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 الإحصائى 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 يلى :

obeikandl.com

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

(ب) تقدير متوسطات وتباينات المجتمعات وغيرها من الأدلة .

(جـ) الكشف عما إذا كانت الفروق المشاهدة في العينات هى فروق راجعة إلى الصدفة أو تقلبات العينات أو هى 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 :

فى نظرية العينات نميز بين ثلاثة أنواع من التوزيعات .

(١) توزيع المجتمع Population Distribution

(٢) توزيع العينة Sample Distribution

(٣) توزيع المعاينة Sampling Distribution

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 الحقيقي** 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 $\bar{س}$ 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** ن 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 ن . ويمكننا أن نحسب الوسط الحسابي والانحراف المعيارى ومعامل الالتواء ... لهذا التوزيع . وبالمثل يمكن أن نتصور توزيع المعاينة للتباين للعينات ذوات الحجم ن أو توزيع المعاينة لأى مقياس آخر .

في هذا المثال يستحيل عملياً إيجاد توزيع المعاينة للوسط الحسابي بالطريقة المذكورة لأن عدد العينات الممكنة هو عدد فلكى نعجز عن الحصول عليه . وفي المعتاد نحصل على توزيعات المعاينة بطرق رياضية ، إلا أنه لتوضيح مفهوم توزيع المعاينة نضرب المثال البسيط الآتي .

مثال (٦ – ١) :

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

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

(ثانيا) أحسب الوسط الحسابى 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 ٦ × ٦ = ٣٦ 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 هى :

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

الأوساط الحسابية لهذه العينات هى :

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

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 الحسابى لأوزان ٦ أرانب للعينات من الحجم ٢

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

(ثانيا) الوسط الحسابى لتوزيع المعاينة $= \mu_{\bar{س}} = \frac{١}{ن} \text{مج } ك_ر \bar{س}_ر$

أى أن $\mu_{\bar{س}} = \frac{١}{٣٦}$ (١×١١ + ٢×١١,٥ +...+٤×١٦)

$= \frac{٥٠٤}{٢٦} = ١٤$ أوقية

تباين توزيع المعاينة = $\frac{1}{ن}$ [$\sum ك_ر \bar{س}_ر^2 - \frac{(\sum ك_ر \bar{س}_ر)^2}{ن}$]

أى أن $\sigma_{\bar{س}}^2$ = $\frac{1}{36}$ [(١١×١)^2 + ١١,٥×٢)^2 +...+ ١٦×٤)^2 - $\frac{٥٠٤^2}{36}$]

= $\frac{11}{6}$

الوسط الحسابى للمجتمع μ = $\frac{1}{6}$ (١١+١٦+١٢+١٥+١٦+١٤) = ١٤

تباين المجتمع σ^2 = $\frac{1}{6}$ [(١١^2 + ١٦^2 +...+ ١٤^2) - $\frac{٨٤^2}{6}$] = $\frac{11}{3}$

بالمقارنة نجد ما يلى :

(أ) $\mu_{\bar{س}} = \mu$

أى أن الوسط الحسابى لتوزيع المعاينة = الوسط الحسابى للمجتمع .

(ب) $\sigma_{\bar{س}}^2 = \frac{\sigma^2}{ن}$

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

وهاتان النتيجتان 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 $\bar{س}_ر$ لتقدير الوسط الحسابى μ للمجتمع يكون هناك خطأ فى التقدير قدره $\bar{س}_ر - \mu$ ففى هذا المثال لدينا μ = ١٤ وإذا قدرنا هذا المتوسط من متوسط العينة الأولى وهو ١١ يكون هناك خطأ قدره ١١-١٤ = -٣ وحين نقدره من متوسط العينة الثانية وهو ١١,٥

obeikandl.com

يكون هناك خطأ قدره ٢ (١١,٥ - ١٤) = ٢ × - ٢,٥ = -٥ (مع ملاحظة أن هناك عينتين متوسط كل منهما ١١,٥) 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) .

(٦ - ٢ - ١) الإحصاءة A statistic

فى المثال السابق كان لدينا ٣٦ وسطا حسابيا $\bar{س}_1$ ، $\bar{س}_2$ ، ... ، $\bar{س}_{36}$ وسمينا التوزيع التكرارى لهذه المتوسطات بتوزيع المعاينة للوسط الحسابى للعينات ذوات الحجم ٢.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 $\bar{س}$ ونسمى هذا المتغير حينئذ بالإحصاءة $\bar{س}$ ويكون التوزيع سابق الذكر هو توزيع المعاينة لهذه الإحصاءة .

كذلك إذا كانت $ع^2_1$ ، $ع^2_2$ ، ... هى تباينات جميع العينات التى من نفس الحجم التى يمكن أن تؤخذ من مجتمع ما فإننا ننظر إليها كقيم متغير عشوائى $ع^2$ ويكون توزيع هذه القيم هو توزيع المعاينة للإحصاءة $ع^2$. وبالمثل لأى مقياس إحصائى آخر .

(٦ - ٢ - ٢) الخطأ المعيارى Standard Error

يسمى الانحراف المعيارى لتوزيع المعاينة لإحصاءة ما بالخطأ المعيارى لهذه الإحصاءة . ففى المثال السابق الخطأ المعيارى للوسط الحسابى أو للاحصاءة $\bar{س}$ هو $\sigma_{\bar{س}} = \sqrt{\frac{١١}{٦}} = ١,٨٣$ تقريبا . والأخطاء المعيارية تلعب دورا رئيسيا فى نظرية العينات كما سنتبين بعد .

(٦ – ٢ – ٣) التقدير غير المتحيز Unbiased Estimate

نظرا لأن الوسط الحسابى $\mu_{\bar{س}}$ للإحصاءة $\bar{س}$ يساوى الوسط الحسابى μ للمجتمع ، نقول إن $\bar{س}$ هو مُقدِّر غير متحيز للوسط الحسابى للمجتمع unbiased estimator كما نقول للوسط الحسابى $\bar{س}_ر$ لأى عينة عشوائية بأنه تقدير غير متحيز لمتوسط المجتمع ، وأى فرق $\bar{س}_ر - \mu$ يعتبر كما سبق القول خطأ فى تقدير متوسط المجتمع من متوسط العينة ، ويقاس مدى دقة هذا التقدير بالخطأ المعيارى $\sigma_{\bar{س}} = \frac{\sigma}{\sqrt{ن}}$ لتوزيع المعاينة للاحصاءة $\bar{س}$. لاحظ أن هذا الخطأ يتوقف على كل من σ ، ن ويكون هذا الخطأ صغيرا إذا كانت σ صغيرة أو كان حجم العينة ن 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 نقدره ، ومن الواضح أن هذه الصفة مطلوبة فى التقدير . وكما وجدنا أن $\bar{س}$ تقدير غير متحيز للبارامتر μ نجد أن $ع^٢ = \frac{١}{ن - ١}$ مجـ (س – $\bar{س}$)^٢ تقدير غير متحيز للتباين $\sigma^٢$ 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 $\bar{س}$ للعينات ذوات الحجم ن المأخوذة من هذا المجتمع يكون توزيعا **معتدلا** وسطه الحسابى μ وانحرافه المعيارى $\frac{\sigma}{\sqrt{ن}}$ (أى الخطأ المعيارى) ، وبالتالى فإن الإحصاءة

$$ص_{١} = \frac{\mu - \bar{س}}{\sigma / \sqrt{ن}} \qquad (١)$$

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

كما أن الإحصاءة

$$ص_{٢} = \frac{\mu - \bar{س}}{ع / \sqrt{ن}} \qquad (٢)$$

يكون توزيعها مطابقاً لتوزيع ت بدرجات حرية عددها ن - ١ . وتستخدم هذه الإحصاءة حينما تكون $\sigma^{٢}$ مجهولة القيمة وهذا ما يحدث في أغلب الحالات ، ونضطر حينئذ لاستخدام تباين العينة وهو

$ع^{٢} = \frac{١}{ن-١} (مجس^{٢} - \frac{(مجس)^{٢}}{ن})$ كتقدير غير متحيز للتباين $\sigma^{٢}$ 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 $\bar{س}_1$ ترمز إلى الوسط الحسابي لعينة عشوائية حجمها $ن_1$ مأخوذة من المجتمع الأول وأن الإحصاءة $\bar{س}_2$ ترمز إلى الوسط الحسابي لعينة عشوائية حجمها $ن_2$ مأخوذة من المجتمع الثاني . إذا كانت هاتان العينتان مستقلتين فإن توزيع المعاينة للإحصاءة $(\bar{س}_1 - \bar{س}_2)$ يكون توزيعاً معتدلا وسطه الحسابي $(\mu_1 - \mu_2)$ وانحرافه المعيارى $\sqrt{\frac{\sigma_1^2}{ن_1} + \frac{\sigma_2^2}{ن_2}}$ وبالتالى فإن الإحصاءة :

$$صـ_3 = \frac{(\bar{س}_1 - \bar{س}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{ن_1} + \frac{\sigma_2^2}{ن_2}}} \qquad (٣)$$

يكون توزيعها مطابقاً للتوزيع المعتدل المعيارى .

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

كما أن الإحصاءة

$$صـ_4 = \frac{(\bar{س}_1 - \bar{س}_2) - (\mu_1 - \mu_2)}{ع\sqrt{\frac{1}{ن_1} + \frac{1}{ن_2}}} \qquad (٤)$$

يكون توزيعها مطابقاً لتوزيع ت بدرجات حرية عددها $ن_1 + ن_2 - ٢$

$$\text{حيث } ع_{ر}^{2} = \frac{(ن_{1} - 1) ع_{1}^{2} + (ن_{2} - 1) ع_{2}^{2}}{ن_{1} + ن_{2} - 2} \qquad (أ)$$

$$= \frac{مج(س_{1ر} - \bar{س}_{1})^{2} + مج(س_{2ر} - \bar{س}_{2})^{2}}{ن_{1} + ن_{2} - 2} \qquad (أ)$$

وحيث $\sigma_{1} = \sigma_{2}$

أما إذا كان $\sigma_{1} \neq \sigma_{2}$ فتستخدم الاحصاءة

$$\text{صـ}_{0} = \frac{(\bar{س}_{1} - \bar{س}_{2}) - (\mu_{1} - \mu_{2})}{\sqrt{ع_{1}^{2}/ن_{1} + ع_{2}^{2}/ن_{2}}} \qquad (٥)$$

التى لها توزيع ت بدرجات حرية $\left(\frac{ع_{1}^{2}}{ن_{1}} + \frac{ع_{2}^{2}}{ن_{2}}\right)^{2} \div$

$$\left[\left(\frac{ع_{1}^{2}}{ن_{1}}\right)^{2} \div (ن_{1} - 1) + \left(\frac{ع_{2}^{2}}{ن_{2}}\right)^{2} \div (ن_{2} - 1)\right]$$

ملاحظة :

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

(أ) يقترب توزيع الإحصاءة صـ$_{1}$ المعرفة في (١) من التوزيع المعتدل المعيارى حين يقترب حجم العينة من اللانهاية .

obeikandl.com

(ب) يقترب توزيع الإحصاءة صـ٣ المعرفة في (٣) من التوزيع المعتدل المعيارى حين تقترب كل من ن١ ، ن٢ من اللانهاية .

إن المعني 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 صـ١ ، صـ٣ هو بالتقريب توزيع معتدل معيارى ، وكلما كبر حجم العينات كلما قل الخطأ الناشيء عن هذا التقريب .

(ثالثاً) توزيع المعاينة لخارج قسمة تباينين :

إذا كانت الإحصاءتان ع٢١ ،ع٢٢ ترمزان إلى تبايني عينتين عشوائيتين مستقلتين مأخوذتين من مجتمعين معتدلين لهما نفس التباين ($\sigma^2_1 = \sigma^2_2$) فإن الإحصاءة

$$صـ٦ = \frac{ع٢١}{ع٢٢} \qquad (٦)$$

يكون توزيعها مطابقاً لتوزيع ف بدرجتي حرية ن١ - ١ ، ن٢ - ١ حيث ن١ ، ن٢ هما حجما العينتين. ويلاحظ أنه إذا كانت ع٢١ ، ع٢٢ قيمتين مأخوذتين من عينتين عشوائيتين مستقلتين من مجتمع واحد أو من مجتمعين لهما نفس التباين فإن النسبة ع٢١/ ع٢٢ تكون قريبة من الواحد لأن كلا من ع٢١ ، ع٢٢ 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 ن من التجارب هو توزيع ذى الحدين دليلاه ن ، ح،وسطه الحسابى ن ح وانحرافه المعيارى $\sqrt{\text{ن ح ك}}$ ويكون التكرار النسبي لوقوع هذا الحدث هو خارج قسمة عدد مرات وقوع الحدث على العدد ن وعلى ذلك فإن متوسط هذه النسبة يساوى :

$$\frac{\text{ن ح}}{\text{ن}} = \text{ح} \text{ وانحرافها المعيارى } \frac{\sqrt{\text{ن ح ك}}}{\text{ن}} = \sqrt{\frac{\text{ح ك}}{\text{ن}}}$$

وإذا أخذنا من هذا المجتمع جميع العينات ذوات الحجم ن وحسبنا في كل منها القيمة س لعدد مرات وقوع ذلك الحدث ثم حسبنا نسبة هذا العدد وهو ر = $\frac{\text{س}}{\text{ن}}$ فإننا نحصل على توزيع المعاينة لهذه النسبة .

وقد رأينا فى البند (٤ – ٦) أنه إذا كانت ن كبيرة ولم تكن أى من ح أو ك قريبة من الصفر فإن :

(أ) توزيع الاحتمال لعدد مرات وقوع الحدث في العينات ذوات الحجم ن يقترب من توزيع معتدل وسطه الحسابي ن ح ، وانحرافه المعيارى $\sqrt{\text{ن ح ك}}$ وبالتالى فإن توزيع الإحصاءة

$$\text{ص}_{\text{٧}} = \frac{\text{س} - \text{ن ح}}{\sqrt{\text{ن ح ك}}} \qquad (٧)$$

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

(ب) وبالمثل فإن توزيع الإحصاءة :

$$\text{ص}_{\text{٨}} = \frac{\text{ر} - \text{ح}}{\sqrt{\frac{\text{ح ك}}{\text{ن}}}} \qquad (٨)$$

يقترب من التوزيع المعتدل المعيارى .

(٦ – ٥) اختبارات الفروض : TESTS OF HYPOTHESES

الفرض الإحصائى : 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 وهى 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 تؤدى إلى اتخاذ قرار برفض أو قبول الفرض فور حصولنا على قيم مشاهدة في عينة عشوائية . وفي المعتاد نضع فرضاً نرمز له بالرمز ف . نسميه **بالفرض الصفرى** أو بفرض العدم null hypothesis نصوغه بحيث يعبر عن انعدام الفرق بين شيئين ، مثلا $\mu_1 = \mu_2$ (أى $\mu_1 - \mu_2 = ٠$) أو $\mu = أ$ (أى $\mu - أ = ٠$) أو لا توجد علاقة بين متغيرين . وأى فرض يراد اختباره ضد الفرض الصفرى يسمى بالفرض الآخر أو بالفرض البديل alternative hypothesis ويرمز له بالرمز ف$_1$ وقد يأخذ هذا الفرض الشكل . $\mu_1 \neq \mu_2$ أو $\mu_1 > \mu_2$ أو $\mu_1 < \mu_2$.

ولاختبار الفرض الصفرى ف$_0$ ضد الفرض الآخر ف$_1$ نبدأ باختيار إحصاءة تناسب الفرض المختبر بشرط أن يكون توزيع المعاينة لهذه الإحصاءة معلوماً لنا ثم نحدد في هذا التوزيع منطقة م يكون احتمال وقوع القيم المشاهدة لهذه الإحصاءة فيها هو احتمال صغير α ، مثلا ٠,٠٥ ، ٠,٠١ ، ٠,٠٠١ ، ... وتسمى هذه المنطقة بالمنطقة الحرجة أو بمنطقة الرفض critical region or region of rejection أما الاحتمال α فيسمى بمستوى الدلالة للاختبار level of significance وتأخذ القاعدة التي يعطيها الاختبار الصورة الآتية :

« إذا وقعت قيمة مشاهدة لهذه الإحصاءة ، محسوبة على أساس صحة الفرض الصفرى ف$_0$ ، داخل المنطقة الحرجة نرفض ف$_0$ عند مستوى الدلالة α وإلا نقبل ف$_0$ » .

وذلك على أساس أن احتمال وقوع القيمة المشاهدة داخل المنطقة الحرجة هو احتمال ضئيل يجعلنا نشك في صحة الفرض الصفرى . ونظراً لأن القيم المشاهدة يمكن أن تقع في أى جزء من التوزيع فإن هذه القاعدة تعنى أننا نغامر برفض الفرض الصفرى (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 $١ - \alpha$.

وجدير بالملاحظة أنه بينما يمكننا الاختبار من هدم الفرض الصفرى 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 دليلا كافيا 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$_١$.

واختبار الفرض ف. عن مجتمع ضد فرض آخر ف$_١$ يستلزم بطبيعة الحال الحصول على عينة عشوائية من هذا المجتمع ، ويفضل اختيار قيمة α حتي قبل اختيار العينة . ويتوقف تحديدنا لقيمة α على طبيعة المشكلة التي نتناولها ودرجة المغامرة التي نقبلها لتحمل مسئولية الخطأ المحتمل . وبصفة عامة يمكن تلخيص خطوات إجراء اختبار الفروض فيما يلى :

(أولا) نحدد الفرض الصفرى ف. والفرض الآخر ف$_١$ من واقع المشكلة التي نتناولها .

(ثانياً) نحدد الإحصاءة التى تناسب الفرض المختبر ونحدد المنطقة الحرجة م فى 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 الصفرى صحيح .

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

ويجدر الإشارة هنا إلى أن اختبارات الفروض ما هى إلا وسيلة حسابية تستخدم البيانات التى حصلنا عليها من العينات لإلقاء الضوء على صحة أو خطأ الفرض الصفرى ، وينبغى 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 دورا هاما فى عملية الاستدلال الاحصائى . وحين تكون العينة صغيرة فإن اختبار الدلالة لا يؤدى إلى رفض الفرض الصفرى إلا إذا كان هذا الفرض خاطئا بدرجة كبيرة ، وبالعكس حين تكون العينة كبيرة فإن أى انحراف صغير عن الفرض الصفرى يسهل اكتشافه كانحراف ذى دلالة . وبصفة عامة يفضل أن يكون حجم العينة كبيرا بدرجة كافية منعا للوقوع فيما يسمى بالخطأ من النوع الثانى ، وهذا ما سوف نتناوله فى الفصل السابع .

في البنود (٦ – ٦) ، (٦ – ٧) (٦ – ٨) الآتية نتناول ثلاثة من أشهر اختبارات الفروض تعرف باختبار ت واختبار χ^2 (كا^2) واختبار ف .

(٦ – ٦) اختبار ت : THE t-TEST

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 هى :

$$ت = \frac{\text{القيمة المشاهدة للاحصاءة} - \text{الوسط الحسابى للإحصاءة}}{\text{تقدير للخطأ المعيارى للاحصاءة}}$$

بشرط أن يكون للإحصاءة توزيع معتدل وأن تكون القيمة المشاهدة هى 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 بيولوجى 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 معيارى 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 الذى تنتجه الشركة يساوى الجرعة القياسية التي حددتها التعليمات الحكومية وهى ٦٠٠ وحدة نشاط / سم^{٣}. نتبع الخطوات الأربع المشار إليها فى البند السابق .

(١) الفرض الصفرى ف$_{٠}$: $\mu = ٦٠٠$

الفرض الآخر ف$_{١}$: $\mu \neq ٦٠٠$

$\alpha = ٠,٠٥$ ،

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

يدعونا إلى عدم الثقة في القيمة المفروضة للدليل μ . (هذا على أساس أن العينة هى عينة عشوائية ممثلة للمجتمع تمثيلا جيداً وبالتالى فإن وسطها الحسابي $\bar{س}$ هو عدد نثق فيه) . هذا التحليل يشير مباشرة إلى أن أنسب إحصاءة لقياس صغر أو كبر هذا الفرق هى الإحصاءة (٢) وهى :

$$ت = \frac{\mu - \bar{س}}{ع / \sqrt{ن}}$$

التي نعلم أن توزيعها يطابق توزيع ت بدرجات حرية عددها $\nu = ن - ١$ ، مع ملاحظة أن هذه الإحصاءة تنعدم عندما تتساوى قيمتي μ ، $\bar{س}$ وتكبر قيمتها المطلقة كلما كبر الفرق بينهما .

المنطقة الحرجة $م = \{ت : |ت| > ت_{[١ - \frac{\alpha}{٢}]}\}$

الشكل (٦ -١)

وهى تتألف من جزءين واقعين أسفل جانبي منحني ت واحتمال كل منهما $\frac{\alpha}{٢}$ وبالتالى فإن احتمال وقوع قيمة الإحصاءة في هذه المنطقة هو احتمال صغير α .

(٣) نحسب قيمة هذه الإحصاءة من بيانات العينة وعلى أساس صحة الفرض الصفرى . وإذا رمزنا لهذه القيمة بالرمز $ت_{ى}$ نجد أن :

$$ت_{ى} = \frac{٥٩٢,٥ - ٦٠٠}{١١,٢ / \sqrt{١٠}} = -٢,١٢ \text{ حيث } \nu = ١٠ - ١ = ٩$$

فإذا كنا قد اخترنا $\alpha = ٠,٠٥$ نجد من جدول ت أن القيمة الحرجة التى تحدد منطقة الرفض هى : $ت_{٠,٠٥ ,١٩|} = ٢,٢٦٢$

(٤) الاستنتاج : بما أن $ت_{ى} = -٢,١٢$ لا تقع في المنطقة الحرجة لا يكون لدينا دليل ضد الفرض الصفرى عند مستوى الدلالة ٠,٠٥ أى أننا نقبل الفرض أن $\mu = ٦٠٠$ ونقرر بأن إنتاج الشركة يتمشى مع التعليمات الحكومية .

الاختبار ذو الجانب الواحد وذو الجانبين :

في المثال السابق اختبرنا الفرض الصفرى ف$_{٠}$: $\mu = ٦٠٠$ ضد الفرض ف$_{١}$: $\mu \neq ٦٠٠$ وعلى ذلك كان لنا أن نرفض ف$_{٠}$ في حالتين هما : أن تكون $\mu < ٦٠٠$ أو تكون $\mu > ٦٠٠$ ولذلك شملت المنطقة الحرجة كلا جانبي توزيع ت . ونقول حينئذ إن الاختبار ذو جانبين أو إنه اختبار غير اتجاهى .

ولكن إذا كانت زيادة وحدات النشاط في المستحضر البيولوجى عن المستوى القياسي لا ينتج عنها ضرر بينما النقصان عنه يفقد المستحضر فعاليته فإن اهتمامنا يكون منصباً على ما إذا كان متوسط العينة أصغر صغراً ذا دلالة من المستوى القياسي . في هذه الحالة يكون المطلوب اختبار الفرض الصفرى ف$_{٠}$: $\mu = ٦٠٠$ ضد الفرض ف$_{١}$: $\mu < ٦٠٠$ وتكون المنطقة الحرجة فى جانب واحد فقط من التوزيع هو الجانب الأيسر ، أى أن المنطقة الحرجة تكون في هذه الحالة على الصورة ،

$$م = \{ت : ت < -ت_{٢\alpha\,[ن-١]}\}$$

الشكل (٦-٢)

ونقول حينئذ إن الاختبار ذو جانب واحد أو إنه اختبار اتجاهى . فإذا كان مستوى الدلالة هو α = ٠,٠٥ فإن القيمة التي تحد المنطقة الحرجة من اليمين وتجعل المساحة بينها وبين المنحنى ٠,٠٥ يمكن إيجادها من جدول ت باستخدام الاحتمال $\acute{\alpha}$ = ٢ α = ٠,١٠ .

من الجدول نجد أن ت $_{٠,١}$ [٩] = ١,٨٣٣

إذن القيمة التي تحد المنطقة الحرجة من اليمين هى – ١,٨٣٣٠

بما أن –٢,١٢ < –١,٨٣٣ فإن القيمة المشاهدة –٢,١٢ تقع في المنطقة الحرجة وبالتالى نرفض الفرض الصفرى ف عند مستوى الدلالة ٠,٠٥ 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 يجيب على سؤال مختلف هو السؤال الذى تتطلب المشكلة الإجابة عنه .

وبالمثل يمكن أن يكون اهتمامنا منصباً على الجانب الأيمن فقط من التوزيع حيث تكون المنطقة الحرجة على الصورة م = {ت : ت> ت $_{٢\alpha[١-\nu]}$} . وفي تناول أى مشكلة علينا أن نفكر جيداً قبل أن نحدد ما إذا كانت تتطلب اختباراً ذا جانب واحد أو ذا جانبين تحسباً من الوقوع في خطأ في عملية الاستدلال . وهذا التحديد ينبغى أن يتقرر عند تصميم التجربة وقبل جمع البيانات وبحسب التساؤل الذى تطرحه المشكلة . فمثلا 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 جانبين . أما إذا كنا نقارن نوعا جديدا من الدواء بنوع تقليدى فإن اهتمامنا يكون منصبا على معرفة ما إذا كان الدواء الجديد أفضل من التقليدى وهنا يجب أن يكون الاختبار ذا جانب واحد .

ملاحظة (١) :

يتطلب استخدام اختبار ت بهذه الصورة تحقق الشرطين الآتيين :

(أ) أن تكون العينة عشوائية بسيطة .

(ب) أن يكون المجتمع معتدلا (أو يمكن اعتباره معتدلا) .

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

CONFIDENCE INTERVALS

إذا كان الوسط الحسابي μ لمجتمع ما مجهولا ، يمكن تقديره بواسطة الوسط الحسابي $\bar{س}$ 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 **معتدلا** وسحبنا منه عينة عشوائية بسيطة حجمها ن ووسطها الحسابي $\bar{س}$ وانحرافها المعيارى ع فإن الفترة :

$$\left(\bar{س} - \frac{ع}{\sqrt{ن}}\ ت_{[\alpha - ١]\nu}\ ،\ \bar{س} + \frac{ع}{\sqrt{ن}}\ ت_{[\alpha - ١]\nu}\right) \qquad (١٠)$$

تسمى بفترة ثقة بدرجة $(١ - \alpha)$ لمتوسط المجتمع μ . وهذه العبارة لا تعني أن احتمال وقوع μ في هذه الفترة هو $(١ - \alpha)$ 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 $(١ - \alpha)$ لكل منها فإن $(١ - \alpha)$ من هذه الفترات تشمل البارامتر μ » ، وهذا ما يمكن إثباته رياضياً .

ويسمى العددان $\bar{س} \pm \frac{ع}{\sqrt{ن}}$ ت$_{[ن - ١]\alpha}$ اللذان يحدان فترة الثقة بحدى الثقة بدرجة $(١ - \alpha)$ أما العدد $\frac{ع}{\sqrt{ن}}$ فهو تقدير للخطأ المعيارى للإحصاءة $\bar{س}$ ، كما تسمى α بمعامل الثقة .

مثال (٦ - ٣) :

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

الحل :

لدينا ن = ١٠ ، $\bar{س}$ = ٥٩٢,٥ ، ع = ١١,٢ ، ١ - α = ٠,٩٥ ومنها α = ٠,٠٥

، ت$_{[٩]٠,٠٥}$ = ٢,٢٦٢ ، $\frac{ع}{\sqrt{ن}} = \frac{١١,٢}{\sqrt{١٠}}$ = ٣,٥٤

بالتعويض في (١٠) نجد أن :

الحد الأدني للفترة = ٥٩٢,٥ - ٣,٥٤ × ٢,٢٦٢ = ٥٨٤,٤٩

الحد الأعلى للفترة = ٥٩٢,٥ + ٣,٥٤ × ٢,٢٦٢ = ٦٠٠,٥١

إذن الفترة (٥٨٤,٤٩ ، ٦٠٠,٥١) هى فترة ثقة بدرجة ٩٥٪ لمتوسط المجتمع .

ملاحظة (٢) :

في المثال (٦ – ٢) قبلنا الفرض أن μ = ٦٠٠ ضد الفرض $\mu \neq$ ٦٠٠ 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 متوسطى مجتمعين معتدلين :

(أو اختبار دلالة الفرق بين متوسطى 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 Internediate White وقد نتجت البيانات الآتية من عينتين عشوائيتين مستقلتين .

ذات الزهور الحمراء : $ن_1$ = ٤٣ ، $\bar{س}_1$ = ١٣٤,٧٦ ، $ع_1$ = ٣٢,٣٥

ذات الزهور البيضاء : $ن_2$ = ٢١ ، $\bar{س}_2$ = ١٦٩,٥٠ ، $ع_2$ = ٦٢,٩٣

اختبر ما إذا كان الفرق المشاهد بين متوسطى العينتين يرجع إلى تقلبات العينات أو إلى وجود فرق حقيقي بين المتوسطين μ_1 ، μ_2 للمجتمعين اللذين سحبت منهما العينات . افترض اعتدالية المجتمعين وتساوى تبايينهما وحدد α = ٠,٠٥

الحل :

الفرض الصفرى ف$_{0}$: $\mu_1 = \mu_2$ (لا يوجد فرق بين متوسطى المجتمعين)

الفرض الآخر ف$_{1}$: $\mu_1 \neq \mu_2$ (اختبار ذو جانبين)

الإحصاءة المناسبة هنا هى الإحصاءة (٤) وهى :

$$ت = \frac{(\bar{س}_1 - \bar{س}_2) - (\mu_2 - \mu_1)}{ع \sqrt{\frac{1}{ن_1} + \frac{1}{ن_2}}}$$

التي نعلم أن توزيعها يطابق توزيع ت بدرجات حرية عددها $\nu = ن_1 + ن_2 - 2$

وعلماً بأن $ع^2 = \frac{(ن_1 - 1)\, ع_1^2 + (ن_2 - 1) ع_2^2}{ن_1 + ن_2 - 2}$

على فرض أن تباينى المجتمعين متساويان

المنطقة الحرجة م = $\{ت : |ت| > ت_{\alpha[ن_1 + ن_2 - 2]}\}$

نحسب ت$_{ى}$ من بيانات العينة وعلى أساس صحة الفرض الصفرى .

$$ع^2 = \frac{٤٢ \times ٣٢,٣٥^2 + ٢٠ \times ٦٢,٩٣^2}{٦٢} = ١٩٨٦,٤١٣٥$$

$$= ١٩٨٦,٤١٣٥$$

$$\therefore ع = ٤٤,٥٦٩$$

$$\text{تقدير الخطأ المعيارى} = ع \sqrt{\frac{1}{ن_1} + \frac{1}{ن_2}} = ٤٤,٥٦٩ \sqrt{\frac{1}{٤٣} + \frac{1}{٢١}} = ١١,٨٥٥$$

$$\therefore ت_{ى} = \frac{١٣٤,٧٦ - ١٦٩,٥٠}{١١,٨٥٥} = -٢,٩٣ \text{ حيث } \nu = ٤٣ + ٢١ - ٢ = ٦٢$$

من جدول ت نجد أن ت$_{٠,٠٥[٦٢]}$ = ٢,٠٠٠

بما أن ٢,٩٣- < ٢- فإن القيمة ت$_{ى}$ = ٢,٩٣- تقع في المنطقة الحرجة (في الجانب الأيسر) وإذن نرفض الفرض الصفرى ف$_{.}$: $\mu_{١}$ = $\mu_{٢}$ عند مستوى الدلالة ٠,٠٥ ونقرر أن العينتين مأخوذتان من مجتمعين مختلفين في وسطهما الحسابي .

ملاحظة (٣)

يتطلب استخدام اختبار ت بهذه الصورة تحقق الشروط الآتية :

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

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

فترات الثقة للفرق بين متوسطى مجتمعين معتدلين :

يمكن إثبات أنه تحت الشروط المذكورة ، يصلح العددان :

$$(\bar{س}_{١} - \bar{س}_{٢}) \pm \text{خ.م.ت}_{\alpha[\nu_{١,٠} + \nu_{٢}-٢]} \qquad (١١)$$

كحدين لفترة ثقة بدرجة (١ - α) للفرق ($\mu_{١}$ - $\mu_{٢}$) بين متوسطى المجتمعين،

حيث خ . م = ع $\sqrt{\frac{١}{ن_{١}} + \frac{١}{ن_{٢}}}$ هو تقدير للخطأ المعيارى للإحصاءة $\bar{س}_{١}$ - $\bar{س}_{٢}$.

ففي المثال (٦ – ٤) نجد أن :

الحد الأدني لفترة الثقة = (١٦٩,٥ – ١٣٤,٧٦) – ١١,٨٥٥ × ٢ = ١١,٠٣

الحد الأعلى لفترة الثقة = (١٦٩,٥ – ١٣٤,٧٦) + ١١,٨٥٥ × ٢ = 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 α = ٠,٠٥ ما لم يذكر غير ذلك .

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

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

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

(أ) اختبر الفرض أن الوسط الحسابي للوزن في مجتمع هذه الحشرة هو μ = ١ ضد كل من الفرضين الآتيين (١) $\mu \neq 1$ ، (٢) $\mu > 1$

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

(٢) باستخدام عينة حجمها ن = ١٥ ووسطها الحسابي $\bar{س}$ = ٨٥,٣ وتباينها ع^2 = ١٣ اختبر الفرض أن متوسط المجتمع μ = ٨٠ وأوجد فترة ثقة بدرجة ٩٩٪ لهذا المتوسط .

(٣) في عينة عشوائية من ٢٠ شخصاً يعالجون من مرض معين وجد أن عدد الكرات الدموية البيضاء مقدرة بالآلاف كالآتي . أوجد فترة ثقة بدرجة ٩٥٪ لمتوسط عدد الكرات البيضاء في مجتمع هؤلاء المرضى .

٨ ١٣ ٧ ٨ ١٣ ٦ ٤ ١٢ ٧ ٩ ١٠ ٧ ٨

١٠ ١٢ ١٣ ١٥ ١٠ ٥ ١١

(٤) في بحث ثبات reliability إحدى طرق القياس في تجارب التبريد وجدت القيم الآتية لدرجات الحرارة في عينتين 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 العينتين ذو دلالة عند المستوى ٠,٠٥ ؟

(لتسهيل حساب التباين اطرح ١٠٠ من جميع الأعداد) .

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

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 $\mu < ٥٠$

(جـ) 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 بواسونيا بشرط أن تقع مستقلة عن بعضها بمعني 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 ن ووسطها الحسابي $\bar{س}$ وتباينها $ع^{٢}$ ونضع الفرض الصفرى أن هذه الأحداث تقع عشوائياً مستقلة عن بعضها وبالتالى تتوزع بواسونياً وتكون النسبة بين التباين والوسط الحسابي في المجتمع هى ح = ١ فإذا كان هذا الفرض صحيحاً فإن النسبة $ح_{ى} = ع^{٢} / \bar{س}$ المشاهدة في العينة ينبغى 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 ح = ١ نستخدم الإحصاءة

$$ت = \frac{\frac{ع^{٢}}{\bar{س}} - ١}{\sqrt{\frac{٢}{ن - ١}}} \qquad (١٢)$$

التي يمكن إثبات أن توزيعها يطابق توزيع ت بدرجات حرية عددهًا ن – ١ . 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 :

$$(\text{أ})\ \text{ح}_{\text{ع}} = \text{ع}^{2}/\overline{\text{س}} = \frac{١,٩٦٥}{١,٨} = ١,٠٩٢$$

الفرض الصفرى ف$_{٠}$: ح = ١

الفرض الآخر ف$_{١}$: ح > ١

من (١٢) نجد أن :

$$\text{ت}_{\text{ع}} = \frac{١,٠٩٢ - ١}{\sqrt{\frac{٢}{٣٩٩}}} = \frac{٠,٠٩٢}{٠,٠٧١} = ١,٢٩٦$$

من الجدول (٧) نجد أن $\text{ت}_{٠,١٠[٣٩٩]} = ١,٦٤٥$

بما أن ١,٢٩٦ < ١,٦٤٥ لا يكون لدينا دليل ضد الفرض الصفرى ح=١ ولا يسعنا إلا قبوله . وهذا يعني أن الأحداث تتوزع بواسونياً وبالتالى تقع عشوائياً مستقلة عن بعضها البعض .

$$(\text{ب})\ \text{ح}_{\text{ع}} = \text{ع}^{2}/\overline{\text{س}} = \frac{٢,١٥}{٣,٣} = ٠,٦٥٢$$

obeikandl.com

$ف_{.}$: ح = ١

$ف_{١}$: ح > ١

$$ت_{ى} = \frac{١ - ٠,٦٥٢}{٠,٠٧١} = -٤,٩٠١$$

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

(٦ - ٧) اختبار $\chi^{٢}$ THE CHI—SQUARE TEST

في كثير من الأبحاث نتناول مجتمعاً صنفت عناصره بحسب خاصة أو صفة ما إلى م من الأقسام المنفصلة $ا_{١}$ ، $ا_{٢}$ ، ... ، أم ويكون المطلوب اختبار ما إذا كان هذا المجتمع له نموذج معين من حيث هذه الخاصة ، 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 $ا_{١}$ هو $ح_{١}$. واحتمال انتمائه إلى القسم $ا_{٢}$ هو $ح_{٢}$ ... واحتمال انتمائه إلى القسم $ا_{م}$ هو $ح_{م}$. وإذا سحبنا عينة عشوائية حجمها ن فإننا نتوقع أن يكون ن $ح_{١}$ من عناصرها منتمياً إلى القسم $أ_{١}$ ، ن $ح_{٢}$ منها منتمياً إلى القسم $أ_{٢}$ ... ، ن $ح_{م}$ منتمياً إلى القسم $أ_{م}$. نفرض أننا قمنا فعلاً بسحب عينة عشوائية حجمها ن ووجدنا أن عدد العناصر التي انتمت إلى الأقسام $أ_{١}$ ، $أ_{٢}$ ... ، $أ_{م}$ هو على الترتيب $ك_{١}$ ، $ك_{٢}$ ... ، $ك_{م}$. إن هذه الأعداد تسمى بالتكرارات المشاهدة ، أما الأعداد ن $ح_{١}$ ، ن $ح_{٢}$ ... ، ن $ح_{م}$ فهي التكرارات النظرية المتوقعة لها ، ولذلك نجد

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

والإحصاءة المناسبة لذلك هى :

$$صـ_{٩} = مـ_{ر=١} \frac{(ك_ر - ف_ر)^٢}{ف_ر} \quad \text{حيث } ف_ر = ع_ر ن \qquad (١٣)$$

$$\text{وحيث } مـ\, ك_ر = مـ\, ف_ر = ن \qquad (١٤)$$

التي يمكن إثبات أن توزيعها يقترب من توزيع $\chi^٢$ بدرجات حرية (م – ١) كلما اقتربت ن من اللانهاية .

ولصحة استخدام هذه الإحصاءة ينبغى أن تتحقق الشروط الآتية :

(أ) أن تكون العينة عشوائية .

(ب) ألا يقل حجم العينة عن ٤٠ (لأن توزيع الإحصاءة صـ٩ تقريبى) .

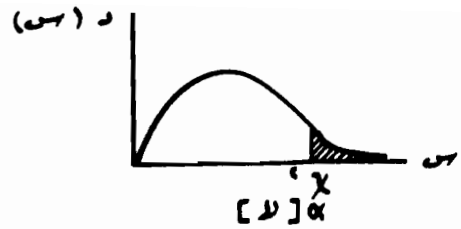
(جـ) ألا تقل أى ق_ر عن خمسة ، على أنه إذا وجدت قيمة ف_ر تقل عن خمسة تضم هذه القيمة إلى قيمة مجاورة لها ، ونفعل ذلك لقيم ك_ر المناظرة .

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

إن الفرض الصفرى ف_٠ هنا هو عدم وجود فرق بين التكرارات المشاهدة ك_ر والتكرارات ف_ر المتوقعة لها . أما الفرض الآخر ف_١ فهو أن هنالك تفاوتاً بين هذين النوعين من التكرارات يجعل القيمة المشاهدة للإحصاءة (١٣) كبيرة كبراً ذا دلالة . وعلى ذلك تأخذ المنطقة الحرجة الصورة الآتية :

$$\left\{ {}_{[\nu]\alpha}\chi^٢ < \chi^٢ : \chi^٢ \right\} = م$$

حيث ν 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 $ك_ر$ ، $ق_ر$ (حيث ر = ١ ، ٢ ، ... ، م) ونظراً لأن حجم العينة ن هو مقدار ثابت معروف مقدماً فإن التكرارات $ك_ر$ (أو $ق_ر$) لا تكون جميعها مستقلة إذ نستطيع إذا عرفنا (م – ١) منها أن نعرف من المتساوية (١٤) التكرار الباقي . وعلى هذا فإن عدد المقارنات المستقلة هو م – ١ . فإذا كانت القيمة المشاهدة للإحصاءة (١٣) قد حسبت اعتماداً على حجم العينة فقط فإن ν = م – ١ .

وكقاعدة عامة كل قيد خطى تتقيد 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 وبالتالى فإن كلا من هذه التقديريات يشكل قيداً على التوزيع ينقص واحداً من درجات الحرية . وكطريقة إجرائية سهلة ، نحسب عدد درجات الحرية كالآتي :

ν = عدد المقارنات المستقلة – عدد الأدلة التي قدرت من العينة (١٥)

فيما يلى 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 أصفرا ، ١٠٨ مستديراً أخضرا ، ١٠١ مجعدا أصفرا ، ٣٢ مجعداً أخضرا . 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 .

الحـــل :

لدينا ٣١٥ + ١٠٨ + ١٠١ + ٣٢ = ٥٥٦ نباتاً (حجم العينة) .

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

مستدير أصفر = ٥٥٦ × $\frac{٩}{١٦}$ = ٣١٢,٧٥

مستدير أخضر = ٥٥٦ × $\frac{٣}{١٦}$ = ١٠٤,٢٥

مجعد أصفر = ٥٥٦ × $\frac{٣}{١٦}$ = ١٠٤,٢٥

مجعد أخضر = ٥٥٦ × $\frac{١}{١٦}$ = ٣٤,٧٥

التكرارات المشاهدة $ك_ر$:	٣١٥	١٠٨	١٠١	٣٢
التكرارات المتوقعة $ق_ر$:	٣١٢,٧٥	١٠٤,٢٥	١٠٤,٢٥	٣٤,٧٥

الفرض الصفرى $ف_٠$: نظرية مندل صحيحة .

الفرض الآخر $ف_١$: نظرية مندل غير صحيحة وبالتالى تكون القيمة المشاهدة للإحصاء (١٣) كبيرة كبراً ذا دلالة .

$$\chi^{٢}_{حـ} = \frac{(٣١٥-٣١٢,٧٥)^{٢}}{٣١٢,٧٥} + \frac{(١٠٨-١٠٤,٢٥)^{٢}}{١٠٤,٢٥} + \frac{(١٠١-١٠٤,٢٥)^{٢}}{١٠٤,٢٥} + \frac{(٣٢-٣٤,٧٥)^{٢}}{٣٤,٧٥} = ٠,٤٧$$

حيث $\nu = ٤ - ١ = ٣$

من جدول χ^2 نجد أن $\chi^2_{٠,٠٥}{}_{[٣]} = ٧,٨٢$

العدد ٠,٤٧ يقل كثيراً عن ٧,٨٢ (لا يقع في منطقة الرفض) وإذن لا يمكن رفض الفرض الصفرى عند مستوى الدلالة ٠,٠٥ وبالتالى 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 ، ... ، 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

الحـــل :

مجموع الأرقام التي بالصفحة = ١٧ + ٣١ + ... + ٣٦ = ٢٥٠ رقماً

إذا كانت الأرقام العشرة تظهر باحتمالات متساوية فإن كلا منها يجب أن يظهر ٢٥٠ ÷ ١٠ = ٢٥ مرة .

أى أن التكرارات المتوقعة على أساس صحة هذا الفرض هى ٢٥ لكل من الأعداد العشرة .

$$\chi^2_{حـ} = \frac{(٢٥-١٧)^2}{٢٥} + \frac{(٢٥-٣١)^2}{٢٥} + \ldots\ldots + \frac{(٢٥-٣٦)^2}{٢٥}$$

$$= ٢٣,٣ \qquad \text{حيث } \nu = ١٠ - ١ = ٩$$

القيمة الحرجة $\chi^2_{[٩]٠,٠١}$ = ٢١,٦٦٦

بما أن ٢٣,٣ أكبر من ٢١,٦٦٦ نرفض الفرض الصفرى عن تساوى احتمالات ظهور الأرقام العشرة عند مستوى الدلالة ٠,٠١ 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 نظرى 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	$س_ر$:	٠	١	٢	٣	٤
عـــدد العـــــائلات	$ك_ر$:	١٣	٤٣	٩٢	١١٢	٦٠

الحل :

على فرض صحة الفرض الصغرى أن إنجاب الذكور وإنجاب الإناث متساويا الاحتمال يكون احتمال إنجاب الذكور ح = $\frac{١}{٢}$ ويكون توزيع عدد الذكور في العائلات

ذوات الأربعة الأطفال هو توزيع ذى الحدين دليلاه ٤ ، $\frac{1}{2}$. وقد وجدنا في المثال (٣ – ٥) أن هذا التوزيع هو :

عدد الذكور في العائلة $س_ر$: ٠ ١ ٢ ٣ ٤

عـــــدد العـــائلات $ف_ر$: ٢٠ ٨٠ ١٢٠ ٨٠ ٢٠

$$\chi^2_{ح} = \frac{(٢٠-١٣)^2}{٢٠} + \frac{(٨٠-٤٣)^2}{٨٠} + \frac{(١٢٠-٩٢)^2}{١٢٠} + \frac{(٨٠-١١٢)^2}{٨٠} + \frac{(٢٠-٦٠)^2}{٢٠}$$

$$= ١١٨,٨٩ \qquad \text{حيث } \nu = ٥ - ١ = ٤$$

$$\chi^2_{[٤]٠,٠١} = ١٣,٢٧٧$$

بما أن ١١٨,٨٩ > ١٣,٢٧٧ 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 معتدلا للتوزيع المعطى ووجدنا ما يلى :

التكرارات المشاهدة $ك_ر$ ٦ ٦ ١١ ١٤ ١٦ ١٥ ٨

التكرارات المتوقعة $ق_ر$ ٦,٨١ ٦,٥٤ ٩,٦١ ١٢,٩٨ ١٣,٦٦ ١٥,٥٧ ١٢,٤٧

التكرارات المشاهدة $ك_ر$: ١٠ ٨ ٦

التكرارات المتوقعة $ق_ر$: ٩,٤٤ ٦,٣٧ ٥,٠٠

obeikandl.com

$$\chi^2_{حـ} = \frac{(٦-٦,٨١)^2}{٦,٨١} + \frac{(٦-٦,٥٤)^2}{٦,٥٤} + \ldots\ldots + \frac{(٨-٦,٣٧)^2}{٦,٣٧} + \frac{(٦-٦,٥٥)^2}{٦,٥٥} = ٢,٩٤٢٧$$

بما أننا كنا قد قدرنا من العينة اثنين من أدلة المجتمع هما الوسط الحسابي والانحراف المعيارى فإن عدد درجات الحرية ν = (١٠ − ١) − ٢ = ٧ .

$$\chi^2_{٠,٠٥\,[٧]} = ١٤,٠٦٧$$

بما أن ٢,٩٤٢٧ < ١٤,٠٦٧ لا نستطيع رفض الفرض الصفرى عن اعتدالية المجتمع عند مستوى الدلالة ٠,٠٥

ملاحظة (١) :

إن عيب اختبار χ^2 لحسن المطابقة أنه لا يهتم بإشارات الفروق بين التكرارات المشاهدة $ك_ر$ والتكرارات النظرية $ه_ر$ (لأنه يأخذ مربعات هذه الفروق) ففي توفيق توزيع معتدل مثلا قد يحدث أن تكون قيم $ه_ر$ أصغر من قيم $ك_ر$ في الجزء الأوسط من التوزيع وتكون أكبر منها في الطرفين وهذا يعني أن التوزيع المشاهد أكثر انبساطاً من التوزيع المعتدل أى يأخذ شكلا يختلف عن شكل التوزيع المعتدل . ومع هذا قد تكون قيمة $\chi^2_{حـ}$ ليست بذات دلالة وتدعونا إلى اعتبار أن المجتمع معتدل . ولهذا لا يكون الاختبار مناسباً إلا إذا لم يكن هناك نمط معين للفروق بين $ه_ر$ ، $ك_ر$.

(٦ − ٧ − ٣) اختبار استقلال خاصتين :

TEST OF INDEPENDENCE

في بعض الدراسات نتناول مجتمعاً صنفت عناصره بحسب خاصة ما إلى م من الأقسام المنفصلة $ا_١$ ، $ا_٢$ ، ... ، $ا_م$ وصنفت من ناحية ثانية بحسب خاصة أخرى إلى هـ من الأقسام المنفصلة $ب_١$ ، $ب_٢$ ، ... ، $ب_هـ$ ويكون المراد اختبار ما إذا كانت هاتان الخاصتان مستقلتين ، بمعني أن يكون توزيع إحداهما 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 بجدول الاقتران contingency table وهو عبارة عن مصفوفة ذات م صفا تمثل أقسام الخاصة الأولى ، هـ عموداً تمثل أقسام الخاصة الثانية ويوصف الجدول حينئذ بأنه جدول اقتران م × هـ . نضع الفرض الصفرى أن الخاصتين مستقلتان ، وعلى أساس صحة هذا الفرض نحسب التكرارات النظرية المناظرة للتكرارات المشاهدة ونضعها في جدول اقتران م × هـ أيضا . بمقارنة التكرارات المشاهدة بالتكرارات النظرية نستطيع الحكم على استقلال الخاصتين بواسطة الإحصاءة 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 ٢ × ٢ الآتي :

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

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 الكلى للمتوفين والعدد الكلى للأحياء وهى 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 فمثلا :

التكرار النظرى لعدد المتوفين من المدخنين $= ٤٠٢ \times \frac{١٧١}{١٤٦٩} = ٤٦,٨$ رجلا

التكرار النظرى لعدد الأحياء من المدخنين $= ٤٠٢ \times \frac{١٢٩٨}{١٤٦٩} = ٣٥٥,٢$رجلا

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

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

$$\chi^٢ = \frac{(٤٦,٨-٥٤)^٢}{٤٦,٨} + \ldots + \ldots + \frac{(٩٤٢,٨-٩٥٠)^٢}{٩٤٢,٨} = ١,٧٣$$

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 .

إذن ν = ١ − (٤−٢) = ١

$\chi^2_{[1]0,05}$ = ٣,٨٤١

بما أن ١,٧٣ < ٣,٨٤١ لا نستطيع رفض الفرض الصفرى عند المستوى ٠,٠٥ ويمكننا القول بأنه في حدود مجموعة العمر التي درست ، عادة التدخين والممات مستقلتان ، أى أن التدخين لا يؤثر في الوفاة في حدود هذه التجربة .

مثال (٦ − ١١) :

جدول الاقتران ٢ × ٣ الآتي يحمل التكرارات المشاهدة في عينة عشوائية من ٩٩ طفلا حديث الولادة من حيث طول الطفل وطول محيط رأسه بالسنتيمترات ساعة الولادة . ابحث استقلال طول المولود وطول محيط رأسه .

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

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 ينبغى أن نقسم كلا من الأعداد ٤٠ ، ٥٠ ، ٩ بالنسبة ٧٨ : ٢١ لنحصل على جدول التكرارات النظرية الآتي :

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

$$\chi^2_{ى} = \frac{(٣١,٥-٤٠)^2}{٣١,٥} + \ldots + \frac{(١,٩-٧)^2}{١,٩} = ٢٩,٣٧٧$$

$$\nu = (٣-٦) - ١ = ٢$$

$$\chi^2_{[٢]\ ,٠٥} = ٥,٩٩١$$

نرفض الفرض الصفرى 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 فى خلايا القطر الرئيسى أى فى الخلايا العلوية اليمنى وفى الوسط وفى الخلايا السفلية اليسرى كما فى هذا المثال .

(ب) وإذا كانت هناك علاقة سالبه تظهر تكرارات كبيرة نسبيا فى الخلايا السفلية اليمنى وفى الوسط وفى الخلايا العلوية اليسرى .

(جـ) 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 التى تظهر فى المجاميع الهامشية .

ملاحظة (٢) :

هناك قاعدة سهلة لإيجاد عدد درجات الحرية في أى جدول اقتران م × هـ حيث م > ١ ، هـ > ١ وهى :

$$\nu = (م - ١)(هـ - ١) \qquad (١٦)$$

وتفسير هذه القاعدة أنه مادامت المجاميع الهامشية للأعمدة ثابتة فإننا في ملأ خانات جدول الاقتران يمكننا استنتاج واحد من الأعداد التي بأى عمود من الـ هـ − ١ الأعداد الأخرى التي بهذا العمود ، وبالمثل يمكن استنتاج واحد من الأعداد التي بأى صف من الـ م − ١ الأعداد الأخرى التي بهذا الصف . وبهذا نكون أحراراً فقط في ملأ (م − ١) (هـ − ١) من خلايا الجدول .

ملاحظة (٣) :

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

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

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

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

(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 $\chi^٢$ ، إذ يمكن إثبات أنه على أساس عينة عشوائية حجمها ن وتباينها ع^٢ مأخوذة من مجتمع معتدل تباينه $\sigma^٢$ فإن المتباينة :

$$\frac{(ن - ١)ع^٢}{\chi^٢_{(١-\frac{\alpha}{٢})}\ [ن-١]} < \sigma^٢ < \frac{(ن - ١)\ ع^٢}{\chi^٢_{\frac{\alpha}{٢}}\ [ن-١]} \qquad (١٧)$$

تشكل فترة ثقة بدرجة (١ – α) لتباين المجتمع . وليس من الضرورى 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 :

لدينا $\alpha = ٠,٠٥$ $\quad \frac{\alpha}{٢} = ٠,٠٢٥$ $\quad ١ - \frac{\alpha}{٢} = ٠,٩٧٥$ $\quad ن = ١٢$

من جدول $\chi^٢$ نجد أن $\chi^٢_{\frac{\alpha}{٢}\,[ن-١]} = \chi^٢_{[١١]٠,٠٢٥} = ٢١,٩$

، $\chi^٢_{(١-\frac{\alpha}{٢})\,[ن-١]} = \chi^٢_{[١١]٠,٩٧٥} = ٣,٨٢$

بالتعويض فى (١٧) نجد أن :

الحد الأدنى للفترة $= \frac{٩,٧٣ \times ١١}{٢١,٩} = ٤,٨٨$

الحد الأعلى للفترة $= \frac{٩,٧٣ \times ١١}{٣,٨٢} = ٢٨,٠١٨$

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

تمارين (٦ - ٢)

(فى 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 . اختبر 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 مصايد فيران في أماكن مختلفة من غابة ما . في فترة ثلاثة أشهر سجلت أعداد الفيران المصيدة كالآتى :

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

obeikandl.com

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	ابريل	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 مثلا) 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 :

	يحتوى على مقومات الحياة	لا يحتوى على مقومات الحياة
الجوال أ :	٦٤	١٦
الجوال ب :	٣٤	٤٦

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) بجدول 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_1 وكان لمتغير آخر **مستقل** عن المتغير الأول توزيع χ^2 بدرجات حرية ν_2 فإن المتغير الذى ينشأ من ضم هذين المتغيرين معا يكون له توزيع χ^2 بدرجات حرية $\nu_1 + \nu_2$ » 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 ٢ × ٢ ووجد أن χ^2 = ٤,١ 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 $\chi^2_{٠,٠٥\,[١]}$ = ٣,٨٤١ وهذا يعنى أن المصل كان له تأثير فى الحماية من المرض .

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

المجموع = ٤,١ + ٣,٩ + ٤,٢ = ١٢,٢ بدرجات حرية عددها ثلاث ونظرا لأن القيمة الحرجة $\chi^2_{٠,٠٥\,[٣]}$ = ٧,٨١٥ 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 تباييني مجتمعين معتدلين σ^2_1 ، σ^2_2 عن طريق تبايني عينتين عشوائيتين مستقلتين $ع^2_1$ ، $ع^2_2$ مأخوذتين منهما . وقد رأينا أن اختبار ت عن دلالة الفرق بين متوسطى عينتين بالصورة المبينة بالبند (٦ ــ ٦ ــ ٣) يستلزم التأكد من تساوى تبايني المجتمعين ، ولذا ينبغى 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 .

ويبني هذا الاختبار على الإحصاءة :

$$صـ_{١٠} = \frac{ع^2_1}{ع^2_2} \qquad (١٨)$$

التي تسمى بنسبة التباين variance ratio والتي يتطابق توزيعها مع توزيع المتغير ف السابق دراسته بدرجتي حرية $\nu_1 = ن_1 - ١$ ، $\nu_2 = ن_2 - ١$ تحت الافتراضات الآتية :

(أ) أن تكون كل من العينتين عشوائية بسيطة .

(ب) أن تكون العينتان مستقلتين .

(حـ) أن تكون العينتان مأخوذتين من مجتمع معتدل واحد أو من مجتمعين معتدلين لهما نفس التباين .

ملاحظة :

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

أكبر من الواحد أى بحيث تكون ع$^{٢}_{١}$ > ع$^{٢}_{٢}$ مع ملاحظة أن $\nu_{١}$ تحدد العمود ، $\nu_{٢}$ تحدد الصف .

مثال (٦ - ١٤) :

أخذت عينتان عشوائيتان حجماهما 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 :

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

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

الحل :

ف$_{حـ}$ = $\frac{٢٥,٧}{٧,٠٦}$ = ٣,٦٤ بدرجتي حرية ٤ ، ٥

(أولا) الفرض الصفرى ف$_{٠}$: $\sigma^{٢}_{١} = \sigma^{٢}_{٢}$

الفرض البديل ف$_{١}$ $\sigma^{٢}_{١} > \sigma^{٢}_{٢}$ (اختبار ذو جانب واحد)

من الجدول (٩) : ف$_{٠,٠١ [٥،٤]}$ = ١١,٤

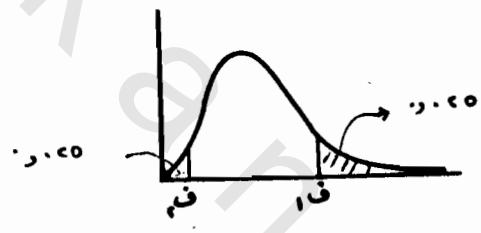
بما أن ٣,٦٤ < ١١,٤ لا نستطيع رفض ف عند المستوى ٠,٠١

(ثانياً) الفرض الصفرى ف : $\sigma^{٢}_{١} = \sigma^{٢}_{٢}$

الفرض الآخر ف : $\sigma^{٢}_{١} \neq \sigma^{٢}_{٢}$ (اختبار ذو جانبين مساحة كل منهما ٠,٠٢٥

نريد إيجاد ف$_{١}$ ، ف$_{٢}$ حيث ل (ف > ف$_{١}$) = ٠,٠٢٥
، ل (ف > ف$_{٢}$) = ٠,٩٧٥ مع ملاحظة أن الاحتمال ٠,٩٧٥ لا يوجد بالجدول .

من الجدول $ف_{٠,٠٢٥ [٥ ، ٤]}$ = ٧,٣٩
إذن القيمة الحرجة اليمني ف$_{١}$ = ٧,٣٩
المتساوية ل (ف > ف$_{٢}$) = ٠,٩٧٥ تكافىء
المتساوية ل $\left(\frac{١}{ف} > \frac{١}{ف_{٢}} \right)$ = ٠,٠٢٥



الشكل (٦-٤)

من الجدول $ف_{٠,٠٢٥ [٥ ، ٤]}$ = ٩,٣٦ راجع ملاحظة البند (٥-٣)

$$\therefore \frac{١}{ف_{٢}} = ٩,٣٦$$

∴ القيمة الحرجة اليسرى ف$_{٢}$ = ١ ÷ ٩,٣٦ = ٠,١٠٧
بما أن ٧٣,٩ > ٣,٦٤ > ٠,١٠٧ لا نستطيع رفض ف عند المستوى ٠,٠٥
أى أننا نقبل أن $\sigma_{١}^{٢} = \sigma_{٢}^{٢}$ عند هذا المستوى .

obeikandl.com

(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 س الذى يعبر عن عدد مرات وقوع الحدث أ في العينات ذوات الحجم ن توزيع ذى الحدين دليلاه ن ، ح وسطه الحسابي ن ح وتباينه ن ح ك حيث ك = ١ – ح . ونعلم من البند (٦ – ٤) أنه إذا كان المتغير العشوائي ر = $\frac{س}{ن}$ يعبر عن النسبة المشاهدة في العينة فإن توزيع الإحصاءة :

$$ص = \frac{ر - ح}{\sqrt{\frac{ح\ ك}{ن}}}$$

يقترب من التوزيع المعتدل المعيارى بشرط أن تكون ن كبيرة وألا تكون ح أو ك قريبة من الصفر .

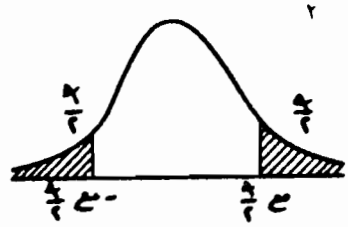
في هذه الحالة يمكن إثبات أن الفترة :

$$\left(ر - \sqrt{\frac{ر\ (١ - ر)}{ن}} \text{ مع } \frac{\alpha}{٢} \text{ ، } ر + \sqrt{\frac{ر(١-ر)}{ن}} \text{ مع } \frac{\alpha}{٢} \right) \qquad (١٩)$$

هى فترة ثقة بدرجة (١ − α) للنسبة ح .

حيث مع $\frac{\alpha}{٢}$ هى القيمة الحرجة في التوزيع المعتدل المعيارى التي تحقق المعادلة

$$ل\ (\ -مع_{\frac{\alpha}{٢}} < ع < مع_{\frac{\alpha}{٢}}\ ) = ١ - \alpha \qquad (٢٠)$$



الشكل (٦ − ٥)

فمثلا حين α = ٠,٠٥ فإن مع $\frac{\alpha}{٢}$ = ١,٩٦ راجع المثال (٤ − ٣).

وحين α = ٠,٠١ فإن مع $\frac{\alpha}{٢}$ = ٢,٥٨ راجع المثال (٤ ــ 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 :

$$ن = \frac{٢٣}{٤٠٠} = ٠,٠٥٧٥ \qquad \text{نسبة الإصابة في العينة}$$

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

$$\text{تقدير الخطأ المعيارى} = \sqrt{\frac{ن(١-ن)}{ن}} = \sqrt{\frac{٠,٠٥٧٥ \times ٠,٩٤٢٥}{٤٠٠}} = ٠,٠١١٦$$

من (١٩) وبأخذ α = ٠,٠٥ نجد أن

الحد الأدني للفترة = ٠,٠٥٧٥ - ٠,٠١١٦ × ١,٩٦ = ٠,٠٣٥

الحد الأعلى للفترة = ٠,٠٥٧٥ + ٠,٠١١٦ × ١,٩٦ = ٠,٠٨٠٢

إذن الفترة (٠,٠٣٥ ، ٠,٠٨٠) هى فترة ثقة بدرجة ٩٥٪ لنسبة المصابين بالبول السكرى في المجتمع .

(٦ - ٩ - ١) اختبار دلالة الفرق بين نسبتي 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 نسبتي وقوع الحدث فيهما ر$_١$ ، ر$_٢$ على الترتيب . يمكن إثبات أن توزيع المعاينة للإحصاءة .

$$صـ = \frac{(ر_١ - ر_٢) - (ح_١ - ح_٢)}{\sqrt{\frac{ح_١ ك_١}{ن_١} + \frac{ح_٢ ك_٢}{ن_٢}}} \qquad (٢١)$$

يقترب من التوزيع المعتدل المعيارى كلما كبرت كل من ن$_١$ ، ن$_٢$ ولم تكن أى من ح$_١$ ، ك$_١$ ، ح$_٢$ ، ك$_٢$ قريبة من الصفر .

وإذا فرضنا أن ح$_١$ = ح$_٢$ = ح ، ك = ١ - ح فإن هذه الإحصاءة تكتب .

$$صـ = \frac{ر_١ - ر_٢}{\sqrt{ح\ ك\ (\frac{١}{ن_١} + \frac{١}{ن_٢})}} \qquad (٢١)$$

وفى هذه الحالة تقدر ح من العينتين كالآتي :

$$ح = \frac{ن_١ ر_١ + ن_٢ ر_٢}{ن_١ + ن_٢}$$

وعلى ذلك يمكن استخدام الإحصاءة (٢١) للكشف عما إذا كان الفرق المشاهد بين $س_١$ ، $س_٢$ في العينتين هو فرق صغير لا دلالة له أم فرق يدل على وجود فرق حقيقي بين النسبتين $ح_١$ ، $ح_٢$ في المجتمعين .

مثال (٦ - ١٦) :

لمعرفة تأثير طُعم 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 و٥٠٠ حالة في المجموعة الثانية . اختبر ما إذا كان لهذا الطعم أثر في الوقاية من هذا المرض مستخدماً مستوى الدلالة ٠,٠١

الحل :

الفرض الصفرى ف$_٠$: $ح_١ = ح_٢$ (لا يوجد تأثير للطعم المعطى)

الفرض الآخر ف$_١$: $ح_١ < ح_٢$ (اختبار ذو جانب واحد) .

نسبة الإصابة في المجموعة الأولى $س_١ = \frac{١٠٠}{١٠٠٠} = ٠,١٠$

نسبة الإصابة في المجموعة الثانية $س_٢ = \frac{٥٠٠}{١٥٠٠} = ٠,٣٣$

من (٢١) : تقدير ح $= \frac{٠,١ \times ١٠٠٠ + ٠,٣٣ \times ١٥٠٠}{١٥٠٠ + ١٠٠٠} = ٠,٢٣٨$

تقدير الخطأ المعيارى $= \sqrt{ح ك \left(\frac{١}{ن_١} + \frac{١}{ن_٢}\right)} = \sqrt{٠,٢٣٨ \times ٠,٧٦٢ \left(\frac{١}{١٠٠٠} + \frac{١}{١٥٠٠}\right)}$

$= ٠,٠١٧$

من (٢١) : صـ$_{س}$ $= \frac{٠,١٠ - ٠,٣٣}{٠,٠١٧} = ١٣,٥٢$

بما أن −١٣,٥٢ < −٢,٣٣ نرفض الفرض الصفرى عند المستوى ٠,٠١ 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 :

$$\left\{ (س_١ - س_٢) - خ.م. \text{ مع } \frac{\alpha}{٢} \; ، (س_١ - س_٢) + خ.م. \text{ مع } \frac{\alpha}{٢} \right\} \qquad (٢٢)$$

حيث $خ.م. = \sqrt{ح ك \left(\frac{١}{ن_١} + \frac{١}{ن_٢} \right)}$

هى فترة ثقة بدرجة $(١ - \alpha)$ للفرق $ح_١ - ح_٢$ بين نسبتي المجتمعين . ففي المثال (٦ − ١٥) الأخير نجد أن :

الحد الأدني للفترة = (٠,٣٣ − ٠,١) − ٠,٠١٧ × ٢,٥٨ = ٠,١٨٦

الحد الأعلى للفترة = (٠,٣٣ − ٠,١) + ٠,٠١٧ × ٢,٥٨ = ٠,٢٧٤

أى أن الفترة (٠,١٨٦ ، ٠,٢٧٤) هى فترة ثقة بدرجة ٩٩٪ للفرق بين النسبتين في المجتمعين .

تمارين (٦ − ٣)

(١) في عينة عشوائية من ٤٠٠ شخص حقنوا بمصل ما تأثر ١٣٦ تأثيراً ضاراً . أوجد فترة ثقة بدرجة ٩٥٪ 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 . أنشيء فترة ثقة بدرجة ٩٥٪ لاحتمال أنه إذا صيدت سمكة من هذه البحيرة تكون غير صالحة للأكل .

obeikandi.com

(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	٦	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 المرضي الذين يتلقون أقراص الحساسية لا تختلف عنها في المرضي 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 الصفرى أنه لا يوجد فرق بين المجتمعين ضد الفرض الآخر أن تعب منتصف الصباح أكثر تفشياً 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 $\bar{س}$ 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 $\bar{س}$ ؟ أى ما هو الحد الأعلى الذى يمكن التجاوز عنه لانحراف $\bar{س}$ عن القيمة الحقيقيةμ » ؟ 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 الحسابي μ وانحرافه المعيارى σ فإن الإحصاءة $\bar{س}$ للعينات ذوات الحجم ن يكون لها توزيع معتدل وسطه الحسابى μ وانحرافه المعيارى $\frac{\sigma}{\sqrt{ن}}$ وتتحقق هذه النظرية أيضا (بالتقريب) حين لا يكون المجتمع معتدلا بشرط أن يكون حجم العينة كبيرا (ن $\geq$ ٣٠) .

وإذا كانت $\bar{س}$ هى الوسط الحسابي لعينة عشوائية ما فإن الفترة :

$$(\bar{س} - \frac{\sigma}{\sqrt{ن}} \text{ مع } \frac{\alpha}{2} \text{ ، } \bar{س} + \frac{\sigma}{\sqrt{ن}} \text{ مع } \frac{\alpha}{2})$$

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

$$ل(\text{مع}_{\frac{\alpha}{2}} > ع > \text{مع}_{\frac{\alpha}{2}}) = 1 - \alpha$$

وبذلك تكون أكبر قيمة للانحراف $|\bar{س} - \mu|$ هى $\frac{\sigma}{\sqrt{ن}}$ 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 خ وأردنا أن نكون على ثقة بدرجة $(1 - \alpha)$ بألا يتعدى الخطأ الذى تقع فيه القيمة خ فإن حجم العينة المطلوب يجب أن يحقق المعادلة $\frac{\sigma}{\sqrt{ن}}$ مع $\frac{\alpha}{2}$ = خ . وبحل هذه المعادلة في ن نجد أن :

$$ن = (\frac{\sigma \text{ مع } \frac{\alpha}{2}}{خ})^2 \qquad (٢٣)$$

وهذه هى القيمة المطلوبة لحجم العينة ن الذى يكفي لتحقيق الغرض المطلوب .

كما يمكن أن نأخذ العدد ن حيث

$$ن = \frac{\sigma^2}{خ^2 \alpha} \qquad (٢٤)$$

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

ويلاحظ أن إيجاد قيمة ن من أى من المعادلتين (٢٣) أو (٢٤) يتطلب معرفة ولو تقريبية بالانحراف المعيارى 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 المعيارى هما ٢,٦ و٠,٣ على الترتيب . ما حجم العينة اللازم لإعطائنا ثقة بدرجة ٩٥٪ بألا يزيد الخطأ في تقدير الوسط الحسابي μ للمجتمع عن ٠,٠٦ ؟

الحل :

لدينا خ = ٠,٠٦ ، $1 - \alpha$ = ٠,٩٥ ومنها α = ٠,٠٥ ،

$م_{\frac{\alpha}{2}}$ = ١,٩٦

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

$$ن = \left(\frac{١,٩٦ \times ٠,٣}{٠,٠٦}\right)^{٢} = ٩٦,٠٤$$

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

نلاحظ أنه بأخذ ن= ٩٧ فإن $\frac{ع}{\sqrt{ن}} \times ١,٩٦ = \frac{١,٩٦ \times ٠,٣}{\sqrt{٩٧}} = ٠,٠٥٩٧$

وهذا أصغر من ٠,٠٦ .

(ثانيا) عند تقدير الفرق بين متوسطى مجتمعين معتدلين :

بنفس الطريقة يمكن إيجاد حد أعلى لحجم العينة اللازم لتقدير الفرق $\mu_١ - \mu_٢$ بين متوسطى مجتمعين معتدلين بحيث نكون على ثقة بدرجة $١ - \alpha$ بألا يزيد المقدار $\bar{س}_١ - \bar{س}_٢$ المشاهد فى عينتين عن المقدار $\mu_١ - \mu_٢$ للمجتمعين ، عن قيمة معينة خ مفروضة مسبقا ، وما علينا إلا استخدام الخطأ المعيارى لتوزيع المعاينة للإحصاءة $\bar{س}_١ - \bar{س}_٢$ وهو $\sqrt{\frac{\sigma_١^٢}{ن_١} + \frac{\sigma_٢^٢}{ن_٢}}$ بدلا من الخطأ المعيارى $\frac{\sigma}{\sqrt{ن}}$ لتوزيع المعاينة للإحصاءة $\bar{س}$. وإذا كانت العينتان متساويتى الحجم : $ن_١ = ن_٢ = ن$ وكان المجتمعان متساويين فى التباين : $\sigma_١^٢ = \sigma_٢^٢ = \sigma^٢$ فإن ذلك الخطأ المعيارى يأخذ الصورة $\frac{\sigma \sqrt{٢}}{\sqrt{ن}}$ ويكون الحد الأعلى المطلوب لحجم كل من العينتين هو

$$ن = \left(\frac{\text{مع}_{\frac{\alpha}{٢}} \sigma \sqrt{٢}}{\text{خ}} \right)^٢ \qquad (٢٥)$$

ويلاحظ أن هذه الصيغة ضعف الصيغة (٢٣) .

مثال (٦ - ١٨) :

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

٦٠ قطة قسمت عشوائيا إلى مجموعتين متكافئين فى 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 فى وزن المجموعة التجريبية ١٠ وحدات ومتوسط النقص فى وزن المجموعة الضابطة ١٣,٣ وحدة . على فرض أن تباين كل من المجموعتين ٤٠ أوجد حدا أعلى لحجم كل من العينتين الذى يعطينا ثقة بدرجة ٩٥٪ بألا يزيد الفرق المشاهد فى متوسطى العينتين عن الفرق الحقيقى بين متوسطى المجموعتين عن ٣ وحدات .

الحل :

لدينا $\sigma^2 = ٤٠$ ، خ $= \pm ٣$ ، $\alpha = ٠,٠٥$ وإذن $مع_{\frac{\alpha}{٢}} = ١,٩٦$

من الصيغة (٢٥) : $ن = \frac{٢ \times ٤٠}{٩} (١,٩٦)^2 = ٣٤,١٤٨ \sim ٣٥$

(ثالثا) عند تقدير نسبة وقوع حدث في مجتمع :

نفرض أن نسبة وقوع حدث معين في مجتمع هو مقدار ثابت مجهول ح ونفرض أننا نرغب في معرفة حجم العينة المناسب لتقدير هذه النسبة عن طريق النسبة حَ التي تظهر في عينة عشوائية ، بحيث نكون على ثقة بدرجة ١ − α ألا يزيد الخطأ الناشيء عن هذا التقدير عن مقدار معين خ .

من البند (٦ − ٤) نعلم أنه إذا كان حجم العينة كبيراً (ن ≥ ٣٠) فإن توزيع المعاينة للنسبة حَ لوقوع هذا الحدث في العينات ذوات الحجم ن يقترب من توزيع معتدل وسطه الحسابي ح وانحرافه المعيارى $\sqrt{\frac{ح ك}{ن}}$ حيث ك = ١ − ح . في هذه الحالة يكون حد الخطأ أى أكبر قيمة للمقدار | حَ − ح | هو $\sqrt{\frac{ح ك}{ن}}$. $مع_{\frac{\alpha}{٢}}$ ،

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

بحل المعادلة $\sqrt{\frac{ح ك}{ن}}$ مع$_{\frac{\alpha}{٢}}$ = خ نحصل على

$$ن = ح ك \left(\frac{مع_{\frac{\alpha}{٢}}}{ح}\right)^{٢} \qquad (٢٦)$$

وهذه هى القيمة المطلوبة لحجم العينة . غير أن هذا الحل غير قابل للاستخدام لأنه يشتمل على البارامتر ح وهو الذى نبحث عن تقديره . ولكن نظراً لأن أكبر قيمة لحاصل الضرب ح ك هى $\frac{١}{٤}$ أى أن ح ك $\leqslant \frac{١}{٤}$ دائماً ،

[لأن ٤ ح ك − ١ = ٤ ح (١ − ح) − ١ = − (٤ ح^{٢} − ٤ ح + ١)

= − (٢ ح − ١)^{٢} $\leqslant$ ٠]

فإن الصيغة (٢٦) يمكن أن تكتب كالآتي :

$$ن = \frac{١}{٤}\left(\frac{مع_{\frac{\alpha}{٢}}}{خ}\right)^{٢} \qquad (٢٧)$$

واختيار الحد الأقصي لحجم العينة من هذه المعادلة يؤدى الغرض المنشود بغير أى معرفة مسبقة لقيمة البارامتر ح .

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

$$ن = ح ك \left(\frac{مع_{\frac{\alpha}{٢}}}{خ}\right)^{٢} \qquad (٢٨)$$

وهذه القيمة تقل عن تلك التي تعطيها الصيغة (٢٧) لأنها مبنية على معلومات عن القيمة المحتملة للدليل ح .

مثال (٦ – ١٩) :

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

(أ) في حالة عدم وجود أى معرفة لقيمة ح

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

الحل :

لدينا خ = ± ٠,٠٥ ، ١ – α = ٠,٩٨ إذن α = ٠,٠٢ ، $\frac{\alpha}{٢}$ = ٠,٠١

القيمة هى قيمة المتغير المعتدل المعيارى أ التى تحقق المعادلة

ل (أ > ع > – أ) = ٠,٩٨

ل (أ > ع > ٠) = ٠,٩٨ ÷ ٢ = ٠,٤٩

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

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

إذن الحجم المطلوب للعينة هو ٥٤٣ .

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

إذن الحجم المطلوب للعينة هو ٤٥٦ .

تمارين (٦ – ٤)

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

obeikandl.com

تكاد تكون ثابتة عند $\sigma = ٤$. ما عدد عينات الماء التي ينبغى للباحث تحليلها لكى 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 المعيارى لمدة الجفاف هما ٣٢ دقيقة و٤ 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) لكى 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 المعيارى 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 ؟

(٦ – ١١) مراقبة الانتاج QUALITY CONTROL

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

ومن المعروف أن الوحدات المنتجة فى أى مصنع لا يمكن أن تكون جميعها متشابهة تماما مهما تقدمت التكنولوجية الصناعية بل توجد دائما انحرافات صغيرة عن المواصفات الموضوعة للسلعة تنشأ عن عدد كبير من العوامل الصغيرة التى لا يمكن التحكم فيها وتعتبر عوامل عشوائية ، ولا مفر للمصنع من أن يسمح بالتجاوز عن هذه الانحرافات طالما كانت فى حدود معقولة يقبلها المنتج والمستهلك ، أما إذا خرجت الانحرافات عن هذه الحدود فإن المصنع يشتبه فى وجود خلل ما إما فى أجزاء آلة المصنع أو فى سلوك العمال أو فى 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 الصفرى $\mu = ٦$ سم ضد الفرض $\mu \neq ٦$ سم ويجرى هذا الاختبار فى فترات منتظمة من الزمن ، مثلا كل نصف ساعة أو كل ساعة ، وفى كل مرة يأخذ عينة عادة من ٣ إلى ١٠ وحدات ويقيس متوسط الطول $\bar{س}$ فيها و يطبق الاختبار . فإذا أدى الاختبار إلى قبول الفرض الصفرى فإن هذا يعنى أن الإنتاج يسير سيرا طبيعيا ، أما إذا أدى إلى رفض الفرض الصفرى فإن هذا يعنى أن انحرافات أطوال الوحدات المنتجة عن الطول المطلوب هى 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 فى الوحدات المنتجة **معتدل** متوسطه μ وانحرافه المعيارى σ فإن توزيع المعاينة

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

$$(\bar{س} - ٢,٥٨ \times \frac{\sigma}{\sqrt{ن}} \ ، \ \bar{س} + ٢,٥٨ \times \frac{\sigma}{\sqrt{ن}})$$

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

$$\bar{س} + ٢,٥٨ \times \frac{\sigma}{\sqrt{ن}} > \mu > \bar{س} - ٢,٥٨ \times \frac{\sigma}{\sqrt{ن}}$$

ومن هذه المتباينة المزدوجة ينتج أن

$$\mu + ٢,٥٨ \times \frac{\sigma}{\sqrt{ن}} > \bar{س} > \mu - ٢,٥٨ \times \frac{\sigma}{\sqrt{ن}} \qquad (٢٩)$$

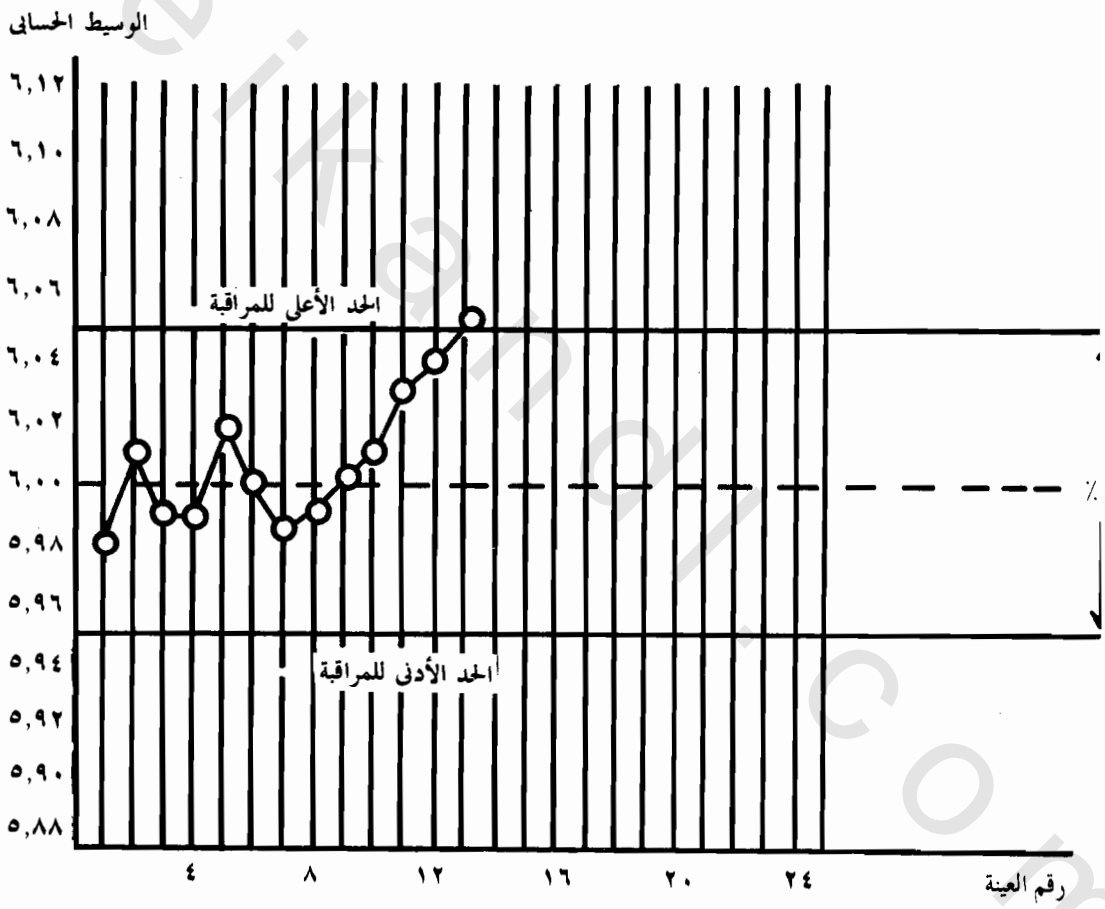
ويسمى الطرف الأيسر من هذه المتباينة **بحد المراقبة الأدنى** ، ويسمي الطرف الأيمن منها **بحد المراقبة الأعلى** lower and upper control limits وعند المراقبة إذا أخذت عينة ووجد أن متوسطها $\bar{س}$ يقع بين هذين الحدين يقبل الفرض الصفرى ويستمر الإنتاج دون تعديل ، أما إذا ساوت $\bar{س}$ أحد الحدين أو تعدته فيرفض الفرض الصفرى 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 $\bar{س}$ لكل عينة مختبرة بالترتيب فى خريطة تسمى **بخريطة المراقبة** control chart . فإذا فرضنا أن μ = ٦ سم ،

σ = ٠,٠٦ سم ، ن = ١٠ فان حدى المراقبة يكونان ٦ ± $\frac{٠,٠٦}{\sqrt{١٠}}$ × ٢,٥٨

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



الشكل (٦-٦) : خريطة المراقبة للوسط الحسابى

وتبين هذه الخريطة الخطين الممثلين للحدين 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