

الفصل السادس

نظرية العينات

THEORY OF SAMPLING

(٦-١) نظرية العينات :

تبحث نظرية العينات في العلاقات بين المجتمعات والعينات المأخوذة من هذه المجتمعات ، وقد حوت من النظريات والتوزيعات والأساليب ما يمكننا من تقدير أدلة المجتمعات أو مقارنتها أو إصدار قرارات أو تنبؤات بشأنها عن طريق دراسة وتحليل عينات مأخوذة منها مما يدخل تحت موضوع الاستدلال الإحصائي .

. STATISTICAL INFERENCE

ويقصد بالاستدلال الإحصائي أى إجراء يستخدم نظرية الاحتمال في إصدار قرارات عن مجتمع أو عدة مجتمعات عن طريق عينات مأخوذة منها مع تحديد درجة الثقة في هذه القرارات . ولعملية الاستدلال الإحصائي مجالان رئيسيان هما مجال تقدير بارامترات المجتمعات ومجال اختبار الفروض الإحصائية ، على أنه من الشروط الأساسية في هذه العملية أن تكون العينات عشوائية لأن جميع نظريات الاحتمال التي تعتمد عليها مؤسسة على فرض العشوائية . كما أنه كلما كبر حجم العينة كلما كان الاستدلال أكثر دقة .

ومن بين المسائل التي تبرز في الأبحاث التطبيقية وتحتاج إلى عملية الاستدلال الإحصائي ما يلي :

(أ) اختبار صواب أو خطأ فروض مطروحة عن أدلة المجتمعات .

(ب) تقدير متوسطات وتباينات المجتمعات وغيرها من الأدلة .

(ج) الكشف عما إذا كانت الفروق المشاهدة في العينات هي فروق راجعة إلى الصدفة أو تقلبات العينات أو هي فروق جوهرية تدل على وجود اختلاف حقيقي بين المجتمعات التي أخذت منها هذه العينات .

(د) الكشف عن تأثير واحد أو أكثر من العوامل أو المعالجات على متغير ما أو ظاهرة معينة .

(هـ) تقدير درجة ونوع العلاقة أو الارتباط بين المتغيرات وإصدار تنبؤات عنها .

وسوف نتدارس هذه المسائل وغيرها في هذا الفصل وما يليه من فصول ، بعد تقديم بضعة مفاهيم ونظريات تتطلبها الدراسة الواعية لتلك المسائل .

(٦-٢) توزيعات المعاينة :

في نظرية العينات نميز بين ثلاثة أنواع من التوزيعات .

Population Distribution

(١) توزيع المجتمع

Sample Distribution

(٢) توزيع العينة

Sampling Distribution

(٣) توزيع المعاينة

ولبيان الفرق بين هذه التوزيعات نفرض أننا نرغب في إيجاد الوسط الحسابي لدخل العائلة في مجتمع مدينة مؤلفة من ٢٠٠٠ عائلة . إذا أمكننا معرفة دخول جميع عائلات المدينة ووضعنا هذه الدخول في توزيع تكرارى فإن هذا التوزيع يسمى بتوزيع المجتمع للدخول . ونستطيع بالطبع أن نحصل مباشرة على الوسط الحسابى الحقيقى μ لدخل العائلة في مجتمع المدينة ، أو على أى دليل آخر يخص هذا المجتمع .

أما إذا اخترنا عينة من مجتمع هذه المدينة فإن التوزيع التكرارى لدخول العائلات في هذه العينة يسمى بتوزيع العينة . والوسط الحسابي $\bar{x}$ لهذا التوزيع لا يكون عادة مساوياً للوسط الحسابي الحقيقي μ للمجتمع ، إلا أننا قد نأخذ هذا الوسط الحسابي تحت شروط معينة ، كتقدير للوسط الحسابي الحقيقي للمجتمع . وبطبيعة الحال تختلف الأوساط الحسابية للعينات المأخوذة من نفس المجتمع حتي ولو كانت من نفس الحجم .

أما إذا فرضنا أننا حصلنا من المجتمع على جميع العينات التي من نفس الحجم n وأوجدنا الوسط الحسابي لكل من هذه العينات ثم وضعنا هذه الأوساط في توزيع تكرارى فإن هذا التوزيع يسمى حينئذ بتوزيع المعاينة للوسط الحسابي (أو للأوساط الحسابية) للعينات ذوات الحجم n . ويمكننا أن نحسب الوسط الحسابي والانحراف المعياري ومعامل الالتواء ... لهذا التوزيع . وبالمثل يمكن أن نتصور توزيع المعاينة للتباين للعينات ذوات الحجم n أو توزيع المعاينة لأى مقياس آخر . في هذا المثال يستحيل عملياً إيجاد توزيع المعاينة للوسط الحسابي بالطريقة المذكورة لأن عدد العينات الممكنة هو عدد فلكي نعجز عن الحصول عليه . وفي المعتاد نحصل على توزيعات المعاينة بطرق رياضية ، إلا أنه لتوضيح مفهوم توزيع المعاينة فنضرب المثال البسيط الآتي .

مثال (٦ - ١) :

يتألف مجتمع من ٦ أرناب أوزانها بالأوقيات ١١ ، ١٦ ، ١٢ ، ١٥ ، ١٦ ، ١٤ .

(أولاً) أوجد توزيع المعاينة للوسط الحسابي للعينات ذوات الحجم ٢ (اعتبر أن المعاينة مع الإرجاع) .

(ثانياً) أحسب الوسط الحسابي والتباين للتوزيع الناتج وقارنهما بالوسط الحسابي والتباين للمجتمع .

الحل :

(اولا) بما أن المعاينة مع الإرجاع أى مع رد كل عدد يؤخذ إلى المجتمع قبل أخذ عدد آخر فإن عدد العينات ذوات الحجم ٢ هو $6 \times 6 = 36$ لأن أى عدد من الأعداد الستة يمكن أن يقترن (في عينة حجمها ٢) بأى عدد من الأعداد الستة (بما في ذلك نفسه) وتكون جميع العينات الممكنة هي :

(١١،١١) ، (١٦،١١) ، (١٥،١١) ، (١٢،١١) ، (١٦،١١) ، (١٤،١١)
 (١١،١٦) ، (١٦،١٦) ، (١٥،١٦) ، (١٢،١٦) ، (١٦،١٦) ، (١٤،١٦)
 (١١،١٢) ، (١٦،١٢) ، (١٥،١٢) ، (١٢،١٢) ، (١٦،١٢) ، (١٤،١٢)
 (١١،١٥) ، (١٦،١٥) ، (١٥،١٥) ، (١٢،١٥) ، (١٦،١٥) ، (١٤،١٥)
 (١١،١٦) ، (١٦،١٦) ، (١٥،١٦) ، (١٢،١٦) ، (١٦،١٦) ، (١٤،١٦)
 (١١،١٤) ، (١٦،١٤) ، (١٥،١٤) ، (١٢،١٤) ، (١٦،١٤) ، (١٤،١٤)

الأوساط الحسابية لهذه العينات هي :

١٢,٥	،	١٣,٥	،	١٣	،	١١,٥	،	١٣,٥	،	١١
١٥	،	١٦	،	١٥,٥	،	١٤	،	١٦	،	١٣,٥
١٣	،	١٤	،	١٣,٥	،	١٢	،	١٤	،	١١,٥
١٤,٥	،	١٥,٥	،	١٥	،	١٣,٥	،	١٥,٥	،	١٣
١٥	،	١٦	،	١٥,٥	،	١٤	،	١٦	،	١٣,٥
١٤	،	١٥	،	١٤,٥	،	١٣	،	١٥	،	١٢,٥

ويمكن تلخيص هذه الأوساط في التوزيع التكرارى المبين بالعمودين الأول والثانى من الجدول (٦ - ١) وهذا هو توزيع المعاينة المطلوب .

الجدول (٦ - ١)

توزيع المعاينة للوسط الحسابي لأوزان ٦ أرناب للمينات من الحجم ٢

خطأ التقدير $K_i (\bar{x}_i - \mu)$	K_i	$\bar{x}_i$
٣ -	١	١١
٥ -	٢	١١,٥
٢ -	١	١٢
٣ -	٢	١٢,٥
٤ -	٤	١٣
٣ -	٦	١٣,٥
٠	٥	١٤
١	٢	١٤,٥
٥	٥	١٥
٦	٤	١٥,٥
٨	٤	١٦
صفر	٣٦	المجموع

(ثانياً) الوسط الحسابي لتوزيع المعاينة $\mu = \frac{1}{N} \sum K_i \bar{x}_i$

$$\text{أى أن } \mu = \frac{1}{36} (16 \times 4 + \dots + 11,5 \times 2 + 11 \times 1)$$

$$= \frac{504}{36} = 14 \text{ أوقية}$$

$$\text{تباين توزيع المعاينة} = \frac{1}{n} \left[\sum_{j=1}^k x_j^2 - \frac{(\sum_{j=1}^k x_j)^2}{n} \right]$$

$$\text{أى أن } \sigma^2 = \frac{1}{36} \left[\frac{504}{36} - (16 \times 4 + \dots + 11,5 \times 2 + 11 \times 1) \right]$$

$$= \frac{11}{9}$$

$$\text{الوسط الحسابى للمجتمع } \mu = \frac{1}{n} (14 + 16 + 15 + 12 + 16 + 11) = 14$$

$$\text{تباين المجتمع } \sigma^2 = \frac{1}{n} \left[\frac{184}{9} - (14 + \dots + 16 + 11) \right] = \frac{11}{9}$$

بالمقارنة نجد ما يلى :

$$(أ) \quad \mu = \bar{\mu}$$

أى أن الوسط الحسابى لتوزيع المعاينة = الوسط الحسابى للمجتمع .

$$(ب) \quad \sigma^2 = \overline{\sigma^2}$$

أى أن تباين توزيع المعاينة = تباين المجتمع مقسوما على حجم العينة .

وهاتان النتيجةتان هما قاعدتان عامتان بالنسبة لتوزيع المعاينة للوسط الحسابى ويمكن إثباتهما رياضيا ، سواء كان المجتمع منتهيا والمعاينة مع الإرجاع أو كان لا نهائيا .

ملاحظة :

الأوساط الحسابية للعينات المأخوذة من مجتمع ما تختلف بطبيعة الحال عن بعضها البعض ، وحين نأخذ أحد هذه الأوساط $\bar{x}$ لتقدير الوسط الحسابى μ للمجتمع يكون هناك خطأ فى التقدير قدره $\bar{x} - \mu$ ففى هذا المثال لدينا $\mu = 14$ وإذا قدرنا هذا المتوسط من متوسط العينة الأولى وهو 11 يكون هناك خطأ قدره $14 - 11 = 3$ وحين نقدره من متوسط العينة الثانية وهو 11,5

يكون هناك خطأ قدره $2(11,5 - 14) = 2,5 - 5 = 0$ (مع ملاحظة أن هناك عيتين متوسط كل منهما $11,5$) وهكذا بالنسبة لأخطاء التقدير من متوسطات العينات الأخرى كما هو مبين بالعمود الثالث من الجدول (٦ - ١)، حيث نلاحظ أيضا أن مجموع أخطاء التقدير هذه يساوى صفرا وهذا هو الذى أدى إلى المساواة (أ).

A statistic

(٦ - ٢ - ١) الإحصاء

في المثال السابق كان لدينا 36 وسطا حسابيا $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_{36}$ وسمينا التوزيع التكرارى لهذه المتوسطات بتوزيع المعاينة للوسط الحسابى للعينات ذوات الحجم 2 . وحين نسحب عينة ما من مجتمع الأرانب فإننا لا نعلم مقدما المتوسط الذى نحصل عليه منها بل يكون ذلك متروكا للصدفة. ولذلك ننظر إلى هذه المتوسطات على أنها قيم مشاهدة من متغير عشوائى $\bar{x}$ ونسمى هذا المتغير حينئذ بالإحصاء $\bar{x}$ ويكون التوزيع سابق الذكر هو توزيع المعاينة لهذه الإحصاءة.

كذلك إذا كانت $\bar{x}_1, \bar{x}_2, \dots$ هى تباينات جميع العينات التى من نفس الحجم التى يمكن أن تؤخذ من مجتمع ما فإننا ننظر إليها كقيم متغير عشوائى $\bar{x}$ ويكون توزيع هذه القيم هو توزيع المعاينة للإحصاءة $\bar{x}$. وبالمثل لأى مقياس إحصائى آخر.

Standard Error

(٦ - ٢ - ٢) الخطأ المعيارى

يسمى الانحراف المعيارى لتوزيع المعاينة لإحصاءة ما بالخطأ المعيارى لهذه الإحصاءة. ففي المثال السابق الخطأ المعيارى للوسط الحسابى أو للإحصاءة $\bar{x}$ هو $\sigma_{\bar{x}} = \sqrt{\frac{11}{4}} = 1,83$ تقريبا. والأخطاء المعيارية تلعب دورا رئيسيا فى نظرية العينات كما ستبين بعد.

نظرا لأن الوسط الحسابي $\bar{x}$ للإحصاءة $\bar{x}$ يساوى الوسط الحسابي μ للمجتمع ، نقول إن $\bar{x}$ هو مُقدَّر غير متحيز للوسط الحسابي للمجتمع unbiased estimator كما نقول للوسط الحسابي $\bar{x}$ لأى عينة عشوائية بأنه تقدير غير متحيز لمتوسط المجتمع ، وأى فرق $\bar{x} - \mu$ يعتبر كما سبق القول خطأ فى تقدير متوسط المجتمع من متوسط العينة ، ويقاس مدى دقة هذا التقدير بالخطأ المعياري $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ لتوزيع المعاينة للإحصاءة $\bar{x}$. لاحظ أن هذا الخطأ يتوقف على كل من σ ، n ويكون هذا الخطأ صغيرا إذا كانت σ صغيرة أو كان حجم العينة n كبيرا ومن ثم كان كبر العينة من العوامل الرئيسية لدقة التقدير .

وبصفة عامة إذا كنا نقدر أحد بارامترات مجتمع من عينة فإن التقدير يوصف بأنه غير متحيز إذا كان متوسط قيم هذا التقدير على جميع العينات العشوائية التى يمكن أخذها من المجتمع يساوى البارامتر الذى نقديره ، ومن الواضح أن هذه الصفة مطلوبة فى التقدير . وكما وجدنا أن $\bar{x}$ تقدير غير متحيز للبارامتر μ نجد أن $E\left(\frac{1}{n} (s - \bar{x})\right) = 0$ تقدير غير متحيز للتباين σ^2 للمجتمع .

(٦ - ٣) المعاينة من مجتمعات معتدلة

Sampling From Normal Populations

إن الهدف الرئيسى من دراسة الإحصاء معرفة كيفية الحكم على المجتمعات وإصدار قرارات عنها عن طريق عينات مأخوذة منها ، وهذا ما سميناه بعملية الاستدلال الإحصائى . ولما كانت هذه العملية تتوقف على معرفتنا بتوزيعات الاحتمال للإحصاءات التى نستخدمها ، وجب علينا أن نحيط بهذه التوزيعات توطئة لتحقيق ذلك الهدف .

ومن أهم توزيعات المعاينة تلك التى تكون فيها المعاينة من مجتمعات معتدلة ،
وسنقدم فى هذا البند عددا من هذه التوزيعات دون التعرض للبراهين الرياضية .

(أولا) توزيع المعاينة للوسط الحسابى

Sampling Distribution of the Mean

إذا كان لدينا مجتمع معتدل وسطه الحسابى μ وانحرافه المعيارى σ فإن توزيع
المعاينة للإحصاءة $\bar{x}$ للعينات ذوات الحجم n المأخوذة من هذا المجتمع يكون
توزيعا معتدلا وسطه الحسابى μ وانحرافه المعيارى $\frac{\sigma}{\sqrt{n}}$ (أى الخطأ المعيارى) ،
وبالتالى فإن الإحصاءة

$$(١) \quad \frac{\mu - \bar{x}}{\sigma / \sqrt{n}} = z_1$$

يكون توزيعها مطابقاً للتوزيع المعتدل المعيارى . راجع الخاصة (د) بالبند
(٤ - ١) . والفروض فى هذه الإحصاءة أن تكون كل من μ ، σ معروفة القيمة .

كما أن الإحصاءة

$$(٢) \quad \frac{\mu - \bar{x}}{\sigma / \sqrt{n}} = z_2$$

يكون توزيعها مطابقاً لتوزيع ت بدرجات حرية عددها $n - ١$. وتستخدم هذه
الإحصاءة حينما تكون σ مجهولة القيمة وهذا ما يحدث فى أغلب الحالات ،
ونضطر حينئذ لاستخدام تباين العينة وهو

$$s^2 = \frac{1}{n-1} (\sum x_i^2 - \frac{(\sum x_i)^2}{n})$$

كتقدير غير متحيز للتباين σ^2 للمجتمع .

(ثانياً) توزيع المعاينة للفرق بين وسطين حسابيين :

نفرض أن لدينا مجتمعين معتدلين وسطاهما الحسابيان μ_1 ، μ_2 وانحرافاهما المعياريان σ_1 ، σ_2 ونفرض أن الإحصاءة $\bar{y}_1$ ترمز إلى الوسط الحسابي لعينة عشوائية حجمها n_1 مأخوذة من المجتمع الأول وأن الإحصاءة $\bar{y}_2$ ترمز إلى الوسط الحسابي لعينة عشوائية حجمها n_2 مأخوذة من المجتمع الثاني . إذا كانت هاتان العينتان مستقلتين فإن توزيع المعاينة للإحصاءة $(\bar{y}_1 - \bar{y}_2)$ يكون توزيعاً

معتدلاً وسطه الحسابي $(\mu_1 - \mu_2)$ وانحرافه المعياري $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ وبالتالي فإن الإحصاءة :

$$(3) \quad \frac{(\bar{y}_1 - \bar{y}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = z$$

يكون توزيعها مطابقاً للتوزيع المعتدل المعياري .

والمفروض هنا أن تكون كل من μ_1 ، μ_2 ، σ_1 ، σ_2 معروفة القيمة .
كما أن الإحصاءة

$$(4) \quad \frac{(\bar{y}_1 - \bar{y}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = z$$

يكون توزيعها مطابقاً لتوزيع ت بدرجات حرية عددها $n_1 + n_2 - 2$

$$(1) \quad \frac{{}_2^2 \sigma_1^2 (1 - \nu_2) + {}_1^2 \sigma_2^2 (1 - \nu_1)}{2 - \nu_2 + \nu_1} = {}_2^2 \sigma_1^2 \text{ حيث}$$

$$(1) \quad \frac{{}_2^2 (\bar{S}_1 - \bar{S}_2) + {}_1^2 (\bar{S}_2 - \bar{S}_1)}{2 - \nu_2 + \nu_1} =$$

$$\text{وحيث } {}_2^2 \sigma_1^2 = {}_1^2 \sigma_2^2$$

أما إذا كان ${}_2^2 \sigma_1^2 \neq {}_1^2 \sigma_2^2$ فتستخدم الاحصاءة

$$(5) \quad \frac{({}_2^2 \mu - {}_1^2 \mu) - ({}_2^2 \bar{S}_1 - {}_1^2 \bar{S}_2)}{\sqrt{{}_2^2 \sigma_1^2 / \nu_2 + {}_1^2 \sigma_2^2 / \nu_1}}$$

التي لها توزيع ت بدرجات حرية $\div {}_2^2 \left(\frac{{}_2^2 \sigma_1^2}{\nu_2} + \frac{{}_1^2 \sigma_2^2}{\nu_1} \right)$

$$\left[(1 - \nu_2) \div {}_2^2 \left(\frac{{}_2^2 \sigma_1^2}{\nu_2} \right) + (1 - \nu_1) \div {}_1^2 \left(\frac{{}_1^2 \sigma_2^2}{\nu_1} \right) \right]$$

ملاحظة :

في المعاينة من أى مجتمع (ليس معتدلاً بالضرورة) نستطيع رياضياً إثبات ما
يلى اعتماداً على النظرية الشهيرة المسماة بنظرية النهاية المركزية central limit
theorem .

(أ) يقترب توزيع الإحصاءة $\bar{S}_1$ المعرفة في (١) من التوزيع المعتدل المعيارى
حين يقترب حجم العينة من اللانهاية .

(ب) يقترب توزيع الإحصاءة $\bar{v}$ المعرفة في (٣) من التوزيع المعتدل المعياري حين تقترب كل من n_1 ، n_2 من اللانهاية .

إن المعنى التطبيقي لهاتين النظريتين أنه عند المعاينة من مجتمع غير معتدل وبشرط أن تكون العينات كبيرة الحجم (أكبر من ٣٠) يجوز عملياً اعتبار أن توزيع كل من $\bar{v}_1$ ، $\bar{v}_2$ هو بالتقريب توزيع معتدل معياري ، وكلما كبر حجم العينات كلما قل الخطأ الناشئ عن هذا التقريب .

(ثالثاً) توزيع المعاينة لخارج قسمة تباينين :

إذا كانت الإحصاءتان $\bar{v}_1$ ، $\bar{v}_2$ ترمزان إلى تباينين عينيتين عشوائيتين مستقلتين مأخوذتين من مجتمعين معتدلين لهما نفس التباين ($\sigma^2 = \sigma^2$) فإن الإحصاءة

$$\frac{\bar{v}_1}{\bar{v}_2} = \bar{v} \quad (٦)$$

يكون توزيعها مطابقاً لتوزيع ف بدرجتي حرية $n_1 - 1$ ، $n_2 - 1$ حيث n_1 ، n_2 هما حجم العينتين . ويلاحظ أنه إذا كانت $\bar{v}_1$ ، $\bar{v}_2$ قيمتين مأخوذتين من عينتين عشوائيتين مستقلتين من مجتمع واحد أو من مجتمعين لهما نفس التباين فإن النسبة $\bar{v}_1 / \bar{v}_2$ تكون قريبة من الواحد لأن كلا من $\bar{v}_1$ ، $\bar{v}_2$ تقدير لنفس التباين σ^2 .

(٦ - ٤) المعاينة من توزيع ذى الحدين :

SAMPLING FROM A BINOMIAL DISTRIBUTION

اعتبر مجتمعاً مقسماً إلى قسمين منفصلين من حيث وقوع أو عدم وقوع حدث معين أ . افرض أن احتمال وقوع هذا الحدث في أى تجربة واحدة هو مقدار ثابت ح واحتمال عدم وقوعه ك = ١ - ح . حسب البند (٣ - ٣) وتحت الشروط

المذكورة يكون توزيع عدد مرات وقوع الحدث في n من التجارب هو توزيع ذي الحدين دليلاً n ، h ، وسطه الحسابي nh وانحرافه المعياري $\sqrt{nh}$ h ويكون التكرار النسبي لوقوع هذا الحدث هو خارج قسمة عدد مرات وقوع الحدث على العدد n وعلى ذلك فإن متوسط هذه النسبة يساوي :

$$\frac{h}{n} = \frac{nh}{n} \quad \text{وانحرافها المعياري} \quad \sqrt{\frac{nh}{n}}$$

وإذا أخذنا من هذا المجتمع جميع العينات ذوات الحجم n وحسبنا في كل منها القيمة $\bar{h}$ لعدد مرات وقوع ذلك الحدث ثم حسبنا نسبة هذا العدد وهو $\bar{h}/n$ فإننا نحصل على توزيع المعاينة لهذه النسبة .

وقد رأينا في البند (٤ - ٦) أنه إذا كانت n كبيرة ولم تكن أى من h أو h/n قريبة من الصفر فإن :

(أ) توزيع الاحتمال لعدد مرات وقوع الحدث في العينات ذوات الحجم n يقترب من توزيع معتدل وسطه الحسابي nh ، وانحرافه المعياري $\sqrt{nh}$ h وبالتالى فإن توزيع الإحصاءة

$$(٧) \quad \frac{\bar{h} - h}{\sqrt{\frac{nh}{n}}} = z_v$$

يقترب من التوزيع المعتدل المعياري مع (٠، ١) .

(ب) وبالمثل فإن توزيع الإحصاءة :

$$(٨) \quad \frac{\bar{h} - h}{\sqrt{\frac{nh}{n}}} = z_v$$

يقترب من التوزيع المعتدل المعياري .

٦ - ٥) اختبارات الفروض : TESTS OF HYPOTHESES

الفرض الإحصائي : هو جملة أو مقولة نذكرها عن مجتمع أو عدة مجتمعات بهدف اختبار صواب أو خطأ هذه الجملة ، ولمعرفة هذا الصواب أو هذا الخطأ نستعين بما يسمى باختبارات الفروض وهي اختبارات تبني على إحصاءات تكون توزيعاتها معروفة لنا مثل تلك التي وردت في البندين (٦ - ٣) و (٦ - ٤) السابقين .

واختبار الفرض هو بكل بساطة قاعدة تؤدي إلى اتخاذ قرار برفض أو قبول الفرض فور حصولنا على قيم مشاهدة في عينة عشوائية . وفي المعتاد نضع فرضاً نرمز له بالرمز F . نسميه **بالفرض الصفري** أو بفرض العدم null hypothesis نصوغه بحيث يعبر عن انعدام الفرق بين شيئين ، مثلاً $\mu_1 = \mu_2$ (أى $\mu_1 - \mu_2 = 0$) أو $\mu = \mu_0$ (أى $\mu - \mu_0 = 0$) أو لا توجد علاقة بين متغيرين . وأى فرض يراد اختباره ضد الفرض الصفري يسمى بالفرض الآخر أو بالفرض البديل $\text{alternative hypothesis}$ ويرمز له بالرمز F وقد يأخذ هذا الفرض الشكل .

$$\mu \neq \mu_0 \text{ أو } \mu < \mu_0 \text{ أو } \mu > \mu_0$$

ولاختبار الفرض الصفري F ضد الفرض الآخر F نبدأ باختيار إحصاءة تناسب الفرض المختبر بشرط أن يكون توزيع المعاينة لهذه الإحصاءة معلوماً لنا ثم نحدد في هذا التوزيع منطقة M يكون احتمال وقوع القيم المشاهدة لهذه الإحصاءة فيها هو احتمال صغير α ، مثلاً 0.05 ، 0.01 ، 0.001 ، 0.0001 وتسمى هذه المنطقة بالمنطقة الحرجة أو بمنطقة الرفض $\text{critical region or region of rejection}$ أما الاحتمال α فيسمى بمستوى الدلالة للاختبار $\text{level of significance}$ وتأخذ القاعدة التي يعطيها الاختبار الصورة الآتية :

« إذا وقعت قيمة مشاهدة لهذه الإحصاءة ، محسوبة على أساس صحة الفرض الصفري F ، داخل المنطقة الحرجة نرفض F عند مستوى الدلالة α وإلا نقبل F . »

وذلك على أساس أن احتمال وقوع القيمة المشاهدة داخل المنطقة الحرجة هو احتمال ضئيل يجعلنا نشك في صحة الفرض الصفري . ونظراً لأن القيم المشاهدة يمكن أن تقع في أى جزء من التوزيع فإن هذه القاعدة تعنى أننا نغامر برفض الفرض الصفري (عند المستوى α) مع علمنا بأننا قد نكون مخطئين في رفضنا هذا ، وإن كان احتمال هذا الخطأ هو احتمال صغير لا يزيد عن α . أما إذا وقعت القيمة المشاهدة في المنطقة المكملة للمنطقة الحرجة ، وتسمى بمنطقة القبول ، فلا يسعنا إلا قبول الفرض ف. على أساس أن احتمال وقوع القيمة المشاهدة فيها هو احتمال كبير $1 - \alpha$.

وجدير بالملاحظة أنه بينما يمكننا الاختبار من هدم الفرض الصفري إلا أنه لا يستطيع أن يثبت صحته وكل ما يستطيع أن يفعله لصالح هذا الفرض هو أن يبين عدم وجود ما يتعارض معه من واقع العينة المستخدمة ، وحين نقول إننا نقبل ف. لا نعنى أننا أثبتنا صحته وإنما نعنى أن ما لدينا من بيانات لا تعطى دليلاً كافياً يدعو إلى رفضه ، وحين نقول إننا نرفض ف. فإننا لا نعنى أننا أثبتنا زيفه وإنما نعنى أن ما لدينا من بيانات لا تدعم صحته بل تشير إلى صحة الفرض الآخر ف.

واختبار الفرض ف. عن مجتمع ضد فرض آخر ف. يستلزم بطبيعة الحال الحصول على عينة عشوائية من هذا المجتمع ، ويفضل اختيار قيمة α حتى قبل اختيار العينة . ويتوقف تحديدنا لقيمة α على طبيعة المشكلة التي نتناولها ودرجة المغامرة التي نقبلها لتحمل مسؤولية الخطأ المحتمل . وبصفة عامة يمكن تلخيص خطوات إجراء اختبار الفروض فيما يلي :

(أولاً) نحدد الفرض الصفري ف. والفرض الآخر ف. من واقع المشكلة التي نتناولها .

(ثانياً) نحدد الإحصاءة التي تناسب الفرض المختبر ونحدد المنطقة الحرجة م في توزيع هذه الإحصاءة بناء على مستوى الدلالة α السابق اختياره .

(ثالثاً) نحسب قيمة الإحصاءة من واقع البيانات المشاهدة في عينة عشوائية وعلى أساس أن الفرض الصفري صحيح .

(رابعاً) الاستنتاج : نتخذ قراراً برفض أو قبول الفرض الصفري ف عند مستوى الدلالة α بحسب وقوع القيمة المحسوبة للإحصاءة داخل أو خارج المنطقة الحرجة .

ويجدر الإشارة هنا إلى أن اختبارات الفروض ما هي إلا وسيلة حسابية تستخدم البيانات التي حصلنا عليها من العينات لإلقاء الضوء على صحة أو خطأ الفرض الصفري ، وينبغي للباحث عند إصدار قراره أن يضيف إلى نتيجة الاختبار جميع ما لديه من معلومات وخبرات وأبحاث سابقة عن موضوع الدراسة .

كما يجدر الإشارة إلى أن حجم العينة يلعب دوراً هاماً في عملية الاستدلال الإحصائي . وحين تكون العينة صغيرة فإن اختبار الدلالة لا يؤدي إلى رفض الفرض الصفري إلا إذا كان هذا الفرض خاطئاً بدرجة كبيرة ، وبالعكس حين تكون العينة كبيرة فإن أى انحراف صغير عن الفرض الصفري يسهل اكتشافه كانحراف ذى دلالة . وبصفة عامة يفضل أن يكون حجم العينة كبيراً بدرجة كافية منعا للوقوع فيما يسمى بالخطأ من النوع الثانى ، وهذا ما سوف نتناوله فى الفصل السابع .

فى البنود (٦ - ٦) ، (٦ - ٧) (٦ - ٨) الآتية نتناول ثلاثة من أشهر اختبارات الفروض تعرف باختبار ت واختبار χ^2 (كاً) واختبار ف .

THE t-TEST

(٦ - ٦) اختبار ت :

هناك عدة إحصاءات تطابق توزيعاتها توزيع ت السابق دراسته . وإذا بني اختبار فروض على أى من هذه الإحصاءات قيل إنه اختبار ت . والصورة العامة لهذه الإحصاءات هي :

ت = $\frac{\text{القيمة المشاهدة للإحصاءة} - \text{الوسط الحسابي للإحصاءة}}{\text{تقدير للخطأ المعياري للإحصاءة}}$

بشرط أن يكون للإحصاءة توزيع معتدل وأن تكون القيمة المشاهدة هي تقدير غير متحيز لمتوسط الإحصاءة .

(٦ - ٦ - ١) اختبار فرض عن الوسط الحسابي لمجتمع معتدل :

مثال (٦ - ٢) :

تقضي التعليمات الحكومية بأن تكون الجرعة القياسية من مستحضر بيولوجي ٦٠٠ وحدة نشاط في السنتيمتر المكعب . اختيرت عينة عشوائية حجمها ١٠ من إنتاج شركة ما ووجد أن متوسطها ٥٩٢,٥ وحدة نشاط في السنتيمتر المكعب بانحراف معياري ١١,٢ وحدة . هل نستطيع القول بأن إنتاج هذه الشركة يتمشي مع التعليمات الحكومية ؟

الحل :

نريد أن نختبر ما إذا كان الوسط الحسابي μ لمجتمع المستحضر الذي تنتجه الشركة يساوي الجرعة القياسية التي حددتها التعليمات الحكومية وهي ٦٠٠ وحدة نشاط / سم^٣. تتبع الخطوات الأربع المشار إليها في البند السابق .

(١) الفرض الصفري $\mu = 600$

الفرض الآخر $\mu \neq 600$

$\alpha = 0,05$ ،

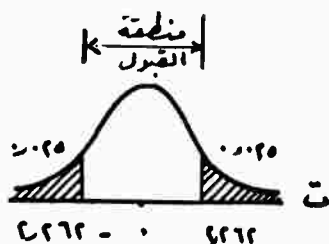
(٢) لنهتدى إلى الإحصاءة المناسبة نذكر أننا نريد اختبار فرض عن قيمة الوسط الحسابي μ للمجتمع عن طريق قيمة الوسط الحسابي $\bar{x}$ لعينة ونتوقع أنه إذا كان هذا الفرض صحيحاً فإن $\bar{x}$ تكون قريبة من μ ، أى أننا نريد اختبار ما إذا كان الفرق بين μ ، $\bar{x}$ هو فرق صغير نغزوه إلى الصدفة أو فرق جوهري (ذو دلالة

يدعونا إلى عدم الثقة في القيمة المفروضة للدليل μ . (هذا على أساس أن العينة هي عينة عشوائية ممثلة للمجتمع تمثيلاً جيداً وبالتالي فإن وسطها الحسابي $\bar{x}$ هو عدد نثق فيه) . هذا التحليل يشير مباشرة إلى أن أنسب إحصاء لقياس صغر أو كبر هذا الفرق هي الإحصاءة (٢) وهي :

$$T = \frac{\bar{x} - \mu}{\sqrt{s^2/n}}$$

التي نعلم أن توزيعها يطابق توزيع ت بدرجات حرية عددها $\nu = n - 1$ ، مع ملاحظة أن هذه الإحصاءة تنعدم عندما تتساوى قيمتي μ ، $\bar{x}$ وتكبر قيمتها المطلقة كلما كبر الفرق بينهما .

$$\{T : |T| > t_{1-\alpha/2}\} = \text{المنطقة المخرجة} = 2$$



الشكل (٦-١)

وهي تتألف من جزئين واقعين أسفل جانبي منحنى ت واحتمال كل منهما $\alpha/2$. وبالتالي فإن احتمال وقوع قيمة الإحصاءة في هذه المنطقة هو احتمال صغير α . (٣) نحسب قيمة هذه الإحصاءة من بيانات العينة وعلى أساس صحة الفرض الصفري . وإذا رمزنا لهذه القيمة بالرمز T_0 نجد أن :

$$T_0 = \frac{592,5 - 600}{\sqrt{10} / 11,2} = -2,12 \quad \text{حيث } \nu = 10 - 1 = 9$$

فإذا كنا قد اخترنا $\alpha = 0,05$ نجد من جدول ت أن القيمة الحرجة التي تحدد منطقة الرفض هي : $t_{0,05} = 2,262$

(٤) الاستنتاج : بما أن $t = -2,12$ لا تقع في المنطقة الحرجة لا يكون لدينا دليل ضد الفرض الصفري عند مستوى الدلالة $0,05$ أى أننا نقبل الفرض أن $\mu = 600$ ونقرر بأن إنتاج الشركة يتمشى مع التعليمات الحكومية .

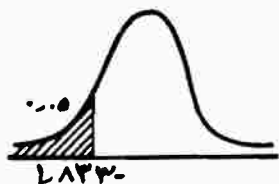
الاختبار ذو الجانب الواحد وذو الجانبين :

في المثال السابق اخترنا الفرض الصفري ف : $\mu = 600$ ضد الفرض ف : $\mu \neq 600$ وعلى ذلك كان لنا أن نرفض ف في حالتين هما : أن تكون $\mu < 600$ أو تكون $\mu > 600$ ولذلك شملت المنطقة الحرجة كلا جانبي توزيع ت . ونقول حينئذ إن الاختبار ذو جانبيين أو إنه اختبار غير اتجاهي .

ولكن إذا كانت زيادة وحدات النشاط في المستحضر البيولوجي عن المستوى القياسي لا ينتج عنها ضرر بينما النقصان عنه يفقد المستحضر فعاليته فإن اهتمامنا يكون منصّباً على ما إذا كان متوسط العينة أصغر صغراً ذا دلالة من المستوى القياسي . في هذه الحالة يكون المطلوب اختبار الفرض الصفري ف : $\mu = 600$ ضد الفرض ف : $\mu > 600$ وتكون المنطقة الحرجة في جانب واحد فقط من التوزيع هو الجانب الأيسر ، أى أن المنطقة الحرجة تكون في هذه الحالة على الصورة ،

$$\{t : t > t_{1-\alpha}\} = \text{منطقة الرفض}$$

→ منطقة القبول ←



الشكل (٦-٢)

ونقول حينئذ إن الاختبار ذو جانب واحد أو إنه اختبار اتجاهى . فإذا كان مستوى الدلالة هو $\alpha = 0,05$ فإن القيمة التي تحد المنطقة الحرجة من اليمين وتجعل المساحة بينها وبين المنحنى $0,05$ يمكن إيجادها من جدول ت باستخدام الاحتمال $\alpha_2 = \alpha = 0,10$.

من الجدول نجد أن ت $1,833 = [0,1]$

إذن القيمة التي تحد المنطقة الحرجة من اليمين هي $0,1,833$ -

بما أن $2,12 > 1,833$ فإن القيمة المشاهدة $-2,12$ تقع في المنطقة الحرجة وبالتالي نرفض الفرض الصفري ف . عند مستوى الدلالة $0,05$ لصالح الفرض الآخر ونقرر أن متوسط إنتاج الشركة يقل عن المستوى الذى تحدده التعليمات الحكومية .

ولا نعجب من اختلاف هذه النتيجة عن النتيجة السابقة لأن كلا منهما يجب على سؤال مختلف هو السؤال الذى تتطلب المشكلة الإجابة عنه .

وبالمثل يمكن أن يكون اهتمامنا منصّباً على الجانب الأيمن فقط من التوزيع حيث تكون المنطقة الحرجة على الصورة $2 = \{ ت : ت > \alpha_2 \}$. وفي تناول أى مشكلة علينا أن نفكر جيداً قبل أن نحدد ما إذا كانت تتطلب اختباراً ذا جانب واحد أو ذا جانبيين تحسباً من الوقوع في خطأ في عملية الاستدلال . وهذا التحديد ينبغى أن يتقرر عند تصميم التجربة وقبل جمع البيانات وبحسب التساؤل الذى تطرحه المشكلة . فمثلاً إذا كنا نقارن أداء مجموعة من الطلاب بمستوى معروف فإن اهتمامنا ينصب على معرفة ما إذا كانت المجموعة أحسن أو أسوأ من هذا المستوى ، وهنا يجب أن يكون الاختبار ذا جانبيين . أما إذا كنا نقارن نوعاً جديداً من الدواء بنوع تقليدى فإن اهتمامنا يكون منصّباً على معرفة ما إذا كان الدواء الجديد أفضل من التقليدى وهنا يجب أن يكون الاختبار ذا جانب واحد .

ملاحظة (١) :

يتطلب استخدام اختبارات بهذه الصورة تحقق الشرطين الآتيين :

- (أ) أن تكون العينة عشوائية بسيطة .
- (ب) أن يكون المجتمع معتدلاً (أو يمكن اعتباره معتدلاً) .

(٦ - ٢ - ٢) فترات الثقة للوسط الحسابي لمجتمع معتدل :

CONFIDENCE INTERVALS

إذا كان الوسط الحسابي μ لمجتمع ما مجهولاً ، يمكن تقديره بواسطة الوسط الحسابي $\bar{x}$ لعينة عشوائية ذات حجم مناسب مأخوذة من هذا المجتمع ، ومثل هذا التقدير يسمى بالتقدير بنقطة point estimation ، ولكن نظراً لأن متوسطات العينات المأخوذة من نفس المجتمع تختلف من عينة إلى أخرى حتي ولو كانت العينات من نفس الحجم فإن هذا التقدير يشوبه بعض الصعوبات خاصة في تحديد مدى الثقة التي نضعها فيه .

ولذلك يفضل في كثير من الأحيان تقدير الوسط الحسابي وغيره من أدلة المجتمع عن طريق ما يسمى بالتقدير بفترة interval estimation حيث تحدد فترة يقع فيها الدليل المجهول بدرجة ثقة معينة .

حينما يكون المجتمع معتدلاً وسحبنا منه عينة عشوائية بسيطة حجمها n ووسطها الحسابي $\bar{x}$ وانحرافها المعياري σ فإن الفترة :

$$(10) \quad \left(\bar{x} - \frac{\sigma}{\sqrt{n}} t_{[1-\alpha/2]}, \bar{x} + \frac{\sigma}{\sqrt{n}} t_{[1-\alpha/2]} \right)$$

تسمى بفترة ثقة بدرجة $(1 - \alpha)$ لمتوسط المجتمع μ . وهذه العبارة لا تعني أن احتمال وقوع μ في هذه الفترة هو $(1 - \alpha)$ لأن μ عدد ثابت لا توزيع له وإنما تفسر هذه العبارة كما يلي :

« إذا سحبنا جميع العينات التي من نفس الحجم وأوجدنا فترة الثقة بدرجة $(1 - \alpha)$ لكل منها فإن $(1 - \alpha)$ من هذه الفترات تشمل البارامتر μ » ، وهذا ما يمكن إثباته رياضياً .

ويسمى العددان $\bar{x} \pm \frac{s}{\sqrt{n}}$ ت $\frac{E}{\sqrt{n}}$ اللذان يحذان فترة الثقة بحدى الثقة

بدرجة $(1 - \alpha)$ أما العدد $\frac{E}{\sqrt{n}}$ فهو تقدير للخطأ المعياري للإحصاءة $\bar{x}$ ، كما تسمى α بمعامل الثقة .

مثال (٦ - ٣) :

أوجد فترة ثقة بدرجة ٩٥٪ للوسط الحسابي μ لمجتمع المستحضر البيولوجي المذكور في المثال (٦ - ٢) من واقع البيانات المعطاة في هذا المثال .

الحل :

لدينا $n = 10$ ، $\bar{x} = 592,5$ ، $E = 11,2$ ، $1 - \alpha = 0,95$ ، ومنها $\alpha = 0,05$

$$t_{[9],0,05} = 2,262 ، \frac{E}{\sqrt{n}} = \frac{11,2}{\sqrt{10}} = 3,54$$

بالتعويض في (١٠) نجد أن :

$$\text{الحد الأدنى للفترة} = 592,5 - 2,262 \times 3,54 = 584,49$$

الحد الأعلى للفترة = $592,5 + 3,54 \times 2,262 = 600,51$ هي فترة ثقة بدرجة ٩٥٪ لمتوسط المجتمع .

ملاحظة (٢) :

في المثال (٦ - ٢) قبلنا الفرض أن $\mu = 600$ ضد الفرض $\mu \neq 600$ عند مستوى الدلالة ٠,٠٥ ، ونلاحظ أن العدد ٦٠٠ يقع في فترة الثقة التي أوجدناها في المثال (٦ - ٣) ، والواقع أن أى عدد نفرضه عن متوسط المجتمع μ يكون مقبولا عند المستوى ٠,٠٥ طالما كان واقعاً في هذه الفترة . تحقق من ذلك باختيار بعض القيم المناسبة واختبارها كما بالبند (٦ - ٦ - ١) .

(٦ - ٦ - ٣) مقارنة متوسطي مجتمعين معتدلين :

(أ) اختبار دلالة الفرق بين متوسطي عينتين مستقلتين مأخوذتين من مجتمعين معتدلين .

مثال (٦ - ٤) :

في دراسة عن نبات حنك السبع كان المطلوب مقارنة الوسط الحسابي للنباتات ذات الزهور الحمراء من النوع Suttons Eclipse بالوسط الحسابي للنباتات ذات الزهور البيضاء من النوع Suttons Intermediate White وقد نتجت البيانات الآتية من عينتين عشوائيتين مستقلتين .

ذات الزهور الحمراء : $\bar{x} = 43$ ، $s^2 = 134,76$ ، $n = 32,35$

ذات الزهور البيضاء : $\bar{x} = 21$ ، $s^2 = 169,50$ ، $n = 62,93$

اختبر ما إذا كان الفرق المشاهد بين متوسطي العينتين يرجع إلى تقلبات العينات أو إلى وجود فرق حقيقي بين المتوسطين μ_1 و μ_2 للمجتمعين اللذين سحبت منهما العينتان . افترض اعتداله احتمليين و تساوى تباينهما وخذ $\alpha = 0,05$

الحل :

الفرض الصفري $F_0: \mu_1 = \mu_2$ (لا يوجد فرق بين متوسطي المجتمعين)
 الفرض الآخر $F_1: \mu_1 \neq \mu_2$ (اختبار ذو جانبيين)
 الإحصاء المناسبة هنا هي الإحصاءة (٤) وهي :

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim \epsilon$$

التي نعلم أن توزيعها يطابق توزيع ت بدرجات حرية عددها $\nu = n_1 + n_2 - 2$

$$\frac{\epsilon^2 (1 - \alpha) + \epsilon^2 (1 - \alpha)}{2 - \alpha + \alpha} = \epsilon^2$$

على فرض أن تبايني المجتمعين متساويان

المنطقة الحرجة $\alpha = \{ |T| : |T| < t_{\alpha/2, \nu} \}$
 نحسب T من بيانات العينة وعلى أساس صحة الفرض الصفري .

$$\epsilon^2 = \frac{62,93 \times 20 + 32,35 \times 42}{62} = 1986,4135$$

$$= 1986,4135$$

$$\therefore \epsilon = 44,569$$

$$T = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim \epsilon = 44,569 = \sqrt{\frac{1}{21} + \frac{1}{43}} \times 11,855$$

$$T = \frac{169,50 - 134,76}{11,855} = 2,93 \text{ حيث } \nu = 2 - 21 + 43 = 62$$

من جدول ت نجد أن $t_{0,0005, 62} = 3,858$

بما أن $-2,93 > -2$ فإن القيمة $t_{\alpha/2} = -2,93$ تقع في المنطقة الحرجة (في الجانب الأيسر) وإذن نرفض الفرض الصفري $F: \mu_1 = \mu_2$ عند مستوى الدلالة $0,05$ ونقرر أن العينتين مأخوذتان من مجتمعين مختلفين في وسطهما الحسابي.

ملاحظة (٣)

يتطلب استخدام اختبار t بهذه الصورة تحقق الشروط الآتية :

- (أ) أن تكون كل من العينتين عشوائية بسيطة .
- (ب) أن تكون كل من العينتين مأخوذة من مجتمع معتدل .
- (ج) أن تكون العينتان مستقلتين .
- (د) أن يكون تباينا المجتمعين متساويان .

إذا لم تكن العينتان مستقلتين فإن اختبار t للمقارنة بين متوسطي المجتمعين يأخذ الصورة التي سترد في البند (٨ - ٧ - ١) القادم .

فترات الثقة للفرق بين متوسطي مجتمعين معتدلين :

يمكن إثبات أنه تحت الشروط المذكورة ، يصلح العدان :

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2} \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}} \quad (١١)$$

كحدين لفترة ثقة بدرجة $(1 - \alpha)$ للفرق $(\mu_1 - \mu_2)$ بين متوسطي المجتمعين،

حيث $t_{\alpha/2}$ = $t_{\alpha/2, n_1+n_2-2}$ هو تقدير للخطأ المعياري للإحصاءة $\bar{x}_1 - \bar{x}_2$.

ففي المثال (٦ - ٤) نجد أن :

$$\text{الحد الأدنى لفترة الثقة} = (١٦٩,٥ - ١٣٤,٧٦) - ٢ \times ١١,٨٥٥ = ١١,٠٣$$

$$\text{الحد الأعلى لفترة الثقة} = (١٦٩,٥ + ١٣٤,٧٦) - ٢ \times ١١,٨٥٥ = ٥٨,٤٥$$

أى أننا يمكن أن نأخذ الفترة (١١,٠٣ ، ٥٨,٤٥) كفترة ثقة بدرجة ٩٥٪ للفرق بين متوسطى المجتمعين .

تمارين (٦ - ١)

في المسائل الآتية اعتبر أن المجتمعات معتدلة وأن العينات عشوائية ومستقلة وأن مستوى الدلالة $\alpha = ٠,٠٥$ ما لم يذكر غير ذلك .

(١) وزنت ١٩ عذراء جرادة فكانت الأوزان كما يلي بالجرامات :

$$٠,٦٧ \quad ٠,٩٥ \quad ١,١٨ \quad ١,٠٥ \quad ١,٢٢ \quad ٠,٨٢ \quad ٠,٩٣ \quad ٠,٦٨ \quad ٠,٥٢ \quad ٠,٧٠$$

$$٠,٥٧ \quad ٠,٥٩ \quad ٠,٣٥ \quad ٠,٩٢ \quad ٠,٦٧ \quad ٠,٥٤ \quad ١,١٢ \quad ٠,٨٩ \quad ٠,٨٤$$

(أ) اختبر الفرض أن الوسط الحسابي للوزن في مجتمع هذه الحشرة هو

$$\mu = ١ \text{ ضد كل من الفرضين الآتين (١) } \mu \neq ١, \text{ (٢) } \mu > ١$$

(ب) أوجد فترة ثقة بدرجة ٩٥٪ للدليل μ .

(٢) باستخدام عينة حجمها $n = ١٥$ ووسطها الحسابي $\bar{x} = ٨٥,٣$ وتباينها

$\sigma^2 = ١٣$ اختبر الفرض أن متوسط المجتمع $\mu = ٨٠$ وأوجد فترة ثقة بدرجة

٩٩٪ لهذا المتوسط .

(٣) في عينة عشوائية من ٢٠ شخصاً يعالجون من مرض معين وجد أن عدد

الكريات الدموية البيضاء مقدرة بالآلاف كالآتي . أوجد فترة ثقة بدرجة ٩٥٪

لمتوسط عدد الكريات البيضاء في مجتمع هؤلاء المرضى .

$$٨ \quad ٧ \quad ١٠ \quad ٩ \quad ٧ \quad ١٢ \quad ٤ \quad ٦ \quad ١٣ \quad ٨ \quad ٧ \quad ١٣ \quad ٨$$

$$١١ \quad ٥ \quad ١٠ \quad ١٥ \quad ١٣ \quad ١٢ \quad ١٠$$

(٤) في بحث ثبات reliability إحدى طرق القياس في تجارب التبريد وجدت القيم الآتية لدرجات الحرارة في عيتين أ ، ب :

أ : ١٠٦,٩ ١٠٦,٣ ١٠٧,٠ ١٠٦,٠ ١٠٤,٩

ب : ١٠٦,٥ ١٠٦,٧ ١٠٦,٨ ١٠٦,١ ١٠٥,٦

هل الفرق بين متوسطى العيتين ذو دلالة عند المستوى ٠,٠٥ ؟
(لتسهيل حساب التباين اطرح ١٠٠ من جميع الأعداد) .

(٥) أخذت عيتان متكافئتان من الأبقار ووضعتا تحت نفس الظروف فيما عدا أنهما أعطيتا نوعين مختلفين من الغذاء ، فكانت الزيادة في الأوزان بالأرطال بعد شهرين كما يلي :

العينة الأولى : ٣٣ ٦٦ ٢٦ ٤٣ ٤٦ ٥٥ ٥٤

العينة الثانية : ٥٣ ٥٣ ٣٧ ٧٣ ٥٨ ٦١ ٣٨

(أ) هل يمكن القول بأن نوعى الغذاء لا يختلفان في متوسط الزيادة في وزن الأبقار ؟

(ب) أوجد فترة ثقة بدرجة ٩٥٪ للفرق بين متوسطى المجتمعين .

(٦) طرح فرض بأن متوسط عدد الزهور الشعاعية ray florets في زهرة ما هو $\mu = ٥٠$. أخذت عينة حجمها ٥١ من هذه الزهرة فوجد أن متوسط عدد الزهور الشعاعية ٤٤,٣٣ بانحراف معيارى ٥,١٦٧

(أ) اختبر الفرض أن $\mu = ٥٠$

(ب) اختبر الفرض أن $\mu > ٥٠$

(ج) أوجد فترة ثقة بدرجة ٩٥٪ للدليل μ

(د) أوجد فترة ثقة بدرجة ٩٩٪ للدليل μ

(٦ - ٦ - ٤) اختبار استقلال الأحداث النادرة :

نعلم أنه في توزيع بواسون يتساوى التباين والوسط الحسابي أى أن النسبة بينهما تساوى الواحد الصحيح - راجع الخاصة (٩) من البند (٣ - ٤) . وفي تناولنا للأحداث النادرة في البندين (٣ - ٤ - ٢) ، (٣ - ٤ - ٣) علمنا أن الأحداث النادرة تتوزع بواسونياً بشرط أن تقع مستقلة عن بعضها بمعنى أن يكون وقوع هذه الأحداث غير متأثر إلا بالعوامل العشوائية وحدها . وفي كثير من التطبيقات يكون من المرغوب فيه الكشف عن هذه الاستقلالية ثم بحث ما قد يوجد من أنماط نتيجة لعدم توفر هذه الخاصة - راجع البند (٣ - ٤ - ٣) . ولتحقيق هذا الغرض نحصل على عينة عشوائية من تلك الأحداث وليكن حجمها n ووسطها الحسابي $\bar{x}$ وتباينها s^2 ونضع الفرض الصفري أن هذه الأحداث تقع عشوائياً مستقلة عن بعضها وبالتالي تتوزع بواسونياً وتكون النسبة بين التباين والوسط الحسابي في المجتمع هي $1 = \frac{s^2}{\bar{x}}$ فإذا كان هذا الفرض صحيحاً فإن النسبة $\frac{s^2}{\bar{x}} = 1$ / $\bar{x}$ المشاهدة في العينة ينبغي أن تكون قريبة من الواحد الصحيح ، أما إذا كان هذا الفرض خاطئاً فإن الفرق بين هذه النسبة والواحد يكون فرقاً ذا دلالة .

ولاختبار صحة الفرض $1 = \frac{s^2}{\bar{x}}$ نستخدم الإحصاء

$$T = \frac{1 - \frac{s^2}{\bar{x}}}{\sqrt{\frac{2}{n-1}}} \quad (12)$$

التي يمكن إثبات أن توزيعها يطابق توزيع T بدرجات حرية عددها $n - 1$. ويجمل بنا أن نستخدم اختباراً ذا جانب واحد لأن المطلوب هنا معرفة ما إذا كانت

ح أكبر من الواحد (أو أقل من الواحد) للمساهمة في تفسير ما قد يوجد من أنماط .

مثال (٦ - ٥) :

اعتبر عينة عشوائية حجمها ٤٠٠ من الأحداث النادرة . اختبر عشوائية (استقلال) هذه الأحداث عند مستوى الدلالة ٠,٠٥ في كل من الحالتين الآتيتين :

(أ) إذا كان الوسط الحسابي للعينة ١,٨ والتباين ١,٩٦٥ .

(ب) إذا كان الوسط الحسابي للعينة ٣,٣ والتباين ٢,١٥ .

الحل :

$$(أ) \chi^2 = \frac{1,965}{1,8} = 1,092$$

الفرض الصفري H_0 : $\chi^2 = 1$

الفرض الآخر H_1 : $\chi^2 < 1$

من (١٢) نجد أن :

$$T = \frac{1 - 1,092}{\sqrt{\frac{1}{399}}} = \frac{0,092}{0,071} = 1,296$$

من الجدول (٧) نجد أن $T_{0,1,399} = 1,645$

بما أن $1,296 < 1,645$ لا يكون لدينا دليل ضد الفرض الصفري H_0 ولا يسعنا إلا قبوله . وهذا يعني أن الأحداث تتوزع بواسونياً وبالتالي تقع عشوائياً مستقلة عن بعضها البعض .

$$(ب) \chi^2 = \frac{2,15}{3,3} = 0,652$$

ف. : ح = ١

ف. : ح > ١

$$ت. = \frac{١ - ٠,٦٥٢}{٠,٠٧١} = ٤,٩٠١$$

بما أن $٤,٩٠١ > ١,٦٤٥$ نرفض الفرض الصفري عند المستوى $٠,٠٥$ ونستنتج أن الأحداث لا تتوزع بواسونياً وبالتالي لا تقع مستقلة عن بعضها البعض ، وأغلب الظن يكون هناك نمط من نوع التنافر .

THE CHI—SQUARE TEST

(٦ - ٧) اختبار χ^2

في كثير من الأبحاث نتناول مجتمعاً صنفنا عناصره بحسب خاصية أو صفة ما إلى م من الأقسام المنفصلة $١, ٢, ٣, \dots$ ، أم ويكون المطلوب اختبار ما إذا كان هذا المجتمع له نموذج معين من حيث هذه الخاصية ، أى يكون المطلوب اختبار الفرض أن التكرارات النسبية في هذه الأقسام (في المجتمع) لها مقادير معينة ح. ١ ، ح. ٢ ، ... ، ح. م على الترتيب .

إذا كان هذا الفرض صحيحاً فإننا لو سحبنا عشوائياً من المجتمع عنصراً واحداً لكان احتمال انتماء هذا العنصر إلى القسم ١ هو ح. ١ . واحتمال انتائه إلى القسم ٢ هو ح. ٢ . واحتمال انتائه إلى القسم ٣ هو ح. ٣ . وإذا سحبنا عينة عشوائية حجمها n فإننا نتوقع أن يكون n ح. ١ من عناصرها منتبهاً إلى القسم ١ ، n ح. ٢ منها منتبهاً إلى القسم ٢ ، n ح. ٣ منها منتبهاً إلى القسم ٣ . نفرض أننا قمنا فعلاً بسحب عينة عشوائية حجمها n ووجدنا أن عدد العناصر التي انتبعت إلى الأقسام $١, ٢, ٣, \dots$ هي $١, ٢, ٣, \dots$ على الترتيب $١, ٢, ٣, \dots$. إن هذه الأعداد نسمي بالتكرارات المشاهدة ، أما الأعداد $١, ٢, ٣, \dots$ نسميها بالتكرارات المتوقعة ، وهي تلك التي نتوقعها إذا كان الفرض صحيحاً .

اختبار الفرض المذكور يمكن أن يؤسس على المقارنة بين هذين النوعين من التكرارات .

والإحصاء المناسبة لذلك هي :

$$ص_1 = \frac{(ك_1 - ح_1)}{ح_1} \quad \text{حيث } ح_1 = ح_2 \quad (١٣)$$

$$\text{وحيث } ح_1 = ح_2 = ح \quad (١٤)$$

التي يمكن إثبات أن توزيعها يقترب من توزيع χ^2 بدرجات حرية (م - ١) كلما اقتربت ح من اللانهاية .

ولصحة استخدام هذه الإحصاء ينبغي أن تتحقق الشروط الآتية :

(أ) أن تكون العينة عشوائية .

(ب) ألا يقل حجم العينة عن ٤٠ (لأن توزيع الإحصاء $ص_1$ تقريبي) .

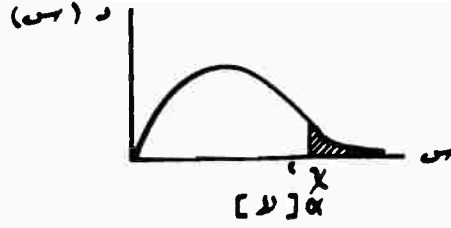
(ج) ألا تقل أى $ك_1$ عن خمسة ، على أنه إذا وجدت قيمة $ك_1$ تقل عن خمسة تضم هذه القيمة إلى قيمة مجاورة لها ، ونفعل ذلك لقيم $ك_1$ المناظرة .

وجدير بالملاحظة أن (١٣) لا تساوى الصفر إلا إذا كانت $ك_1 = ح_1$ لجميع ر وتزيد قيمتها كلما كبرت الفروق بين $ك_1$ ، $ح_1$.

إن الفرض الصفري ف. هنا هو عدم وجود فرق بين التكرارات المشاهدة $ك_1$ والتكرارات $ح_1$ المتوقعة لها . أما الفرض الآخر ف. فهو أن هنالك تفاوتاً بين هذين النوعين من التكرارات يجعل القيمة المشاهدة للإحصاء (١٣) كبيرة كبراً ذا دلالة . وعلى ذلك تأخذ المنطقة الحرجة الصورة الآتية :

$$\{ \chi^2_{[v]\alpha} < \chi^2 : \chi^2 \} = \alpha$$

حيث v ترمز إلى درجات الحرية .



الشكل (٦-٣)

حساب درجات الحرية :

نعلم أن توزيع χ^2 يتوقف كلية على عدد درجات الحرية ν . ولايجاد هذا العدد نحسب عدد المقارنات المستقلة بين $k_1, \dots, k_r$ (حيث $r = 1, 2, \dots$) ونظراً لأن حجم العينة n هو مقدار ثابت معروف مقدماً فإن التكرارات k_r (أو q_r) لا تكون جميعها مستقلة إذ نستطيع إذا عرفنا $(m - 1)$ منها أن نعرف من المتساوية (١٤) التكرار الباقي . وعلى هذا فإن عدد المقارنات المستقلة هو $m - 1$. فإذا كانت القيمة المشاهدة للإحصاءة (١٣) قد حسبت اعتماداً على حجم العينة فقط فإن $\nu = m - 1$.

وكقاعدة عامة كل قيد خطي تفقيد به القيم الداخلة في تركيب الإحصاءة (١٣) ينقص واحداً من درجات الحرية ، ولقد كان ثبات حجم العينة هو أحد هذه القيود . وإذا كنا قد احتجنا لحساب القيم المتوقعة إلى واحد أو أكثر من أدلة المجتمع المجهولة واضطررنا إلى تقديرها من العينة فإننا نكون قد قيدنا أنفسنا بالعينة التي بأيدينا وبالتالي فإن كلا من هذه التقديرات يشكل قيداً على التوزيع ينقص واحداً من درجات الحرية . وكطريقة إجرائية سهلة ، نحسب عدد درجات الحرية كالآتي :

$\nu = \text{عدد المقارنات المستقلة} - \text{عدد الأدلة التي قدرت من العينة}$ (١٥)

فيما يلي ثلاثة تطبيقات تستخدم اختبار χ^2 بالصيغة المبينة في (١٣) .

(٦ - ٧ - ١) اختبار فرض عن توزيع مجتمع :

مثال (٦ - ٦) :

في أحد تجارب مندل الشهيرة في تهجين النباتات نتج من نبات البسلة ما يلي :
٣١٥ نباتاً مستديراً أصفراً ، ١٠٨ مستديراً أخضر ، ١٠١ مجعداً أصفراً ، ٣٢
مجعداً أخضر . وطبقاً لنظرية مندل يجب أن تكون أعداد هذه الأنواع متناسبة
مع ٩ : ٣ : ٣ : ١ . هل هناك دليل يدعو إلى الشك في نظرية مندل من واقع
هذه البيانات ؟ استخدم مستوى الدلالة ٠,٠٥ .

الحل :

لدينا ٥٥٦ نباتاً (حجم العينة) .

إذا كانت نظرية مندل صحيحة ، تتوزع هذه النباتات بالنسب ٩ : ٣ : ٣ : ١ :
كالآتي :

$$\text{مستدير أصفر} = \frac{9}{16} \times 556 = 312,75$$

$$\text{مستدير أخضر} = \frac{3}{16} \times 556 = 104,25$$

$$\text{مجعد أصفر} = \frac{3}{16} \times 556 = 104,25$$

$$\text{مجعد أخضر} = \frac{1}{16} \times 556 = 34,75$$

التكرارات المشاهدة لـ : ٣١٥ ١٠٨ ١٠١ ٣٢

التكرارات المتوقعة لـ : ٣١٢,٧٥ ١٠٤,٢٥ ١٠٤,٢٥ ٣٤,٧٥

الفرض الصفري H_0 : نظرية مندل صحيحة .

الفرض الآخر H_1 : نظرية مندل غير صحيحة وبالتالي تكون القيمة

المشاهدة للإحصاء (١٣) كبيرة كبراً ذا دلالة .

$$\chi^2 = \frac{(34,75-32)^2}{34,75} + \frac{(104,25-101)^2}{104,25} + \frac{(104,25-108)^2}{104,25} + \frac{(312,75-315)^2}{312,75} = 0,47$$

$$\chi^2 = 1 - \epsilon = 0.3$$

$$\chi^2 = 7.82 \text{ نجد أن } \chi^2_{[3], 0.05} = 7.82$$

العدد ٠,٤٧ يقل كثيراً عن ٧,٨٢ (لا يقع في منطقة الرفض) وإذن لا يمكن رفض الفرض الصفري عند مستوى الدلالة ٠,٠٥ وبالتالي نقبله ونقرر أنه لا يوجد دليل يدعو إلى الشك في نظرية مندل وأن نتيجة هذه التجربة تدعم هذه النظرية .

مثال (٦ - ٧) :

المفروض في جداول الأعداد العشوائية أن تظهر الأرقام ٠ ، ١ ، ٢ ، ٣ ، ٤ ، ٥ ، ٦ ، ٧ ، ٨ ، ٩ باحتمالات متساوية . بفحص إحدى صفحات أحد هذه الجداول وجد التوزيع الآتي :

الرقم	٠	١	٢	٣	٤	٥	٦	٧	٨	٩
التكرار	١٧	٣١	٢٩	١٨	١٤	٢٠	٣٥	٣٠	٢٠	٣٦

اختبر الفرض أن الأرقام العشرة تظهر باحتمالات متساوية مستخدماً مستوى الدلالة ٠,٠١

الحل :

مجموع الأرقام التي بالصفحة = ١٧ + ٣١ + ٢٠ + ٣٦ = ١٠٤ رقماً
إذا كانت الأرقام العشرة تظهر باحتمالات متساوية فإن كلا منها يجب أن يظهر ١٠.٤ = ١٠٤ ÷ ١٠ مرة .

أي أن التكرارات المتوقعة على أساس صحة هذا الفرض هي ١٠.٤ لكل من الأعداد العشرة .

$$\frac{(25-36)^2}{25} + \dots + \frac{(25-31)^2}{25} + \frac{(25-17)^2}{25} = \chi^2$$

$$23,3 = \text{حيث } \nu = 10 - 1 = 9$$

$$21,666 = \chi^2_{[9],0,01}$$

بما أن 23,3 أكبر من 21,666 نرفض الفرض الصفري عن تساوى احتمالات ظهور الأرقام العشرة عند مستوى الدلالة 0,01 ويتضمن هذا أن هناك شك في دقة هذا الجدول .

(٦ - ٧ - ٢) اختبار حسن المطابقة

TEST OF GOODNESS OF FIT

يقصد بهذا الاختبار تحديد مدى ملاءمة توزيع نظري معروف مثل ذى الحدين وبواسون والمعتدل ، لتوزيع تكرارى مشاهد في عينة ، بهدف التحقق من صلاحيته كتوزيع للمجتمع الذى أخذت منه العينة .

مثال (٦ - ٨) :

اختبر الفرض القائل أن إنجاب الذكور وإنجاب الإناث متساويا الاحتمال من واقع البيانات المعطاة في المثال (٣ - ٥) عن العائلات ذوات الأربعة الأطفال وهى :

عدد الذكور في العائلة	٥	٤	٣	٢	١	٠	س :
عدد العائلات	١٣	٤٣	٩٢	١١٢	٦٠	ك :	

الحل :

على فرض صحة الفرض الصفري أن إنجاب الذكور وإنجاب الإناث متساويا الاحتمال يكون احتمال إنجاب الذكور $H = \frac{1}{2}$ ويكون توزيع عدد الذكور في العائلات

ذوات الأربعة الأطفال هو توزيع ذى الحدين دليلاه ٤ ، $\frac{1}{4}$. وقد وجدنا في المثال (٣ - ٥) أن هذا التوزيع هو :

عدد الذكور في العائلة سر : ٠ ١ ٢ ٣ ٤
عدد العائلات سر : ٢٠ ٨٠ ١٢٠ ٨٠ ٢٠

$$\chi^2 = \frac{(20-60)^2}{20} + \frac{(80-112)^2}{80} + \frac{(120-92)^2}{120} + \frac{(80-43)^2}{80} + \frac{(20-13)^2}{20} = 118,89$$

$$118,89 = \chi^2 \text{ حيث } \nu = 5 - 1 = 4$$

$$\chi^2_{[4],0,01} = 13,277$$

بما أن $118,89 < 13,277$ نرفض صحة الفرض أن احتمال إنجاب الذكور يساوى احتمال إنجاب الإناث ، وذلك عند مستوى الدلالة ٠,٠١ .

مثال (٦ - ٩) اختبار الاعتدالية : TEST OF NORMALITY

اختبر ما إذا كان من الممكن اعتبار العينة المعطاة في المثال (٤ - ٤) مأخوذة من مجتمع معتدل .

الحل :

في ذلك المثال وفقنا توزيعاً معتدلاً للتوزيع المعطى ووجدنا ما يلي :

٨	١٥	١٦	١٤	١١	٦	٦	التكرارات المشاهدة
١٢,٤٧	١٥,٥٧	١٣,٦٦	١٢,٩٨	٩,٦١	٦,٥٤	٦,٨١	التكرارات المتوقعة
٦	٨	١٠					التكرارات المشاهدة
٦	٣٧	٤٤					التكرارات المتوقعة

$$\chi^2 = \frac{(6,55-6)^2}{6,55} + \frac{(6,37-8)^2}{6,37} + \dots + \frac{(6,54-6)^2}{6,54} + \frac{(6,81-6)^2}{6,81} = 2,9427$$

بما أننا كنا قد قدرنا من العينة اثنين من أدلة المجتمع هما الوسط الحسابي والانحراف المعياري فإن عدد درجات الحرية $\nu = (10 - 1) = 9$.

$$\chi^2_{0.05, 9} = 14,067$$

بما أن $2,9427 < 14,067$ لا نستطيع رفض الفرض الصفري عن اعتدالية المجتمع عند مستوى الدلالة 0,05.

ملاحظة (١) :

إن عيب اختبار χ^2 لحسن المطابقة أنه لا يهتم بإشارات الفروق بين التكرارات المشاهدة h_i والتكرارات النظرية e_i (لأنه يأخذ مربعات هذه الفروق) ففي توفيق توزيع معتدل مثلاً قد يحدث أن تكون قيم h_i أصغر من قيم e_i في الجزء الأوسط من التوزيع وتكون أكبر منها في الطرفين وهذا يعني أن التوزيع المشاهد أكثر انبساطاً من التوزيع المعتدل أي يأخذ شكلاً يختلف عن شكل التوزيع المعتدل. ومع هذا قد تكون قيمة χ^2 ليست بذات دلالة وتدعونا إلى اعتبار أن المجتمع معتدل. ولهذا لا يكون الاختبار مناسباً إلا إذا لم يكن هناك نمط معين للفروق بين h_i و e_i .

(٦ - ٧ - ٣) اختبار استقلال خاصيتين :

TEST OF INDEPENDENCE

في بعض الدراسات نتناول مجتمعاً صنفت عناصره بحسب خاصية ما إلى م من الأقسام المنفصلة $A_1, A_2, \dots, A_m$ وصنفت من ناحية ثانية بحسب خاصية أخرى إلى هـ من الأقسام المنفصلة $B_1, B_2, \dots, B_h$ ويكون المراد اختبار ما إذا كانت هاتان الخاصتان مستقلتين، بمعنى أن يكون توزيع إحدهما غير متأثر بالأخرى.

في هذه الحال تؤخذ عينة عشوائية وتدرس من حيث عدد العناصر التي تظهر في أقسام الخاصتين وتوضع التكرارات الناتجة بحيث تقترن التكرارات في أقسام الخاصة الأولى بالتكرارات في أقسام الخاصة الثانية فيما يسمى بمجدول الاقتران contingency table وهو عبارة عن مصفوفة ذات م صفا تمثل أقسام الخاصة الأولى ، ه عموداً تمثل أقسام الخاصة الثانية ويوصف الجدول حينئذ بأنه جدول اقتران 2×5 . نضع الفرض الصفري أن الخاصتين مستقلتان ، وعلى أساس صحة هذا الفرض نحسب التكرارات النظرية المناظرة للتكرارات المشاهدة ونضعها في جدول اقتران 2×5 أيضا . بمقارنة التكرارات المشاهدة بالتكرارات النظرية نستطيع الحكم على استقلال الخاصتين بواسطة الإحصاء المعرفة في (١٣) .

مثال (٦ - ١٠) :

في تجربة عن الرجال في الأعمار بين ٦٠ ، ٦٤ عاماً صنف الرجال من حيث خاصتي التدخين والوفاة ، وقسمت خاصة التدخين إلى قسمين هما « يدخن ولا يدخن » وقسمت خاصة الوفاة إلى قسمين : رجال لا يزالون على قيد الحياة ورجال توفوا في بحر ٦ سنوات من بدء التجربة . في عينة من ١٤٦٩ رجلا جمعت البيانات في جدول الاقتران 2×2 الآتي :

المجموع	عادة التدخين		الوفاة
	لا يدخن	يدخن	
١٧١	١١٧	٥٤	توفي
١٢٩٨	٩٥٠	٣٤٨	حي
١٤٦٩	١٠٦٧	٤٠٢	المجموع

هل هذه البيانات تشير إلى أن الوفاة مستقلة عن عادة التدخين ؟

الحل :

ف : الوفاة وعادة التدخين خاصتان مستقلتان ، أو لا توجد علاقة بين الخاصتين .

إذا كان هذا الفرض صحيحاً تكون نسبة المتوفين إلى الأحياء واحدة في المجتمع سواء للمدخنين أو لغير المدخنين . ونظراً لأن هذه النسبة غير معروفة نضطر إلى تقديرها من العينة ، ومن المعقول أن نأخذ النسبة التي ظهرت في العينة بين العدد الكلي للمتوفين والعدد الكلي للأحياء وهي ١٧١ : ١٢٩٨ . وباستخدام هذه النسبة نحصل على التكرارات النظرية في الخلايا الأربع وذلك بتقسيم كل من عدد الذين يدخنون وهو ٤٠٢ وعدد الذين لا يدخنون وهو ١٠٦٧ بهذه النسبة فمثلاً :

$$\text{التكرار النظري لعدد المتوفين من المدخنين} = ٤٠٢ \times \frac{١٧١}{١٤٦٩} = ٤٦,٨ \text{ رجلاً}$$

$$\text{التكرار النظري لعدد الأحياء من المدخنين} = ٤٠٢ \times \frac{١٢٩٨}{١٤٦٩} = ٣٥٥,٢ \text{ رجلاً}$$

وبالمثل بالنسبة لغير المدخنين . وبذلك نحصل على جدول التكرارات النظرية الآتي :

الوفاة	عادة التدخين		المجموع
	لا يدخن	يدخن	
توفي	١٢٤,٢	٤٦,٨	١٧١
حي	٩٤٢,٨	٣٥٥,٢	١٢٩٨
المجموع	١٠٦٧	٤٠٢	١٤٦٩

$$\chi^2 = \frac{(٩٤٢,٨ - ٩٥٠)^2}{٩٤٢,٨} + \dots + \dots + \frac{(٤٦,٨ - ٥٤)^2}{٤٦,٨} = ١,٧٣$$

نحسب عدد درجات الحرية كالآتي . لدينا ٤ مقارنات ، وبما أن عدد المدخنين ثابت (وهو ٤٠٢) وكذلك عدد غير المدخنين (وهو ١٠٦٧) فإن درجات الحرية تنقص ٢ ، وبما أننا قدرنا دليل واحد للمجتمع من العينة وهو النسبة ١٧١ : ١٢٩٨ فإن درجات الحرية تنقص واحداً آخر .

$$\text{إذن } \nu = 1 - (2 - 4) = 1$$

$$\chi^2_{0.05, 1} = 3.841$$

بما أن $3.841 > 1.73$ لا نستطيع رفض الفرض الصفري عند المستوى ٠,٠٥ . ويمكننا القول بأنه في حدود مجموعة العمر التي درست ، عادة التدخين والمات مستقلتان ، أى أن التدخين لا يؤثر في الوفاة في حدود هذه التجربة .

مثال (٦ - ١١) :

جدول الاقتران 3×2 الآتي يحمل التكرارات المشاهدة في عينة عشوائية من ٩٩ طفلاً حديث الولادة من حيث طول الطفل وطول محيط رأسه بالسنتيمترات ساعة الولادة . اجث استقلال طول المولود وطول محيط رأسه .

المجموع	طول المولود			محيط الرأس
	٥٥-٥٣	٥٢-٥٠	٤٩-٤٧	
٧٨	٢	٣٦	٤٠	٣٥-٣٢
٢١	٧	١٤	٠	٣٩-٣٦
٩٩	٩	٥٠	٤٠	المجموع

الحل :

كما في المثال السابق ، لو كان طول المولود ومحيط رأسه مستقلين لتوزعت أعداد الأطفال في كل من الأقسام الثلاثة للأطوال بنفس النسبة . ونقدر هذه النسبة من العينة على أنها ٧٨ : ٢١ أى ينبغي أن نقسم كلا من الأعداد ٤٠ ، ٥٠ ، ٩٠ بالنسبة ٧٨ : ٢١ لنحصل على جدول التكرارات النظرية الآتي :

المجموع	طول المولود			محيط الرأس
	٥٥-٥٣	٥٢-٥٠	٤٩-٤٧	
٧٨	٧,١	٣٩,٤	٣١,٥	٣٥-٣٢
٢١	١,٩	١٠,٦	٨,٥	٣٩-٣٦
٩٩	٩	٥٠	٤٠	المجموع

$$\chi^2 = \frac{(1,9-7)^2}{1,9} + \dots + \frac{(31,5-40)^2}{31,5} = 29,377$$

$$\nu = 3 - 1 = 2$$

$$\chi^2_{0,05(2)} = 5,991$$

نرفض الفرض الصفري عن استقلال طول المولود وطول محيط رأسه عند مستوى الدلالة ٠,٠٥ ونحكم بوجود علاقة بين هاتين الخاصتين .

ملاحظة (١)

(أ) إذا كانت هناك علاقة موجبة بين الخاصتين تظهر تكرارات كبيرة نسبيا في خلايا القطر الرئيسى أى في الخلايا العلوية اليمنى وفي الوسط وفي الخلايا السفلية اليسرى كما في هذا المثال .

(ب) وإذا كانت هناك علاقة سالبة تظهر تكرارات كبيرة نسبيا في الخلايا السفلية اليمنى وفي الوسط وفي الخلايا العلوية اليسرى .

(ج) أما إذا كان هناك علاقة صغيرة أو لا توجد علاقة فإن التكرارات تميل إلى أن تكون متناسبة الكثافة بمعنى أن يكون توزيع التكرارات بنفس النسبة التي تظهر في المجاميع الهامشية .

ملاحظة (٢) :

هناك قاعدة سهلة لإيجاد عدد درجات الحرية في أى جدول اقتران $m \times h$ حيث $m < 1$ ، $h < 1$ وهى :

$$\nu = (m - 1)(h - 1) \quad (١٦)$$

وتفسير هذه القاعدة أنه مادامت المجاميع الهامشية للأعمدة ثابتة فإننا في ملأ خانات جدول الاقتران يمكننا استنتاج واحد من الأعداد التي بأى عمود من ال $h - 1$ الأعداد الأخرى التي بهذا العمود ، وبالمثل يمكن استنتاج واحد من الأعداد التي بأى صف من ال $m - 1$ الأعداد الأخرى التي بهذا الصف . وبهذا نكون أحراراً فقط في ملأ $(m - 1)(h - 1)$ من خلايا الجدول .

ملاحظة (٣) :

هناك أيضاً قاعدة سهلة لحساب التكرار النظرى المناظر لتكرار مشاهد في أى خلية ، وذلك بضرب المجموعين الهامشين للصف والعمود الذى تقع فيه الخلية

ثم قسمة الناتج على المجموع الكلى للتكرارات (حجم العينة) ، فمثلا في المثال (٦ - ١١) الأخير :

$$\text{التكرار النظرى المناظر للعدد ٣٦ هو } \frac{٥٠ \times ٧٨}{٩٩} = ٣٩,٤$$

$$\text{، التكرار النظرى المناظر للعدد ٧ هو } \frac{٩ \times ٢١}{٩٩} = ١,٩$$

(٦ - ٧ - ٤) فترات الثقة لتباين مجتمع معتدل :

في البند (٦ - ٦ - ٢) أوجدنا صيغة لفترات الثقة للوسط الحسابي μ لمجتمع معتدل واستخدمنا لذلك توزيع ت . أما فترات الثقة لتباين مجتمع معتدل فتحتاج لاستخدام توزيع χ^2 ، إذ يمكن إثبات أنه على أساس عينة عشوائية حجمها n وتباينها σ^2 مأخوذة من مجتمع معتدل تباينه σ^2 فإن المتباينة :

$$(١٧) \quad \frac{\chi^2 (١ - \alpha)}{[١ - \alpha] \frac{\sigma^2}{n}} < \sigma^2 < \frac{\chi^2 (١ - \alpha)}{[١ - \alpha] \frac{\sigma^2}{n}}$$

تشكل فترة ثقة بدرجة $(١ - \alpha)$ لتباين المجتمع . وليس من الضروري هنا أن تكون العينة كبيرة الحجم كما هو الحال في التطبيقات سابقة الذكر لأن التوزيع الذى بنيت عليه هذه الفترة هو توزيع مضبوط طالما كان المجتمع معتدلا .

مثال (٦ - ١٢) :

أخذت عينة عشوائية حجمها ١٢ من مجتمع معتدل فوجد أن تباينها ٩,٧٣ . أوجد فترة ثقة بدرجة ٩٥٪ لتباين المجتمع .

الحل :

$$\text{لدينا } \alpha = 0,05 \quad \frac{\alpha}{2} = 0,025 \quad 1 - \frac{\alpha}{2} = 0,975 \quad \nu = 12$$

$$\text{من جدول } \chi^2 \text{ نجد أن } \chi^2_{\frac{\alpha}{2}} = [1-\nu] \chi^2_{[0,975,12]} = 21,9$$

$$\chi^2_{\frac{\alpha}{2}} = [1-\nu] \chi^2_{[0,975,12]} = 3,82$$

بالتعويض في (١٧) نجد أن :

$$\text{الحد الأدنى للفترة} = \frac{9,73 \times 11}{21,9} = 4,88$$

$$\text{الحد الأعلى للفترة} = \frac{9,73 \times 11}{3,82} = 28,018$$

إذن الفترة (٤,٨٨ ، ٢٨,٠١٨) هي فترة ثقة بدرجة ٩٥٪ لتباين المجتمع .

تمارين (٦ - ٢)

(في كل المسائل الآتية اذكر الشروط اللازم توفرها لصحة الحل .)

(١) حسب نظرية مندل عن تهجين النباتات تكون نسبة نبات البسلة الأخضر إلى نبات البسلة الأصفر ٣ : ١ . اختر هذه النظرية على ضوء البيانات المشاهدة الآتية :

(أ) وجد في عينة عشوائية أن هناك ٣٥٥ نباتاً أخضر ، ١٢٣ نباتاً أصفر .

(ب) وجد في عينة عشوائية أن هناك ٤٢٨ نباتاً أخضر ، ١٥٢ نباتاً أصفر .

(٢) وضعت خمس مصائد فيران في أماكن مختلفة من غابة ما . في فترة ثلاثة

أشهر سجلت أعداد الفيران المصيدة كالآتي :

المصيدة :	أ	ب	ح	د	هـ
عدد الفيران :	٢٣	٧	٢٥	١٩	٢١

اختبر الفرض أنه لا يوجد فرق بين المصايد الخمس في عدد ما تصطاده من الفيران .

(٣) اختبر الفرض أن عدد المواليد اليومية في مجتمع ما ثابت خلال شهور السنة مستخدماً البيانات الآتية :

الشهر : يناير فبراير مارس إبريل مايو يونيو يوليو أغسطس سبتمبر أكتوبر نوفمبر ديسمبر

عدد المواليد : ٨٠ ٧٨ ٨٦ ٨٢ ٨٣ ٧٨ ٧٩ ٧٦ ٧٨ ٧٦ ٧٢ ٧٦

(٤) لمعرفة ما إذا كان مجتمع ما يفضل نوعاً أ من سلعة ما (معجون أسنان مثلاً) على نوع آخر ب أجرى استفتاء على عينة من ٣٠٠ شخص فوجد أن ١٦٨ شخصاً منهم يفضلون النوع أ وأن ١٣٢ شخصاً يفضلون النوع ب . فهل هذه النتيجة تعني أن المجتمع يفضل النوع أ ؟

(٥) في ٣٦٠ رمية لزهرتين من النرد ظهر ما مجموعة سبعة ٧٤ مرة وظهر ما مجموعة إحدى عشر ٢٤ مرة . اختبر الفرض أن الزهرتين غير متحيزتين .

(٦) في المثال (٣ - ٩) عن توزيع خلايا الخميرة وفقنا توزيع بواسون لتوزيع تكرارى مشاهد في عينة . استخدم المستوى ٠,٠٥ لاختبار حسن المطابقة .

(٧) أخذ جوالان من حبوب الشعير لاختبار تأثير معالجة حرارية معينة على حيوية الحبوب وقد ترك الجوال أ دون معالجة (مجموعة مراقبة) وأعطيت المعالجة للجوال ب ثم أخذت عينة من ٨٠ حبة من كل جوال وفحصت بطريقة ما من حيث حيوية الحبوب فكانت النتيجة كما يلي :

لا يحتوى على

مقومات الحياة

١٦

٤٦

يحتوى على

مقومات الحياة

٦٤

٣٤

الجوال أ :

الجوال ب :

هل هناك فرق ذو دلالة بين الجوالين نتيجة للمعالجة الحرارية ؟ أى هل حيوية الحبوب مستقلة عن المعالجة الحرارية ؟

(٨) أخذت مجموعتان متكافئتان أ ، ب كل منهما تتكون من ١٠٠ شخص مريض بمرض معين وأعطى دواء للمجموعة الأولى ، ولم يعط للمجموعة الثانية (مجموعة المراقبة) فوجد أن عدد الذين شفوا من المجموعتين أ ، ب هما ٧٥ ، ٦٥ على الترتيب . أختبر الفرض أن هذا الدواء يساعد على الشفاء من المرض .
(٩) في عينة من خمس فراشات وجد أن تباين طول الجناح ١٣,٥٢ . إذا فرض أن مجتمع أجنحة هذه الفراشات معتدل (بالنسبة لطول الجناح) فأوجد فترة ثقة بدرجة ٩٥٪ لتباين هذا المجتمع .

(١٠) فى دراسة عن نوع من الضفادع وجد في عينة من ٣٩ ضفدعا أن متوسط فترة نداءات الذكور ١٨٩ وحدة بانحراف معيارى ٣٢ وحدة .
(أ) أوجد فترة ثقة بدرجة ٩٥٪ لمتوسط فترة النداء في المجتمع .
(ب) أوجد فترة ثقة بدرجة ٩٥٪ لتباين فترة النداء في المجتمع .

(١١) بمجدول الاقتران الآتى التوزيع المشترك لمجموعة من ٩٥٠ طالبا من حيث خاصتى الذكاء والتغذية . هل تشير هذه البيانات إلى أن ذكاء الطلاب مستقل عن التغذية ؟

نوع التغذية	نسبة الذكاء				المجموع
	أقل من ٨٠	٨٠-٨٩	٩٠-٩٩	١٠٠ فأكثر	
تغذية حسنة	٢٤٥	٢٢٨	١٧٧	٢١٩	٨٦٩
سوء تغذية	٣١	٢٧	١٣	١٠	٨١
المجموع	٢٧٦	٢٥٥	١٩٠	٢٢٩	٩٥٠

(٦ - ٧ - ٥) خاصة الجمع لتوزيعات χ^2 المستقلة :

يمكن إثبات النظرية الآتية : « إذا كان لمتغير ما توزيع χ^2 بدرجات حرية ν ، وكان لمتغير آخر مستقل عن المتغير الأول توزيع χ^2 بدرجات حرية ν_1 ، فإن المتغير الذى ينشأ من ضم هذين المتغيرين معا يكون له توزيع χ^2 بدرجات حرية $\nu + \nu_1$ » وتمتد هذه النظرية لأى عدد منتهى من المتغيرات المستقلة .

مثال (٦ - ١٣) :

فى دراسة عن تأثير التطعيم بمصل ما فى الوقاية من مرض ما أخذت مجموعتان متكافئتان من الأطفال طعمت إحداهما بالمصل ولم تطعم الأخرى . وبعد فترة محددة حسب عدد الذين أصيبوا والذين لم يصابوا بالمرض فى كلتا المجموعتين ووضعت النتائج فى جدول 2×2 ووجد أن $\chi^2 = 4,1$ بدرجة حرية واحدة . كررت نفس الدراسة مرتين أخريتين بشكل مستقل (فى مكانين آخرين) ووجد أن قيمتى χ^2 هما ٣,٩ ، ٤,٢ بدرجة حرية واحدة لكل منهما . نلاحظ ما يلى :

(أولا) إذا أخذنا كلا من هذه التجارب الثلاث على حدة نجد أن كلا منها يشير إلى فرق ذى دلالة بين المجموعة التى طعمت بالمصل والمجموعة الضابطة عند المستوى ٠,٠٥ وذلك لأن القيمة الحرجة $\chi^2_{[1], 0,05} = 3,841$ وهذا يعنى أن المصل كان له تأثير فى الحماية من المرض .

(ثانيا) إذا ضمنا نتائج التجارب الثلاث معا ، وهذا جائز لأنها مستقلة ، نجد أن :

المجموع $= 4,1 + 3,9 + 4,2 = 12,2$ بدرجات حرية عددها ثلاث ونظرا لأن القيمة الحرجة $\chi^2_{[3], 0,05} = 7,815$ فإن هذه النتيجة تدعم كلا من النتائج الثلاث السابقة وتؤكد أن الفروق بين أزواج المجموعات لم تكن بالصدفة .

(٦ - ٨) اختبار ف :

THE F - TEST

إن الأهمية الرئيسية لهذا الاختبار هو قدرته على اختبار تساوى تبايني مجتمعين معتدلين σ_1^2 ، σ_2^2 عن طريق تبايني عينتين عشوائيتين مستقلتين $\bar{X}_1$ ، $\bar{X}_2$ مأخوذتين منهما . وقد رأينا أن اختبارات عن دلالة الفرق بين متوسطى عينتين بالصورة المبينة بالبند (٦ - ٦ - ٣) يستلزم التأكد من تساوى تبايني المجتمعين ، ولذا ينبغي إجراء اختبار ف بشكل روتيني قبل تطبيق هذا الاختبار . ومن التطبيقات الهامة التي يستخدم فيها اختبار ف ذلك التطبيق المسمى بتحليل التباين الذى سنتناوله في الفصل الثامن حيث نحتاج أيضاً إلى اختبار تساوى تباينات المجتمعات .

ويبنى هذا الاختبار على الإحصاءة :

$$F = \frac{\bar{X}_1^2}{\bar{X}_2^2} \quad (١٨)$$

التي تسمى بنسبة التباين variance ratio والتي يتطابق توزيعها مع توزيع المتغير ف السابق دراسته بدرجةتي حرية $\nu_1 = \nu_2 - 1$ ، $\nu_2 = \nu_1 - 1$ تحت الافتراضات الآتية :

- (أ) أن تكون كل من العينتين عشوائية بسيطة .
- (ب) أن تكون العينتان مستقلتين .
- (ج) أن تكون العينتان مأخوذتين من مجتمع معتدل واحد أو من مجتمعين معتدلين لهما نفس التباين .

ملاحظة :

لاستخدام جدول ف للقيم الخارجة يجب أن توضع نسبة التباين بحيث تكون

أكبر من الواحد أى بحيث تكون $\sigma^2 < \sigma^2$ مع ملاحظة أن σ تحدد العمود ،
 σ تحدد الصف .

مثال (٦ - ١٤) :

أخذت عيتان عشوائيتان حجمهما ٥ ، ٦ من الإناث والذكور من حشرة
 ما وحسبت مدة البقاء على الحياة لكل حشرة بعد حرمانها من الطعام والشراب
 فوجد أن تباين الأولى ٢٥,٧ وتباين الثانية ٠٧,٠٦ على أساس أن المجتمعين
 معتدلين :

(أولاً) اختبر عند مستوى الدلالة ٠,٠١ أن تباين مجتمع الإناث أكبر من تباين
 مجتمع الذكور .

(ثانياً) اختبر عند مستوى الدلالة ٠,٠٥ أن المجتمعين متساويا التباين .

الحل :

$$F = \frac{25,7}{7,06} = 3,64 \text{ بدرجتي حرية } 4, 5$$

(أولاً) الفرض الصفري ف : $\sigma^2 = \sigma^2$

الفرض البديل ف : $\sigma^2 < \sigma^2$ (اختبار ذو جانب واحد)

من الجدول (٩) : ف ٠,٠١ [٥,٤] = ١١,٤

بما أن $3,64 < 11,4$ لا نستطيع رفض ف عند المستوى ٠,٠١

(ثانياً) الفرض الصفري ف : $\sigma^2 = \sigma^2$

الفرض الآخر ف : $\sigma^2 \neq \sigma^2$ (اختبار ذو جانبيين مساحة كل منهما

٠,٠٢٥

نريد إيجاد F_1 ، F_2 حيث $L(F < F_1) = 0.025$ ،
 $L(F < F_2) = 0.975$ مع ملاحظة أن الاحتمال 0.975 لا يوجد
 بالجدول .

$$\begin{aligned} 7.39 &= \text{من الجدول } F_{0.025, [4, \infty]} \\ 7.39 &= \text{إذن القيمة الحرجة اليمنى } F_1 \\ 0.975 &= \text{المتساوية لـ } (F < F_2) \text{ تكافئ} \\ 0.025 &= \left(\frac{1}{F_2} < \frac{1}{F_1} \right) \text{ المتساوية لـ} \end{aligned}$$



الشكل (٤-٦)

من الجدول $F_{0.025, [4, \infty]} = 9.36$ راجع ملاحظة البند (٥-٣)

$$\therefore 9.36 = \frac{1}{F_2}$$

$\therefore$ القيمة الحرجة اليسرى $F_1 = 1 \div 9.36 = 0.107$

بما أن $0.107 < 3.64 < 73.9$ لا نستطيع رفض F_0 عند المستوى 0.05 ،
 أى أننا نقبل أن $\sigma_1 = \sigma_2$ عند هذا المستوى .

(٦ - ٩) فترات الثقة للنسبة في مجتمع :

نفرض أن نسبة وقوع حدث أ في مجتمع ما هو عدد مجهول ح ، فمثلا قد تعبر ح عن نسبة ظهور زهور حمراء من نوع من البذور ، أو نسبة الإصابة بمرض ما في نوع من الماشية ، أو نسبة النيكوتين في نوع من السجائر . يمكننا تقدير النسبة ح من واقع عينة عشوائية كبيرة وذلك بأخذ النسبة المشاهدة ر بين عدد مرات وقوع الحدث أ وعدد وحدات العينة ، وتزداد ثقتنا في التقدير ر كلما زاد حجم العينة . وإذا أردنا إيجاد صيغة لفترات الثقة للنسبة ح فإننا نستخدم التقدير ر كأساس لبناء هذه الصيغة وبحسب المنطق الآتي :

نعلم من البند (٣ - ٣) أنه تحت شروط عشوائية العينة وثبات النسبة ح واستقلال الأحداث يكون للمتغير العشوائي $\frac{r}{n}$ الذي يعبر عن عدد مرات وقوع الحدث أ في العينات ذوات الحجم ن توزيع ذي الحدين دليلا ن ، ح وسطه الحسابي ن ح وتباينه ن ح حيث $k = 1 - ح$. ونعلم من البند (٦ - ٤) أنه إذا كان المتغير العشوائي $\frac{r}{n} = \frac{r}{n}$ يعبر عن النسبة المشاهدة في العينة فإن توزيع الإحصاءة :

$$\frac{r - n\chi}{\sqrt{\frac{n\chi k}{n}}} = \chi$$

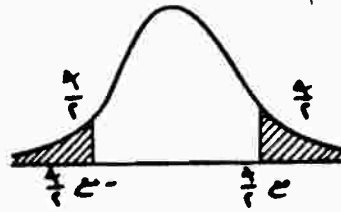
يقرب من التوزيع المعتدل المعياري بشرط أن تكون ن كبيرة وألا تكون ح أو ك قريبة من الصفر .

في هذه الحالة يمكن إثبات أن الفترة :

$$(19) \quad \left(\frac{r}{n} - \sqrt{\frac{(r-1)r}{n}} \right) \text{ مع } \frac{r}{n} , \left(\frac{r}{n} + \sqrt{\frac{(r-1)r}{n}} \right) \text{ مع } \frac{r}{n}$$

هى فترة ثقة بدرجة $(1 - \alpha)$ للنسبة ح .
حيث مع $\frac{\alpha}{2}$ هى القيمة الحرجة في التوزيع المعتدل المعيارى التي تحقق المعادلة

$$(20) \quad L \left(\frac{\alpha}{2} \right) < \bar{c} < U \left(\frac{\alpha}{2} \right) = 1 - \alpha$$



الشكل (٦ - ٥)

فمثلا حين $\alpha = 0.05$ فإن مع $\frac{\alpha}{2} = 1.96$ راجع المثال (٤ - ٣) .
وحيث $\alpha = 0.01$ فإن مع $\frac{\alpha}{2} = 2.58$ راجع المثال (٤ - ٣) .
ويلاحظ أن أى عدد يقع في هذه الفترة يصلح لأن يؤخذ كتقدير للنسبة ح .

مثال (٦ - ١٥) :

أخذت عينة عشوائية حجمها ٤٠٠ شخص من مجتمع ما ووجد أن ٢٣ منهم مصابون بمرض البول السكرى . أوجد فترة ثقة بدرجة ٩٥٪ لنسبة المصابين بهذا المرض في هذا المجتمع .

الحل :

$$r = \frac{23}{400} = 0.0575 \quad \text{نسبة الإصابة في العينة}$$

نأخذ هذه النسبة كتقدير للنسبة ح مع ملاحظة عشوائية العينة وكبير حجمها .

$$\text{تقدير الخطأ المعياري} = \frac{(r-1)r}{n} = \sqrt{\frac{0.9425 \times 0.0575}{400}} = 0.0116$$

من (١٩) وبأخذ $\alpha = 0,05$ نجد أن
الحد الأدنى للفترة $= 0,0575 - 0,116 \times 1,96 = 0,035$
الحد الأعلى للفترة $= 0,0575 + 0,116 \times 1,96 = 0,0802$
إذن الفترة (0,035 ، 0,080) هي فترة ثقة بدرجة ٩٥٪ لنسبة المصابين بالبول
السكرى في المجتمع .

(٦ - ٩ - ١) اختبار دلالة الفرق بين نسبتين عينتين مستقلتين :

نفرض أن لدينا مجتمعين نسبة وقوع حدث ما في أحدهما ح_١ ونسبة وقوعه
في الآخر ح_٢ . ونفرض أننا حصلنا على عينتين عشوائيتين مستقلتين من هذين
المجتمعين ووجدنا أن نسبتين وقوع الحدث فيهما ر_١ ، ر_٢ على الترتيب . يمكن
إثبات أن توزيع المعاينة للإحصاءة .

$$(21) \quad \frac{(r_1 - r_2) - (H_1 - H_2)}{\sqrt{\frac{r_1 H_1}{n_1} + \frac{r_2 H_2}{n_2}}} = z$$

يقترّب من التوزيع المعتدل المعياري كلما كبرت كل من n_1 ، n_2 ولم تكن أى
من ح_١ ، ح_٢ ، ك_١ ، ك_٢ قريبة من الصفر .
وإذا فرضنا أن ح_١ = ح_٢ = ح ، ك_١ = ك_٢ = ك فإن هذه الإحصاءة تكتب .

$$(21) \quad \frac{r_1 - r_2}{\sqrt{H \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} = z$$

وفي هذه الحالة تقدر ح من العينتين كالآتي :

$$H = \frac{n_1 r_1 + n_2 r_2}{n_1 + n_2}$$

وعلى ذلك يمكن استخدام الإحصاءة (٢١) للكشف عما إذا كان الفرق المشاهد بين $\bar{S}_1$ ، $\bar{S}_2$ في العينتين هو فرق صغير لا دلالة له أم فرق يدل على وجود فرق حقيقي بين النسبتين H_1 ، H_2 في المجتمعين .

مثال (٦ - ١٦) :

لمعرفة تأثير طعم ما في الوقاية من وباء الكوليرا اختيرت عينتان عشوائيتان حجم الأولى ١٠٠٠ شخص وحجم الثانية ١٥٠٠ شخص وحقنت أفراد المجموعة الأولى بالطعم ولم يحقن أفراد المجموعة الثانية (مجموعة مراقبة) . وبعد فترة من الزمن ظهرت ١٠٠ حالة مرضية في المجموعة الأولى و ٥٠٠ حالة في المجموعة الثانية . اختبر ما إذا كان لهذا الطعم أثر في الوقاية من هذا المرض مستخدماً مستوى الدلالة ٠,٠١

الحل :

الفرض الصفري F_0 : $H_1 = H_2$ (لا يوجد تأثير للطعم المعطى)
الفرض الآخر F_1 : $H_1 > H_2$ (اختبار ذو جانب واحد) .

$$\text{نسبة الإصابة في المجموعة الأولى } \bar{S}_1 = \frac{100}{1000} = 0,10$$

$$\text{نسبة الإصابة في المجموعة الثانية } \bar{S}_2 = \frac{500}{1500} = 0,33$$

$$\text{من (٢١) : تقدير } H = \frac{1500 \times 0,33 + 1000 \times 0,1}{1500 + 1000} = 0,238$$

$$\text{تقدير الخطأ المعياري} = \sqrt{H \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} = \sqrt{0,238 \times \left(\frac{1}{1000} + \frac{1}{1500} \right)} = 0,017$$

$$\text{من (٢١) : } Z = \frac{0,33 - 0,10}{0,017} = 13,52$$

بما أن $13,52 - > 2,33$ نرفض الفرض الصفري عند المستوى $0,01$ ونستنتج أن للطعم تأثير في الوقاية من المرض .

(٦ - ٩ - ٢) فترات الثقة للفرق بين نسبتي مجتمعين :

يمكن إثبات أنه تحت الشروط سابقة الذكر تكون الفترة :

$$(22) \quad \left\{ \frac{\alpha}{2}, \frac{\alpha}{2} \right\} \text{ مع } (x_1 - x_2) \pm \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

حيث x_1, x_2 هي النسبتان

هي فترة ثقة بدرجة $(1 - \alpha)$ للفرق $x_1 - x_2$ بين نسبتي المجتمعين . ففي المثال (٦ - ١٥) الأخير نجد أن :

$$\text{الحد الأدنى للفترة} = 0,33 - 0,1 = 0,23$$

$$\text{الحد الأعلى للفترة} = 0,33 + 0,1 = 0,43$$

أي أن الفترة $(0,23, 0,43)$ هي فترة ثقة بدرجة 99% للفرق بين النسبتين في المجتمعين .

تمارين (٦ - ٣)

(١) في عينة عشوائية من ٤٠٠ شخص حققوا بمصل ما تأثر ١٣٦ تأثيراً ضاراً . أوجد فترة ثقة بدرجة 95% لنسبة الأشخاص الذين يتأثرون تأثيراً ضاراً من هذا المصل .

(٢) من ١٠٠ سمكة صيدت من بحيرة ما وجد أن ١٦ منها لا تصلح للأكل نتيجة لتلوث كيميائي في بيئة هذه البحيرة . أنشئ فترة ثقة بدرجة 95% لاحتمال أنه إذا صيدت سمكة من هذه البحيرة تكون غير صالحة للأكل .

(٣) في تجربة عن تهجين فيران ذوات الشعر النحيف المتهدل وفيران ذوات الشعر المجعد وجد في ٣٢ ولادة بكل منها ٨ فيران أن عدد الولادات التي تحتوى بالضبط على س فأراً ذا شعر نحيف ومتهدل كالآتي :

س : ٠ ١ ٢ ٣ ٤ ٥ ٦ ٧ ٨
عدد الولادات : ٠ ١ ٢ ٤ ١٢ ٥ ٢ ٠

على فرض أن توزيع ذى الحدين يصلح نموذجاً في هذه التجربة أوجد فترة ثقة بدرجة ٩٩٪ لاحتمال الحدث « ولادة فأر ذى شعر نحيف ومتهدل » - راجع البند (٣ - ٣ - ١) .

(٤) أخذت مجموعتان عشوائيتان بكل منهما ٨٠ مريضاً . أعطى للمجموعة الأولى أقراص تحتوى على دواء ضد الحساسية وأعطى للمجموعة الثانية أقراص تمويه (لا تحتوى على أى دواء) ، فظهرت أعراض الحساسية في ٢٣ من المجموعة الأولى ، ٥١ من المجموعة الثانية . اختبر عند مستوى الدلالة ٠,٠١ ما إذا كانت نسبة ظهور الحساسية في المرضى الذين يتلقون أقراص الحساسية لا تختلف عنها في المرضى الذين يتلقون أقراص التمويه .

(٥) في عينة عشوائية من ٢٠٠ شخص لا يتناولون طعام الإفطار أفاد ٨٢ منهم أنهم يشعرون بتعب منتصف الصباح . وفي عينة عشوائية من ٤٠٠ شخص يتناولون طعام الإفطار أفاد ١١٦ منهم أنهم يشعرون بتعب منتصف الصباح . استخدم مستوى الدلالة ٠,٠١ لاختبار الفرض الصفري أنه لا يوجد فرق بين المجتمعين ضد الفرض الآخر أن تعب منتصف الصباح أكثر تفشيًا بين الأشخاص الذين لا يتناولون طعام الإفطار .

(٦ - ١٠) تحديد حجم العينة :

من الأمور التي تشغل الباحث عند تصميم تجربة لحل مشكلة ما تحديد عدد

وحدات العينة (العشوائية) اللازم لضمان أن تكون أحكامه عن المجتمع الذى يدرسه على درجة كافية من العمومية والدقة . وبطبيعة الحال كلما كانت العينة كبيرة كلما زادت الثقة في هذه الأحكام ، غير أن كبر حجم العينة يحتاج إلى الكثير من الجهد والوقت والتكاليف ، سواء في عملية المعاينة أو في قياس وتحليل البيانات ولذلك فإن كفاءة التصميم تتطلب إيجاد حد أعلى معقول لحجم العينة . وسنبحث هذا الموضوع في الحالات الثلاث الآتية .

(أولا) عند تقدير الوسط الحسابى لمجتمع :

نفرض أننا نتساءل عن حجم العينة المناسب لتقدير متوسط المجتمع μ من المتوسط $\bar{x}$ لعينة عشوائية مأخوذة منه . إن الإجابة عن هذا التساؤل لا تتأتى إلا إذا أجبنا عن السؤال العملى الآتى : « ما مقدار الخطأ الذى يمكن السماح به عند تقدير μ عن طريق $\bar{x}$ ؟ أى ما هو الحد الأعلى الذى يمكن التجاوز عنه لانحراف $\bar{x}$ عن القيمة الحقيقية μ » ؟ وهذا السؤال لا يجاب عنه إحصائيا وإنما هو من اختصاص الباحث التطبيقى وهو الذى يجيب عنه من واقع خبرته بميدان البحث . فإذا رأى الباحث أن الحد الأعلى للخطأ المسموح به هو عدد ما خ ، ورأى فى الوقت نفسه أن يعين درجة ثقة (٩٥ ٪ مثلا) فى عدم تخطى هذا الحد عند التطبيق ، فإن الحجم المناسب للعينة الذى يحقق الفرض المنشود ينتج حسب الأساس الآتى :

من البند (٦ - ٣) نعلم أنه إذا كان لدينا مجتمع معتدل وسطه الحسابى μ وانحرافه المعيارى σ فإن الإحصاءة $\bar{x}$ للعينات ذوات الحجم n يكون لها توزيع معتدل وسطه الحسابى μ وانحرافه المعيارى $\frac{\sigma}{\sqrt{n}}$ وتحقق هذه النظرية أيضا

(بالتقريب) حين لا يكون المجتمع معتدلا بشرط أن يكون حجم العينة كبيرا
($n \leq 30$) .

وإذا كانت $\bar{x}$ هي الوسط الحسابي لعينة عشوائية ما فإن الفترة :

$$(\bar{x} - \frac{\sigma}{\sqrt{n}} \text{ مع } \frac{\alpha}{2}, \bar{x} + \frac{\sigma}{\sqrt{n}} \text{ مع } \frac{\alpha}{2})$$

تكون فترة ثقة بدرجة $1 - \alpha$ للمتوسط μ للمجتمع ، حيث مع $\frac{\alpha}{2}$ هي قيمة المتغير المعتدل المعيارى التي تحقق المعادلة :

$$L(\text{مع } \frac{\alpha}{2} < E < \text{مع } \frac{\alpha}{2}) = 1 - \alpha$$

وبذلك تكون أكبر قيمة للانحراف $|\mu - \bar{x}|$ هي $\frac{\sigma}{\sqrt{n}} \text{ مع } \frac{\alpha}{2}$ ويسمى هذا

العدد بحد الخطأ error bound ، بدرجة ثقة $1 - \alpha$

وإذا اخترنا أن يكون الخطأ المسموح به هو مقدار معين x وأردنا أن نكون على ثقة بدرجة $(1 - \alpha)$ بألا يتعدى الخطأ الذى تقع فيه القيمة x فإن حجم

العينة المطلوب يجب أن يحقق المعادلة $\frac{\sigma}{\sqrt{n}} \text{ مع } \frac{\alpha}{2} = x$. وبحل هذه المعادلة

في n نجد أن :

$$(23) \quad n = \left(\frac{\frac{\sigma}{\sqrt{n}} \text{ مع } \frac{\alpha}{2}}{x} \right)^2$$

وهذه هي القيمة المطلوبة لحجم العينة n الذى يكفي لتحقيق الغرض المطلوب .

كما يمكن أن نأخذ العدد n حيث

$$(24) \quad \frac{\sigma}{\sqrt{n}} \text{ مع } \frac{\alpha}{2} = x$$

كحد أعلى لحجم العينة . وتنتج هذه المعادلة من متباينة تشيشف الشهيرة التي لا تتطلب توزيعاً أو شروطاً معينة ، إلا أنها غالباً ما تعطى قيمة أكبر مما ينبغي لحجم العينة .

ويلاحظ أن إيجاد قيمة n من أى من المعادلتين (٢٣) أو (٢٤) يتطلب معرفة ولو تقريبية بالانحراف المعياري σ للمجتمع . وحين تكون قيمة σ مجهولة تماماً فلا مفر من تقديرها من عينة عشوائية استطلاعية كبيرة لا يقل حجمها عن ٣٠ .

مثال (٦ - ١٧) :

في عينة عشوائية حجمها ٣٦ وجد أن الوسط الحسابي والانحراف المعياري هما ٢,٦ و ٠,٣ على الترتيب . ما حجم العينة اللازم لإعطائنا ثقة بدرجة ٩٥٪ بألا يزيد الخطأ في تقدير الوسط الحسابي μ للمجتمع عن ٠,٠٦ ؟

الحل :

لدينا $x = ٠,٠٦$ ، $١ - \alpha = ٠,٩٥$ ومنها $\alpha = ٠,٠٥$ ،

$$مع \frac{\alpha}{2} = ١,٩٦$$

نأخذ الانحراف المعياري ٠,٣ الناتج من العينة الاستطلاعية كتقدير للانحراف المعياري للمجتمع . ثم نحسب n من الصيغة (٢٣) كالآتي :

$$n = \left(\frac{١,٩٦ \times ٠,٣}{٠,٠٦} \right)^2 = ٩٦,٠٤$$

إذن الحجم المطلوب للعينة هو ٩٧ أى أنه يكفي أن نوجد الوسط الحسابي $\bar{x}$ لعينة عشوائية من هذا الحجم لكي نكون واثقين بدرجة ٩٥٪ أن اختلاف $\bar{x}$ عن μ لا يزيد عن ٠,٠٦ .

$$\therefore 0.0597 = \frac{1.96 \times 0.3}{\sqrt{n}} = 1.96 \times \frac{\epsilon}{\sqrt{n}} \text{ فإن } 97 = n \text{ نلاحظ أنه بأخذ } n = 97 \text{ فإن } \frac{\epsilon}{\sqrt{n}} = 0.0597$$

وهذا أصغر من 0.06 .

(ثانيا) عند تقدير الفرق بين متوسطى مجتمعين معتدلين :

بنفس الطريقة يمكن إيجاد حد أعلى لحجم العينة اللازم لتقدير الفرق $\mu_1 - \mu_2$ بين متوسطى مجتمعين معتدلين بحيث نكون على ثقة بدرجة $1 - \alpha$ ألا يزيد المقدار $\bar{x}_1 - \bar{x}_2$ المشاهد فى عينتين عن المقدار $\mu_1 - \mu_2$ للمجتمعين ، عن قيمة معينة ϵ مفروضة مسبقا ، وما علينا إلا استخدام الخطأ المعيارى

لتوزيع المعاينة للإحصاءة $\bar{x}_1 - \bar{x}_2$ وهو $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ بدلا من

الخطأ المعيارى $\frac{\sigma}{\sqrt{n}}$ لتوزيع المعاينة للإحصاءة $\bar{x}$. وإذا كانت العينتان متساويتى

الحجم : $n_1 = n_2 = n$ وكان المجتمعان متساويين فى التباين : $\sigma_1^2 = \sigma_2^2 = \sigma^2$

فإن ذلك الخطأ المعيارى يأخذ الصورة $\frac{\sqrt{2}\sigma}{\sqrt{n}}$ ويكون الحد الأعلى المطلوب لحجم

$$\text{كل من العينتين هو } n = \left(\frac{\frac{\sigma^2}{\epsilon^2} \cdot 2}{1 - \alpha} \right) \quad (20)$$

ويلاحظ أن هذه الصيغة ضعف الصيغة (٢٣) .

مثال (٦ - ١٨) :

فى تجربة تهدف إلى اختبار دواء تخسيس وتحكم فى زيادة الوزن عند النساء ، رأى البدء باستخدام الدواء على إناث القطط المنزلية ، فاختيرت عينة عشوائية من

٦٠ قطة قسمت عشوائيا إلى مجموعتين متكافئتين في الوزن بكل منهما ٣٠ قطة ، مجموعة تجريبية وأخرى ضابطة ، أعطى الدواء إلى المجموعة التجريبية فقط ووضعت المجموعتان تحت نفس الظروف . وبعد ٦ أسابيع وباستخدام وحدة قياس معينة وجد أن متوسط النقص في وزن المجموعة التجريبية ١٠ وحدات ومتوسط النقص في وزن المجموعة الضابطة ١٣,٣ وحدة . على فرض أن تباين كل من المجموعتين ٤٠ أوجد حداً أعلى لحجم كل من العينتين الذي يعطينا ثقة بدرجة ٩٥٪ بألا يزيد الفرق المشاهد في متوسطي العينتين عن الفرق الحقيقي بين متوسطي المجموعتين عن ٣ وحدات .

الحل :

لدينا $\sigma^2 = 40$ ، $\chi = \pm 3$ ، $\alpha = 0.05$ ، وإذن $\frac{\alpha}{2} = 0.025$ ، 1.96

من الصيغة (٢٥) : $n = \frac{40 \times 2}{9} = 9$ ، $35 \sim 34,148 = 2(1.96)$

(ثالثاً) عند تقدير نسبة وقوع حدث في مجتمع :

نفرض أن نسبة وقوع حدث معين في مجتمع هو مقدار ثابت مجهول π ونفرض أننا نرغب في معرفة حجم العينة المناسب لتقدير هذه النسبة عن طريق النسبة $\hat{\pi}$ التي تظهر في عينة عشوائية ، بحيث نكون على ثقة بدرجة $1 - \alpha$ ألا يزيد الخطأ الناشئ عن هذا التقدير عن مقدار معين χ .

من البند (٦ - ٤) نعلم أنه إذا كان حجم العينة كبيراً ($n \geq 30$) فإن توزيع المعاينة للنسبة $\hat{\pi}$ لوقوع هذا الحدث في العينات ذوات الحجم n يقترب من توزيع معتدل وسطه الحسابي π وانحرافه المعياري $\sqrt{\frac{\pi(1-\pi)}{n}}$ حيث $k = 1 - \alpha$. في

هذه الحالة يكون حد الخطأ أي أكبر قيمة للمقدار $|\hat{\pi} - \pi|$ هو $\sqrt{\frac{\pi(1-\pi)}{n}} \cdot \frac{\chi}{2}$ ،

يلاحظ أن هذا الحد هو نفس الحد الذى وجدناه في (أولا) بوضع $\sqrt{\frac{c}{n}}$ بدلا من $\frac{\sigma}{\sqrt{n}}$

$$\text{بحل المعادلة } \sqrt{\frac{c}{n}} = \frac{mc}{2} = x \text{ نحصل على}$$

$$(26) \quad n = c \left(\frac{mc}{2} \right)^2$$

وهذه هي القيمة المطلوبة لحجم العينة . غير أن هذا الحل غير قابل للاستخدام لأنه يشتمل على البارامتر c وهو الذى نبحث عن تقديره . ولكن نظراً لأن أكبر قيمة لحاصل الضرب c هي $\frac{1}{4}$ أى أن $c \geq \frac{1}{4}$ دائماً ،

$$[\text{لأن } c \geq \frac{1}{4} = 1 - c = 1 - (c - 1) = 1 - (c - 1)^2 = 1 - (c^2 - 2c + 1) = 2c - c^2 = c(2 - c) \geq 0]$$

فإن الصيغة (26) يمكن أن تكتب كالآتي :

$$(27) \quad n = c \left(\frac{mc}{2} \right)^2 \frac{1}{c}$$

واختيار الحد الأقصى لحجم العينة من هذه المعادلة يؤدي الغرض المنشود بغير أى معرفة مسبقة لقيمة البارامتر c .

أما إذا كان لدينا معلومات تفيد بأن هذا البارامتر يساوى بالتقريب قيمة معينة c مثلاً فإن ن يمكن إيجادها من المعادلة .

$$(28) \quad n = c \left(\frac{mc}{2} \right)^2 \frac{1}{c}$$

وهذه القيمة تقل عن تلك التي تعطيها الصيغة (27) لأنها مبنية على معلومات عن القيمة المحتملة للدليل c .

مثال (٦ - ١٩) :

في عملية مسح صحى عن طريق العينة ، يراد تقدير النسبة ح للأشخاص ضعاف البصر . كم شخصاً ينبغي اختبارهم إذا كان المسئولون يريدون أن يتأكدوا بدرجة ٩٨٪ أن الخطأ في التقدير يقع في المدى $\pm ٠,٠٥$.

(أ) في حالة عدم وجود أى معرفة لقيمة ح

(ب) إذا كان المعروف أن ح = ٠,٣ بالتقريب ؟

الحل :

لدينا $\alpha - ١ = ٠,٩٨$ ، $\alpha = ٠,٠٢$ ، $\alpha = \frac{٠,٠١}{٢}$ ، $\pm ٠,٠٥ = \alpha$ ، القيمة هي قيمة المتغير المعتدل المعيارى أ التى تحقق المعادلة

$$ل (أ < ع < أ - أ) = ٠,٩٨$$

$$ل (أ < ع < أ) = ٠,٩٨ \div ٢ = ٠,٤٩$$

من جدول المساحات أسفل المنحني المعتدل نجد أن $أ = ٢,٣٣$

$$(أ) من (٢٧) : \frac{١}{٤} = \frac{٢,٣٣}{٠,٠٥} = ٥٤٢,٨٩$$

إذن الحجم المطلوب للعينة هو ٥٤٣ .

$$(ب) من (٢٨) : ٠,٧ \times ٠,٣ = \frac{٢,٣٣}{٠,٠٥} = ٤٥٦,٠٢٧٦$$

إذن الحجم المطلوب للعينة هو ٤٥٦ .

تمارين (٦ - ٤)

(١) يريد باحث تقدير متوسط محتوى الفسفات في وحدة حجم من ماء لإحدى البحيرات والمعروف من دراسات سابقة أن الانحراف المعيارى لهذا المحتوى ذو قيمة

تكاد تكون ثابتة عند $\sigma = 4$. ما عدد عينات الماء التي ينبغي للباحث تحليلها لكي يثق بدرجة ٩٠٪ أن الخطأ في التقدير لا يتعدى ٠,٨ ؟

(٢) أ - أراد مهندس تقدير الوسط الحسابي للمدة التي تجف فيها خلطة معينة من الأسمنت تستخدم في إصلاح الطرق . وقد جرب هذه الخلطة في ١٠٠ بقعة ووجد أن الوسط الحسابي والانحراف المعياري لمدة الجفاف هما ٣٢ دقيقة و ٤ دقائق على الترتيب . استخدم هذه المعلومات لإيجاد فترة ثقة بدرجة ٩٩٪ لمتوسط مدة جفاف هذه الخلطة .

ب - أراد المهندس أن يحدد حجم العينة (أى عدد البقع التي يجرب فيها الخلطة) لكي يكون متأكداً بدرجة ٩٥٪ أن متوسط مدة الجفاف المحسوب من عينة بهذا الحجم لا يختلف عن المتوسط الحقيقي لهذه المدة إلا بدقيقة واحدة على الأكثر . أوجد هذا الحجم علماً بأن الخبرة السابقة تشير إلى أن الانحراف المعياري لمدة الجفاف هو ٥ دقائق بالتقريب .

(٣) في أحد مراكز القلب والصدر يراد تقدير المعدل ح لوقوع حالات ضيق التنفس بين الذكور من متوسطى العمر ممن كانوا يدخنون أكثر من علبتين من السجائر في اليوم خلال خمس سنوات سابقة . ما حجم العينة التي نختاره بحيث نكون على ثقة بدرجة ٩٥٪ أن الخطأ في تقدير النسبة ح لا يزيد عن ٠,٠٣ ، علماً بأن المعروف أن قيمة ح الحقيقية قريبة من ٠,١٥ ؟

QUALITY CONTROL

(٦ - ١١) مراقبة الانتاج

من تطبيقات نظرية العينات في الصناعة ذلك التطبيق المسمى بمراقبة جودة الإنتاج ومعاينات القبول ، وهو يتعلق بإحدى المشكلات التي تظهر في المصانع التي تنتج سلعا على نطاق واسع . ويتمثل هذا التطبيق في استخدام فكرة فترات الثقة في التفتيش على جودة السلع المنتجة أثناء عملية الإنتاج .

ومن المعروف أن الوحدات المنتجة في أى مصنع لا يمكن أن تكون جميعها متشابهة تماما مهما تقدمت التكنولوجيا الصناعية بل توجد دائما انحرافات صغيرة عن المواصفات الموضوعية للسلعة تنشأ عن عدد كبير من العوامل الصغيرة التي لا يمكن التحكم فيها وتعتبر عوامل عشوائية ، ولا مفر للمصنع من أن يسمح بالتجاوز عن هذه الانحرافات طالما كانت في حدود معقولة يقبلها المنتج والمستهلك ، أما إذا خرجت الانحرافات عن هذه الحدود فإن المصنع يشته في وجود خلل ما إما في أجزاء آلة المصنع أو في سلوك العمال أو في الإدارة أو أية مصادر أخرى للأخطاء ، وعليه حينئذ أن يتحرى عن أسباب هذا الخلل ويعمل على تلافيه . وليس من المعقول أن يفحص المصنع كل وحدة ينتجها والمتبع أن يجرى الآتى .

نفرض مثلا أن مصنعا ينتج يوميا عشرات الألوف من الأنابيب الاسطوانية ذات مواصفات معينة منها أن طول الاسطوانة ٦ سنتيمترات مثلا . ليطمئن المصنع على توفر هذه الصفة يضع اختبارا إحصائيا ليختبر به الفرض أن الوحدات المنتجة تتوفر فيها الخاصة المطلوبة ، أى ليختبر الفرض الصفري $\mu = \mu_0$ ضد الفرض $\mu \neq \mu_0$. ويجرى هذا الاختبار في فترات منتظمة من الزمن ، مثلا كل نصف ساعة أو كل ساعة ، وفى كل مرة يأخذ عينة عادة من ٣ إلى ١٠ وحدات وقيس متوسط الطول $\bar{x}$ فيها و يطبق الاختبار . فإذا أدى الاختبار إلى قبول الفرض الصفري فإن هذا يعنى أن الإنتاج يسير سيرا طبيعيا ، أما إذا أدى إلى رفض الفرض الصفري فإن هذا يعنى أن انحرافات أطوال الوحدات المنتجة عن الطول المطلوب هي انحرافات جوهرية وينبغى حينئذ التحرى عن أسباب هذه الانحرافات . وهذا ما يسمى بمراقبة الإنتاج .

وتوضع قاعدة الاختبار على هيئة فترة ثقة مركزها القيمة μ المطلوبة (أى أن حدى الثقة يكونان على بعدين متساويين من μ) . وإذا افترضنا أن توزيع الأطوال في الوحدات المنتجة معتدل متوسطه μ وانحرافه المعياري σ فإن توزيع المعاينة

للمتوسطات الحسابية للعينات التى من الحجم n يكون معتدلا متوسطه μ وانحرافه
المعيارى $\frac{\sigma}{\sqrt{n}}$ - راجع البند (٦ - ٣ - أولا) - وتكون الفترة

$$(\bar{x} - 2,58 \times \frac{\sigma}{\sqrt{n}}, \bar{x} + 2,58 \times \frac{\sigma}{\sqrt{n}})$$

هى فترة ثقة بدرجة ٩٩٪ للمتوسط μ . أى أنه فى ٩٩٪ من العينات التى نأخذها
تقع μ فى هذه الفترة ، أى تتحقق المتباينة الآتية :

$$\bar{x} - 2,58 \times \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 2,58 \times \frac{\sigma}{\sqrt{n}}$$

ومن هذه المتباينة المزدوجة ينتج أن

$$(29) \quad \bar{x} - 2,58 \times \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 2,58 \times \frac{\sigma}{\sqrt{n}}$$

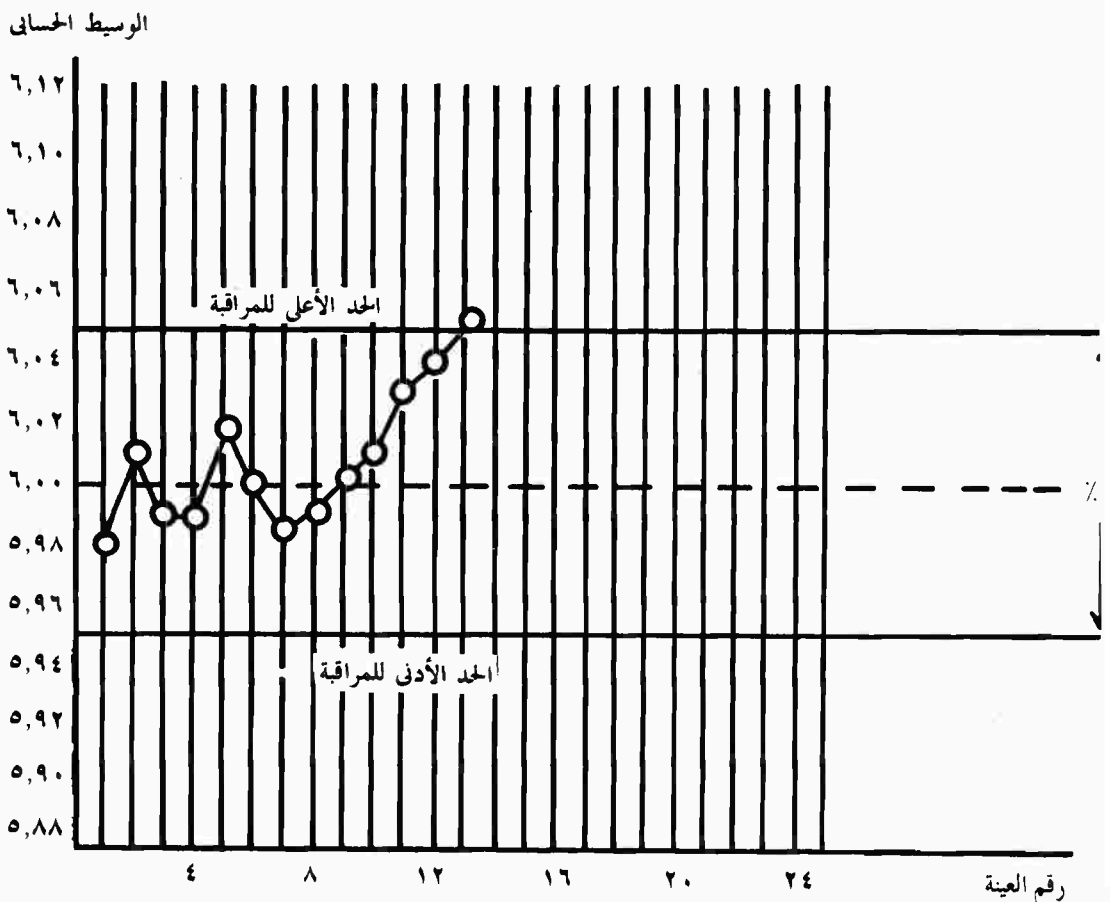
ويسمى الطرف الأيسر من هذه المتباينة **بحد المراقبة الأدنى** ، ويسمى الطرف الأيمن
منها **بحد المراقبة الأعلى** lower and upper control limits وعند المراقبة إذا أخذت عينة
ووجد أن متوسطها $\bar{x}$ يقع بين هذين الحدين يقبل الفرض الصفري ويستمر
الإنتاج دون تعديل ، أما إذا ساوت $\bar{x}$ أحد الحدين أو تعدته فيرفض الفرض
الصفري عند مستوى الدلالة ٠,٠١ . وينبغى حينئذ اتخاذ ما يلزم (وربما إيقاف
الآلة) لتحرى أسباب الخلل وإصلاحه .

وفى المعتاد تستخدم فترة ثقة بدرجة ٩٩٪ أما إذا أردنا استخدام فترة ثقة بدرجة
٩٥٪ فإننا نضع العدد ١,٩٦ بدلا من العدد ٢,٥٨ فى المتباينة (٢٩) - راجع
المثال (٤ - ٣) - أما قيمة σ فيمكن تقديرها من عينة عشوائية كبيرة تؤخذ من
الإنتاج عندما يكون فى حالته الطبيعية .

ولتسهيل عملية المراقبة ترصد قيمة المتوسط $\bar{x}$ لكل عينة مختبرة بالترتيب في خريطة تسمى بخريطة المراقبة control chart . فإذا فرضنا أن $\mu = 6$ سم ،

$$\sigma = 0.6 \text{ سم} , n = 10 \text{ فان حدى المراقبة يكونان } 6 \pm \frac{0.6}{\sqrt{10}} \times 2.58$$

أى بالتقريب ٥,٩٥ ، ٦,٠٥ من السنتيمترات وتكون خريطة المراقبة كالآتى :



الشكل (٦-٦) : خريطة المراقبة للوسط الحسابى

وتبين هذه الخريطة الخططين الممثلين للحددين الأدنى والأعلى للمراقبة ، كما تمثل النقط متوسطات ١٣ عينة (حجم كل منها ١٠) بترتيب اختيارها وهى ، ٥,٩٨١ ، ٦,٠١١ ، ٥,٩٩١ ، ٥,٩٩٢ ، ٦,٠٢٢ ، ٦,٠٠٣ ، ٥,٩٨٥ ، ٥,٩٩٢ ، ٦,٠٠٤ ، ٦,٠١٣ ، ٦,٠٣٢ ، ٦,٠٤٣ ، ٦,٠٥٤ ، ويبدو من قيمة المتوسط الأخير أن الإنتاج فى حاجة إلى تعديل عند الوصول إلى العينة ١٣ . هذا ويمكن بنفس الطريقة متابعة التغير فى تباين الإنتاج أو فى أى بارامتر آخر .

تمارين (٦ - ٥)

أخذت عشر عينات حجم كل منها ٤ وحدات فى فترات منتظمة ووجد أن سمك الوحدات كما فى الجدول الآتى . مثل متوسطات هذه العينات على خريطة مراقبة على فرض أن المجتمع معتدل متوسطه ٥ وانحرافه المعيارى ١,٥٥ .

الزمن	٨/-	٨/٣٠	٩/-	٩/٣٠	١٠/-	١٠/٣٠	١١/-	١١/٣٠	١٢/-	١٢/٣٠
سمك الوحدة	٣	٣	٥	٧	٧	٤	٥	٦	٥	٥
	٤	٦	٢	٥	٣	٤	٦	٤	٥	٢
	٨	٦	٥	٤	٦	٣	٤	٦	٦	٥
	٤	٨	٦	٤	٥	٦	٦	٤	٤	٣