كلية ادارة الاعمال

احصاء 2 Statistics

محاضرة رقم 8

الفصل الثاني للعام الدراسي 2023-2024

الاستاذ الدكتور محمود محمد ديب طيوب

المعاينة

ECHANTILLONNAGE- SAMPLING

1- مقدمة :

من أهم المشكلات التي تواجه الباحث عند إجراء أي بحث هو تحديد نطاق العمل وعلى الرغم من أن لكل بحث ظروفه التي تحدد نطاقه ومستلزماته إلا أنه من الواضح أنه كلما زاد عدد المفردات التي يشملها البحث كلما أصبحت النتائج التي يتوصل إليها الباحث مستندة إلى أساس قوي. إلا أن ذلك أحياناً يتطلب جهوداً وإمكانيات أكبر إضافةً لإجراء العديد من التجارب. إن هذه المعضلات تجعل الأنظار تتوجه نحو إجراء البحوث والتجارب على عدد محدود من المفردات

أن الدراسة الميدانية يمكن أن تجري بالعينة التي تسحب من المجتمع الإحصائي بحيث تكون ممثلة تمثيلاً صادقاً لصفات هذا المجتمع. إن مثل هذه الدراسات لا تجرى من أجل التعرف على صفات وحدات العينة فقط وإنما للاستدلال على المقاييس والمؤشرات المختلفة للمجتمع الذي سحبت منه هذه العينة.

المشكلة التي يواجهها الباحث عند الأخذ بأسلوب العينات هي تحديد حجم العينة المرغوب سحبها وهذا يتعلق بعدة عوامل أهمها: الإمكانيات المادية المخصصة لذلك ـ مقدار الثقة التي يرغب الباحث الحصول عليها وفي الحقيقة فإن هناك علاقة عكسية بين الأخطاء الاحتمالية وبين حجم العينة، بمعنى آخر، كلما ازداد حجم العينة نقصت الأخطاء الاحتمالية والعكس بالعكس. إلا أنه كلما ازداد حجم العينة المطلوبة كلما ازدادت الإمكانيات والمستلزمات الضرورية لإجراء البحث. وبالتالي ازداد مقدار الثقة التي يمكن وضعها في نتائجها.

إضافةً لذلك فإن مخطط المعاينة المرغوب يتعلق أيضاً بالهدف من الدراسة ـ الإمكانيات المادية ـ طرق التحليل الرياضي الممكنة ـ التقنيات المتوفرة ـ المقاييس الزمانية والمكانية التي تؤخذ على اساسها (المفردات) المشاهدات.

نظرية العينات بالتعريف: مجموعة الطرق الرياضية والتنظيمية التي تساعد على إجراء البحوث الإحصائية غير الشاملة وذلك بهدف إيجاد الخصائص العامة للظاهرة المدروسة عن طريق تصميم النتائج المستخلصة من هذه البحوث عن الموضوع ككل.

2- ـ فوائد نظرية العينات:

لنظرية العينات العديد من الفوائد التي تجعلها في بعض الحالات أكثر تطبيقاً من طريقة الحصر الشامل:

- 1 ـ تختصر كثيراً من الوقت والجهد والتكاليف لعمليات البحث.
- 2 ـ تمكن من الحصول على معلومات إحصائية مميزة لوحدات الموضوع المدروس.
- 3 ـ تفيد في تصحيح معلومات البحث الشامل عندما تكون نسبة الأخطاء فيها كبيرة ما تفيد في معرفة الدقة المتوفرة في معلومات هذا البحث.

- 4 ـ تعتبر في بعض الحالات هي الطريقة الوحيدة أو الرئيسية التي يمكن استخدامها وذلك لتعذر أو استحالة المسح الشامل.
- 5 ـ العينات هي الأسلوب الوحيد لدراسة بعض خواص الموضوعات الهامة مثال (ملوحة البحر، كريات الدم . . .الخ).

3- الخطوات الرئيسية لتصميم العينة:

- ✓ 1 ـ تحديد المشكلة والهدف المرجو من الدراسة.
- ✓ 2 ـ تعريف وتحديد المجتمع المراد دراسته.
- ✓ 3 ـ تحديد المعلومات المطلوب الحصول عليها.
- ✓ 4 ـ تحديد طرق جمع البيانات.
- ✓ 5 ـ تحديد الإطار وحجم العينة وتكاليفها.
- ✓ 6 ـ إجراء اختبار سريع وأولي عند بعض القضايا الهامة لإدخالها في موضوع الدراسة.
- ✓ 7 ـ تلخيص وتحليل البيانات.
- ✓ 8 ـ تقدير المؤشرات الإحصائية المطلوبة للمجتمع استناداً على العينة.

تعتبر المرحلة الأخيرة من المسائل الهامة في بحوث العينات وذلك لأن العديد من العوامل يؤثر على صحة هذه التقديرات كتجانس المجتمع أو احتمال سحب واحدات المعاينة أو حجم العينة المسحوبة. حيث ينجم عن ذلك أخطاء في تقدير هذه المعالم.

2 ـ العينات العشوائية المنتظمة Echantillonne Systematique:

يتلائم هذا النوع من المعاينة مع الدراسات التي تنصب على المجتمعات المتجانسة والتي تنتمي مفرداتها إلى نوعية واحدة والمبدأ العام يكمن في تحديد ما يسمى بوحدة الابتداء ومن ثم السحب على مسافة منتظمة تسمى فترة السحب.

مثال:

إذا كان لدينا مجتمعاً مؤلفاً من 5000 فرداً نريد سحب عينة منتظمة بحجم 100 فرد. فيكون لدينا السحب $= \frac{5000}{100} = 50$ فرداً ومن بعد ذلك نحدد وحدة الابتداء ولنفرض أن المجتمع N ترقم أفراده من الرقم 1 حتى رقم N وعلينا في هذه الحالة أن نحدد بشكلٍ عشوائي وحدة الابتداء من مجموعة الأعداد التي تقع من 1 إلى 50 وبإضافة مقدار التمثيل بطريقة منتظمة إلى الرقم الأول على النحو التالي:

وليكن رقم وحدة الامتداد 14 السحب وفق التالي:

المفردة الأولى المفردة الثانية وهكذا

64= 14+50 الثالثة 114 والرابعة 164 . . .الخ.

وبصفةٍ عامة فإذا كان الرقم المختار عشوائياً x وأن كل مفردة من مفردات العينة تمثل عدداً أو مقداراً K من مفردات المجتمع فإن:

x = المفردة الأولى.

$x+K$ = المفردة الثانية.

$x+2.K$ = المفردة الثالثة.

أما المفردة التي ترتيبها n تحمل الرقم $[x+(n-1)(K)]$ أي مثلاً:

المفردة التي ترتيبها 8 تحمل الرقم $14+(8-1)(50)=364$.

أما المفردة التي ترتيبها 100 تحمل الرقم $14+(100-1)(50)=4964$.

مزايا هذه الطريقة:

تتميز هذه الطريقة بالسهولة واختصار الوقت والتكاليف. وتتصف بملائمتها مع الدراسات الاجتماعية التي تتعلق بالسكان من حيث توزع الدخول أو استطلاع الرأي حول طريقة مثلاً جديدة لتحديد النسل لكن خطورة هذه الطريقة تكمن في الاختيار العشوائي لوحدة الابتداء خاصةً إذا ما ظهر للباحث أن هذا الرقم له وضع خاص.

2 ـ العينة العشوائية الطبقية Echantillonne Stratifie:

يتلائم هذا النوع من المعاينة مع الدراسات الخاصة بالمجتمعات غير المتجانسة والتي يمكن تقسيمها من الداخل إلى مجموعات كل مجموعة تتصف بدرجة عالية من التجانس الداخلي، والمبدأ العام للمعاينة الطبقية تقوم على أساس تقسيم المجتمع المدروس إلى طبقات متجانسة داخلياً. وتجري المعاينة داخل كل طبقة بطريقة المعاينة العشوائية البسيطة. أما عدد الوحدات المطلوب سحبها من كل طبقة تتم على النحو التالي:

الجزء من الطبقة = حجم العينة المطلوب X حجم الطبقة

حجم المجتمع الأصلي

مثلاً: إذا فرضنا أن حجم المجتمع الأصلي يساوي10000 وحدة ويقسم إلى أربع طبقات A, B, C, D وحجم هذه الطبقات على الترتيب 1000، 4000، 2000، 3000، وإذا أردنا سحب عينة حجمها 1000 مفردة من هذا المجتمع فإن:

من الطبقة الأولى: وحدة $N_A \geq 1000 \times \frac{1000}{10000} = 100$

من الطبقة الثانية: وحدة $N_B \geq 1000 \times \frac{4000}{10000} = 400$

وحدة $N_C \geq 1000 \times \frac{2000}{10000} = 200$

وحدة $N_D \geq 1000 \times \frac{3000}{10000} = 300$

وبهذا يكون حجم العينة المسحوبة من المجتمع الأصلي تساوي 100 وحدة. هذا النوع من المعاينة يسمح إلى درجة كبيرة بدراسة المجتمعات على حسب درجة تجانسها مما يؤدي إلى تقليل الأخطاء عند تعميم نتائج العينة على كل المجتمع وتقدير بعض خصائص المجتمع من هذه العينة وتقليل إلى حد كبير التباين في المقاييس والمؤشرات الإحصائية المستمدة من هذه العينات.

هام جدا

- 5- توزيعات المعاينة:

لنفرض أن مجتمعاً إحصائياً محدوداً Y. فالوسط الحسابي والتباين لهذا المجتمع يجري حسابهما وفق الصيغتين التاليتين:

$$\bar{Y} = \frac{\Sigma y_i}{N}$$

$$\sigma^2 = \frac{\Sigma(y_i - \bar{Y})^2}{N}$$

حيث أن N تشير إلى عدد مفردات المجتمع.

من هذا يمكننا أن نحسب مجموعة من العينات يتوقف عددها على حجم المجتمع N وعلى حجم العينة المسحوب n وعلى أسلوب سحب العينة أي هل هو مع الإعادة أو بدون إعادة. ولكل عينة يمكننا أن نحدد مؤشراتها كالوسط الحسابي والتباين مثلاً. وبذلك نحصل على عدة متوسطات حسابية وعدة تباينات تعرف باسم مجتمع الأوساط الحسابية أو مجتمع التباينات للعينات المسحوبة.

إن توزع هذه الأوساط وهذه التباينات يسمى توزيعات المعاينة. إن مؤشرات توزيع العينة وتوزيعات المعاينة فهي تتأثر بأسلوب السحب وعليه يمكن تمييز حالتين:

$$M = \begin{cases} C^n_{N+n-1} & \text{السحب مع الإعادة} \\ C^n_{N1} & \text{السحب بدون الإعادة} \end{cases}$$

M عدد العينات:

إذ أن:

$$C^n_N = \frac{N!}{n!(N-n)!}$$

وبما أن كل عينة يعطينا تقديراً للمؤشر المدروس فإنه يمكننا القول أنه يوجد لدينا تقديراً لكل مؤشر من مؤشرات المجتمع. وبالتالي يمكننا اعتبار أي تقدير من التقديرات الممكنة متحولاً عشوائياً خاضع لتوزيع احتمالي معين وتختلف من تقدير لآخر.

ـ 6 ـ الخواص الإحصائية للمقدرات:

تستخدم هذه الخواص لتقرير جودة التقدير وهذه الخواص هي:

ـ عدم التحيز: نقول عن التقدير $\hat{\theta}$ أنه تقدير غير متحيز إذا كان توقعه الرياضي مساوياً المؤشر المطلوب تقديره θ:

$$E(\hat{\theta})=\theta$$

هذا يعني مثلاً: الوسط الحسابي للعينة هو تقدير غير متحيز للوسط الحسابي للمجتمع وأن تقدير غير متحيز لتباين المجتمع σ^2 أي أن التوقع $E(\hat{\theta})$ مأخوذ على جميع العينات الممكنة.

ـ التماسك: نقول عن التقدير $\hat{\theta}$ أنه تقدير متماسك للمؤشر θ إذا كانت قيم $\hat{\theta}$ تنتهي احتمالياً إلى θ عندما تنتهي n إلى $N(\infty)$ أي أن:

$$P(\hat{\theta}\xrightarrow[n\to N]{}\theta)=1$$

ويتعلق هذا المعيار بالصيغة الرياضية التي ستستخدم في حساب $\hat{\theta}$.

ـ الفعالية: تقول عن التقدير $\hat{\theta}$ أنه تقدير فعال للمؤشر θ إذا كان التباين الناتج عنه أصغر من جميع التباينات الناتجة عن التقديرات الأخرى أي أن:

$$\sigma^2_{\hat{\theta}}=Min[\sigma^2_{\hat{\theta}_K}]$$

أي أن:

$$\sigma^2_{\hat{\theta}}=\frac{\sum_{i=1}^{n}\left(E(\hat{\theta}-\theta)\right)^2}{M}=\min imum$$

ـ الكفاية: تقول عن التقديرات $\hat{\theta}$ أنه تقدير كاف للمؤشر θ إذا استخدم جميع المعلومات التي توفرها العينة عن المؤشر ويعتبر أفضل المقدرات لأنه يستخدم جميع بيانات العينة.

7ـ التقدير النقطي لمؤشرات المجتمع:

يعني إمكانية تمثيل العدد الوحيد الذي يمثل التقدير بنقطة واحدة على محور موجه وفي هذه الحالة تستخدم المعلومات المتوفرة في عينة حجمها n للوصول إلى عدد واحد أو نقطة تكون تقديراً للوسيط المراد تقديره. وتجدر الإشارة إلى أمثلاً نستطيع تقييم جودة طريقة في التقدير على أساس تقدير واحد إذ لا بد من مراقبة النتائج عند تطبيق الطريقة بصورة متكررة عدداً كبيراً من المرات، وعندئذٍ نلاحظ مدى تمركز النقاط حول نقطة ما. وبما أن التقديرات هي أعداداً فيمكن تقييم جودة المقدار بإقامة توزيع تكراري للتقديرات التي نحصل عليها من عينات متكررة ونلاحظ مدى قرب مركز ذا التوزيع من قيمة الوسيط أو مدى تمركز هذا التوزيع حول قيمة الوسيط. وسنتحدث باختصار عن أهم التقديرات الممكنة لمؤشرات المجتمع المطلوبة وهي:

1 ـ تقدير متوسط خاصة في مجتمع:

إن الوسط الحسابي للعينة هو تقدير غير متحيز للوسط الحسابي للمجتمع $\overline{Y}$ الذي سحبت منه العينة سواءً كان السحب مع الإعادة أو بدون إعادة نظراً لأن التوقع الرياضي لهذا المؤشر يساوي الوسط الحسابي للمجتمع:

$$\overline{Y}=\frac{\sum_{i=1}^{n} y_i}{N}$$

إذ أن هي قيم خاصة في المجتمع وهي مقادير مجهولة كما أن متوسطها $\overline{Y}$ مجهول ولإيجاد تقدير لـ $\overline{Y}$ تستخدم الوسط الحسابي للعينة $\overline{x}$ أي:

$$\overline{x}=\frac{\sum_{i=1}^{n} x_i}{n}$$

إذ أن قيم نفس الخاصة في العينة ذات الحجم n وبالتالي يكون التقدير يساوي:

$$\overline{\overline{Y}}=\overline{X}=\frac{\sum_{i=1}^{n} x_i}{n}$$

وهذا التقدير غير متحيز بحسب تعريف عدم التحيز لأن التوقع الرياضي يساوي:

$$E(\overline{X})=\frac{1}{N}\sum_{i=1}^{n} y_i=\overline{Y}$$

وهو تقدير متماسك لأنه يساوي:

$$\overline{X}=\frac{\sum_{i=1}^{n} x_i}{n}\xrightarrow[n\to N]{}\frac{\Sigma y_i}{N}=\overline{Y}$$

وهو تقدير فعال لأن بحسب خواص التباين. أي تباين القياسات عن وسطها الحسابي أصغر من تباينها عن أي قيمة أخرى. وبما أنه يستخدم جميع معلومات العينة فهو إذن تقدير كاف.

2ـ تقدير تباين المجتمع:

يعرف تباين المجتمع الإحصائي بالعلاقة التالية:

$$\sigma^2=\frac{1}{N}\sum_{i=1}^{N}(y_i-\overline{Y})$$

ويمكن استخدام تباين العينة كمقدار لتباين المجتمع:

$$S^2=\frac{1}{n-1}\sum_{i=1}^{N}(x_i-\overline{x})$$

وحتى يكون هذا التقدير غير متحيز يجب أن يكون: $E(\bar{S})=\sigma^2$

أي نثبت ذلك بإضافة وبطرح ثابت يساوي $\bar{x}$ إلى التباين العينة ومن ثم يحسب التوقع الرياضي لهذا التباين:

$$E(D^2)=E\left[\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right]$$

$$E(D^2)=E\left[\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})-(\bar{X}-\bar{Y})^2\right]$$

$$=\frac{1}{n}\sum_{i=1}^{n}\left[(x_i-\bar{Y})^2-2(x_i-\bar{Y})(x_i-\bar{Y})+(x_i-\bar{Y})^2\right]$$

وبملاحظة أن:

$$E(x_i-\bar{Y})(x_i-\bar{Y})=(\bar{X}i-\bar{Y})(\Sigma X-\overline{nY})=n(\bar{X}i-\bar{Y})^2$$

ونجد أن:

$$E(D^2)=\frac{1}{n}E\left[\sum_{0i=1}^{n}(X-\bar{Y})-n(\bar{X}-\bar{Y})^2\right]$$

$$E(D^2)=\frac{1}{n}E\sum_{0i=1}^{n}(x_i-\bar{Y})^2-\frac{n}{2}E(\bar{X}-\bar{Y})^2$$

$$E(D^2)=\frac{1}{n}\sum_{i=1}^{n}\sigma^2-\sigma^2_{\bar{x}}-\sigma^2_{\bar{x}}$$

إذ أن التوقع الرياضي : $E(\bar{X}_i-\bar{Y})^2$ يساوي تباين المجتمع σ^2 ورمزنا $\sigma^2_{\bar{X}}$ ، ولإيجاد هذا التقدير نجد أن $E(D^2)\neq\sigma^2$ أي أن تباين العينة ويمثل تقدير غير متحيز لتباين المجتمع σ^2 ولإيجاد هذا التقدير غير المتحيز لـ σ^2 يجب حساب قيمة تباين التوسطات عن الوسط الحسابي العام $\sigma^2_{\bar{X}}$ بدلالة σ^2 وهذا يساوي

حالة السحب:

$$\sigma^2_{\bar{X}}=\begin{cases}\dfrac{\sigma^2}{n} & \text{حالة السحب مع حالة الإعادة}\\[2ex] \dfrac{N-n}{N-1}\cdot\dfrac{\sigma^2}{n} & \text{حالة السحب بدون الإعادة}\end{cases}$$

وبالتعويض بالعلاقة التالية: $E(D^2)=\frac{1}{n}\Sigma\sigma^2-\sigma^2_{\bar{x}}-\sigma^2_{\bar{x}}$

نجد أنه في حالة السحب مع الإعادة: $E(D^2)=\sigma^2-\frac{\sigma^2}{n}=\frac{n-1}{n}\sigma^2$

وبضرب $E(D^2)$ بالمقدار $\frac{n-1}{n}$ نجد أن:

$$E\left(\frac{n-1}{n}.D^2\right)=\frac{n-1}{n}E(.D)^2=\frac{n-1}{n}.\frac{n-1}{n}\sigma^2=\sigma^2$$

وبعد الإصلاح نجد أن: $S^2=\frac{n-1}{n}.D^2=\frac{n}{n-1}\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$

وبالتالي يكون التقدير في حالة السحب مع الإعادة: $(X_i-\overline{X}_c)$

$$\delta^2=S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$$

أما في حالة السحب بدون الإعادة:

فنجد أنه لو عوضنا σ_X^2 في العلاقة التالية:

$$E(.D)^2=\frac{1}{n-1}\sum_{i=1}^{n}\sigma^2-\sigma_{\overline{x}}^2-\sigma_{\overline{x}}^2$$

نحصل على: $E(.D)^2=\sigma^2-\frac{n-N}{n-1}.\frac{\sigma^2}{N}$

$$E(.D)^2=\frac{\sigma^2}{N}\left[\frac{n(N-1)-(n-N)}{n-1}.\right]$$

ومنه: $E(.D)^2=\frac{n-1}{n}.\frac{N}{N-1}.\sigma^2$

وإذا أخذنا نفس المقدار: $\frac{n}{N-1}D^2$

نجد أن: $E\left(\frac{n}{N-1}D^2\right)=E(S)^2=\frac{N}{N-1}\sigma^2$

وإذا أخذنا المقدار $\frac{N}{N-1}$ فإننا نجد أن التقدير في حالة السحب بدون إعادة يساوي:

$$\delta^2\,\frac{N-1}{N}=S^2=\frac{N-1}{N(n-1)}\sum_{i=1}^{n}(X_i-\overline{X})^2$$

وبما أن المقدار $\frac{N-1}{N}$ قريب جداً من الواحد فإنه يمكننا اعتبار تباين العينة S^2 تقديراً لـ σ^2 في كلتا حالتي السحب والصيغة هي:

$$\sigma^2 = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

ويعود لسبب لتقسيم مجموع مربعات انحرافات القياسات عن وسطها الحسابي على $n-1$ بدلاً من n في حالة التقدير من العينة للأسباب التالية:

أ ـ عندما تتعامل مع بيانات عينة مسحوبة من مجتمع إحصائي نجد أن:

$$\Sigma(X_i - \overline{Y}) > \Sigma(x_i - \overline{X})^2$$

أي أن مجموع مربعات الاختلافات الكلية في المجتمع الإحصائي أكبر من مجموع مربع الاختلافات الكلية في العينة لذلك فإن القسمة على يحصل تباين العينة بضرب من التباين الحقيقي للمجتمع أي أن متوسط توزيع البيانات للعينات يعطي تقديراً متحيزاً لتباين المجتمع الإحصائي. ما لم يتم القسمة على $n-1$:

ب ـ عندما توجد التباين لمجموعة من البيانات فإننا نخسر درجة حرية واحدة بسبب القيد المفروض على الانحرافات وهو أن مجموع القيم حول الوسط يساوي الصفر. أما درجة الحرية فيقصد بها عدد الحدود التي يمكن أن تتحرك بحرية في مجموعة من البيانات.

3ـ تقدير تباين تقدير المتوسط $\hat{\sigma}_{\overline{x}}^2$:

إن $\hat{\sigma}_{\overline{x}}^2$ عبارة عن تباين متوسطات العينات الممكنة $\overline{X}_i$ عن متوسط المجتمع $\overline{Y}$ ويساوي:

$$\sigma_{\overline{X}}^2 = E(\overline{X}i - \overline{Y})^2$$

وبما أن $\sigma_{\overline{x}}^2$ معطى بدلالة التباين σ^2 وبحسب حالة السحب نجد أن:

$$\sigma_{\overline{x}}^2 = \begin{cases} \dfrac{\hat{\sigma}^2}{n} & \text{حالة السحب مع حالة الإعادة} \\[2ex] \dfrac{N-n}{N-1}\cdot\dfrac{\hat{\sigma}^2}{n} & \text{حالة السحب بدون الإعادة} \end{cases}$$

وباستبدال التقدير σ^2 في العلاقتين:

في حالة السحب مع الإعادة $\delta^2 = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$

في حالة السحب بدون الإعادة $\delta^2 = S^2 = \frac{N-1}{N}S^2 = \frac{N-1}{N(n-1)}\sum_{i=1}^{n}(X_i - \overline{X})^2$

نحصل على التقدير التالي للمقدار $\sigma_{\bar{X}}^2$ وذلك بحسب حالة السحب نجد أن:

$$\hat{\sigma}_{\bar{X}}^2 = \begin{cases} \dfrac{S^2}{n} & \text{حالة السحب مع حالة الإعادة} \\ \dfrac{N-n}{N-1} \cdot \dfrac{S^2}{n} & \text{حالة السحب بدون الإعادة} \end{cases}$$

وإذا كان حجم المجتمع N كبيراً فإن $\dfrac{N-1}{N} \approx 1$ وبذلك يمكننا اعتبار المقدار $\dfrac{S^2}{n}$ في هذه الحالة وفي كلتا حالتي السحب قديراً للتباين $\sigma_{\bar{X}}^2$.

4 ـ تقدير تباين تقدير إجمالي المجتمع:

إن تقدير الإجمالي يعطى بالعلاقة التالية: $\hat{Y} = N\bar{X}$

فإن تباين التقدير لـ $\hat{Y}$ يعرف بالعلاقة التالية: $\sigma_{\hat{Y}}^2 = N^2 \sigma_{\bar{X}}^2$ وإن تقديره يعرف بالعلاقة التالية وبحسب حالة السحب:

$$\sigma_{\hat{Y}}^2 = N^2 \sigma_{\bar{X}}^2 = \begin{cases} N^2 \dfrac{S^2}{n} & \text{حالة السحب مع حالة الإعادة} \\ N^2 \dfrac{N-n}{N} \cdot \dfrac{S^2}{n} & \text{حالة السحب بدون الإعادة} \end{cases}$$

5 ـ تقدير الفرق بين متوسطي مجتمعين وتباينهما:

لنفرض لدينا مجتمعين Y_1 و Y_2 ومتوسطهما $\overline{Y}_1$ و $\overline{Y}_2$ على التوالي فإذا سحبنا عينتين n_1 و n_2 على الترتيب وكان متوسطي الخاصتين في العينتين على التوالي $\overline{x}_1$ و $\overline{x}_2$ فيكون لدينا:

$$\overline{Y}_1 = \overline{x}_1$$

$$\overline{Y}_2 = \overline{x}_2$$

وبذلك نستنتج أن تقدير الفرق بين $\overline{Y}_1 - \overline{Y}_2$ هو الفرق بين $\overline{x}_1 - \overline{x}_2$ وبذلك يكون:

$$(\overline{Y}_1 - \overline{Y}_2) = (\overline{x}_1 - \overline{x}_2)$$

وبما أن العينتين مستقلتان فإن تباين الفرق $(\overline{x}_1 - \overline{x}_2)$ يساوي إلى مجموع تباينهما أي أن:

$$\sigma^2_{(\overline{x}_1 - \overline{x}_2)} = \sigma^2_{\overline{x}_1} + \sigma^2_{\overline{x}_2}$$

وبذلك نجد أن تقدير تباين الفرق بين المتوسطين يساوي:

$$\sigma^2_{(\overline{x}_1 - \overline{x}_2)} = \sigma^2_{\overline{x}_1} + \sigma^2_{\overline{x}_2}$$

وهنا نميز حالتين بحسب حالة السحب وفق التالي:[1]

$$\sigma^2_{(\overline{X}_1 - \overline{X}_2)} = \begin{cases} \dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2} & \text{حالة السحب مع حالة الإعادة} \\[2ex] \dfrac{N_1 - n_1}{N_1} \cdot \dfrac{S_1^1}{n_1} + \dfrac{N_2 - n_2}{N_2} \cdot \dfrac{S_2^1}{n_2} & \text{حالة السحب بدون الإعادة} \end{cases}$$

تقدير نسبة مؤشر نوعي:

لتقدير نسبة خاصة لابد من تحديد عدد عناصر المجتمع التي تتصف بها. فإذا كانت تتصف بتلك الخاصة نقوم بإضافة (I) إلى عدد العناصر المتصفة بها ولا نضيف شيئاً في الحالة المعاكسة. وبالقياس إلى ما برهناه في حالة تقدي الوسط في المجتمع فإننا نجد أن تقدير النسبة R في المجتمع يعطى بواسطة النسبة r في العينة أي أن:

$$R = r = \frac{m}{n}$$

وهذا التقدير هو تقدير غير متميز ومتماسك وفعال وكاف :

وإن تقدير التباين يساوي: $S^2 = r.q$

حيث أن: $r = \frac{m}{n}$

$.q = 1 - r$

وتقدير تباين التقدير R وبحسب حالة السحب يعرف بالعلاقة التالية:

$$\sigma_X^2 = \begin{cases} \frac{r.q}{n} \approx \frac{r.q}{n} & \text{حالة السحب مع حالة الإعادة} \\ \frac{N-n}{N} \cdot \frac{r.q}{n} & \text{حالة السحب بدون الإعادة} \end{cases}$$

7 ـ تقدير الفرق بين نسبتين في مجتمعين وتباينهما:

بطريقة مشابهة لتقدير الفرق بين المتوسطين نجد أن الفرق بين النسبتين R_1 و R_2 لخاصة ما في المجتمعين يقدر بواسطة الفرق بين النسبتين في العينتين r_1 و r_2 وفق التالي: $(R_1 - R_2) = r_1 - r_2$ وتباين الفرق $(r_1 - r_2)$ يقدر بواسطة العلاقة وبحسب في حالة السحب نجد أن:

$$\sigma_{(\bar{R}1-\bar{R}2)}^2 = \sigma_{(\bar{R}1)}^2 + \sigma_{(\bar{R}2)}^2 \begin{cases} \frac{r_1 \cdot q_1}{n_1} \approx \frac{r_2 \cdot q_2}{n_2} & \text{حالة السحب مع حالة الإعادة} \\ \frac{N_1 - n_1}{N_1} \cdot \frac{r_1 \cdot q_1}{n_1} + \frac{N_2 - n_2}{N_2} \cdot \frac{r_2 \cdot q_2}{n_2} & \text{حالة السحب بدون الإعادة} \end{cases}$$

6 ـ8 ـ مجالات الثقة Intervales de Confiances:

1 ـ مقدمة عامة:

لقد ذكرنا أن من مهمات النظرية الإحصائية تقويم وتقدير معايير المجتمع انطلاقاً من معايير العينة. وهذا بحد ذاته ما هو إلا مؤشر لإحدى المقادير المجهولة، لذلك سيبقى هذا المؤشر عديم الفائدة ما دام غير مصحوب بمجال أو فترة (الأمان) الثقة الذي يعرف المجال الذي يتراوح بين الوسط الحسابي وواحدات معينة من الانحراف المعياري بالزيادة والنقصان.

وعلى الرغم من أن معظم التقديرات غير متحيزة وفعالة ومتماسكة وكافية للمؤشر المجهول إلا أن أي منهما لا يعطينا أية درجة من الثقة فنحن لا نعرف إذا كانت القيمة قريبة من القيم الحقيقية أم بعيدة عنها لذلك لابد من إيجاد وسيلة تتضمن لنا وباحتمال معين أن تكون القيمة الحقيقية المجهولة واقعة في مجال معين.

فإذا كان θ مؤشراً في المجتمع (ا, الإجمالي . . .الخ) وكان تقديره θ من العينة فكيف يمكننا إنشاء مجالاً يحتوي على القيمة الحقيقية θ باحتمال قدره β أو يحتوي دلالة قدره α إذ أن: $\alpha = 1-\beta$.

فإذا استطعنا تحديد مركز المجال المطلوب والذي سنرمز له بالرمز C وتحديد نصف طوله والذي سنرمز له بـ l فإن المجال المطلوب:

$$P(C \pm \ell) \Rightarrow P(C-1<0<C+1)=\beta=1-\alpha$$

وحتى نستطيع تحديد كل من C و l فإننا نفترض أن التقديرات الممكنة $\hat{\theta}$ خاضعة للتوزيع الطبيعي والذي توقعه الرياضي $E(\hat{\theta})=\theta$

وانحرافه المعياري يقدر بـ σ_{θ}^{2} فإذا كان حجم العينة n كبيراً أي $N \geq 30$ فإننا نحسب المتحول المعياري: $t=\dfrac{\hat{\theta}-\theta}{\sigma_{\hat{\theta}}^{2}}$

أي يكون خاضعاً للتوزيع الطبيعي المعياري $N(0,1)$ الذي متوسطه = 0 وانحرافه المعياري = 1.

علماً بأن قانون التوزيع الطبيعي المعياري يعطى بالعلاقة التالية: $\varphi(t)=\dfrac{1}{\sqrt{2\pi}} e^{-t^2/2}$.

إننا نريد أن ننشئ مجالاً يحوي القيمة الحقيقية المجهولة θ باحتمال قدره β واستناداً إلى العلاقة التالية:

$t=\dfrac{X-\overline{X}}{\sigma}$ المتحول العشوائي المعياري.

نجد أن:

$$P(-Z_B \le t \le Z_B) = Z\phi(Z_b) - 1 = \beta$$

إذ أن Z_B هو نصف طول المجال المعياري الذي يجعل ذلك الاحتمال مساوياً للمقدار β.

فإذا كان الاحتمال β معلوماً (وغالباً ما نضعه مساوياً لـ 0.95 أو 0.99 فإنه يمكننا حساب Z_B المقابلة له وذلك باستخدام جدول (1) تابع الاحتمالات للتوزيع الطبيعي المعياري وذلك بطريقة معاكسة.

نجد أن:

$$Z_B = \phi^{-1}\left(\frac{1-\beta}{2}\right) = \phi^{-1}\left(1-\frac{\alpha}{2}\right)$$

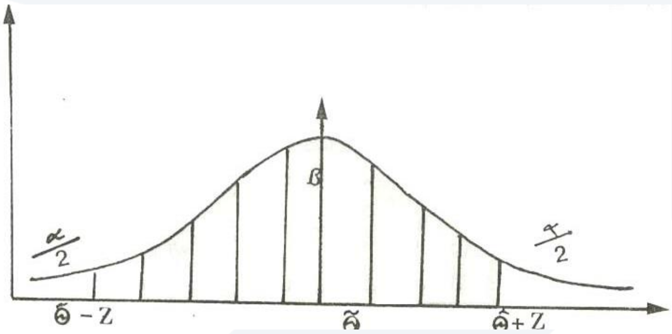
حيث أن $\phi^{-1}\left(1-\frac{\alpha}{2}\right)$ هو قيمة المتحول Z المقابلة للاحتمال $\left(1-\frac{\alpha}{2}\right)$ لذلك نستبدل الرمز Z_B بالرمز $Z_{1-\frac{\alpha}{2}}$ وهي قيمة Z المقابلة للاحتمال $\left(1-\frac{\alpha}{2}\right)$ وبتعويض t بقيمتها في العلاقة:

$$P\left(-Z_{1-\frac{\alpha}{2}} \le \frac{\hat{\theta}-\theta}{\hat{\sigma}_{\hat{\theta}}} \le Z_{1-\frac{\alpha}{2}}\right) = b = 1-\alpha$$

وبما أننا نريد إنشاء مجال للثقة يحتوي القيمة الحقيقية لـ $\hat{\theta}$ فإننا نضرب أطراف المتراجحة بالخطأ المعياري $\hat{\sigma}_{\hat{\theta}}$ فنحصل على العلاقة التالية:

$$P\left(\hat{\theta} - Z_{1-\frac{\alpha}{2}}\hat{\sigma}_{\hat{\theta}} \le \theta \le \hat{\theta} + Z_{1-\frac{\alpha}{2}}\hat{\sigma}_{\hat{\theta}}\right) = b = 1-\alpha$$

وهي العلاقة التي تعطينا مجال عادي يحوي المؤشر θ باحتمال β مركزه هو التقدير θ ونصف طوله $l = Z_{1-\frac{\alpha}{2}}$



أن قانون توزيع المقدار $\frac{\theta-\theta}{\sigma_{\theta}^{2}}$ يختلف باختلاف حجم العينة.

a - **عندما يكون حجم العينة صغيراً** $n<30$ **: في** هذه الحالة يكون المقدار $\frac{\hat{\theta}-\theta}{\sigma_{\theta}^{2}}$ خاضع لتوزيع ستودينت من أجل t_{n-1} مع $n-1$ درجة حرية ومجال الثقة للمؤشر θ يكتب على الشكل التالي:

$$P\left(\theta-t_{(1-\frac{\alpha}{2},,n-1)}\sigma_{\hat{\theta}}\leq\theta\leq\theta+t_{(1-\frac{\alpha}{2},,n-1)}\sigma_{\hat{\theta}}\right)=\beta$$

حيث أن $\hat{\theta}$: التقدير النقطي لـ θ ، σ_{θ} الانحراف المعياري للتقدير النقطي.

$t_{(1-\frac{\alpha}{2},,n-1)}$: القيمة النظرية (الجدولية) التي نحصل عليها من جدول ستودينت (2) والتي تحدد بحسب درجة الثقة المعتمدة والمقابلة لـ $n-1$ درجة حرية.

B - **عندما يكون حجم العينة كبيراً** $n>30$:

في هذه الحالة يخضع المقدار $\frac{\hat{\theta}-\theta}{\sigma}$ للتوزيع الطبيعي المعياري $N(0,1)$ وبالتالي يكتب مجال الثقة لـ θ في العينات الكبيرة كما يلي:

$$P\left(\hat{\theta}-Z_{1-\frac{\alpha}{2}}\hat{\sigma}_{\hat{\theta}}\leq\theta\leq\theta+Z_{1-\frac{\alpha}{2}}\hat{\sigma}_{\hat{\theta}}\right)=\beta$$

حيث أن: $Z_{(1-\frac{\alpha}{2},n-1)}$: القيمة الجدولية التي يمكن الحصول عليها من جدول التوزيع الطبيعي المعياري والتي تحدد طبقاً لدرجة الثقة المعتمدة.

2 ـ التقدير المجالي لبعض مؤشرات المجتمع الإحصائي:

1 ـ مجال الثقة لمتوسط مجتمع طبيعي:

لنفرض أن متحول عشوائي يخضع للتوزيع الطبيعي المعياري قيمته المتوقعة $\overline{Y}$ مجهولة وتباينه σ^2 معلوم أو يمكن تقديره من معطيات العينة المدروسة. إن متوسط العينة $\overline{x}$ هو تقدير متوسط المجتمع $\overline{Y}$ وبحسب حجم العينة يمكننا إنشاء مجال الثقة وفق التالي:

a ـ حالة عينة صغيرة الحجم يكون مجال الثقة:

$$P\left(\overline{x}-Z_{(1-\frac{\alpha}{2},n-1)}\frac{S}{\sqrt{n}}\leq\overline{Y}\leq\overline{x}+Z_{(1-\frac{\alpha}{2},n-1)}\frac{S}{\sqrt{n}}\right)=\beta$$

حيث أن: S: الانحراف المعياري للعينة والذي نحصل عليه من العلاقة:

$$S=\sqrt{\frac{\Sigma(x_i-x)^2}{n-1}}$$

مثال تطبيقي:

نسحب عينة مؤلفة من 26 شخصاً بشكلٍ عشوائي وبعد قياس معدل البول في الدم تبين لنا أن متوسط معدل البول $\overline{x}=0,27G/L$ في الدم وتباين مقداره $S^2=0,0025$ وبفرض أن معدل البول في الدم متحول عشوائي يتبع قانون التوزيع الطبيعي. المطلوب إيجاد مجال الثقة لمتوسط معدل البول في الدم وذلك بمستوى الدلالة 5%.

الحل:

ليكن متحول عشوائي يمثل معدل البول في الدم ويخضع للتوزيع الطبيعي أي $x\approx N(0,1)$ ، وبما أن $\overline{x}$ هو تقدير مفضل لـ $\overline{Y}$ وبما أن حجم العينة المدروسة صغيراً $n<30$ في مجال الثقة. هو من الشكل:

$$P\left(\overline{x}-t_{(1-\frac{\alpha}{2},n-1)}\frac{S}{\sqrt{n}}\leq\overline{Y}\leq\overline{x}+Z_{(1-\frac{\alpha}{2},n-1)}\frac{S}{\sqrt{n}}\right)=0,95$$

ومن جدول ستودبينت نحصل على القيمة الجدولية

$$t_{(1-\frac{\alpha}{2},,n-1)}=t_{(0,95,,25)}=2,060$$

ومنه نجد:

$$P\left(0,27-2,060\frac{0,05}{\sqrt{26}}\leq\overline{Y}\leq0,27+2,060\frac{0,05}{\sqrt{26}}\right)=0,95$$

$$P(0,27 \pm 0,004205) \Rightarrow P(0,228 \le \overline{Y} \le 0,31205) = 0,95$$

ـ حالة عينة كبيرة الحجم $n > 30$ الثقة من العلاقة التالية:

$$P\left(\bar{x} - Z_{(1-\frac{\alpha}{2},)}\frac{S}{\sqrt{n}} \le \overline{Y} \le \bar{x} + Z_{(1-\frac{\alpha}{2},)}\frac{S}{\sqrt{n}}\right) = \beta$$

مثال تطبيقي:

تبين بعد دراسة عينة مؤلفة من 100 طفل أن متوسط الوزن $\bar{x} = 30{,}5\,Kg$ وأن الانحراف المعياري للمجتمع الذي سحبنا منه العينة $S = 9\,Kg$ والمطلوب تحديد مجال الثقة المتوسط وزن الأطفال في هذا المجتمع دلالة 5%.

الحل:

ليكن متحول عشوائي يمثل وزن الطفل ويخضع للتوزيع الطبيعي المعياري $N(0,1)$ وبما أن $\bar{x}$ هو تقدير مفضل لـ $\overline{Y}$ وبما أن حجم العينة المدروسة كبيراً فمجال الثقة. هو من الشكل:

$$P\left(\bar{x} - Z_{1-\frac{\alpha}{2},}\frac{S}{\sqrt{n}} \le \overline{Y} \le \bar{x} + Z_{1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right) = \beta$$

ومن جدول ستودبينت نحصل على القيمة الجدولية

$$t_{(1-\frac{\alpha}{2},,n-1)} = t_{(0,95,,25)} = 2{,}060$$

ومنه نجد:

$$P\left(0,27 - 2{,}060\frac{0{,}05}{\sqrt{26}} \le \overline{Y} \le 0,27 + 2{,}060\frac{0{,}05}{\sqrt{26}}\right) = 0{,}95$$

$$P(0,27 \pm 0,004205) \Rightarrow P(0,228 \le \overline{Y} \le 0,31205) = 0,95$$

وبالتعويض في المعادلة نجد أن:

$$P\left(30,5-1,96\frac{9}{\sqrt{100}}\leq \bar{Y}\leq 30,5-1,96\frac{9}{\sqrt{100}}\right)=0,95$$

وبعد الإصلاح يكون مجال الثقة:

$$P(30\pm l,1,764)\Rightarrow P(28,736\leq \bar{Y}\leq 32,264)=0,95$$

مجال الثقة للفرق بين متوسطي مجتمعين طبيعيين:

رغم أن دراسة المجتمع الواحد هو تطبيق مهم في الحياة العملية إلا أنه في كثير من الأحيان تنشأ الحاجة لدراسة ومقارنة مجتمعين ولمعرفة ما إذا كان هذان المجتمعان مختلفان أم لا. وقد يكون المعيار للاختلاف أو عدمه هو معرفة الفرق بين المتوسطين لهذين المجتمعين فإذا كانا مختلفين دل ذلك على أنهما مختلفان وإن كان المتوسطان غير مختلفين دل ذلك أن المجتمعين أيضاً غير مختلفين. لنفرض لدينا مجتمعين Y_1 و Y_2 ويخضعان للتوزيعن الطبيعيين $N_1(\bar{Y}_1,\sigma_1^2)$ و $N_2(\bar{Y}_2,\sigma_2^2)$ على الترتيب. فإذا سحبنا من المجتمعين عينتين بسحبتين n_1 و n_2 وكان متوسطا العينتان $\bar{x}_1$ و $\bar{x}_2$ على التوالي وتباينهما S_1^2 و S_2^2 فالمطلوب إنشاء مجال للثقة للفرق بين $(\bar{Y}_1,\bar{Y}_2)$ ونعلم أن تقدير الفرق $(\bar{Y}_1,\bar{Y}_2)$ هو:

$$\Delta=(\bar{Y}1,\bar{Y}2)=\bar{x}1-\bar{x}2$$

وبتباين هذا التقدير يعطى بالعلاقة التالية:

$$\sigma_{\Delta}^2(\bar{x}1-\bar{x}2)=\sigma_{\bar{x}1}^2=\sigma_{\bar{x}2}^2=\frac{\sigma_1^2}{n_1}.\frac{\sigma_2^2}{n_2}$$

ولإنشاء مجال الثقة للفرق بين متوسطين نميز بين حالتين:

a ـ حالة حجم العينة الكلية صغيراً $n_1+n_2>30$:

نعلم في هذه الحالة أن تقدير σ_{Δ}^2 يساوي: $\sigma_{\Delta}^2=\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}$

بما أنه لدينا حالة $\sigma_1^2=\sigma_2^2=\sigma^2$ وأن تباين العينتين S_1^2 و S_2^2 هما تقديران لكل من σ_1^2 و σ_2^2 على التوالي ومن ثم لـ σ^2 فإنه يمكننا أن نقدر σ^2 من تباين العينتين S_1^2 و S_2^2 وبالتالي يمكننا أن σ^2 بواسطة العلاقة التالية:

$$\sigma^2=S_2\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-n}$$

وهذا التقدير غير متحيز لـ σ^2 وضمن هذه الشروط فإن التباين σ_{Δ}^2 يعرف بالعلاقة التالية:

$$\sigma_{\Delta}^{2}=\operatorname{var}(\bar{x}1-\bar{x}2)=\sigma_{\bar{x}1}^{2}+\sigma_{\bar{x}2}^{2}=\frac{\sigma_{1}^{2}}{n_{1}}+\frac{\sigma_{2}^{2}}{n_{2}}$$

يصبح على النحو التالي:

$$\sigma_{\Delta}^{2}=\frac{\sigma_{1}^{2}}{n_{1}}+\frac{\sigma_{2}^{2}}{n_{2}}=\sigma^{2}\left(\frac{1}{n_{1}}+\frac{1}{n_{2}}\right)$$

وتقدير غير المتحيز يعطى بالعلاقة التالية:

$$\sigma_{\Delta}^{2}=S^{2}\left(\frac{1}{n_{1}}+\frac{2}{n_{2}}\right)\frac{(n_{1}-1)S_{1}^{2}+(n_{2}-1)S_{2}^{2}}{n_{1}+n_{2}-2}\left(\frac{1}{n_{1}}+\frac{2}{n_{2}}\right)$$

وبما أن حجم العينتين صغيراً ويخضع لتوزيع ستودبينت ذي n_1+n_2-2 درجة حرية وبالتالي نصف طول المجال المطلوب يساوي:

$$t_{(\frac{\alpha}{2},,n_1+n_2-2)}.S\sqrt{\frac{1}{n_1}+\frac{1}{n2}}.$$

وبالتالي يصبح مجال الثقة على الشكل التالي:

$$P\left(\bar{x}_1-\bar{x}_2-t_{(\frac{\alpha}{2},,n_1+n_2-2)}.S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\le\bar{Y}_1-\bar{Y}_2\le\bar{x}_1-\bar{x}_2-t_{(\frac{\alpha}{2},,n_1+n_2-2)}.S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right).=\beta$$

حيث أن:

$$S=\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}$$

وهو المجال المطلوب والشروط بأن يكون: $\sigma_1^2=\sigma_2^2=\sigma^2$

a ـ حالة تكون العينة الكلية كبيرة $n_1+n_2>0$:

في هذه الحالة لا يشترط تساوي التباينين σ_1^2 و σ_2^2 ولكن نعتبر المقدار $\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}$ قيمته تقريبية للمقدار $\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}$

وبما أن حجم العينة الكلية كبيراً تكون خاضعة للتوزيع الطبيعي المعياري $N(0,1)$ وبالتالي يعطى مجال الثقة بالعلاقة التالية:

$$P\left(\bar{x}_1-\bar{x}_2-Z_{1\frac{\alpha}{2}}.\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}\le\bar{Y}_1-\bar{Y}_2\le\bar{x}_1-\bar{x}_2+Z_{1\frac{\alpha}{2}}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}\right).=\beta=1-\alpha$$

مثال تطبيقي:

لنفرض قمنا بدراسة تتعلق بأوزان عينتين من الأطفال فور ولادتهم $n_1 = n_2 = 10$ فوجدنا أن متوسط وزن الطفل في العينتين هو على الترتيب $\bar{x}_1 = 3,6$ و $\bar{x}_2 = 3,2$

وتباينهما على التوالي $S_1^2 = 22$ و $S_2^2 = 20$ والمطلوب إيجاد مجال لثقة للفرق بين متوسطي الوزن في هذين المجتمعين مفترضين أن باحتمال مقداره 95%.

الحل:

بما أن حجم العينة الكلية $n_1 + n_2 > 30$ فإنها خاضعة لتوزيع ستودبينت من $\left(t_{1-\frac{\alpha}{2}}, n_1 + n_2 - 2.\right)$ وبالتالي مجال الثقة يعطى بالعلاقة التالية:

$$P\left(\bar{x}_1 - \bar{x}_2 - t_{1-\frac{\alpha}{2}}.S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \le \bar{Y}_1 - \bar{Y}_2 \le \bar{x}_1 - \bar{x}_2 - t_{1-\frac{\alpha}{2}}.S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right). = \beta$$

علماً بأن:

$$S = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

$$S = \sqrt{\frac{(10-1)22 + (10-1)20}{10+10-2}} = 4.59$$

فالقيمة الجدولية تساوي

$$t_{0,95\left(1-\frac{\alpha}{2},,10+10-2\right)}. = 2,101$$

$$P(3,6 - 2,101)\left((4,59\sqrt{\frac{1}{10} + \frac{1}{10}}) \le \bar{Y}_1 - \bar{Y}_2 \le 3,6 - 3,2 + (2,101)(4,59)\sqrt{\frac{1}{10} + \frac{1}{10}}\right). = 0,95$$

$$P(3,6 - 2,3) - (2,101) \le \bar{Y}_1 - \bar{Y}_2 \le (3,6 - 3,2 + (2,101) = 0,95$$

$$P(0,4 - 4,308 \le \bar{Y}_1 - \bar{Y}_2 \le 0,4 + 4,308) = 0.95$$

و هو المجال المطلوب: $P(-3,908;\ 4,708)$

مثال تطبيقي:

سحبنا من بين الطالبات عينة بحجم $n=15$ ومن بين الطلاب عينة بحجم $n=20$ ودرسنا متوسط نفقاتهم في الشهر فوجدنا أن:

متوسط نفقات الطالبات في العينة: $\bar{x}_1=1200$ ل.س

متوسط نفقات الطلاب في العينة: $\bar{x}_2=1000$ ل.س.

وتباينهما على التوالي $S_1^2=600$ ل.س و $S_2^2=500$ ل.س والمطلوب إيجاد مجال لثقة للفرق بين متوسطي النفقات للطلاب والطالبات باحتمال مقدره 95%..

وبما أن حجم العينة الكلية كبيراً تكون خاضعة للتوزيع الطبيعي المعياري $N(0,1)$ والقيمة الجدولية تساوي: $Z_{(1-\frac{\alpha}{2},0{,}95)}$. ومجال الثقة المطلوب يعطى بالعلاقة التالية:

$$P\left(\bar{x}_1-\bar{x}_2-Z_{1\frac{\alpha}{2}}.\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}\leq \bar{Y}_1-\bar{Y}_2\leq \bar{x}_1-\bar{x}_2+Z_{1\frac{\alpha}{2}}.\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}\right).=\beta$$

$$P\left(1200-1000-1{,}96\sqrt{\frac{600}{15}+\frac{500}{20}}\leq \bar{Y}_1-\bar{Y}_2\leq 1200-1000-1{,}96\sqrt{\frac{600}{15}+\frac{500}{20}}\right).=0{,}95$$

$$P(200-15{,}80\leq \bar{Y}_1-\bar{Y}_2\leq 200-15{,}80).=0{,}95$$

وهو مجال الثقة المطلوب: $P(184{,}2;215{,}8)$

ـ مجال الثقة للنسبة في مجتمع طبيعي (مؤشر نوعي):

تشبه هذه الحالة تماماً حالة التقدير المجال (مجال الثقة) لمتوسط طبيعي. إلا أنه نستبدل المتوسط بنسبة الأفراد الذين يتمتعون بالخاصة أو الصفة المدروسة في المجتمع ونرمز لها بالرمز R ونرمز لها بالرمز r في حالة تقدير العينة:

الأفراد المتمتعة بالخاصة المدروسة في العينة تساوي: $r=R=\frac{m}{n}=\bar{X}$

وتباين النسبة في العينة يساوي: $S^2=\frac{r.q}{n}$

$q=1-r$

ونميز حالتين بحسب العينة:

ـ مجال الثقة لعينة صغيرة الحجم $n<30$: مجال الثقة يعطى بالعلاقة التالية:

$$P\left(r-t_{1-\frac{\alpha}{2}}\sqrt{\frac{r.q}{n}}\leq R\leq r+t_{1-\frac{\alpha}{2}}\sqrt{\frac{r.q}{n}}\right)=1-a$$

مثال تطبيقي:

لتقدير نسبة لمدخنين لنوع A من التبغ سحبنا عينة بدون إعادة بحجم $n=25$ شخصاً فوجدنا أن نسبة المدخنين للنوع A في العينة كانت $r=0{,}20$. أوجد مجال الثقة لنسبة المدخنين في المجتمع باحتمال قدره %95.

الحل: بما أن حجم العينة صغيراً، نبحث عن القيمة الجدولية في جدول ستودبينت المقابلة للاحتمال $t=1-\frac{a}{2}$ تساوي $t_{\beta}=b_{(0,95;24)}$ ومنه مجال الثقة يساوي:

$$P\left(r-t_{1-\frac{\alpha}{2}}\sqrt{\frac{r.q}{n}}\leq R\leq r+t_{1-\frac{\alpha}{2}}\sqrt{\frac{r.q}{n}}\right)=\beta$$

$$P\left(0{,}20-1{,}711\sqrt{\frac{0{,}20\times 0{,}80}{25}}\leq R\leq 0{,}20+1{,}711\sqrt{\frac{0{,}20\times 0{,}80}{25}}\right)=0{,}95$$

$$P(0{,}20-(1{,}711)(0{,}08)\leq R\leq 0{,}20-1{,}711(0{,}08)=0{,}95$$

وهو المجال المطلوب.

b ـ مجال الثقة لعينة كبيرة الحجم $n>30$ **:**

مجال الثقة يعطى بالعلاقة التالية:

$$P\left(r-Z_{1-\frac{\alpha}{2}}\sqrt{\frac{r.q}{n}}\leq R\leq r+Z_{1-\frac{\alpha}{2}}\sqrt{\frac{r.q}{n}}\right)=1-a$$

مثال تطبيقي:

حرصاً على سلامة المواطنين، أصدرت إدارة المرور تعليمات جديدة لتنظيم حركة السير. وقد اختير 400 شخصاً بشكلٍ عشوائي وسئلوا عن مدى موافقتهم أو اعتراضهم على هذه التعديلات فأجاب 250 شخصاً بالارتياح لهذه التعديلات.

المطلوب: أوجد مجال الثقة لنسبة الموافقين على هذه التعديلات في المجتمع. باحتمال قدره 95%.

الحل:

نسبة الموفقين على العينة: $r = \frac{m}{n} = \frac{250}{400} = 0,625$

نسبة غير الموفقين $q = 1 - r = 1 - 0,625 = 0,375$

بما أن حجم العينة كبيراً، فإن القيمة الجدولية عند مستوى الدلالة المطلوب تساوي

$$Z_{(1-\frac{\alpha}{2}, 0,95)}$$

مجال الثقة المطلوب:

$$P\left(r - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{r.q}{n}} \leq R \leq r + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{r.q}{n}}\right). = 1 - a$$

وبالتعويض نجد أن:

$$P\left(0,625 - 1,96\sqrt{\frac{0,625 \times 0,375}{400}} \leq R \leq 0,625 + 1,96\sqrt{\frac{0,625 \times 0,375}{400}}\right). = 0,95$$

$$P(0,625 - 1,96 \times 0,024 \leq R \leq 0,625 - 1,96 \times 0,024) = 0,95$$

$$P(0,578; 0,672 = 0,95$$

أي أن نسبة الموافقين على هذه التعديلات تقع ما بين 58% و 67% تقريباً.

ـ مجال الثقة للفرق بين نسبيتين R_1 و R_2 لخاصة واحدة في مجتمعين طبيعيين:

لنفرض لدينا مجتمعين Y_1 و Y_2 نسحب منهما عينتين n_1 و n_2 فإن r_1 و r_2 عبارة عن نسبة عدد الأفراد المتمتعة بالخاصة المدروسة في العينتين على الترتيب: وهما أيضاً تقديران R_1 و R_2.

وبطريقةٍ مشابهة عند دراسة مجال الثقة للفرق بين متوسطين نميز بين الحالتين التاليتين:

a ـ حالة حجم العينة الكلية صغيراً $n_1 + n_2 = n$ أي $n_1 + n_2 < 30$:

لتحديد مجال الثقة مع اعتبار أن $\sigma_1^2 = \sigma_2^2 = \sigma^2$ وهذا شرط محقق في المجتمعين المدروسين. وفي هذه الحالة نقوم بتقدير نقطي لـ σ^2 وفق العلاقة التالية:

$$\sigma^2 = S^2 = \frac{r_1 q_1 n_1 + r_2 q_2 n_2}{n_1 + n_2 - 2}$$

بما أن حجم العينتين صغيراً فإن العينة الكلية تخضع لتوزيع ستودبينت n_1+n_2-2 درجة حرية وبالتالي مجال الثقة المطلوب يعطى بالعلاقة التالية:

$$P\left(r_1-r_2-t_{1-\frac{\alpha}{2}}.S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\leq R_1-R_2\leq r_1-r_2-t_{1-\frac{\alpha}{2}}.S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right).=\beta$$

مثال تطبيقي:

لدراسة نسبة المدخنين في مجتمعين متميزان كان فيهما $\sigma_1^2=\sigma_2^2$ سحبنا منهما عينتين بحجمين $n_1 10$ و $n_2=15$ فكان عدد المدخنين في العينة الأولى 4 وفي العينة الثانية 5 أوجد مجال لثقة للفرق بين نسبتي المدخنين في المجتمعين باحتمال قدره 95%.

الحل:

نلاحظ أن مجموع العينة الكلية n_1+n_2 صغيراً فيمكننا أن نحدد مجال الثقة وفق العلاقة التالية:

$$P\left(r_1-r_2-t_{1-\frac{\alpha}{2}}.S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\leq R_1-R_2\leq r_1-r_2-t_{1-\frac{\alpha}{2}}.S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right).=\beta$$

حيث أن:

$$S=\sqrt{\frac{r_1q_1n_1+r_2q_2n_2}{n_1+n_2-2}}$$

ومنه نجد أن:

$$S=\sqrt{\frac{(10)(0,4)(0,6)+(0,33)(0,67)(10)}{10+15-2}}=\sqrt{\frac{4,611}{25-2}}=0,45$$

$$P\left(0,4-0,33-(1,714)(0,45)\sqrt{\frac{1}{10}+\frac{1}{15}}\leq R_1-R_2\leq 0,4-0,33+(1,714)(0,45)\sqrt{\frac{1}{10}+\frac{1}{15}}\right).=\beta$$

$$P(0,4-0,33-0,3148\leq R_1-R_2\leq 0,4-0,33+0,3148).=0,45$$

$$P(-0,2448;0,3848).=0,45$$

و هو المجال المطلوب.

b ـ حالة حجم العينة الكلية كبيراً $n_1+n_2>30$:

في هذه الحالة لا يشترط تساوي التباينين $\sigma_1^2=\sigma_2^2$ ولكن المقدار r_1q_1 قيمته تقريبية لـ σ_2^2 و المقدار r_2q_2 قيمته تقريبية لـ σ_2^2 وبذلك يصبح مجال الثقة في شكله النهائي كما يلي:

$$P\left(r_1-r_2-Z_{1\frac{\alpha}{2}}\sqrt{\frac{r_1q_1}{n_1}+\frac{r_2q_2}{n_2}}\le R_1-R_2\le r_1-r_2+Z_{1\frac{\alpha}{2}}\sqrt{\frac{r_1q_1}{n_1}+\frac{r_2q_2}{n_2}}\right).=\beta$$

وفي هذا التوزيع خاضع للتوزيع الطبيعي المعياري $N(0,1)$.

مثال تطبيقي:

في دراسة ميدانية لتأثر نوع من اللقاح ضد انتشار مرض معين يقصد معرفة مدى الفرق بين نسبة المصابين في الريف وفي الحضر (المدينة). فكان حجم العينة المأخوذة من المدينة $n_1=1000$ شخص منهم 300 مصاب، وحجم العينة المأخوذة من الريف $n_2=800$ بينهم وجد 220 شخص مصاب بالمرض. أوجد مجال لثقة للفرق بين نسبة المصابين في الريف والمدينة باحتمال قدره 95%.

الحل:

- التقدير النقطي لنسبة المصابين في المدينة: $r_1=\frac{300}{1000}=0,30$

- نسبة غير المصابين: $q_1=1-r=1-0,30=0,70$

- التقدير النقطي لنسبة المصابين في الريف: $r_2=\frac{280}{1000}=0,275$

- نسبة غير المصابين: $q_2=1-r=1-0,725=0,725$

وبما أن حجم العينة كبيراً فالقيمة الجدولية المقابلة للاحتمال 95% تساوي $Z_{1\frac{\alpha}{2}}=1,96$

ومنه نجد أن مجال الثقة يساوي:

وبالتعويض بالصيغة $$P\left(r_1-r_2-Z_{1\frac{\alpha}{2}}\sqrt{\frac{r_1q_1}{n_1}+\frac{r_2q_2}{n_2}}\le R_1-R_2\le r_1-r_2+Z_{1\frac{\alpha}{2}}\sqrt{\frac{r_1q_1}{n_1}+\frac{r_2q_2}{n_2}}\right).=1-a$$

نجد أن:

$$P\left(0,30-0,275+1,96\sqrt{\frac{0,3\times0,7}{1000}+\frac{0,275\times0,725}{800}}\le R_1-R_2\right).=\beta$$

$$\left(\le 0,30-0,275+1,96\sqrt{\frac{0,3\times0,7}{1000}+\frac{0,275\times0,725}{800}}\right)$$

$$P(0,025-0,041\le R_1-R_2\le 0,025-0,041)=0,95$$

$$P(-0,016;0,66)=0,95$$

وهو المجال المطلوب.

ومن هذا يتضح أن الفرق بين النسبتين يتراوح بين 16 بالألف لصالح أهل الريف و 66 بالألف لصالح أهل المدينة.

ـ مجال الثقة لتباين مجتمع طبيعي:

لإنشاء مجال الثقة لتباين مجتمع طبيعي سنقوم بتحديد الحدين الأيسر (الأدنى) والحد الأيمن للمجال بحيث يكون ذلك المجال يحتوي التباين σ^2 باحتمال مقداره β. والسبب في ذلك هو التعامل مع التوزيع الاحتمالي x^2 الذي يتصف بعدم التناظر وبالتالي يصبح المطلوب البحث عن حدي المجال $]D_2, D_1[$ حيث أن $D_1 > D_2$ وأن $D_1 > 0$ وبحيث يتحقق لدينا الشرط التالي:

$$P(D_1 > \sigma^2 < D_2) = \beta$$

وبما أن التقدير يتم على أساس عينة حجمها n وتباينها S^2 وبضرب مقلوب المتراجحة السابقة بالمقدار $(n-1)S^2$ نحصل على: $P\left[\frac{(n-1)S^2}{D_2} < \frac{(n-1)S^2}{\sigma_2} < \frac{(n-1)S^2}{D_1}\right]$

وبما أن المتحول $\frac{(n-1)S^2}{\sigma^2}$ خاضع لتوزيع كاي مربع x^2_{n-1} ذي $n-1$ درجة حرية. ولتسهيل العمليات الحسابية نرمز للمقدار $\frac{(n-1)S^2}{\sigma^2}$ بالرمز x وبالتالي نكتب الصيغة السابقة على الشكل التالي:

$$P\left\{\left[x < \frac{(n-1)S^2}{D_1}\right] - p\left[x \leq \frac{(n-1)S^2}{D_2}\right]\right\} = \beta$$

وبمقارنة هذه العلاقة مع العلاقة التي تعطينا قيم x_p الحدية في الجداول الملحقة والتي لها الشكل التالي:

$$p(x < x_p) = \int_0^{xp} x_n^2(x).dx = p$$

فإنه يمكننا أن نستنتج أن:

$$p_1 = \left(\chi < \frac{(n-1)S^2}{D_1}\right)$$

$$p_2 = \left(\chi \leq \frac{(n-1)S^2}{D_2}\right)$$

وبذلك نحصل على المعادلة التالية: $p_1 - p_2 = 1 - \alpha = \beta$

وهذه المعادلة تحتوي على مجهولين P_1 و P_2 لا بد إذن من إيجاد حل المجهولين ثم حساب الآخر. وحتى نترك على جانبي مجال الثقة احتمالين متساويين يساويان $\frac{\alpha}{2}$ فإننا نضع الاحتمال الثاني مساوياً $\frac{\alpha}{2}$ أي:

$$p_2 = \left(\chi \leq \frac{(n-1)S^2}{D2} \right) = \frac{\alpha}{2}$$

وبالتعويض في المعادلة $P_1 - P_2 = 1 - \alpha = \beta$ نجد أن:

$$p_1 = \left(\chi < \frac{(n-1)S^2}{D_1} \right) = 1 - \frac{\alpha}{2}$$

وبحل هاتين المعادلتين نجد أن: $\frac{(n-1)S^2}{D_2} = \chi \frac{a}{2}$

$$\frac{(n-1)S^2}{D_2} = \chi_1 \frac{a}{2}$$

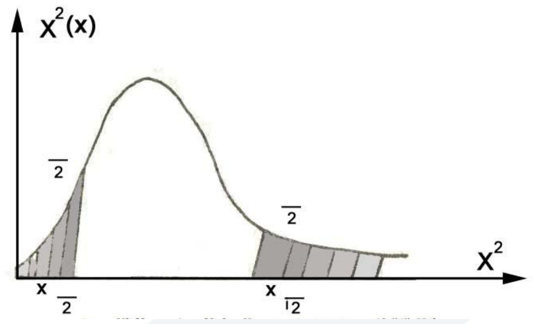
حيث أن $\chi \frac{a}{2}$ هي القيمة الجدولية للمتحول χ^2 المقابلة للاحتمالات $\frac{a}{2}$ ذي $n-1$ درجة حرية.

وإن $\chi_1 - \frac{a}{2}$ هي القيمة الجدولية للمتحول χ^2 المقابلة للاحتمال $1-\frac{a}{2}$ ذي $n-1$ درجة حرية.

ومن المعادلتين السابقتين نحسب كل من $D_1 - D_2$ وفق التالي: $D_2 = \frac{(n-1)S^2}{\chi \frac{a}{2}}$

وكذلك : $D_1 = \frac{(n-1)S^2}{\chi \frac{a}{2}}$

والشكل التالي يوضح ذلك:



وبالتعويض بالمعادلة التالية: $P(D_1 < \sigma^2 < D_2) = \beta$ نحصل على مجال الثقة المطلوب :

$$p = \left[\frac{(n-1)S^2}{\chi \frac{a}{2}} < \sigma^2 < \frac{(n-1)S^2}{\chi \frac{a}{2}} = \beta \right]$$

مثال تطبيقي:

أوجد مجال الثقة لتباين طول الطالب σ^2 علماً بأن تباين العينة التي حجمها n=10 كان يساوي $S^2=24.48$ وذلك باحتمال مقداره 95%.

الحل: نقوم بحساب المجال D_2,D_1 ولذلك نبحث عن قيمة Xa_{2} و $X_{1-\frac{a}{2}}$ عند درجة حرية n-1=(10-1)=9 من جداول X^2 فنجد أن:

$$X_{(1-\frac{a}{2},9)} = X_{(0.95.9)} = 16.9$$

$$X_{(1-\frac{a}{2},9)} = X_{(0.95.9)} = 3.33$$

وبالتعويض بالعلاقة التالية نجد أن:

$$p = \left[\frac{(10-1)24.28}{16.92} < \sigma^2 < \frac{(10-1)24.28}{3.33} = 0.95 \right]$$

$$p = \left[\frac{(218.52}{16.92} < \sigma^2 < \frac{218.52}{3.33} = 0.95 \right]$$

$$p = (12.93 < \sigma^2 < 65.62) = 0.95$$

وهو مجال الثقة المطلوب.

ـ مجال الثقة لنسبة تبايني مجتمعين طبيعيين: لنفترض لدينا مجتمعين طبيعيين تباينهما $\sigma_1^2,\ \ \sigma_2^2,$ على الترتيب. ونريد إنشاء مجال ثقة لنسبة $\frac{\sigma_1^2}{\sigma_2^2},$ أي مقارنة التباينين $\sigma_1^2,$ و $\sigma_2^2,$ نظراً لأهمية ذلك في البحوث العلمية مثلاً لمعرفة الدقة التقنية لنوعين من الأجهزة أو مدى تجانس مجتمعين . . .الخ.

فإذا كانت النسبة $\frac{\sigma_1^2}{\sigma_2^2},$ أو تقديرهما يساوي $\frac{S_1^2}{S_2^2},$ مساوية أو قريبة من الواحد فإن ذلك يعني أن المجتمعين متشابهان من حيث التباين. أما إذا كانت النسبة أصغر من الواحد أو أكبر منه فإن ذلك يدل على أن تبايني المجتمعين مختلفان. ولإنشاء مجال الثقة للنسبة $\frac{\sigma_1^2}{\sigma_2^2},$ نفرض أننا سحبنا عينتين عشوائيتين بجمعين n_1 و n_2 على التوالي وتباينهما S_1^2 و S_2^2 على التوالي ويجب مراعاة أن يكون $S_1^2 \geq S_2^2$ أي يجب أن يكون التباين الأكبر في الصورة. فمجال الثقة يعطى بالعلاقة التالية:

$$p = \left[\frac{S_1^2}{S_2^2} \cdot \frac{1}{F_{1-\frac{a}{2}(n_1-1,n_2-1)}} < \frac{\sigma_1^2}{\sigma_1^2} < \frac{S_1^2}{S_2^2} . F_{1-\frac{a}{2}(n_1-1,n_2-1)} \right] \beta$$

ولحساب حدي المجال يتم العلاقتين التاليتين:

$$d_1 = \frac{S_1^2}{S_2^2} \cdot \frac{1}{F_{1-\frac{a}{2}(n_1-1,n_2-1)}}$$

$$d_2 = \frac{S_1^2}{S_2^2} \cdot \frac{1}{F_{1-\frac{a}{2}(n_2-1,n_1-1)}}$$

ويتم حساب القيم الجدولية من جداول توزيع فيشر F من أجل k_1-n_{1}-1 درجة k_2-n_{1}-1 درجة حرية عند مستوى دلالة معين.

مثال تطبيقي:

للتفضيل بين جهازين لقياس شدة التيار في مخبر للتحليل الطبي أجريت بواسطة الأول 10 قياسات وبواسطة الثاني 8 قياسات فوجدنا أن تباين هاتين العينتين من القياسات يساويان $S_1^2 = 4.14$ و $S_2^2 = 3.21$.

أوجد مجال الثقة النسبية الجهازين $\frac{\sigma_1}{\sigma_2}$ باحتمال مقداره 0.95.

الحل:

- نحدد حدي المجال d_2,d_1 من العلاقتين التاليتين:

$$d_1 = \frac{S_1^2}{S_2^2} \cdot \frac{1}{F_{1-\frac{a}{2}(n_1-1,n_2-1)}}$$

$$d_1 = \frac{S_1^2}{S_2^2} \cdot \frac{1}{F_{1-\frac{a}{2}(n_2-1,n_1-1)}}$$

وكذلك :

نلاحظ أن عدد درجات الحرية: $n_1 - 1 = 10 - 9 = 9 \;,\; n_2 - 1 = 8 - 1 = 7$

وبما أن $\frac{\sigma_1}{\sigma_2} = 0.05$ فإننا من جداول F نجد أن : $F_{1-\frac{a}{2}(n_1-1,n_2-1)} = F_{0.95(79)=3.29}$

و

$$F_{1-\frac{a}{2}(n_2-1,n_1-1)} = F_{0.95(9,7)=3.68}$$

$$p = \left[\frac{4.14}{3.21} \cdot \frac{1}{3.68} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{4.14}{3.21} .3.29\right] = 0.95$$

$$p = \left[0.35 < \frac{\sigma_1^2}{\sigma_2^2} < 4.254\right] = 0.95$$

وهو يميل إلى اليمين ويتضمن القيمة واحد التي تدل تساوي التباينين ([1]).

9- اختيار حجم العينة:

تصميم التجربة في الأساس خطة للحصول على كمية من المعلومات. فبعض القياسات يمكن أن تحتوي كمية كبيرة من المعلومات حول المؤشر المدروس. بينما تحتوي قياسات أخرى كمية أقل أو أنها لا تحتوي أية معلومات إطلاقاً. تؤثر طريقة أخذ العينة أو نوع التصميم التجريبي في كمية المعلومات التي نحصل عليها لقاء ملاحظة واحدة وهي تتحكم مع حجم العينة n في الكمية الإجمالية للمعلومات المتوفرة في العينة.

وفي طليعة ما يواجه الباحث العلمي عند تخطيط تجربته هو اختيار حجم العينة، وربما كان أكثر التساؤلات تواتراً في المسائل الإحصائية هو " كم قياساً يجب أن تتضمن العينة" لكن من المؤكد أن الكمية الإجمالية التي تحتويها العينة من المعلومات ستؤثر في مقياس جودة الاستقرار الإحصائي و لابد من أن يحددها المجرب إضافة لمعرفة مدى الدقة التي يرغب المجرب للتقدير المطلوب ويمكن عندئذ التعبير عن ذلك بوضع حد لخطأ التقدير.

حجم العينة اللازم لتقدير متوسط مؤشر ما:

يمكن تحديد حجم العينة n اللازم من العلاقة التالية:

في حالة السحب مع الإعادة:
$$n \geq \frac{Z^2_{1-\frac{a}{2}} S^2}{d^2}$$

حيث أن S^2: تباين العينة، d^2: مقدار الدقة المطلوبة.

في حالة السحب بدون إعادة:
$$n \geq \frac{NZ^2_{1-\frac{a}{2}} S^2}{Nd^2 + S^2_{1-\frac{a}{2}} S^2} =$$

مثال تطبيقي:

أوجد حجم العينة اللازم سحبه لتقدير متوسط استهلاك الأسرة من النفقات الصحية في الشهر وبدقة d=200 وباحتمال مقداره 95% علماً بأن عينة اختيارية أعطتنا تباين مقداره $S^2=16000$ وأن السحب مع الإعادة نجد أن:

$$n \geq \frac{Z^2_{1-\frac{a}{2}} S^2}{d^2} = \frac{(1.96)^2 .16000}{(20)^2} \approx 154$$ أسرة

أن يكفي 154 أسرة لتحقيق الدقة المطلوبة.

حجم العينة اللازم لتقدير النسبة في مجتمع طبيعي: عندما نريد تقدير نسبة خاصة في مجتمع R وبدقة d وباحتمال مقداره β وبحسب حالة السحب نجد أن:

في حالة السحب مع الإعادة:
$$n \geq \frac{Z^2_{1-\frac{a}{2}} .r.q}{\delta^2}$$

في حالة السحب بدون إعادة:
$$n \geq \frac{NZ^2_{1-\frac{a}{2}} .r.q}{Nd + Z^2_{1-\frac{a}{2}} .r.q}$$

مثال تطبيقي:

نريد تقدير تكرار انتشار مرض معين في مجتمع ما مع العلم أن تكرار هذا المرض بحدود10% والدقة المطلوبة 2% وباحتمال مقداره95%. فما هو حجم العينة اللازم ؟

وبتطبيق الصيغة التالية نجد أن:

$$n \geq \frac{NZ^2_{1-\frac{a}{2}}.r.q}{\delta^2} = \frac{(1.96)^2(0.9)(0.10)}{(0.02)^2} = 865$$
شخصاً

ملاحظة هامة:

إن حجم العينة الذي نحصل عليه هو الحد الأدنى لحجم العينة اللازم ويمكن أن يكون حجمها أكبر من ذلك ونشير هنا إلى أن حجم العينة لا بد و أن يكون عدداً صحيحاً لذلك يجب تقريبه إلى الأعلى دائماً.

تمارين غير محلولة

1. **1 ـ لتكن لدينا القياسات التالية: درجة حرارة 15 مريض في أحد المشافي موزعة على النحو التالي: 37.3.; 37.2; 37.8; 37.1; 37; 36.7; 37.7; 38.5; 38.9; 36.2 38.1; 38.5; 39; 36.8; 37.7;**
2. **ـفما هو تقدير متوسط درجة الحرارة ثم أوجد مجال الثقة لمتوسط درجة الحرارة باحتمال مقداره 95%.**
3. **2 ـ سحبت عينة مؤلفة من n=50 من طلاب كلية الصيدلية فكان متوسط علاماتهم في مقرر الإحصاء الحيوي $\bar{X}$ =55 علامة وبتباين مقداره $S^2=100$. فما هو مجال الثقة الذي يحوي متوسط علامات الطلاب وذلك باحتمال مقداره99%.**
4. **3 ـ عينة مؤلفة من 150 لمبة ميكروسكوب من النوعية A متوسط عمر اللمبة 1400 ساعة وانحراف معياري مقداره 120 ساعة وعينة أخرى مؤلفة من 100 لمبة من النوعية B متوسط عمر اللمبة 1200 ساعة وانحراف معياري مقداره 80 ساعة.**
5. **أوجد مجال للثقة بين متوسطي عمر اللمبات للنوعين B,A باحتمال مقداره 95%.**
6. **4 ـ لتقدير نسبة الأشخاص المصابين بمرض القلب، سحبنا عينة بحجم n=100 شخص من مجتمع مؤلف من 6000 شخص فوجدنا أن 40 منهم مصاب بمرض القلب. أوجد تقدير لنسبة المصابين بمرض القلب ثم حدد مجال الثقة باحتمال 99% لنسبة المصابين بمرض القلب.**
7. **5 ـ اخترنا عينتين عشوائيتين مستقلتين من مجتمعي البالغين والمراهقين لمعرفة رأيهم في برنامج تلفزيوني ما. العينة الأولى حجمها $n_1=400$ بالغ والثانية $n_2=600$ مراهق فتبين أن 100 من الراشدين و 300 فرد من عينة المراهقين أيدوا البرنامج المذكور. أوجد مجال الثقة باحتمال 95% للفرق بين نسبتي الراشدين والمراهقين الذين يتابعون البرنامج المشار إليه.**
8. **6 ـ يمكن أن يأخذ رد فعل شخص لمنشط في تجربة نفسية أحد شكلين A وB. فإذا رغب المجرب في تقدير الاحتمال P بأن رد الفعل سيكون من النوع. فكم شخصاً يجب أن تشتمل التجربة. ولنفرض أننا سنرضى بخطأ تقدير لا يتجاوز % سا وذلك باحتمال يبلغ 90% كما أننا نتوقع أن تكون النسبة في مكان ما إلى جوار 0.6.**

9. 7- أوجد حجم العينة اللازمة سحبها لتقدير نسبة المدخنين بدقة 10% وباحتمال مقداره 99% علماً بأن السحب سيجري مع الإعادة وأن عينة اختيارية سريعة أعطتنا أن نسبتهم 4%.