس العينة المعروفة ؟ أو كيف نعمم نتائج العينة على المجتمع ؟

واستنتاج معلمات المجتمع من التوابع الإحصائية للعينة، أي الاستدلال على معلمات المجتمع هو الذي يسمى "الاستدلال الإحصائي" والاستدلال الإحصائي له مجالان وهما :

المجال الأول : التقدير .

المجال الثاني : اختبارات الفروض .

وفي المجال الأول : يتم تقدير معلمات المجتمع المجهولة بالاعتماد على إحصائيات العينة .

وفي المجال الثاني : يتم اختبار فرض ما عن بعض معلمات المجتمع (أو توزيعه) باستخدام إحصائيات العينة .

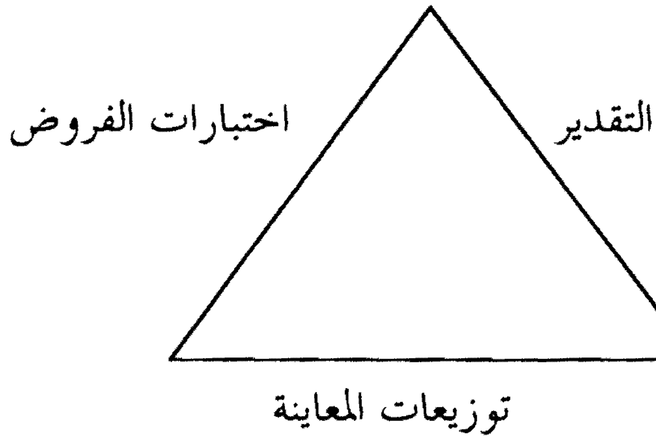
وفي الاستدلال على معلمات المجتمع عن طريق التقدير يتم حساب مؤشرات إحصائية من بيانات العينة باستخدام دالة أو صيغة تسمى إحصائية . وهذه الإحصائية إما أن تستخدم في تقدير المعلمة المجهولة مباشرة ، وهذا ما يعرف بتقدير النقطة ، أو تستخدم في إنشاء مدى أو فترة لها حد أدنى وحد أعلى يرجح الباحث وقوع المعلمة بينهما بدرجة ثقة معينة ، وهذا ما يعرف بتقدير الفترة .

وفي الاستدلال عن طريق اختبارات الفروض يتم حساب ما يسمى "إحصائية الاختبار" لاختبار مدى صحة فرض معين عن بعض معلمات المجتمع (أو توزيع المجتمع ذاته) ليصل الباحث إلى قرار إحصائي إما برفض الفرض المقترح أو عدم رفضه .

وفي الحالتين (التقدير واختبارات الفروض) نحتاج إلى ما يسمى "بتوزيعات المعاينة" أي توزيعات الإحصائيات المستخدمة في عملية الاستدلال .

ويمكن تمثيل الاستدلال الإحصائي بمثلث : قاعدته هي توزيعات المعاينة وضلعاه هما : التقدير واختبارات الفروض .

الاستدلال الإحصائي



والجدول التالي يوضح أهم المعلمات (مقاييس المجتمع) والإحصائيات (مقاييس العينة) التي سوف نتطرق لها - إن شاء الله - خلال هذا الفصل .

المؤشر الإحصائي	المعلمة	الإحصائية
المتوسط	μ	$\bar{X}$
الانحراف المعياري	σ	S
النسبة	R	r
الفرق بين متوسطين	$\mu_1 - \mu_2$	$\bar{X}_1 - \bar{X}_2$
الفرق بين نسبتين	$R_1 - R_2$	$r_1 - r_2$

(١١-١-١) المجتمع الإحصائي :

يتكون المجتمع الإحصائي من جميع المشاهدات الممكنة (ولو نظرياً) للظاهرة أو المشكلة محل البحث مثل ظاهرة الطول ، وظاهرة الوزن للإنسان و ظاهرة الدخل الشهري للأفراد في مدينة الرياض و ظاهرة المبيعات اليومية لإحدى الشركات التجارية والجانب المهم في المجتمع الإحصائي المدروس أن يكون قابلاً للتعريف أو الحصر بطريقة واضحة سواء بشكل فعلي (مثل طلاب إحدى الجامعات) أو بشكل نظري (مثل عدد الأسماك في أحد المحيطات) وتكمن أهمية التمييز بين المجتمعات المنتهية وغير المنتهية في معرفة خصائص توزيعات المعاينة للعينات المسحوبة من المجتمع الإحصائي والتي سنتكلم عنها خلال هذا الفصل .

(١١-١-٢) العينة الإحصائية :

هي جزء من المشاهدات الممكنة لظاهرة معينة مثل أطوال مجموعة من الطلاب في إحدى المدارس الثانوية أو دخول مجموعة من الموظفين في أحد القطاعات الحكومية .

(١١-١-٣) المجتمع المنتهي :

ويتكون من عدد محدود أو منتهٍ من المشاهدات مثل عدد طلاب إحدى الجامعات في عام دراسي معين .

(١١-١-٤) المجتمع غير المنتهي :

وهو يتكون من عدد غير محدود من المشاهدات مثل عدد الأسماك في المحيط الهندي ، أو كمية الاحتياطي النفطي في إحدى الدول الخليجية .

(١١-١-٥) المعاينة العشوائية :

عدد العينات المختلفة ذات الحجم n المسحوبة من مجتمع منتهي ذي حجم N عندما يكون السحب بدون إرجاع هو$\binom{N}{n}$ عينة .ومثال على ذلك يوجد $\binom{12}{2}$ عينة أي 66 عينة مختلفة ذات حجم 2 من مجتمع يتكون من 12 عنصراً ويكون احتمال سحب كل عينة متماثلاً ويساوي $\frac{1}{66}$. أما إذا كان السحب بإرجاع فإن عدد العينات المختلفة هو N^n ، فيكون عدد العينات المسحوبة بإرجاع في مثالنا السابق هو 122 أي 144 عينة مختلفة ذات حجم 2 من العينات المختلفة التي لها الاحتمال نفسه - وهو $\frac{1}{144}$ - في الاختيار .

مثال (١) :

أوجد جميع العينات الممكنة التي حجمها 3 والمسحوبة بدون إرجاع من المجتمع الإحصائي$\{ a , b , c , d , e \}$ثم بين كيف يمكن اختيار عينة عشوائية من هذه العينات.

الحل :

عدد العينات الممكنة هو $\binom{5}{3}$ ويساوي 10 وهذه العينات هي :

abc , abd , abe , acd , ace , ade , bcd , bce , bde , cde

فإذا اخترنا واحدة من هذه العينات بحيث يكون احتمال اختيار أي منها يساوي $\frac{1}{10}$ فإننا نسمي هذه العينة (عينة عشوائية) .

ولاختيار عينة بطريقة عشوائية هناك عدة طرق نذكر منها :

١) نكتب كل العينات الممكنة والتي عددها $\binom{N}{n}$ إذا كان السحب بدون إرجاع على بطاقات ، ثم نخلطها جيداً ثم نسحب منها واحدة بدون النظر إليها. ولكن هذه الطريقة غير مناسبة عندما تكون كل من N،n كبيرتين .

٢) تكتب جميع المشاهدات التي عددها N على بطاقات ، ثم نختار عينة عشوائية حجمها n بحيث يكون احتمال اختبار أي بطاقة من البطاقات متماثلاً.

٣) استخدام جدول الأرقام العشوائية لاختيار العينة العشوائية .

وقد سبق بحث هذه الطرق بالتفصيل في الفصل الأول .

وسوف ندرس فيما يلي بعض توزيعات المعاينة وهي تمثل القاعدة التي نتطلق منها لدراسة نظرية التقدير ونظرية اختبارات الفروض الإحصائية .

(١١-٢) توزيعات المعاينة :

عند اختيار عينة وحساب إحصائية (تابع إحصائي) لها وليكن الوسط الحسابي $\overline{X}$ فإن قيمة الوسط الحسابي سوف تختلف عند سحب عينة ثانية ، وثالثة ، ... أي سوف تختلف قيمته من عينة إلى أخرى . والسؤال هو : ماذا يكون الوضع لو تم سحب كل العينات الممكنة ؟ وما هو التوزيع الاحتمالي الذي تأخذه الإحصائية (التابع الإحصائي) المحسوبة لكل هذه العينات والذي يسمى (توزيع المعاينة للإحصائية) ؟ وفيما يلي نحاول الإجابة على هذا السؤال حيث نتناول بعض توزيعات المعاينة المهمة من حيث تعريفها وشكلها

والعلاقات التي تربط بين الوسط الحسابي والتباين (الانحراف المعياري) لهذه التوزيعات وما يناظرها من معلمات المجتمع المسحوبة منه العينات .

(١١-٢-١) تعريف :

توزيع المعاينة للإحصائية (التابع الإحصائي) هو التوزيع الاحتمالي لهذه الإحصائية المحسوب لكل العينات الممكنة والتي حجمها n والمأخوذة من المجتمع الإحصائي المدروس أياً كان حجمه وأياً كانت طريقة السحب .

(١١-٢-٢) توزيع المعاينة للوسط الحسابي ($\bar{X}$) :

هو التوزيع الاحتمالي للوسط الحسابي $\bar{X}$ المحسوب لكل العينات الممكنة والتي حجمها n والمأخوذة من المجتمع الإحصائي المدروس أياً كان حجمه وأيا كانت طريقة السحب .

ولتوضيح توزيع المعاينة للمتوسط $\bar{X}$ ندرس المثال التالي :

مثال (٢) :

إذا كان لدينا المجتمع الإحصائي التالي :

1 , 2 , 6

أي أن حجم المجتمع $N = 3$ (المقصود هو الإيضاح بمثال بسيط !!) سحبنا جميع العينات الممكنة والتي حجمها $n = 2$.

لاحظ أن متوسط المجتمع μ هو : $\mu = \frac{1+2+6}{3} = 3$

وتباين المجتمع هو :

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} x_i^2 - \mu^2$$

$$\Rightarrow \sigma^2 = \frac{1+4+36}{3} - 9 = \frac{41}{3} - \frac{27}{3} = \frac{14}{3} = 4.67$$

$$\Rightarrow \sigma = \sqrt{4.67} = 2.16$$

وفيما يلي سنقوم بحساب المتوسط والانحراف المعياري لتوزيع المعاينة للوسط والذي يسمى الخطأ المعياري للتوزيع في حالتي السحب بدون إرجاع والسحب بإرجاع .

أ – السحب بدون إرجاع :

حيث إن السحب بدون إرجاع فإن عدد العينات الممكنة هو :

$$\binom{N}{n} = \binom{3}{2} = \frac{3!}{2!\,1!} = 3$$

وهذه العينات هي : $(1, 2) \quad , \quad (1, 6) \quad , \quad (2, 6)$

ومتوسطات هذه العينات هي : $1.5 \quad , \quad 3.5 \quad , \quad 4$

وإذا كانت المعاينة عشوائية بسيطة فإن كل عينة من العينات السابقة يكون لها احتمال متساوي وقدره $\frac{1}{3}$. وبالتالي يكون توزيع المعاينة للمتوسط $\bar{X}$ بالشكل التالي :

$\bar{x}$	1.5	3.5	4	Σ
$f(\bar{x})$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1
$\bar{x} \,.\, f(\bar{x})$	$\frac{1.5}{3}$	$\frac{3.5}{3}$	$\frac{4}{3}$	$\frac{9}{3} = 3$
$\bar{x}2 \,.\, f(\bar{x})$	$\frac{2.25}{3}$	$\frac{12.2}{3}$	$\frac{16}{3}$	$\frac{30.5}{3} = 10.17$

وبالتالي نجد أن :

$$\mu_{\bar{X}} = E(\bar{x}) = \sum_{i=1}^{3} \bar{x}_i \,.\, f(\bar{x}_i)$$

$$\Rightarrow E(\bar{x}) = (\frac{1}{3} \,.\, 1.5) + (\frac{1}{3} \,.\, 3.5) + (\frac{1}{3} \,.\, 4)$$

$$\Rightarrow E(\bar{x}) = \frac{1}{3}(1.5 + 3.5 + 4) = \frac{1}{3} \,.\, 9 = 3 = \mu$$

$$\sigma_{\bar{x}}^{2} = \sum_{i=1}^{3} \bar{x}_i^{2} \cdot f(\bar{x}_i) - (\mu_{\bar{x}})^2$$

$$\because E(\bar{x}^2) = \frac{1}{3}(2.25 + 12.25 + 16) = \frac{30.5}{3} = 10.17،$$

$$\Rightarrow \sigma_{\bar{x}}^{2} = 10.17 - 3^2 = 10.17 - 9 = 1.17$$

ونستنتج من المثال أن (يمكن إثبات ذلك رياضياً) :

$$E(\bar{x}) = \mu_{\bar{x}} = \mu = 3 \quad , \sigma_{\bar{x}}^{2} = \frac{\sigma^2}{n} \cdot \frac{N - n}{N - 1} \quad , \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N - n}{N - 1}}$$

$$\sigma_{\bar{x}}^{2} = \frac{\sigma^2}{n} \frac{N - n}{N - 1} = \frac{4.67}{2} \cdot \frac{3 - 2}{3 - 1} = 2.335 \times \frac{1}{2} = 1.17$$

ب) السحب بإرجاع :

حيث إن السحب بإرجاع فإن عدد العينات الممكنة هو :

$$N^n = 3^2 = 9$$

وهذه العينات هي :

(1 , 1), (1 , 2), (1 , 6), (2 , 1), (2 , 2), (2 , 6), (6 , 1), (6 , 2), (6 , 6)

ومتوسطات هذه العينات هي على التوالي :

1 , 1.5 , 3.5 , 1.5 , 2 , 4 , 3.5 , 4 , 6

ويكون توزيع المعاينة للمتوسط $\bar{X}$ بالشكل التالي :

$\bar{x}$	1	1.5	2	3.5	4	6	$\sum$
$f(\bar{x})$	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{2}{9}$	$\frac{1}{9}$	1
$\bar{x} \cdot f(\bar{x})$	$\frac{1}{9}$	$\frac{3}{9}$	$\frac{2}{9}$	$\frac{7}{9}$	$\frac{8}{9}$	$\frac{6}{9}$	$\frac{27}{9} = 3$
$\bar{x}2 \cdot f(\bar{x})$	$\frac{1}{9}$	$\frac{4.5}{9}$	$\frac{4}{9}$	$\frac{24.5}{9}$	$\frac{32}{9}$	$\frac{36}{9}$	$\frac{102}{9}$

لاحظ ما يلي :

$$E(\bar{X}) = \mu_{\bar{x}} = \sum_{i=1}^{N} \bar{x} \; . f(\bar{x}_i) = 3 = \mu$$

$$\sigma_{\bar{x}}^2 = \sum_{i=1}^{N} \bar{x}_i^2 . f(\bar{x}_i) - [E(\bar{x})]^2 = \frac{102}{9} - \left(\frac{9}{3}\right)^2 = \frac{21}{9} = 2.33$$

$$\Rightarrow \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} = \frac{4.67}{2} = 2.33$$

وانطلاقاً من المثال السابق نعطي التعريف التالي :

(١١ – ٢ –٣) تعريف توزيع المعاينة للمتوسط $\bar{X}$:

هو التوزيع الاحتمالي للمتوسطات الحسابية $\bar{X}$ المحسوبة لكل العينات الممكنة ذات الحجم n والمأخوذة من مجتمع إحصائي حجمه N وله متوسط μ وتباين σ2 . ومتوسط وتباين توزيع المعاينة للمتوسط $\bar{X}$ يعطى بالصورة التالية:

$$E(\bar{X}) = \mu_{\bar{x}} = \mu \longrightarrow \quad (1)$$

$$\sigma_{\bar{x}}^2 = \begin{cases} \dfrac{\sigma^2}{n} & \longrightarrow \text{كبيرة مهما كانت طريقة السحب } N \\ & \longrightarrow \text{صغيرة والسحب بإرجاع } N \\ \dfrac{\sigma^2}{n}\,\dfrac{N-n}{N-1} & \text{صغيرة والسحب بدون إرجاع } N \end{cases} \longrightarrow \quad (2)$$

ويسمى الانحراف المعياري لتوزيع المعاينة للوسط بالخطأ المعياري للوسط وهو الجذر التربيعي لتباين الوسط .

(١١–٢–٤) ملاحظات مهمة :

(١) السحب بإرجاع (بإحلال) من مجتمع منته يساوي تماماً السحب (المعاينة) بدون إرجاع (بدون إحلال) من مجتمع غير منته .

(٢) المقدار $\sqrt{\dfrac{N-n}{N-1}}$ يسمى (معامل التصحيح) ويقترب من الواحد الصحيح كلما كبرت قيمة N (حجم المجتمع) مقارنة بـــ n (حجم العينة).

(٣) إذا كان المجتمع الإحصائي المسحوبة منه العينات له توزيع طبيعي بمتوسط μ وانحراف معياري σ أي أن :

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

فإن متوسط العينة $\bar{X}$ له أيضاً توزيع طبيعي بمتوسط μ وانحراف معياري $\sigma_{\bar{x}}$ أي أن :

$$\bar{X} \sim N(\mu , \sigma_{\bar{x}})$$

والانحراف (الخطأ) المعياري للوسط $\sigma_{\bar{x}}$ هو الجذر التربيعي لتباين الوسط $\sigma^2_{\bar{x}}$ (انظر العلاقة رقم (٢)) .

(٤) إذا كان المجتمع الإحصائي المسحوبة منه العينات لا يتبع التوزيع الطبيعي (أو توزيعه غير معروف) ، فإن متوسط العينة $\bar{X}$ له أيضاً توزيع طبيعي (تقريباً) بشرط أن يكون حجم العينة n كبيراً إذا كان أكبر أو يساوي ثلاثين $(n \geq 30)$ ، (هذه النتيجة المهمة مستفادة من نظرية النهاية المركزية التي سندرسها في نهاية هذا الفصل) .

مثال (٣) :

إذا كان لدينا المجتمع الإحصائي الطبيعي التالي :

$$1 , 2 , 3 , 4 , 5 , 6$$

وسحبنا جميع العينات الممكنة ذات الحجم 3 ، أوجد توزيع المعاينة للمتوسط $\bar{X}$ مع حساب متوسطه وتباينه (مع العلم أن السحب بإرجاع) .

الحل :

لاحظ أن :

$$N=6 \quad , \quad n=3 \ , \ \mu= \frac{1+2+3+4+5+6}{6}=\frac{21}{6}=3.$$

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} X^2 - \mu^2 = \frac{91}{6} - (3.5)^2 = 2.92$$

وباستخدام تعريف توزيع المعاينة نجد أن $\overline{X}$ يتبع توزيع طبيعي ، بمتوسط $\mu_{\bar{x}}$ وتباين $\sigma_{\bar{x}}^2$ يعطى بالصورة التالية :

$$\mu_{\bar{x}} = \mu = 3.5 \qquad , \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$$

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} = \frac{2.92}{3} = 0.97 \quad \Rightarrow \quad \sigma_{\bar{x}} = \sqrt{0.97} = 0.99$$

أي أن :

$$\overline{X} \sim N(3.5 , 0.99)$$

أي أن الوسط الحسابي للعينة $\overline{X}$ له توزيع طبيعي بوسط حسابي يساوي 3.5 وانحراف معياري يساوي 0.99 .

(١١– ٢–٥) ملاحظة :

يتضح من المثال السابق أهمية وفائدة توزيعات المعاينة حيث استطعنا الحصول على الوسط الحسابي والتباين والانحراف المعياري لتوزيع المعاينة بسهولة وسرعة بتطبيق تعريف توزيع المعاينة ، ولو أردنا معرفة ذلك بالطريقة التي استخدمناها في مثال رقم (٢) لتطلب ذلك جهداً كبيراً ووقتاً طويلاً .

مثال (٤) :

إذا كانت درجات الطلاب في الجامعة تتبع التوزيع الطبيعي بمتوسط 70 درجة وانحراف معياري 6 درجات أي أن $X \sim N(70 , 6)$

أوجد توزيع المعاينة للمتوسط $\overline{X}$ وذلك لجميع العينات المسحوبة من طلاب الجامعة ذات الحجم 36 طالباً إذا كان السحب بدون إرجاع .

الحل :

حيث إن توزيع المجتمع طبيعي ، لذا فإن توزيع المعاينة للمتوسط $\overline{X}$ لجميع العينات المسحوبة من هذا المجتمع يتبع أيضاً التوزيع الطبيعي بمتوسط وتباين كما يلي :

$$\mu_{\bar{x}} = \mu \;\; = 70$$

$$\sigma^2_{\bar{x}} = \frac{\sigma^2}{n} = \frac{36}{36} = 1 \;\; \Rightarrow \;\; \sigma_{\bar{x}} = \sqrt{1} = 1$$

$$\Rightarrow \bar{X} \sim N(70\,,\,1)$$

(١١-٢-٦) توزيع المعاينة للفرق بين متوسطي عينتين $(\bar{X}_1 - \bar{X}_2)$:

إذا كان لدينا مجتمعان مستقلان متوسط كل منهما μ_2 , μ_1 بانحراف معياري σ_2 ,σ_1 لكل منهما على التوالي . فإذا أخذنا عينتين عشوائيتين مستقلتين من المجتمعين حجمها n_2 , n_1 ($n_2 \geq 30$ ، $n_1 \geq 30$) على التوالي ، فوجدنا أن متوسط العينتين هو $\bar{X}1$ ، $\bar{X}2$ على التوالي .

فإن توزيع المعاينة للإحصائية $\bar{X}_1 - \bar{X}_2$ يقترب من التوزيع الطبيعي بمتوسط وتباين يعطي بالصورة التالية :

$$\mu_{\bar{X}_1 - \bar{X}_2} = E(\bar{X}1 - \bar{X}2) = \mu_1 - \mu_2 \longrightarrow (3)$$

$$\sigma^2_{\bar{x}_1 - \bar{x}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \longrightarrow (4)$$

أي أن : $(\bar{X}1 - \bar{X}2) \sim N\left(\mu_1 - \mu_2 \;\;,\;\; \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$

أي أن الفرق بين متوسطي عينتين عشوائيتين مستقلتين مسحوبتين من مجتمعين مستقلين له توزيع طبيعي (تقريباً) بمتوسط $\mu_1 - \mu_2$ وانحراف معياري يساوي $\sqrt{\frac{\sigma_1^2}{n1} + \frac{\sigma_2^2}{n2}}$ ، وذلك بشرط أن حجم العينتين كبير ($n_1 \geq 30$ ، $n_2 \geq 30$) .

لاحظ أنه إذا كان توزيع المجتمعين طبيعياً ، فإن توزيع المعاينة للإحصائية $(\bar{X}1 - \bar{X}2)$ يكون طبيعياً (تماماً) ، حتى وإن كان حجم العينتين n_1 ، n_2 صغيراً.

(١١-٢-٧) توزيع المعاينة لنسبة العينة r :

أحياناً يكون المجتمع الإحصائي ذا صفتين فقط ، فمثلاً عند دراسة إنتاج مصنع معين فإن الوحدات المنتجة تنقسم إلى نوعين سليمة ومعيبة ، فإذا كان حجم المجتمع محل الدراسة N وعدد العناصر التي لها الخاصية الأولى هو A فإن عدد العناصر التي لها الخاصية الثانية هو $N - A$.

نقوم بسحب كل العينات الممكنة ذات الحجم n ، ونوجد من كل عينة النسبة r حيث :

$$r = \frac{a}{n} \quad : \quad a = \text{عدد وحدات العينة التي لها الخاصية الأولى}$$

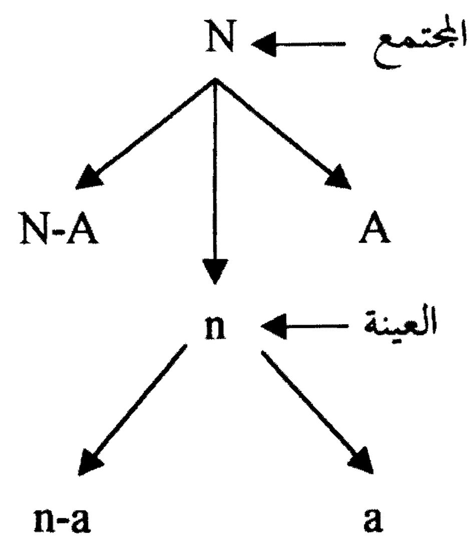
$$\Rightarrow q = \frac{n - a}{n} = 1 - r$$

تسمى أحياناً النسبة r بنسبة النجاح أو احتمال النجاح ، أما النسبة q $(q = 1 - r)$ فتسمى نسبة أو احتمال الفشل .

لاحظ أن نسبة المجتمع R هي :

$$R = \frac{A}{N}$$

$$\Rightarrow Q = \frac{N - A}{N} = 1 - R$$



فإذا سحبنا كل العينات الممكنة ذات الحجم n وكان حجم العينة n كبيراً، فإن توزيع المعاينة لنسبة العينة r يقترب من التوزيع الطبيعي (كلما زادت قيمة n) بمتوسط وتباين يعطى بالصورة التالية :

$$E(r) = \mu r = R \longrightarrow (5)$$

$$\sigma_r^2 = \frac{R \cdot Q}{n} \longrightarrow (6)$$

أي أن :
$$r \sim N(R \ , \sqrt{\frac{R \cdot Q}{n}})$$

مع ملاحظة أنه يتم ضرب الخطأ المعياري في معامل التصحيح إذا كان المجتمع محدوداً والسحب بدون إرجاع .

أي أن نسبة العينة لها تقريباً توزيع طبيعي بوسط حسابي يساوي نسبة المجتمع R وخطأ معياري يساوي $\sqrt{\frac{R \cdot Q}{n}}$.

(١١-٢-٧-١) ملاحظة :

لاحظ أن توزيع المعاينة للنسبة r يقترب من التوزيع الطبيعي فقط إذا كان حجم العينة n كبيراً ، ويمكن اعتبار حجم العينة n كبيراً إذا كان : $n \cdot r \geq 5$ ، $n \cdot q \geq 5$.

مثال (٥) :

مصنع به 100 عامل منهم 20 يدخنون (80 لا يدخنون) سحبت كل العينات الممكنة من العمال والتي حجمها n = 36 مع الإرجاع ، أحسب الوسط الحسابي والخطأ المعياري لتوزيع المعاينة لنسبة المدخنين .

الحل :

نسبة العمال المدخنين بالمصنع هي :
$$R = \frac{20}{100} = 0.20$$

نسبة العمال غير المدخنين بالمصنع شي : $Q = \frac{80}{100} = 0.80$

الوسط الحسابي لتوزيع المعاينة لنسبة المدخنين هو : $\mu_r = R = 0.20$

الخطأ المعياري لنسبة المدخنين هو: $\sigma_r = \sqrt{\frac{R \cdot Q}{n}} = \sqrt{\frac{0.20 \times 0.80}{36}} = 0.067$

(١١-٢-٨) توزيع المعاينة للفرق بين نسبتي عينتين $(r_1 - r_2)$:

إذا كان لدينا مجتمعان مستقلان ، وسحبنا منهما عينتين مستقلتين كبيرتين حجمهما n_1 ، n_2 على التوالي ، وأوجدنا منهما النسبتين r_1 ، r_2 على التوالي، فإن توزيع المعاينة للفرق بين نسبتين $(r_1 - r_2)$ يقترب من التوزيع الطبيعي بمتوسط (توقع) وتباين كما هو موضح فيما يلي : (إذا كانت كل من n_1 ، n_2 كبيرة) .

$$E(r_1) = R_1 , \quad E(r_2) = R_2 \Rightarrow E(r_1 - r_2) = R_1 - R_2 \longrightarrow (7)$$

$$\sigma^2_{r_1 - r_2} = \frac{R_1 \cdot Q_1}{n_1} + \frac{R_2 \cdot Q_2}{n_2} \longrightarrow (8)$$

أي أن :

$$(r_1 - r_2) \sim N\left(R_1 - R_2 , \sqrt{\frac{R_1 \cdot Q_1}{n_1} + \frac{R_2 \cdot Q_2}{n_2}}\right)$$

(١١-٢-٨-١) ملاحظة :

لاحظ أن توزيع المعاينة للإحصائية $(r_1 - r_2)$ تقترب من التوزيع الطبيعي فقط إذا كان حجم العينتين n_1 ، n_2 كبيراً ، وذلك إذا تحقق أن :

$$n_1 \cdot r_1 \geq 5 , n_1 \cdot q_1 \geq 5 , n_2 \cdot r_2 \geq 5 , n_2 \cdot q_2 \geq 5$$

مثال (٦) :

إذا أضفنا إلى المثال السابق مصنعاً به 200 عاملاً منهم 30 يدخنون ، وسحبت كل العينات الممكنة من المصنع الثاني والتي حجمها $n_2 = 40$ ، أحسب الوسط الحسابي والخطأ المعياري للفرق بين نسبتي العمال المدخنين .

الحل :

$$\mu_{r_1-r_2} = R_1 - R_2 = \frac{20}{100} - \frac{30}{200} = 0.20 - 0.15 = 0.05$$

$$\sigma_{r_1-r_2} = \sqrt{\frac{R_1 . Q_1}{n_1} + \frac{R_2 . Q_2}{n_2}} = \sqrt{\frac{0.2 \times 0.8}{36} + \frac{0.15 \times 0.85}{40}} = 0.087$$

(١١-٢-٩) نظرية النهاية المركزية :

وتعتبر هذه النظرية من أهم النظريات في علم الإحصاء والاحتمالات ، وتعطي هذه النظرية توزيع المعاينة لمتوسط العينة $\bar{X}$. ويمكن إثبات هذه النظرية رياضياً ولكن سنكتفي بذكرها بدون إثبات للاستفادة منها في نظرية التقدير واختبارات الفروض .

نظرية النهاية المركزية :

إذا كان $\bar{X}$ متوسط عينة عشوائية حجمها n مسحوبة من مجتمع له متوسط μ وانحراف معياري σ $(\sigma < \infty)$ ، فإن توزيع $\bar{X}$ يقترب من التوزيع الطبيعي بمتوسط μ وانحراف معياري $\frac{\sigma}{\sqrt{n}}$ ويزداد هذا الاقتراب كلما زادت قيمة n (حجم العينة) . أي أن :

$$\bar{X} \sim N\left[\mu , \frac{\sigma}{\sqrt{n}}\right] \quad \text{عندما} \quad n \longrightarrow \infty$$

(١١-٢-١٠) ملاحظة مهمة :

١) إذا كان حجم العينة كبيراً $(n \geq 30)$ فإن توزيع $\bar{X}$ يكون طبيعياً (تقريباً) مهما كان توزيع المجتمع .

٢) إذا كان توزيع المجتمع طبيعياً ، فإن توزيع $\bar{X}$ يكون طبيعياً (تماماً) مهما كان حجم العينة n .

لاحظ أن توزيعات المعاينة (للمتوسط وللفرق بين وسطين وللنسبة وللفرق بين نسبيتين) عبارة عن نتيجة مباشرة لنظرية النهاية المركزية ولذلك اشترطنا في توزيعات المعاينة أن يكون حجم العينة كبيراً .

وسوف نستخدم النظرية السابقة في اشتقاق فترات ثقة للمعلمات التالية :

μ (متوسط المجتمع) ، $\mu_1 - \mu_2$ ، R (نسبة المجتمع) ، $R_1 - R_2$

أما في حالة اختبارات الفروض فسوف نستفيد من النتيجة التالية :

(١١-٢-١١) نتيجة :

إذا كان $\bar{X}$ متوسط عينة عشوائية حجمها n مسحوبة من مجتمع له متوسط μ وانحراف معياري σ ، فإن المتغير العشوائي :

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

يتبع التوزيع الطبيعي المعياري أي أن :

$$Z \sim N(0 \ , \ 1)$$

مع الأخذ في الاعتبار الملاحظة المهمة (١١-٢-١٠) .

مثال (٧) :

إذا كانت درجات الطلاب في الجامعة تتبع التوزيع الطبيعي بمتوسط 70 درجة وانحراف معياري 8 درجات . سحبنا عينة عشوائية حجمها 16 طالباً من الجامعة .

ما هو توزيع متوسط العينة $\bar{X}$ ؟ ثم أوجد الاحتمالات التالية :

$$P(\bar{X} \le 75) \quad , \quad P(60 \le \bar{X} \le 71) \quad , \quad P(\bar{X} \ge 72)$$

الحل :

نطبق نظرية النهاية المركزية حيث :

$$\mu = 70 \quad , \quad \sigma = 8 \quad , \quad n = 16$$

فنجد أن المتغير العشوائي $\bar{X}$ يتبع التوزيع الطبيعي (تماماً) بالرغم من أن حجم العينة صغيراً $(n < 30)$ ، وذلك بسبب أن المجتمع (درجات الطلاب في الجامعة) يتبع التوزيع الطبيعي ، أي أن :

$$\bar{X} \sim N\left(70 \;,\; \frac{8}{\sqrt{16}}\right)$$

$$\Rightarrow \bar{X} \sim N(70 \;,\; 2)$$

وهذا يعني أن متوسط العينة $\bar{X}$ يتبع توزيع طبيعي بمتوسط 70 درجة وانحراف معياري درجتين .

أما الاحتمالات المطلوبة فيمكن إيجادها بسهولة لأن المسألة أصبحت متعلقة بالتوزيع الطبيعي كما يلي : (وذلك باستخدام جدول التوزيع الطبيعي) .

$$P(\bar{X} \le 75) = P\left(Z \le \frac{75-70}{2}\right) = P(Z \le 2.5) = 0.99379 \approx 0.99$$

$$P(68 \le \bar{X} \le 71) = P(\bar{X} \le 71) - P(\bar{X} \le 68) = P(Z \le 0.5) - P(\bar{X} \le -1)$$

$$= 0.6915 - 0.1587 = 0.5328 \approx 0.53$$

$$P(\bar{X} \ge 72) = 1 - P(\bar{X} \le 72)$$

$$= 1 - P(Z \le 1) = 1 - 0.8413 = 0.1587 \approx 0.16$$

مثال (٨) :

إذا كانت مرتبات الموظفين الشهرية في إحدى الوزارات لها متوسط5000 ريال وانحراف معياري 600 ريال . سحبنا عينة عشوائية حجمها 36 موظف من هذه الوزارة ، أوجد توزيع متوسط العينة $\bar{X}$ ثم أحسب الاحتمالات التالية:

$$P(\bar{X} \le 4800) \quad , \quad P(\bar{X} \ge 5100)$$

الحل :

نطبق نظرية النهاية المركزية حيث :

$$\mu = 5000 \quad , \quad \sigma = 600 \quad , \quad n = 36$$

فنجد أن المتغير العشوائي $\bar{X}$ يتبع التوزيع الطبيعي (تقريباً) بمتوسط5000 ريال وانحراف معياري 100 ريال $\left(\frac{600}{\sqrt{36}}\right)$. لاحظ أن المجتمع الأصلي المسحوبة منه العينة (مرتبات الموظفين في الوزارة) قد لا يكون طبيعياً ولكن

نظراً إلى أن حجم العينة كبيراً (n = 36) فإن نظرية النهاية المركزية قابلة للتطبيق كما أوضحنا في الملاحظة المهمة (١١-٢-١٠) ، أي أن :

$$\bar{X} \sim N(5000 \ , \ \frac{600}{\sqrt{36}})$$

$$\Rightarrow \bar{X} \sim N(5000 \ , \ 100)$$

وبالتالي نجد أن :

$$P(\bar{X} \leq 4800) = P(Z \leq -2) = 0.02275$$

$$P(\bar{X} \geq 5100) = 1 - (Z \leq 1) = 1 - 0.8413 = 0.1587$$

(١١-٣) نظرية التقدير :

كل شخص يقوم بعملية التقدير في حياته اليومية ، فمثلاً عندما تريد أن تعبر الشارع فإنك تقوم بعملية تقدير سريعة لما يلي :

سرعة السيارات القادمة ، المسافة بينك وبين السيارات ، عرض الشارع ، وبناءً على ما سبق تقرر أحد القرارات التالية : الانتظار ، العبور مشياً ، العبور جرياً

وهكذا في علم الاقتصاد والإدارة مثلاً فإن العاملين في هذه المجالات يستخدمون التقدير كثيراً مثل تقدير العرض والطلب لسلعة معينة ، تقدير نسبة الموظفين المتأخرين عن العمل ، نسبة تعطل الآلات في اليوم ، ومتوسط المبيعات الشهرية ، الانحراف المعياري لأطوال الطلاب ... الخ .

وللحصول على تقديرات للمؤشرات الإحصائية والاقتصادية المذكورة أعلاه نقوم بسحب عينة عشوائية من المجتمع محل الدراسة ونحسب منها الإحصائيات (التوابع الإحصائية) المطلوبة ثم نستخدم بأسلوب إحصائي هذه الإحصائيات لتقدير معلمات المجتمع مثل تقدير متوسط المجتمع μ ونسبة المجتمع R والانحراف المعياري للمجتمع σ .

(١١- ٣-١) تعريفات مهمة :

١- التقدير بنقطة : هو تقدير معلمة المجتمع بقيمة الإحصائية المقابلة لها من العينة، مثل تقدير متوسط المجتمع μ بمتوسط العينة $\bar{X}$ ، ويمتاز التقدير بنقطة بدقته ولكن يعاب عليه أن احتمال الخطأ فيه كبير .

٢- التقدير بفترة : هو تقدير معلمة المجتمع بمدى من القيم مثل تقدير متوسط المجتمع μ بالفترة : $\bar{X}\pm5$ ، ويمتاز تقدير الفترة بأنه يضم بين حدية الأدنى والأعلى عدداً غير محدود من القيم كما يمكن حساب احتمال صحة التقدير ، أو درجة الثقة فيه ، لذلك تسمى فترات التقدير بفترات الثقة .

٣- المقدر : هو إحصائية العينة التي تستخدم لتقدير معلمة المجتمع ، مثل $\bar{X}$ يعتبر مقدر لمتوسط المجتمع μ .

٤- التقدير : هو قيمة رقمية معينة للمقدر مثل $\bar{X} = 5$ يستخدم كتقدير لمتوسط المجتمع μ ، أي أن التقدير حالة خاصة من المقدر ، وبالتالي فإن التقدير قد يختلف من عينة لأخرى باستخدام المقدر نفسه .

لاحظ أن تقدير النقطة - وإن كان احتمال الخطأ فيه كبيراً - فإنه مفيد في بعض الحالات كما يمكن دراسة خصائصه الإحصائية ، بالإضافة إلى كونه نقطة البداية في تقدير الفترة . ومن دراستنا السابقة لتوزيعات المعاينة يمكن وضع أفضل تقدير نقطة لبعض المعلمات والخطأ المعياري للتقدير كما في الجدول التالي وذلك للمجتمعات اللانهائية :

جدول لأفضل تقدير نقطة لبعض معلمات المجتمع والخطأ المعياري للتقدير

المعلمة المراد تقديرها	الرمز	أفضل تقدير نقطة	الرمز	الخطأ المعياري للتقدير
الوسط الحسابي للمجتمع	μ	الوسط الحسابي للعينة	$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$	$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$
النسبة في المجتمع	R	النسبة في العينة	$r = \frac{a}{n}$	$\sigma_r = \sqrt{\frac{R \cdot Q}{n}}$
الفرق بين متوسطي مجتمعين	$\mu_1 - \mu_2$	الفرق بين متوسطي العينتين	$\overline{X}_1 - \overline{X}_2$	$\sigma_{\overline{x}_1 - \overline{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
الفرق بين نسبتي مجتمعين	$R_1 - R_2$	الفرق بين نسبتي عينتين	$r_1 - r_2$	$\sigma_{r_1 - r_2} = \sqrt{\frac{R_1 Q_1}{n_1} + \frac{R_2 Q_2}{n_2}}$
تباين المجتمع	σ^2	تباين العينة S^2	$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{X})^2$	$\sigma_{S^2}^2 = \sigma^2 \sqrt{\frac{2}{n-1}}$

(١١– ٣–٢) كيفية استخدام فترة الثقة :

عندما درسنا التوزيع الطبيعي أوضحنا أن معظم البيانات تتركز حول متوسط التوزيع μ ، (احتمال تركز البيانات) كما يلي :

$$P(\mu - \sigma \le X \le \mu + \sigma) = 0.6827 = 68.27\ \%$$

$$P(\mu - 2\sigma \le X \le \mu + 2\sigma) = 0.9545 = 95.45\ \%$$

$$P(\mu - 3\sigma \le X \le \mu + 3\sigma) = 0.9973 = 99.73\ \%$$

أي أننا على ثقة مقدارها 95.45 % أن قيمة المتغير العشوائي X [حيث $X \sim N(\mu\ ,\sigma)$] تقع في الفترة ($\mu + 2\sigma$ ، $\mu - 2\sigma$) .

وبالأسلوب نفسه نجد أن :

$$P(\mu - 1.64\,\sigma \le X \le \mu + 1.64\,\sigma) = 0.90 = 90\,\%$$

$$P(\mu - 1.96\,\sigma \le X \le \mu + 1.96\,\sigma) = 0.95 = 95\,\%$$

$$P(\mu - 2.58\,\sigma \le X \le \mu + 2.58\,\sigma) = 0.99 = 99\,\%$$

وهذه تسمى فترة ثقة للمتغير العشوائي X . كما تسمى القيم 90 % ، 95 %، 99 % بدرجات الثقة ويرمز لها – عادة – بالرمز اللاتيني $1 - \propto$ حيث $\propto$ هي المكمل لدرجة الثقة وتسمى مستوى المعنوية (أو الدلالة) . وسوف يتم تعريفها عند تناول اختبار الفروض .

(١١-٣-٣) نتيجة مهمة :

١) إذا كان لدينا متغير عشوائي X حيث : $X \sim N(\mu\ ,\ \sigma)$

فإنه يكفي معرفة قيمة μ ، σ للحصول على فترة ثقة للمتغير العشوائيX بالشكل التالي :

$$P(\mu - Z_{\frac{\propto}{2}}\ .\ \sigma \le X \le \mu + Z_{\frac{\propto}{2}}\ .\ \sigma) = 1 - \propto$$

٢) معامل σ في فترة الثقة هو الذي يحدد درجة الثقة ولذلك يسمى (معامل الثقة) ، وفيما يلي أشهر درجات الثقة ومعاملاتها للتوزيع الطبيعي :

درجة الثقة (*) $1 - \propto$	معامل الثقة $Z_{\frac{\propto}{2}}$
90 %	1.64
95 %	1.96
99 %	2.58

* عند استخدام درجات ثقة أخرى، نقوم باستخراج معاملات الثقة من جدول التوزيع الطبيعي.

(١١-٣-٤) تقدير فترة الثقة لمتوسط المجتمع (μ) :

إذا كان لدينا مجتمع إحصائي طبيعي ، وسحبنا منه عينة عشوائية حجمها n فإن توزيع المعاينة لمتوسط العينة $\bar{X}$ (بحسب نظرية النهاية المركزية) هو :

$$\bar{X} \sim N\left(\mu \ , \ \frac{\sigma}{\sqrt{n}}\right)$$

وبتطبيق النتيجة المهمة (١١-٣-٣) نجد أن :

$$P\left(\mu - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq \mu + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

بإضافة المقدار ($\bar{X}$ - μ -) للطرفين نجد أن :

$$P\left(-\bar{X} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq -\mu \leq -\bar{X} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

وبضرب الطرفين بالمقدار (1 -) وإعادة ترتيب المتباينة نجد أن :

$$P\left(\bar{X} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

وتمثل فترة ثقة لمتوسط المجتمع μ بدرجة $1 - \alpha$ (($1 - \alpha$) %) ويمكن كتابة فترة الثقة السابقة بالصيغة المختصرة التالية :

$$\mu = \bar{X} \pm Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

لاحظ أن فترة الثقة تعتمد على توزيع المعاينة للإحصائية $\bar{X}$ ، وكما درسنا في توزيعات المعاينة ونظرية النهاية المركزية فإن توزيع المعاينة للإحصائية $\bar{X}$ يعتمد على الانحراف المعياري للمجتمع σ وحجم العينة العشوائية n كما نوضح ذلك بالشكل التالي :

σ

معلوم ، المجتمع طبيعي — مجهول

الحالة (1)

n كبيرة $(n \geq 30)$ — n صغيرة $(n < 30)$

المجتمع طبيعي (تقريباً)

الحالة (2) — الحالة (3)

وسنتطرق فيما يلي لهذه الحالات الثلاث :

الحالة الأولى : المجتمع طبيعي ، σ معلوم ، n كبيرة أو صغيرة

في هذه الحالة $\bar{X}$ يتبع التوزيع الطبيعي بمتوسط μ وانحراف معياري $\frac{\sigma}{\sqrt{n}}$ أي أن :

$$\bar{X} \sim N\left(\mu\ ,\ \frac{\sigma}{\sqrt{n}}\right)$$

وبالتالي تكون فترة الثقة لمتوسط المجتمع μ بالشكل التالي :

$$\mu = \bar{X} \pm Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \longrightarrow (9)$$

الحالة الثانية : σ مجهول ، n كبيرة $(n \geq 30)$:

بتطبيق نظرية النهاية المركزية (حيث n كبيرة) فإن $\bar{X}$ يتبع التوزيع الطبيعي وبالتالي تكون فترة الثقة لـــ μ بدرجة ثقة $(1 - \alpha)\%$ بالشكل التالي :

$$\mu = \bar{X} \pm Z_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}} \longrightarrow (10)$$

حيث : S هو الانحراف المعياري للعينة .

الحالة الثالثة : المجتمع طبيعي (أو قريب من الطبيعي) ؛ σ مجهول ، n صغيرة ($n < 30$) :

في هذه الحالة $\bar{X}$ يتبع توزيع t بدرجات حرية n - 1 (*) ، وتكون فترة الثقة لمتوسط المجتمع μ بدرجة ثقة (∝ - 1)% بالشكل التالي :

$$\mu = \bar{X} \pm t_{\frac{\alpha}{2}, n-1} \cdot \frac{S}{\sqrt{n}} \longrightarrow (11)$$

أي أن شروط استخدام توزيع **t** في تقدير الوسط μ هي :

١ - العينة مسحوبة من مجتمع له توزيع طبيعي (أو قريب من التوزيع الطبيعي) .

٢ - تباين المجتمع (أو انحرافه المعياري) مجهول .

٣ - العينة صغيرة ($n < 30$) .

مثال (٩) :

عينة عشوائية حجمها 40 طالباً سحبت من إحدى الكليات فكان الوسط الحسابي والانحراف المعياري لدرجات الطلاب في العينة في أحد الامتحانات هما:

$\bar{X} = 75$ درجة ، $S = 4.8$ درجة

أوجد فترة ثقة للوسط الحسابي لدرجات جميع الطلاب بالكلية في هذا الامتحان بدرجة ثقة 95 % .

---

* درجات الحرية لمجموعة من المشاهدات هو عدد المشاهدات المستقلة من بينها . أي أنها بتعبير آخر عدد المشاهدات (حجم العينة) مطروحاً منه عدد القيود المستقلة (ويرتبط عادةً بمجاميع المربعات) . وفي هذه الحالـة فإن درجات الحرية تساوي ($n - 1$) لأن عدد المشاهدات يساوي n وطرحنا واحد مقابل القيد وهو تقدير المعلمة σ (الانحراف المعياري للمجتمع) بالإحصائية S (الانحراف المعياري للعينة) .

الحل :

لاحظ أن n كبيرة (n = 40) ، $Z_{\frac{\alpha}{2}}$ التي تقابل درجة ثقة 95 % هي :

$$S = 4.8 \quad , \quad \bar{X} = 75 \quad , Z_{0.025} = 1.96$$

فترة الثقة للوسط الحسابي في هذه الحالة هي :

$$\mu = \bar{X} \pm Z_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}$$

$$= 75 \pm 1.96 \ \frac{4.8}{\sqrt{40}} = 75 \pm 1.49 = (73.5 \ , \ 76.5)$$

نحن على ثقة مقدارها 95% أن الوسط الحسابي لدرجات جميع الطلاب لا يقل عن 73.5 درجة ولا يزيد عن 76.5 درجة .

مثال (١٠) :

عينة عشوائية حجمها n = 10 سحبت من مجتمع طبيعي فكان الوسط الحسابي والانحراف المعياري للعينة هما :

$$S = 3.9 \quad , \quad \bar{X} = 32$$

أوجد فترة ثقة للوسط الحسابي للمجتمع بدرجة ثقة 95 % .

الحل :

نلاحظ أن الحالة الثالثة منطبقة هنا حيث المجتمع طبيعي وتباينه غير معروف والعينة صغيرة . لذلك فإن فترة ثقة للوسط الحسابي في هذه الحالة هي :

$$\mu = \bar{X} \pm \ t_{0.025,9} \ \cdot \frac{S}{\sqrt{n}}$$

ونحصل على قيمة **t** من جدول **t** عند درجات حرية تساوي 9 (1 - 10) ، وتحت مستوى معنوية $\frac{\alpha}{2} = \frac{0.05}{2} =$ 0.025 ومن الجدول نجد أن قيمة **t** تساوي 2.262 أي أن :

$$\mu = 32 \pm 2.262 \frac{3.9}{\sqrt{10}} = 32 \pm 2.79 = (29.21\ ,\ 34.79\ )$$

نحن على ثقة مقدارها 95 % أن الوسط الحسابي للمجتمع μ لا يقل عن 29.21 درجة ولا يزيد عن 34.79 درجة .

(١١-٣-٤-١) ملاحظة مهمة :

عندما يكون توزيع المجتمع غير طبيعي (ولا يشابه التوزيع الطبيعي) ، وتباينه غير معروف ، وكذلك حجم العينة n صغيراً $(n < 30)$ ، فإنه لا يمكن معرفة توزيع المعاينة لمتوسط المجتمع $\bar{X}$ وبالتالي لا يمكن تكوين فترة ثقة لمتوسط المجتمع μ كما سبق . (وتستخدم بعض الطرق التي تسمى الطرق اللامعلمية).

(١١-٣-٥) تقدير فترة الثقة للفرق بين متوسطي مجتمعين $(\mu_1 - \mu_2)$:

إذا كان لدينا مجتمعان كبيران وسحبنا منهما عينتين عشوائيتين مستقلتين حجمها n_1 ، n_2 وكان متوسط العينتين $\bar{X}_1$ ، $\bar{X}_2$ بانحراف معياري S_1 ، S_2 على التوالي . سنوضح كيفية تكوين فترة ثقة للمعلمة $(\mu 1 - \mu 2)$ في حالتين كما يلي :

الحالة الأولى :

أ) المجتمعان طبيعيان ، σ_1 ، σ_2 معلوم ، n_1 ، n_2 كبيرة أو صغيرة

ب) المجتمعان غير طبيعيين ، σ_1 ، σ_2 معلومان ، والعينتان كبيرتان فإن الإحصائية $(\bar{X}_1 - \bar{X}_2)$ تتبع التوزيع الطبيعي بمتوسط $(\mu_1 - \mu_2)$ وانحراف معياري $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ أي أن :

$$(\bar{X}_1 - \bar{X}_2) \sim N\ (\mu_1 - \mu_2)\ \sqrt{\frac{\sigma_1^2}{n_2} + \frac{\sigma_2^2}{n_2}}$$

وبالتالي تكون فترة الثقة للمعلمة $(\mu_1 - \mu_2)$ بدرجة ثقة $(1 - \propto)\%$ بالشكل التالي :

$$\mu_1 - \mu_2 = (\bar{X}_1 - \bar{X}_2) \pm Z_{\frac{\propto}{2}} \cdot \sqrt{\frac{\sigma_1^2}{n_2} + \frac{\sigma_2^2}{n_2}} \longrightarrow (12)$$

الحالة الثانية :

σ_1 ، σ_2 مجهولان ، n_1 ، n_2 كبيرة $(n_1 \geq 30$ ، $n_2 \geq 30)$:

نستخدم فترة الثقة السابقة (١٢) ، ولكن باستخدام S_1^2 ، S_2^2 بدلاً من σ_1^2 ، σ_2^2 .(*)

مثال (١١) :

إذا كان مراقب الجودة في أحد مصانع الصلب يختار عينة عشوائية كل شهر لاختبار قوة تحمل القضبان المنتجة فإذا سحب عينة تتكون من 16 قضيباً فوجد أن متوسط قوة تحملها هو 6200 كجم بانحراف معياري 180 كجم .

أوجد فترة ثقة لمتوسط قوة تحمل القضبان المنتجة بدرجة ثقة 95 % في الحالتين التاليتين (بافتراض أن قوة تحمل القضبان تتبع التوزيع الطبيعي) :

أ- $\sigma = 200$ ب- σ مجهول

الحل :

أ-

$$\mu = \bar{X} \pm Z_{\frac{\propto}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

$$\Rightarrow \mu = 6200 \pm 1.96 \cdot \frac{200}{\sqrt{16}} = 6200 \pm 98$$

$$\Rightarrow \mu = (6102 , 6298)$$

---

* وهناك حالات أخرى لن نتعرض لها هنا ، مثل افتراض أن التباينين متساويان ومعلومان أو مجهولان ، وحجم العينتين n_1 ، n_2 صغير.

نحن على ثقة مقدارها 95 % أن متوسط قوة تحمل القضبان المنتجة في هذا المصنع لا يقل عن 6,102 كجم ولا يزيد عن 6,298 كجم .

ب -
$$\mu = \bar{X} \pm t_{\frac{\alpha}{2}, n-1} \cdot \frac{S}{\sqrt{n}}$$

$$\Rightarrow \mu = 6200 \pm 2.131 \cdot \frac{180}{\sqrt{16}} = 6200 \pm 96$$

$$\Rightarrow \mu = (6104 , 6296)$$

نحن على ثقة مقدارها 95 % أن متوسط قوة تحمل القضبان المنتجة في هذا المصنع لا يقل عن 6,104 كجم ولا يزيد عن 6,296 كجم .

مثال (١٢) :

تملك إحدى الشركات مصنعين للمرطبات في الرياض والدمام وتحاول الشركة دوماً أن تحافظ على مستوى متماثل للإنتاج في المصنعين وحتى تعرف مدى تماثل المصنعين في الإنتاج سحبت عينة عشوائية حجمها 36 علبة من كل من الرياض والدمام فوجدت أن متوسط محتوى العلبة في المصنعين هو :

325 مل ، و322 مل بانحراف معياري 4 مل و 3 مل على التوالي .

أوجد فترة ثقة للفرق بين متوسطي المجتمعين بدرجة ثقة 88% في الحالتين التاليتين :

أ- $\sigma_1 = 3 \ , \sigma_2 = \sqrt{7}$ ب- σ_1 ، σ_2 مجهولان

الحل :

أ- من جدول التوزيع الطبيعي نجد أن $Z_{\frac{\alpha}{2}}$ والتي تقابل درجة الثقة 88% هي $Z_{\frac{\alpha}{2}} = 1.55$ وبالتالي فإن :

$$\mu_1 - \mu_2 = (\bar{X}_1 - \bar{X}_2) \pm Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma_1^2}{n_2} + \frac{\sigma_2^2}{n_2}}$$

$$\therefore\ \mu_1 - \mu_2 = (325 - 322) \pm 1.55 \sqrt{\frac{9}{36} + \frac{7}{36}}$$

$$\Rightarrow \mu_1 - \mu_2 = 3 \pm 1.55 \times 0.67$$

$$\Rightarrow \mu_1 - \mu_2 = 3 \pm 1 = (2\ , 4)$$

نحن على ثقة مقدارها 88 % أن متوسط محتوى علبة مصنع الرياض يزيد عن متوسط محتوى علبة مصنع الدمام بما لا يقل عن 2 مل ولا يزيد عن 4 مل.

ب- وحيث إن العينتين كبيرتان ، σ_1 ، σ_2 مجهولان ، فإنه يستخدم S_1^2، S_2^2 بدلاً من σ_1^2 ، σ_2^2 وتكون فترة الثقة للفرق بين متوسطي المجتمعين هي:

$$\mu_1 - \mu_2 = (\bar{X}_1 - \bar{X}_2) \pm Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma_1^2}{n_2} + \frac{\sigma_2^2}{n_2}}$$

$$\therefore \mu_1 - \mu_2 = 3 \pm 1.55 \sqrt{\frac{16}{36} + \frac{9}{36}}$$

$$\Rightarrow \mu_1 - \mu_2 = 3 \pm 1.3 = (1.7\ \ , 4.3)$$

نحن على ثقة مقدارها 88 % أن متوسط محتوى علبة مصنع الرياض يزيد عن متوسط محتوى علبة مصنع الدمام بما لا يقل عن 1.7 مل ولا يزيد عن 4.3 مل .

(١١-٣-٦) ملاحظة (١) :

يمكن تكوين فترة ثقة لأي معلمة في المجتمع (باستثناء التباين والانحراف المعياري للمجتمع) ولتكن B بدرجة ثقة $(1 - \alpha)\%$ بالصيغة التالية :

$$B = b \pm Z_{\frac{\alpha}{2}} \quad \sigma b \quad \longrightarrow (13)$$

أي أن فترة الثقة للمعلمة B تكون بالصيغة التالية :

المعلمة (B) = الإحصائية (b) ± [معامل الثقة × الخطأ المعياري للإحصائية]

(١١-٣-٧) ملاحظة (٢) :

معنى العبارة التي نذكرها في تفسير فترات الثقة وهي على سبيل المثال :
"نحن على ثقة مقدارها 95% أن متوسط المجتمع μ لا يقل عن 60 ولا يزيد عن 80".

معنى ذلك هو ما يلي :

عند تكوين فترات ثقة كثيرة ولتكن 100 فترة محسوبة من 100 عينة عشوائية مأخوذة من المجتمع الإحصائي محل الدراسة ، فإن 95 فترة ثقة من هذه الفترات سوف تشمل قيمة μ الحقيقية .

(١١-٣-٨) تقدير فترة الثقة لنسبة المجتمع (R) :

إذا كان لدينا مجتمع إحصائي تتمتع مفرداته بإحدى خاصيتين (نجاح أو فشل) باحتمال R ، $Q = 1 - R$ على التوالي ، أي له توزيع ذي الحدين بنسبة R ، إذا سحبنا عينة عشوائية كبيرة حجمها n وحسبنا للعينة (احتمال النجاح) r ، فإن توزيع المعاينة للنسبة r هو :

$$r \sim N(R \; , \sqrt{\frac{R \cdot Q}{n}})$$

وبالتالي تكون فترة الثقة لنسبة المجتمع R بدرجة ثقة $(1 - \propto)\%$ هي :

$$R = r \pm Z_{\frac{\propto}{2}} \; . \; \sqrt{\frac{r \cdot q}{n}} \longrightarrow (14)$$

حيث $q = 1 - r$

(١١-٣-٩) ملاحظات مهمة :

(١) لاحظ أن فترة الثقة للنسبة R صالحة فقط إذا كان حجم العينة n كبيراً ، ويمكن اعتبار حجم العينة n كبيراً إذا تحقق الشرط التالي :

$$n \; . \; r \geq 5 \quad , n \; . \; q \geq 5$$

فإذا لم يتحقق هذا الشرط فإن فترة الثقة قد تعطى نتائج غير صحيحة .

(٢) عند حساب الحد الأدنى والأعلى في فترة الثقة للنسبة نلاحظ أن الحد الأدنى للنسبة يجب ألا يقل عن الصفر وأن الحد الأعلى للنسبة لا يزيد عن الواحد الصحيح .

(١١-٣-١٠) تقدير فترة الثقة للفرق بين نسبتي مجتمعين ($R_1 - R_2$):

إذا كان لدينا مجتمعان لهما توزيع ذي الحدين بنسبة R_1 ، R_2 على التوالي، وسحبنا عينتين عشوائيتين مستقلتين كبيرتين من هذين المجتمعين حجمهما n_1 ، n_2 على التوالي ، فوجدنا أن نسبة النجاح في العينتين هما r_1 ، r_2 على التوالي . فإن فترة الثقة للفرق بين النسبتين $(R_1 - R_2)$ تعطى بالصورة التالية :

$$R_1 - R_2 = (r_1 - r_2) \pm Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{r_1 \cdot q_1}{n_1} + \frac{r_2 \cdot q_2}{n_2}} \longrightarrow (15)$$

(١١-٣-١١) ملاحظة :

لاحظ أن فترة الثقة للمعلمة $(R_1 - R_2)$ صالحة فقط إذا كان حجم العينتين n_1 ، n_2 كبيراً ، ويكون بتحقق الشرط التالي :

$$n_1 . r_1 \geq 5 , \quad n_1 . q_1 \geq 5 , \quad n_2 . r_2 \geq 5 , \quad n_2 . q2 \geq 5$$

وإذا لم يتحقق هذا الشرط فإن فترة الثقة السابقة قد تعطى نتائج غير صحيحة .

مثال (١٣) :

قامت إدارة التعليم في إحدى المناطق بسحب عينتين عشوائيتين مستقلتين من مدرستي غرناطة وقرطبة حجمهما 100 ، 81 طالب على التوالي ، فوجد أن عدد الناجحين هو 80 ، 61 على التوالي .

أوجد ما يلي :

أ- فترة ثقة لنسبة الناجحين في مدرسة غرناطة بدرجة ثقة 95 % .

ب- فترة ثقة لنسبة الراسبين في مدرسة قرطبة بدرجة ثقة 80 % .

ج- فترة ثقة للفرق بين نسبتي الناجحين في المدرستين بدرجة ثقة 99% .

الحل :

لاحظ أن : $r_1 = \frac{80}{100} = 0.80 \quad , \quad r_2 = \frac{61}{81} = 0.75$

$q_1 = 0.20 \quad , \quad q_2 = 0.25$

كما أن : $n_1 . r_1 \geq 5 \; , \; n_1 . q_1 \geq 5 \; , \; n_2 . r_2 \geq 5 \; , \; n_2 . q_2 \geq 5$

أ-

$$R_1 = r \pm Z_{\frac{\propto}{2}} \cdot \sqrt{\frac{r_1 \cdot q_1}{n_1}}$$

$$\Rightarrow R_1 = 0.80 \pm 1.96 \sqrt{\frac{0.8 \times 0.2}{100}} = 0.80 \pm 0.08$$

$$\Rightarrow R_1 = (0.72 \quad , \quad 0.88)$$

نحن على ثقة مقدارها 95 % أن نسبة الناجحين في مدرسة غرناطة لا تقل عن 72 % ولا تزيد عن 88 % .

ب - لاحظ أن :

$$1 - \propto = 0.80 \quad \Rightarrow \propto = 0.20 \Rightarrow \frac{\propto}{2} = 0.10$$

نجد أن :

$$Z_{\frac{0.20}{2}} = Z_{0.10} = 1.28$$

وبالتالي تكون فترة الثقة كما يلي :

$$\mathbf{Q}_2 = q_2 \pm Z_{\propto/2} \cdot \sqrt{\frac{r_2 \cdot q_2}{n_2}}$$

$$\Rightarrow Q_2 = 0.25 \pm 1.28 \ . \sqrt{\frac{0.75 \times 0.25}{81}}$$

$$\Rightarrow Q_2 = 0.25 \pm 0.06 = (0.19 \ , \ 0.31)$$

نحن على ثقة مقدارها 80 % أن نسبة الراسبين في مدرسة قرطبة لا تقل عن 19 % ولا تزيد عن 31 % .

ج-
$$R_1 - R_2 = (r_1 - r_2) \pm Z_{\propto/2} \ . \sqrt{\frac{r_1 \cdot q_1}{n_1} + \frac{r_2 \cdot q_2}{n_2}}$$

$$= (0.80 - 0.75) \pm 2.58 \sqrt{\frac{0.8 \times 0.2}{100} + \frac{0.75 \times 0.25}{81}}$$

$$= 0.05 \pm (2.58 \times 0.063)$$

$$= 0.05 \pm 0.16 = (-0.11 \ , \ 0.21)$$

نحن على ثقة مقدارها 99 % أن الفرق بين نسبتي النجاح في مدرستي غرناطة وقرطبة لا يقل عن 11 % ولا يزيد عن 21 % .

ملحوظة :

فترة الثقة لتباين المجتمع والانحراف المعياري له سيتم تناولها في الفصل الثاني عشر والخاص بتطبيقات توزيع $\chi 2$ إن شاء الله .

(١١-٣-١٢) تقدير حجم العينة :

من الموضوعات المهمة جداً في الدراسات الميدانية هو كيفية تحديد حجم العينة المناسب . وهنا سوف نميز بين حالتين :

الأولى : عند تقدير الوسط الحسابي للمجتمع باستخدام التوزيع الطبيعي .

الثاني : عند تقدير نسبة المجتمع باستخدام التوزيع الطبيعي أيضاً .

(١١-٣-١٢- ١) أولاً: تحديد حجم العينة المناسب عند تقدير الوسط :

ذكرنا عند إنشاء فترة ثقة للوسط الحسابي للمجتمع باستخدام التوزيع الطبيعي أنها تأخذ الشكل التالي :

$$\mu = \overline{X} \pm Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

وبنقل $\overline{X}$ إلى الطرف الأيسر نحصل على :

$$\mu - \overline{X} = \pm Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

ويسمى الطرف الأيسر (الخطأ في تقدير الوسط) ونرمز له بالرمز E . أي أن :

$$E = \pm Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

حيث $(E = \mu - \overline{X})$

وبتربيع الطرفين نحصل على : $E^2 = Z^2 \dfrac{\sigma^2}{n}$

ومنها فإن حجم العينة n يأخذ الشكل التالي في حالة تقدير الوسط :

$$n = \frac{Z^2 . \sigma^2}{E^2} \longrightarrow (16)$$

(١١-٣-١٢- ٢) ملاحظات مهمة :

١) يمكن أن يكتب حجم العينة السابق بالشكل التالي :

$$n = (\frac{Z . \sigma}{E})^2 \longrightarrow (17)$$

٢) ينبغي ألا يقل حجم العينة عن الحجم الذي يحصل عليه الباحث باستخدام المعادلة السابقة (١٦) أو (١٧) . وإذا كان الرقم غير صحيح (أي به كسور) فيؤخذ الرقم الصحيح الأكبر أياً كانت قيمة الكسر (أي مهما كان

الكسر صغيراً) . فمثلاً : إذا كان حجم العينة الذي تم الحصول عليه هو :

$$n = 64.2$$

فإن حجم العينة ينبغي ألا يقل عن 65 مفردة .

٣) يحدد الباحث مسبقاً (الخطأ في تقدير الوسط) والذي يرمز له بالرمز E فمثلاً : يقول أن الخطأ في تقدير الوسط الحسابي لدرجات الطلاب يساوي (أو لا يزيد) عن 3 درجات . أي أن : $E = 3$

٤) تحسب قيمة Z من جدول التوزيع الطبيعي المعياري المقابلة لدرجة الثقة المطلوبة . فمثلاً : إذا كان المطلوب هو التقدير بدرجة ثقة 95 % فإن $Z=1.96$ وإذا كانت درجة الثقة المطلوبة هي 99 % فإن $Z = 2.58$.

٥) σ^2 هو تباين المجتمع والذي يكون عادة معلوماً من دراسات أو خبرات سابقة ، وإذا لم يكن معلوماً يستخدم بدلاً منه أفضل تقدير نقطة لتباين المجتمع S^2 (حيث $S^2 = \frac{1}{n-1}\sum(X_i - \bar{X})^2$) .

مثال (١٤) :

عند تقدير الوسط الحسابي لعمر نوع معين من البطاريات ، ما هو حجم العينة المناسب بدرجة ثقة 99 % بحيث لا يزيد الخطأ في تقدير الوسط الحسابي لعمر البطاريات عن 3 ساعات عمل إذا علمت أن تباين عمر البطاريات هو 45؟.

الحل :

ينبغي ألا يقل حجم العينة عند تقدير الوسط عن : $n = \frac{Z^2 . \sigma^2}{E^2}$

حيث : $Z = 2.58$ (والتي تقابل درجة ثقة 99 %) .

، $\sigma^2 = 45$ ، $E = 3$ (الخطأ في تقدير الوسط)

وبالتعويض نحصل على :

$$n = \frac{(2.58)^2 \times 45}{(3)^2} = 33.28 \approx 34$$

أي ينبغي ألا يقل عدد البطاريات (العينة) عن 34 بطارية حتى لا يزيد الخطأ في تقدير متوسط عمر البطاريات عن ثلاث ساعات وذلك بدرجة ثقة 99 % .

(١١-٣-١٢-٣) ثانياً : تحديد حجم العينة المناسب عن تقدير النسبة :

تأخذ فترة الثقة للنسبة الشكل التالي :

$$R = r \pm Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{RQ}{n}}$$

حيث : $Q = 1 - R$

وبنقل r إلى الطرف الأيسر نحصل على : $R - r = \pm Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{RQ}{n}}$

ويسمى الطرف الأيسر (الخطأ في تقدير النسبة) ونرمز له أيضاً بالرمز E.

أي أن $(E = R - r)$ وبالتالي فإن : $E = \pm Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{RQ}{n}}$

وبتربيع الطرفين نحصل على : $E^2 = \frac{Z^2 . R . Q}{n}$

ومنها فإن حجم العينة n يأخذ الشكل التالي في حالة تقدير النسبة :

$$n = \frac{Z^2 . R . Q}{E^2} \longrightarrow (18)$$

(١١-٣-١٢-٤) ملاحظات مهمة :

بالإضافة إلى الملاحظات السابقة عن حجم العينة عند تقدير الوسط نذكر ما يلي:

١) إذا كانت النسبة في المجتمع غير معروفة - وهو الوضع المفترض عادة حيث يكون المطلوب تقديرها - فإن الباحث يعتمد على الخبرة السابقة في تقديرها

أو على الحالات المشابهة بحيث يحصل على أفضل تقدير ممكن للنسبة في المجتمع ومن ثم يستخدمها في حساب حجم العينة كما هو موضح بالمعادلة رقم (١٨) .

٢) لاحظ الفرق بين حجم العينة في الحالتين : حيث يوجد Z^2 في البسط ، E^2 في المقام في الحالتين ، والاختلاف يكمن فقط أنه في حالة تقدير الوسط يتم الضرب في التباين σ^2 (أو S^2) ، وفي حالة تقدير النسبة يتم الضرب في النسبة والنسبة المكملة لها $R \cdot Q$ (أو $r \cdot q$) .

مثال (١٥) :

عند تقدير نسبة النجاح في امتحان الثانوية العامة ، ما هو حجم العينة المناسب من الطلاب بحيث لا يتعدى الخطأ في تقدير نسبة النجاح عن 0.04 وذلك بدرجة ثقة 95 % إذا علمت من خبرتك السابقة (أو بافتراض أن) نسبة النجاح تساوي 80 % ؟

الحل :

حجم العينة المناسب عند تقدير النسبة ينبغي ألا يقل عن :

$$n = \frac{Z^2 . R . Q}{E^2} .$$

حيث :

$Z = 1.96$ (والتي تقابل درجة الثقة 95 %)

$E = 0.04$ (الخطأ في تقدير نسبة)

$R = 0.80$ ؛ $Q = 1 - R = 0.20$

وبالتعويض نحصل على :

$$n = \frac{(1.96)^2 \times 0.80 \times 0.20}{(0.04)^2}$$
$$= 384.16$$
$$\approx 385$$

أي ينبغي ألا يقل حجم العينة عن 385 طالباً حتى لا يزيد الخطأ في تقدير نسبة النجاح عن 4 % وذلك بدرجة ثقة 95 % .

(١١-٤) نظرية اختبارات الفروض :

نحتاج كثيراً في الحياة العملية إلى التحقق من صحة فرض أو إدعاء ما ، وقد يكون هذا التحقق سهلاً ومؤكداً مثل إدعاء شخص ما أن السماء تمطر فإنه للتحقق من هذا الإدعاء يكفي أن نذهب خارج البناية ونتأكد هل السماء تمطر أم لا ؟ وأحياناً أخرى يكون التحقق من صحة الإدعاء صعباً مثل إدعاء شخص ما أنه ذهب إلى أحد الأسواق التجارية في مدينة الرياض . وأحياناً أخرى يكون التحقق مستحيلاً مثل إدعاء شخص ما أنه قد أضمر شيئاً معيناً في نفسه !!

وفي علم الإحصاء نسمي الإدعاء (فرض العدم) ونسمي التحقق منه (اختبار فرض العدم) ونحتاج إلى عملية اختبارات الفروض في كثير من الحالات مثل اختبار الفرض بأن متوسط الدخل للأفراد في مدينة الرياض هو $\mu = 4{,}000$ أو اختبار إدعاء أحد مديري المصانع بأن نسبة التالف في مصنعه R لا تتجاوز 2 % ، أو إدعاء أحد الأطباء أن نسبة الشفاء في مستوصفه أو عيادته لا تقل عن 90 % ...... وهكذا .

(١١-٤-١) تعريفات مهمة :

١- الفرض الإحصائي : هو إدعاء أو ظن حول معلمة أو عدة معلمات في المجتمع الإحصائي (مثل إدعاء حول قيمة μ ، σ ، R) أو حول توزيع المجتمع بذاته .

٢- فرض العدم أو الفرض العدمي : هو الفرض الذي سوف نقوم باختباره أو التحقق من صحته ويرمز له بالرمز Ho مثل : $Ho : \mu = 12$.

٣ - الفرض البديل : وهو الفرض الذي نقبله إذا رفضنا فرض العدم Ho ويرمز له بالرمز H_1 مثل $H_1 : \mu > 12$.

٤- الخطأ من النوع الأول هو رفض Ho بينما هو صحيح ، واحتمال الوقوع بهذا الخطأ يرمز له بالرمز α ويسمى " مستوى المعنوية " أو " الدلالة " أي أن مستوى المعنوية هو احتمال الوقوع في الخطأ من النوع الأول ، أي احتمال رفض فرض العدم بينما هو صحيح .

ودرجة الثقة - وهي المكمل لمستوى المعنوية بمعنى أن مجموعهما يساوي الواحد الصحيح - هي احتمال قبول فرض العدم عندما يكون صحيحاً ويرمز له بالرمز α - 1 .

٥- الخطأ من النوع الثاني هو قبول Ho بينما هو خطأ ، واحتمال الوقوع بهذا الخطأ يرمز له بالرمز β .

٦- الاختبار الأمثل : هو الاختبار الذي يكون فيه الاحتمالان α ، β أصغر ما يمكن .

٧- نظرية اختبارات الفروض تعتمد على استخدام إحصائيات العينة مثل ($\bar{x}$ r, S,) للتحقق من معلمات المجتمع مثل (μ ، σ ، R) أو توزيعه .

٨- القرار الإحصائي هو رفض Ho وقبول H_1 أو قبول Ho ورفض H_1 .

والفكرة الأساسية في اختبارات الفروض الإحصائية هي تقسيم المساحة تحت منحنى التوزيع (سواء كان طبيعياً أو t أو غيرهما) إلى منطقتين غير متداخلتين : الأولى تسمى منطقة القبول ، أي منطقة قبول فرض العدم Ho والثانية تسمى منطقة الرفض أو المنطقة الحرجة ، أي منطقة رفض فرض العدم Ho (وقبول الفرض البديل H_1) .

وسوف يكتشف الطالب أن منطقة القبول هي المنطقة التي تمثلها درجة الثقة α - 1 ، وأن منطقة الرفض يمثلها مستوى المعنوية α .

والأداة الأساسية التي تساعد الباحث في اتخاذ قرار بقبول الفرض العدمي أو رفضه تسمى (إحصائية الاختبار) والتي تأخذ في كثير من الحالات شكلاً ثابتاً

(أو تعتمد على الفكرة نفسها) حيث تتكون من بسط ومقام :

فالبسط عبارة عن الفرق بين قيمة الإحصائية المحسوبة من العينة والقيمة المفترضة للمعلمة المراد اختبارها في الفرض العدمي فكلما كان هذا الفرق كبيراً فإن الباحث يميل إلى رفض فرض العدم ، والعكس صحيح . ويتم الحكم على ما إذا كان الفرق كبيراً أو صغيراً بقسمة هذا الفرق على الخطأ المعياري للإحصائية .

أي أن إحصائية الاختبار تأخذ في كثير من الحالات شكلاً ثابتاً عبارة عن الفرق بين قيمتي الإحصائية والمعلمة المراد اختبارها مقسوم على الخطأ المعياري للإحصائية . فمثلاً لاختبار الوسط الحسابي للمجتمع فإن إحصائية لاختبار تأخذ الشكل التالي :

$$\frac{\bar{X}-\mu_0}{\sigma_{\bar{x}}}$$

فإذا وقعت قيمة إحصائية الاختبار في منطقة القبول فإن القرار يكون قبول فرض العدم Ho ، وإذا وقعت في منطقة الرفض فإن القرار يكون رفض فرض العدم Ho وقبول الفرض البديل H_1 .

(١١-٤-٢) ملاحظة :

يمكن تمثيل العلاقة بين فرض العدم Ho والقرار الإحصائي بالشكل التالي :

Ho / القرار	صحيح	خطأ
قبول	3	خطأ من النوع الثاني
رفض	خطأ من النوع الأول	3

مثال (١٦) :

إذا كانت أطوال الطلاب في الجامعة تتبع التوزيع الطبيعي بمتوسط μ وأردنا أن نختبر فرض العدم Ho : μ = 170 Cm مقابل الفرض البديل :

الحالة الأولى : $H_1 : \mu \neq 170$

الحالة الثانية : $H_1 : \mu > 170$

الحالة الثالثة : $H_1 : \mu < 170$

نقوم بوضع القاعدة التالية لاتخاذ القرار :

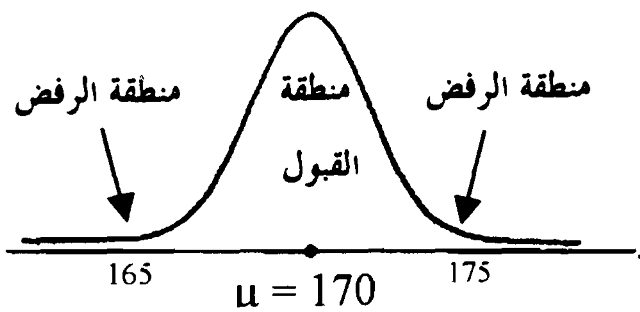
نسحب عينة عشوائية حجمها n طالب من الجامعة ثم نحسب متوسط الطول لهذه العينة $\overline{X}$ ويكون القرار كما يلي :

الحالة الأولى :

نقبل Ho إذا كان $165 \leq \overline{X} \leq 175$

نرفض Ho إذا كان $\overline{X} < 165$ أو $\overline{X} > 175$

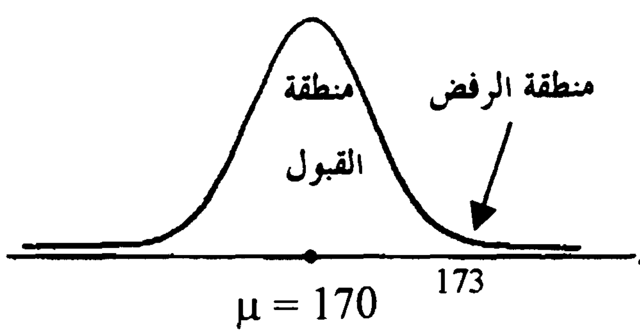
ويمكن تمثيلها على المنحنى الطبيعي كما يلي (بافتراض صحة Ho)



الحالة الثانية :

نقبل Ho إذا كان $\overline{X} \leq 173$

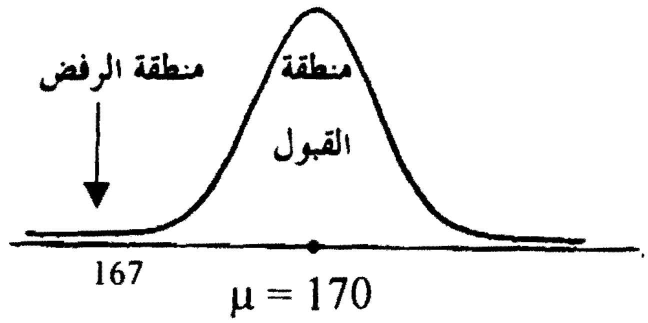
نرفض Ho إذا كان $\overline{X} > 173$



الحالة الثالثة :

نقبل Ho إذا كان $\overline{X} \geq 167$

نرفض Ho إذا كان $\overline{X} < 167$



لاحظ أننا حددنا منطقة الرفض (المنطقة الحرجة) بطريقة اجتهادية شخصية، وبالتالي تختلف منطقة الرفض من شخص إلى آخر ، ولذا سوف نوضح فيما يلي كيفية تحديد منطقة الرفض بطريقة إحصائية موضوعية يتفق عليها الجميع !!.

(١١–٤–٣) أنواع اختبارات الفروض :

إذا أردنا اختبار فرض العدم $Ho : F = Fo$ حيث F إحدى معلمات المجتمع، فإنه يكون لدينا ثلاثة أنواع من الاختبارات بحسب صيغة الفرض البديل H_1 كما يلي:

١– اختبار ذو طرفين (ذيلين) : وذلك عندما تكون منطقة الرفض موزعة على طرفي التوزيع بالتساوي ، وذلك يحدث عندما يكون الفرض البديل "لا يساوي" أي بالصورة التالية:

$H_1 : F \neq Fo$ (أنظر الحالة الأولى في المثال السابق)

٢– اختبار ذو طرف أيمن (علوي) : وذلك عندما تكون منطقة الرفض على يمين التوزيع أي مركزة بالكامل في الطرف الأيمن للمنحنى، وذلك يحدث عندما يكون الفرض البديل "أكبر من"، أي بالصورة التالية : $H_1 : F > Fo$ (أنظر الحالة الثانية في المثال السابق) .

٣- **اختبار ذو طرف أيسر (سفلي)** : وذلك عندما تكون منطقة الرفض على يسار التوزيع أي مركزه بالكامل في الطرف الأيسر للمنحنى ، وذلك يحدث عندما يكون الفرض البديل "أقل من" ، أي بالصورة التالية :

$H_1 : F < Fo$ (أنظر الحالة الثالثة في المثال السابق)

(١١-٤-٤) توضيح كيفية تحديد الفرض البديل :

يمثل الفرض البديل خطوة مهمة في اختبارات الفروض لأنه هو الذي يحدد نوع الاختبار المستخدم ، وعادة يعطي الفرض البديل في السؤال بالنسبة للطلاب ، ولكن ماذا عن الوضع في الحياة العملية حيث إن الباحث هو الذي عليه اختيار الشكل المناسب للفرض البديل (إما يساوي أو أكبر من أو أقل من). وبصفة عامة نقول إذا لم يكن لدى الباحث انطباع أو اتجاه معين (نحو الزيادة أو النقصان) فإن الفرض البديل يكون في الغالب (لا يساوي) ، أي يستخدم الباحث في هذه الحالات اختبار ذا طرفين . أما إذا كان لدى الباحث انطباع أو اتجاه نحو الزيادة : مثل أن يقوم مصنع ما بدورة تدريبية للعمال على الطرق الحديثة للإنتاج بهدف زيادة متوسط الإنتاجية ، أو باتباع أسلوب جديد بهدف زيادة نسبة التوزيع (أو المبيعات) فإن الفرض البديل في مثل هذه الحالات يكون أكبر من ويستخدم الباحث اختبار ذا طرف أيمن . وكذلك إذا كان لدى الباحث انطباع أو اتجاه نحو النقصان (أو التخفيض) كأن يتبع المصنع طريقة جديدة لتخفيض متوسط التكاليف ، أو تتبع محطة الخدمة أسلوب جديد لتقليل متوسط وقت الانتظار ، أو تقليل معدّل الفاقد فإن الفرض البديل في مثل هذه الحالات يكون (أقل من) ويستخدم الباحث اختبار ذا طرف أيسر.

(١١-٤-٥) خطوات الاختبار الإحصائي :

يمكن تلخيص خطوات الاختبار الإحصائي فيما يلي :

١) وضع أو تحديد فرض العدم Ho ، والذي يأخذ عادة شكل معادلة أو مساواة (وهو ما يتعلق بمعلمات المجتمع) .

٢) وضع أو تحديد الفرض البديل H_1 ، والذي يأخذ أحد أشكال ثلاثة : إما لا يساوي ، أو أكبر من ، أو أقل من وهو الذي يحدد نوع الاختبار المستخدم إما اختبار ذو طرفين أو اختبار ذو طرف أيمن أو اختبار ذو طرف أيسر على التوالي .

٣) حساب إحصائية الاختبار : وهي Z الحسابية أو T الحسابية بحسب الحالات التي سبق ذكرها (أو أي إحصائية أخرى) .

٤) تحديد مستوى المعنوية α .

٥) تحديد المنطقة الحرجة (منطقة الرفض والمتبقي هو منطقة القبول) وذلك بتحديد قيم Z (أو t) المعيارية أو الجدولية التي بناءً عليها نحدد المنطقة الحرجة (منطقة الرفض) ، وكما ذكرنا سابقاً فإن الفرض البديل H_1 هو الذي يحدد موقع منطقة الرفض (طرفين أو طرف أيمن أو طرف أيسر).

٦) اتخاذ القرار الإحصائي : وذلك بمقارنة قيم Z (أو T) الحسابية بقيم Z (أو t) الجدولية ، فإذا وقعت Z (أو T) الحسابية في منطقة الرفض فإن القرار هو رفض Ho وقبول H_1 (رفض فرض العدم وقبول الفرض البديل) . أما إذا وقعت Z (أو T) الحسابية في منطقة القبول فإن القرار هو قبول Ho ورفض H_1 (*) .

(١١-٤-٦) اختبارات الفروض حول متوسط المجتمع (μ) :

إذا كان لدينا مجتمع إحصائي كبير وسحبنا منه عينة عشوائية حجمها n

---

* عبارة "قبول Ho ورفض H_1" ليست دقيقة من الناحية الإحصائية النظرية ، والأدق أن نقول "لا نستطيع رفض Ho لعدم توفر معلومات كافية لرفض فرض العدم Ho وهذا يعني رفض الفرض البديل H_1" ، ونظراً لعدم وجود فارق جوهري من الناحية التطبيقية (العملية) بين العبارتين لذا سوف نستخدم دائماً العبارة الأولى "قبول Ho ورفض H_1" حرصاً على الاختصار .

ومتوسطها $\overline{X}$ وانحرافها المعياري S ، ونريد اختبار فرض العدم $Ho : \mu = \mu o$ مقابل الفرض البديل $H_1 : \mu \neq \mu o$.

نميز بين ثلاث حالات كما درسنا في نظرية التقدير ، وهي كما يلي :

الحالة الأولى : المجتمع طبيعي ، σ معلوم ، n صغيرة أو كبيرة

نحسب إحصائية الاختبار Z كما يلي :

$$Z = \frac{\overline{X} - \mu o}{\frac{\sigma}{\sqrt{n}}} \longrightarrow \quad (19)$$

فإذا كانت Z داخل منطقة الرفض نرفض Ho ونقبل H_1 والعكس صحيح.

اصطلاح : سوف نسمي إحصائية الاختبار Z السابقة بـــ Z الحسابية .

الحالة الثانية : σ مجهول ، n كبيرة $(n \geq 30)$

تصبح Z الحسابية كما يلي :

$$Z = \frac{\overline{X} - \mu o}{S / \sqrt{n}} \longrightarrow \quad (20)$$

لاحظ أن إحصائية الاختبار Z في الحالتين (الأولى والثانية) . لها توزيع طبيعي معياري ، وذلك بافتراض صحة فرض العدم Ho (انظر النتيجة (١١-٢-١٠)) .

الحالة الثالثة : المجتمع طبيعي تقريباً ، σ مجهول ، n صغيرة $(n < 30)$

والإحصائية هي T الحسابية حيث :

$$T = \frac{\overline{X} - \mu o}{S / \sqrt{n}} \longrightarrow \quad (21)$$

والإحصائية T لها توزيع t بدرجات حرية n - 1 ، وذلك بافتراض صحة فرض العدم Ho .

(١١-٤- ٧) كيفية تحديد المنطقة الحرجة (منطقة الرفض) :

يعتمد تحديد المنطقة الحرجة على صورة الفرض البديل H_1 كما سبق :

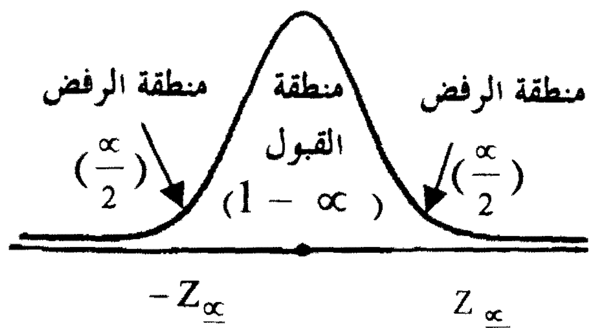
أولاً : الحالة الأولى والثانية :

١- $H_1 : \mu \neq \mu o$

منطقة الرفض تكون بالصورة التالية :

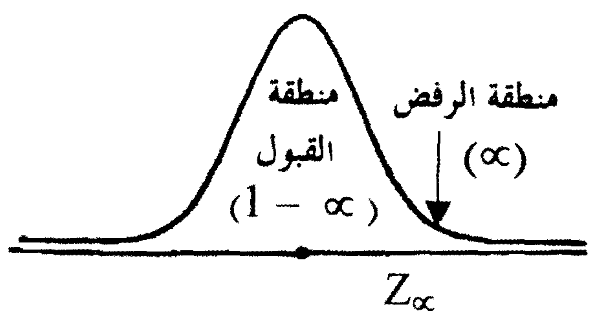
نرفض Ho عندما $Z > Z_{\frac{\alpha}{2}}$ أو $Z < -Z_{\frac{\alpha}{2}}$

اصطلاح : سوف نسمي $Z_{\frac{\alpha}{2}}$ بـــ Z المعيارية أو الجدولية .



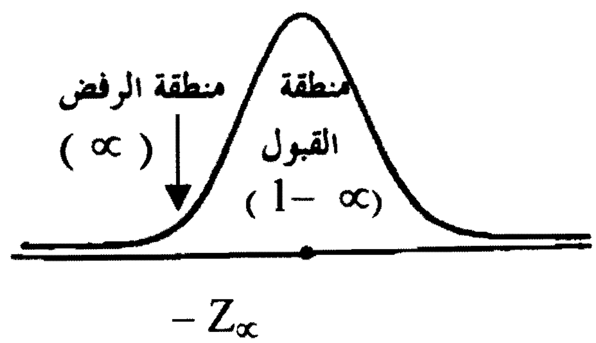
٢- $H_1 : \mu > \mu o$

نرفض Ho عندما $Z > Z_{\alpha}$



٣- $H_1 : \mu < \mu o$

نرفض Ho عندما $Z < -Z_\alpha$



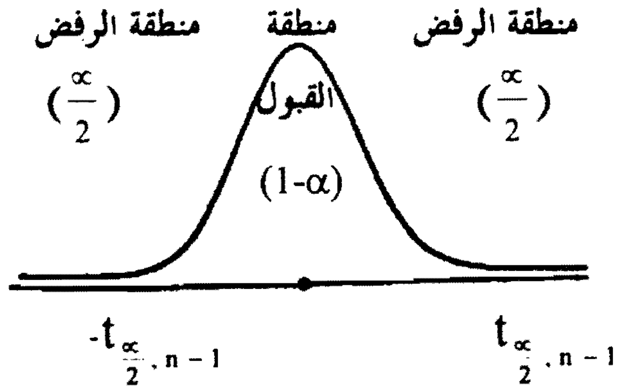
ثانياً : الحالة الثالثة :

حيث إن توزيع المعاينة لـــ $\overline{X}$ في هذه الحالة يتبع توزيع t فإن منطقة الرفض تكون كما يلي :

١ - $H_1 : \mu \neq \mu o$

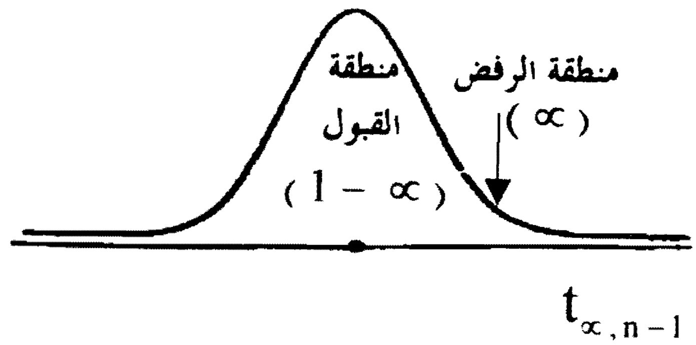
منطقة الرفض تكون بالصورة التالية :

نرفض Ho عندما $T > t_{\frac{\alpha}{2},n-1}$ أو $T < -t_{\frac{\alpha}{2},n-1}$



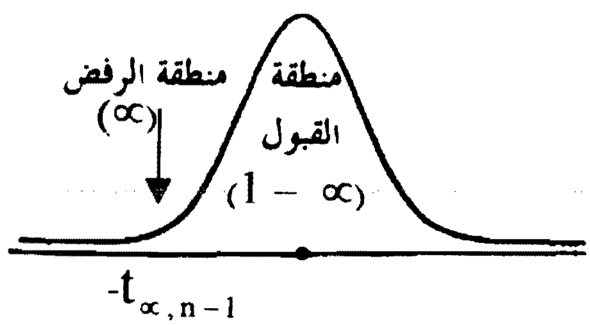
٢- $H_1 : \mu > \mu o$

نرفض Ho عندما $T > t_{\alpha, n-1}$



٣- $H_1 : \mu < \mu o$

نرفض Ho عندما $T < -t_{\alpha, n-1}$



مثال (١٧) :

إذا كان عمر أحد أنواع الساعات يتبع التوزيع الطبيعي بمتوسط μ وانحراف معياري σ . سحبنا عينة عشوائية من هذه الساعات حجمها n فوجدنا أن متوسط عمر الساعات في هذه العينة هو $\overline{X} = 4.5$ بانحراف معياري $S = 0.8$ سنة . اختبر فرض العدم $Ho : \mu = 5$ بمستوى معنوية $\alpha = 0.05$ في الحالات التالية :

أ - $H_1 : \mu \neq 5$, $n = 16$, $\sigma = 1$.

ب - $H_1 : \mu > 5$, $n = 16$.

ج - $H_1 : \mu < 5$, $n = 36$.

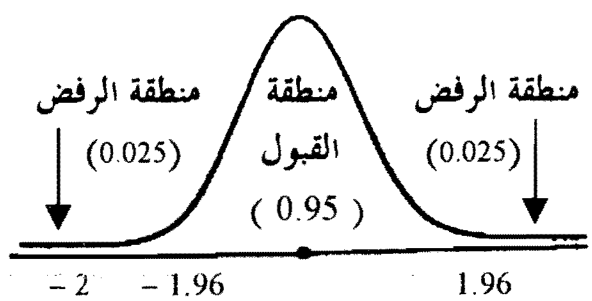
الحل :

أ) $H_1 : \mu = 5 \quad VS. \quad H_1 : \mu \neq 5 \quad , Z_{\frac{0.05}{2}} = 1.96$

نوجد قيمة Z الحسابية :

$$Z = \frac{\overline{X} - \mu o}{\frac{\sigma}{\sqrt{n}}} = \frac{4.5 - 5}{\frac{1}{\sqrt{16}}} = -0.5 \times 4 = -2$$

لاحظ أن الاختبار ذو طرفين كما يلي :



$\because Z = -2 < -1.96 = -Z_{\frac{0.05}{2}}$ ،

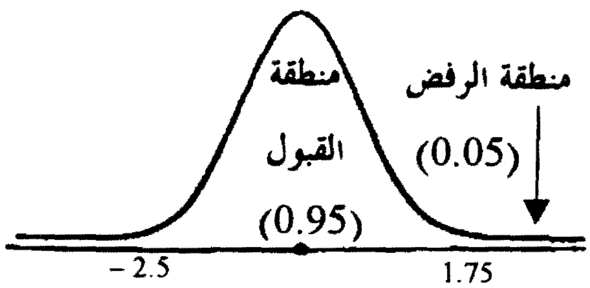
$\therefore$ نرفض Ho ونقبل H_1 بمستوى معنوية 5 % ونقول أن متوسط عمر هذا النوع من الساعات لا يساوي 5 سنوات بمستوى معنوية 5 % .

ب) $Ho : \mu = 5 \quad VS. \quad H_1 : \mu > 5 \quad , n = 16$

نحسب قيمة T الحسابية : $T = \frac{\overline{X} - \mu_0}{\frac{S}{\sqrt{n}}} = \frac{4.5 - 5}{\frac{0.8}{\sqrt{16}}} = -2.5$

لاحظ أن الاختبار ذو طرف أيمن كما يلي :

من الجدول نجد أن : $t_{0.05,15} = 1.75$



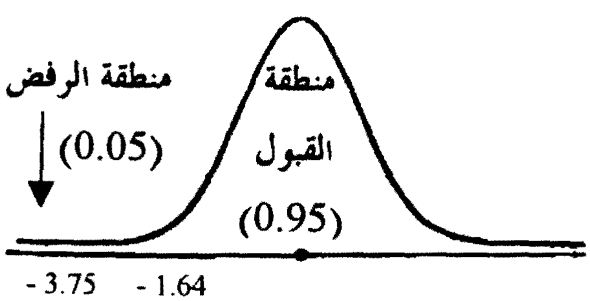
$$\because \quad T = -2.5 < 1.75 = t_{0.05,15}$$ ،

$\therefore$ نقبل Ho ونرفض H_1 بمستوى معنوية 5 % ونقول أن متوسط عمر هذا النوع من الساعات لا يزيد عن 5 سنوات .

ج) نحسب قيمة Z الحسابية :

لاحظ أن الاختبار ذو طرف أيسر كما يلي :

$$Z = \frac{4.5 - 5}{0.8 / \sqrt{36}} = -3.75$$



$$\because Z = -3.75 < -1.64 = -Z_{0.05}$$ ،

$\therefore$ نرفض Ho ونقبل H_1 بمستوى معنوية 5 % ونقول أن متوسط عمر هذا النوع من الساعات أقل من 5 سنوات بمستوى معنوية 5 % .

(١١-٤-٨) ثانياً : اختبارات الفروض حول الفرق بين متوسطي مجتمعين ($\mu_1 - \mu_2$) :

إذا كان لدينا مجتمعان لهما متوسطان μ_1 ، μ_2 وانحراف معياري σ_1 ، σ_2 على التوالي . سحبنا منهما عينتين عشوائيتين مستقلتين حجمهما n_1 ، n_2 بمتوسط $\overline{X}_1$ ، $\overline{X}_2$ وانحراف معياري S_1 ، S_2 على التوالي. اختبر فرض العدم :

$$Ho : \mu_1 - \mu_2 = Do \qquad : \qquad Do \text{ مقدار ثابت}$$

وغالباً ما يكون المقدار الثابت Do مساوياً للصفر ، فيكون فرض العدم Ho أن المتوسطين متساويان .

الحالة الأولى :

أ) $\sigma1$ ، $\sigma2$ معلوم ، n_1 ، n_2 صغيرة أو كبيرة ، المجتمعان طبيعيان.

ب) $\sigma1$ ، $\sigma2$ معلوم ، n_1 ، n_2 كبيرة ($n_2 \geq 30$ ، $n_2 \geq 30$)، المجتمعان غير طبيعيين .

نحسب إحصائية الاختبارZ كما يلي :

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - Do}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \longrightarrow \quad (22)$$

فإذا كانت Z الحسابية داخل منطقة الرفض نرفض Ho ونقبل H_1 .

الحالة الثانية :

$\sigma1$ ، $\sigma2$ مجهول، n_1 ، n_2 كبيرة ($n_2 \geq 30$ ، $n_1 \geq 30$) نفس الحالة السابقة ولكن نستخدم S_1^2 ، S_2^2 بدلاً من σ_1^2 ، σ_2^2 .

لاحظ أن إحصائية Z في الحالتين تتبع التوزيع الطبيعي المعياري ، وذلك بافتراض صحة فرض العدم Ho (انظر النتيجة (١١ - ٢ - ١٠)) (*) .

---

* وكما سبق ذكره في نظرية التقدير فإن هناك حالات أخرى لم نتعرض لها في هذا الكتاب .

مثال (١٨) :

إذا كان لدينا نوعان من المصابيح ، سحبنا عينة عشوائية من النوع الأول حجمها 36 مصباح ومثلها من النوع الثاني ، فوجدنا أن متوسط عمر العينة الأولى 85 ساعة بانحراف معياري 4 ساعات ، متوسط عمر العينة الثاني 81 ساعة بتباين 25 ساعة . والمطلوب :

أ- هل نستطيع أن نستنتج بمستوى معنوية 5 % أن متوسط عمر النوع الأول يزيد عن متوسط النوع الثاني بساعتين على الأكثر مع العلم أن :

$$\sigma_1 = \sigma_2 = 4$$

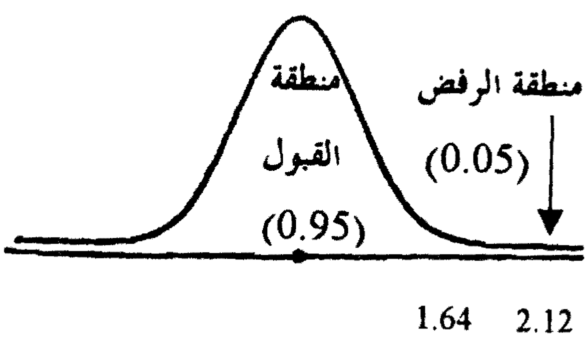
ب- هل نستطيع أن نستنتج بمستوى معنوية 5 % أن الفرق بين عمر النوعين هو ساعتان .

الحل :

أ- $Ho : \mu_1 - \mu_2 = 2 \qquad VS. \qquad H_1 : \mu_1 - \mu_2 > 2$

$$Z = \frac{(85-81)-2}{\sqrt{\frac{16}{36}+\frac{16}{36}}} = \frac{2}{\sqrt{\frac{32}{36}}} = 2.12$$

لاحظ أن الاختبار ذو طرف أيمن ، والشكل التالي يوضح منطقة الرفض :



$\because Z_{0.05} = 1.64 < 2.12 = Z$ ،

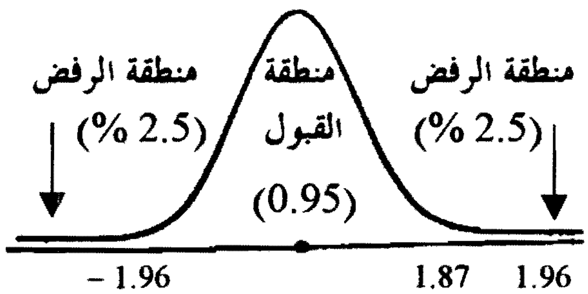
∴ نرفض Ho ونقبل H_1 بمستوى معنوية 5 % ونقول أن متوسط عمر النوع الأول من المصابيح يزيد عن متوسط عمر النوع الثاني بما لا يقل عن ساعتين بمستوى معنوية 5 % .

ب- $Ho : \mu_1 - \mu_2 = 2 \quad VS. \quad H_1 : \mu_1 - \mu_2 \neq 2$

نستخدم S_1^2 ، S_2^2 بدلاً من σ_1^2 ، σ_2^2 (لأنهما مجهولان) فتكون قيمة Z الحسابية كما يلي :

$$Z = \frac{(85-81)-4}{\sqrt{\frac{16}{36}+\frac{25}{36}}} = \frac{2}{\sqrt{\frac{41}{36}}} = 1.87$$

لاحظ أن الاختبار ذو طرفين ، فتكون منطقة الرفض كما يلي :



∴ نقبل Ho ونرفض H_1 بمستوى معنوية 5% ونقول أن متوسط عمر النوع الأول من المصابيح يزيد عن متوسط عمر النوع الثاني بساعتين فقط بمستوى معنوية 5%.

(١١–٤–٩) ثالثاً : اختبارات الفروض حول نسبة المجتمع (R) :

إذا كان لدينا مجتمع كبير يتمتع بخاصيتين (أي له توزيع ذي الحدين بنسبة R) ، وسحبنا منه عينة عشوائية حجمها n ، وحسبنا منها r ، فإذا أردنا أن نختبر فرض العدم : $Ho : R = Ro$ فيجب التأكد أولاً أن :

$$n \, . \, Ro \geq 5 \quad , \quad n\,(1 - Ro) \geq 5$$

ثم نحسب إحصائية الاختبار :

$$Z = \frac{r - Ro}{\sqrt{\frac{Ro \cdot Qo}{n}}} \longrightarrow \quad (23)$$

حيث $Qo = 1 - Ro$

لاحظ أن Z تتبع التوزيع الطبيعي المعياري ، وذلك بافتراض صحة فرض العدم Ho.

فإذا كانت Z داخل منطقة الرفض نرفض Ho ونقبل H_1 والعكس بالعكس .

مثال (١٩) :

قررت الإدارة الهندسية في إحدى الجامعات تنفيذ مواقف سيارات إضافية للطلبة، ولتحديد عدد المواقف قدرت أن نسبة الطلاب الذين يأتون بسيارهم لا تزيد عن 60 % من الطلاب . سُحب عينة عشوائية حجمها 250 طالباً ، فوجدنا 165 منهم يأتون بسيارهم فهل تقدير الإدارة الهندسية صحيح ؟ استخدم مستوى معنوية 5 % ثم 1 % .

الحل :

$$r = \frac{165}{250} = 0.66 \qquad \Rightarrow \qquad q = 0.34$$

لاحظ أن : $Ho \;: R = 0.60 \qquad VS. \quad H1 \;:\; R > 0.60$

$$n \,.\, Ro \;\geq 5 \qquad , \quad n \,.\, Qo \geq 5$$

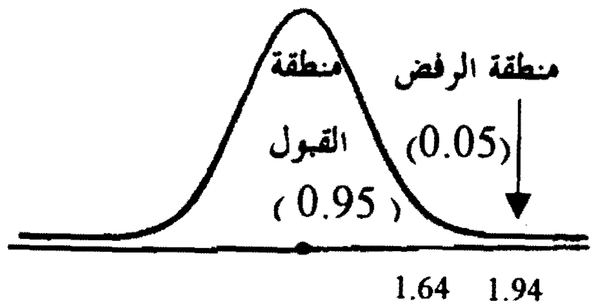
$$Z = \frac{0.66 - 0.60}{\sqrt{\frac{0.6 \times 0.4}{250}}} = 1.94$$

نوجد قيمة Z المعيارية :

عندما $\propto = 0.05$ نجد من الجدول أن :

$$z_{0.05} = 1.64$$

فتكون منطقة الرفض بالشكل التالي :



$$\because Z_{0.05} = 1.94 > 1.64 = Z_{0.05} \quad ،$$

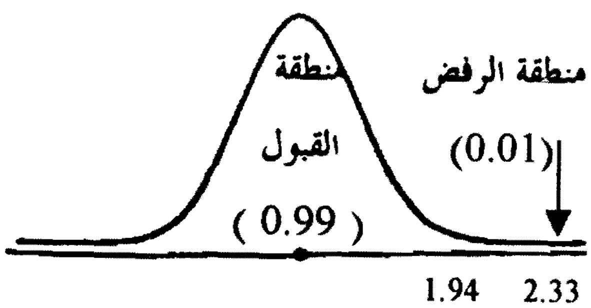
∴ نرفض Ho ونقبل H_1 ، ونقول أن نسبة الطلاب الذين يأتون بسيارهم تزيد عن 60 % بمستوى معنوية 5 % .

عندما $\propto = 0.01$ نجد من الجدول أن :

$$z_{0.01} = 2.33$$

$$\because z = 1.94 < 2.33 = z_{0.01} \quad ،$$

فتكون منطقة الرفض بالشكل التالي :



∴ نقبل Ho ونرفض H_1 ، ونقول أن نسبة الطلاب الذين يأتون بسيارهم لا تزيد عن 60 % بمستوى معنوية 1 % .

(١١-٤-١٠) رابعاً: اختبارات الفروض حول الفرق بين نسبتي مجتمعين ($R_1 - R_2$) :

إذا كان لدينا مجتمعان لهما توزيع ذي الحدين بنسبة R_1 ، R_2 على الترتيب، ونرغب في اختبار الفرض :

مقدار ثابت Do : $Ho : R_1 - R_2 = Do$

لاختبار Ho نقوم بسحب عينتين مستقلتين من المجتمعين المدروسين ونحسب منهما نسبة العينة r_1 ، r_2 على التوالي ثم نتأكد من تحقق الشرط التالي:

$$n_1 . r_1 \geq 5 \;,\quad n_1 . q_1 \geq 5 \;,\quad n_2 . r_2 \geq 5 \;,\quad n_2 . q_2 \geq 5$$

$$Z = \frac{(r_1 - r_2) - Do}{\sqrt{\frac{rq}{n_1} + \frac{rq}{n_2}}} \longrightarrow \quad (24)$$

لاحظ أن الإحصائية Z تتبع التوزيع الطبيعي المعياري ، وذلك بافتراض صحة فرض العدم Ho .

حيث $$r = \frac{n_1 r_1 + n_2 r_2}{n_1 + n_2} = \frac{X_1 + X_2}{n_1 + n_2} \quad ، \quad q = 1 - r$$

فإذا كانت Z داخل منطقة الرفض (المنطقة الحرجة) نرفض Ho ونقبل H_1 والعكس بالعكس .

مثال (٢٠) :

يقدر موظف الإحصاءات العامة أن نسبة الذين يملكون سيارة في الدمام تزيد عن أولئك الذين في الإحساء بنسبة 8 % على الأكثر . للتحقق من صحة هذا التقرير سحبنا عينتين مستقلتين من الدمام والإحساء فكانت النتائج كما يلي :

ملاك السيارة	حجم العينة	المدينة
120	150	الدمام
62	100	الاحساء

اختبر صحة ذلك التقدير بمستوى معنوية 1 % ثم 12 % .

الحل :

واضح أن شرط الاختبار متحقق . لاحظ أن :

$$r_1 = \frac{120}{150} = 0.80 \quad \Rightarrow \quad q_1 = 0.20 \ ,$$

$$r_2 = \frac{62}{100} = 0.62 \quad \Rightarrow \quad q_2 = 0.38 \ ,$$

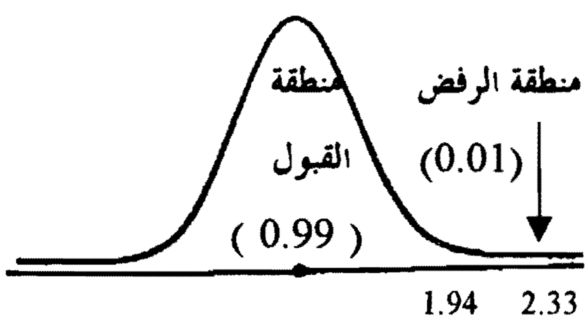
$$Ho = R_1 - R_2 = 0.08 \qquad VS. \qquad H_1 : R_1 - R_2 > 0.08$$

$$r = \frac{X_1 + X_2}{n_1 + n_2} = \frac{120+62}{150+100} = \frac{182}{250} = 0.73$$

$$\because Z = \frac{(0.80 - 0.62) - 0.08}{\sqrt{\frac{0.73 \times 0.37}{150} + \frac{0.73 \times 0.37}{100}}} = 1.49$$

عندما $\propto = 0.01$ نجد من الجدول أن :

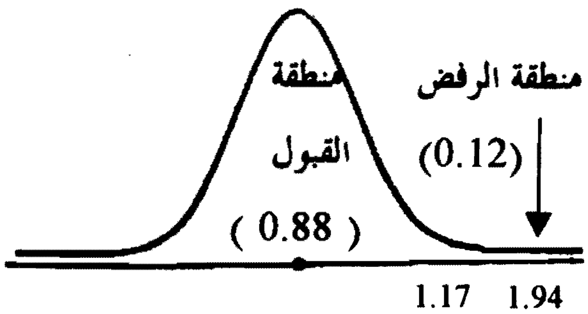
$$Z_{0.01} = 2.33 > 1.49 = Z$$



∴ نقبل Ho ونرفض H_1 ، ونقول أن الفرق بين نسبتي الذين يملكون سيارات في الدمام وأولئك الذين في الاحساء يساوي 8 % بمستوى معنوية 1 %.

عندما $\propto = 0.12$ نجد من الجدول أن :

$$Z_{0.12} = 1.17 < 1.49 = Z$$



∴ نرفض Ho ونقبل H_1 ، ونقول أن الفرق بين نسبتي الذين يملكون سيارات في الدمام والاحساء يزيد عن 8 % بمستوى معنوية 12 % .

ملحوظة :

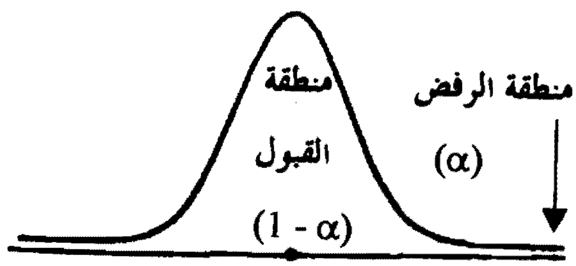
اختبارات الفروض حول تباين المجتمع (وانحرافه المعياري) سيتم تناولها في الفصل الثاني عشر والخاص بتطبيقات توزيع χ^2 إن شاء الله .

جدول مختصر لاختبارات الفروض الإحصائية

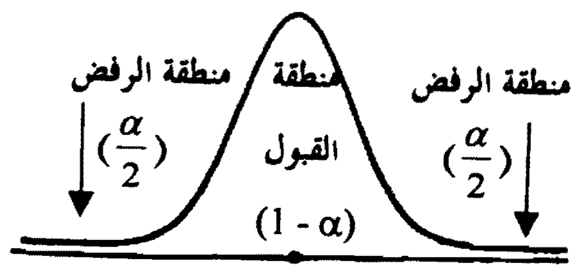
Ho	الحالة	الحسابية	H_1	المنطقة الحرجة (الرفض)
$\mu = \mu o$	الأولى ، الثانية	$Z = \frac{\bar{X} - \mu o}{\sigma / \sqrt{n}}$ $Z = \frac{\bar{X} - \mu o}{S / \sqrt{n}}$	$\mu \neq \mu o$ $\mu > \mu o$ $\mu < \mu o$	$Z > Z_{\alpha/2}$ أو $Z < -Z_{\alpha/2}$ $Z > Z\alpha$ $Z < -Z\alpha$
$\mu = \mu o$	الثالثة	$T = \frac{\bar{X} - \mu o}{S / \sqrt{n}}$	$\mu \neq \mu o$ $\mu > \mu o$ $\mu < \mu o$	$T > t_{\alpha/2,\, n-1}$ أو $T < -t_{\alpha/2,\, n-1}$ $T > t\alpha\,, n-1$ $T < -t\alpha\,, n-1$
$\mu_1 - \mu_2 = Do$	الأولى ، الثانية	$Z = \frac{(\bar{X}_1 - \bar{X}_2) - Do}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ استخدام S_1^2 ، S_2^2 بدلاً من σ_1^2 ، σ_2^2	$\mu_1 - \mu_2 \neq \mu o$ $\mu_1 - \mu_2 > \mu o$ $\mu_1 - \mu_2 < \mu o$	$Z > Z_{\alpha/2}$ أو $Z < -Z_{\alpha/2}$ $Z > Z\alpha$ $Z < -Z\alpha$
$R = Ro$	$n.Ro \geq 5$ $n.Qo \geq 5$	$Z = \frac{r - Ro}{\sqrt{\frac{Ro \cdot Qo}{n}}}$	$R \neq Ro$ $R > Ro$ $R < Ro$	$Z > Z_{\alpha/2}$ أو $Z < -Z_{\alpha/2}$ $Z > Z\alpha$ $Z < -Z\alpha$
$R_1 - R_2 = Do$	$n_1.r_1 \geq 5$ $n_1.q_1 \geq 5$ $n_2.r_2 \geq 5$ $n_2.q_2 \geq 5$	$Z = \frac{(r_1 - r_2) - Do}{\sqrt{\frac{rq}{n_1} + \frac{rq}{n_2}}}$ $r = \frac{n_1 r_1 + n_2 r_2}{n_1 + n_2}, q = 1 - r$	$R_1 - R_2 \neq Do$ $R_1 - R_2 > Do$ $R_1 - R_2 < Do$	$Z > Z_{\alpha/2}$ أو $Z < -Z_{\alpha/2}$ $Z > Z\alpha$ $Z < -Z\alpha$

رسم إيضاحي للمنطقة الحرجة (الرفض) على المنحنى الطبيعي المعياري

اختيار ذو طرف أيمن



اختيار ذو طرفين



اختيار ذو طرف أيسر