٥

المعاينة Sampling

١- مقدمة

٢- الخطوات الرئيسية التى يجب إتباعها فى المعاينة

٣- طرق المعاينة

٤- المعاينة العشوائية البسيطة

٥- مصادر الخطأ فى المعاينة

٥-١ مقدمة

يهتم علم الإحصاء بدراسة العشائر بما تتسم به من معالم بغرض الدراسة أو التخطيط كتقدير حجم القوى العاملة والبطالة أو الإنتاج الزراعى أو الصناعى أو استهلاك الكهرباء أو الغاز أو التركيب العمرى للسكان أو نسبة عدد الذكور إلى الإناث فى كل مرحلة عمرية وغيرها الكثير. ويمكن الحصول على تلك المعلومات عن طريق:

١- الحصر الشامل كإجراء حصر شامل أو تعداد لمجتمع سكان مصر (مثلا) فى توقيت معين أو عند الرغبة فى الحصول على معلومات حول العديد من التقسيمات الفرعية فى العشيرة حيث يكون حجم العينة كبيراً وضخماً بحيث يمثل الحصر الشامل أفضل حل أو أن يكون الباحث ليس على علم تام بطبيعة مفردات العشيرة التى يرغب فى دراستها.

٢- سحب عينة ما بطريقة ما (تسمى تقنية المعاينة sampling technique) كأن تؤخذ عينة من دم مريض لفحصها حيث إن سحب كل دم المريض يؤدى إلى الوفاة، أى أن استخدام العينة فى مثل تلك الحالة ضرورة.

٣ - الحصر الشامل والعينة معاً حيث يتيح استخدام البيانات والنتائج المتحصل عليها مقارنة نتائج الحصر التام بالعينة وقياس مدى دقة الأخيرة.

والعينة لها فوائد كثيرة منها:

١ - أنها أقل تكلفة حيث يستخدم جزء صغير من العشيرة المطلوب دراستها ولكنه ممثل لها.

٢ - يستغرق الحصول على البيانات المطلوبة عن طريق العينة وقتاً أقل مقارنة بالحصر الشامل وخاصة عندما يكون هناك ضرورة ملحة فى سرعة الحصول على البيانات لكى يتسنى السرعة فى اتخاذ القرار.

٣ - يتطلب الحصول على بيانات تفصيلية، وقد تكون معقدة بعض الشىء، إلى استخدام عدادين مدربين جيداً قد لا تكفى الأعداد المتاحة منهم لتغطية العشيرة ككل.

٤ - ونظراً لأنه يمكن استخدام أفراد عددهم أقل فى حالة العينات عنه فى حالة التعداد التام فيمكن تدريبهم التدريب الجيد وبذلك فإن كفاءتهم تكون أعلى وبالتالى تكون البيانات المتحصل عليها أكثر دقة.

والسؤال الذى يتبادر إلى الذهن: ما هو حجم العينة الأمثل optimum sample size الذى يمكن اختياره؟ فمعروف أن كلما زاد حجم العينة أدى ذلك إلى زيادة دقة التقدير accuracy ولكن ذلك يتطلب أيضاً إنفاقاً أكثر ووقتاً أطول، ومن هنا كان لابد من إجراء موازنة بين دقة التقدير والاقتصاد economy عند تحديد حجم العينة.

٥-٢ الخطوات الرئيسية التى يجب إتباعها فى المعاينة

١- تحديد السؤال (أو الأسئلة) المراد الإجابة عليها فيما يتعلق بالعشيرة المراد دراستها.

٢- تحديد العشيرة التى ستسحب منها العينة تحديداً دقيقاً ومعرفة كل مفرداتها بدقة. وقد لا تكون العشيرة الخاضعة للمعاينة study population هى نفس العشيرة المطلوب دراستها target population ولكن تستخدم الأولى لتعميم نتائجها على الثانية، كأخذ عينة من فئران التجارب لإجراء دراسة عليها ثم تعميم نتائج تلك الدراسة على الإنسان مثلاً.

٣- تحديد البيانات المراد جمعها أو قياسها بدقة وقد يكون ذلك بالمقابلة الشخصية أو بالفحص الطبى مثلاً أو بإجراء تجربة أو بتوجيه أسئلة باستخدام التليفون أو البريد أو بعض منها أو كلها. ويجب قبل اختيار العينة تقسيم العشيرة إلى أجزاء تسمى وحدات المعاينة sampling units منفصلة عن بعضها تماماً بحيث تشمل العشيرة بأكملها. وقد تكون الوحدة مصباحاً كهربائياً أو حقلاً أو مساحة معينة من الأرض محددة الأبعاد وهذه القائمة من جميع وحدات المعاينة يطلق عليها "إطار frame".

٤- تحديد حجم العينة الأمثل طبقاً لتكاليف وحدة المعاينة ودرجة الدقة المطلوبة.

٥- إجراء اختبار مبدئى pre-test حيث يمكن عن طريق تحليل مبدئى pre-analysis أن تعدل طريقة توجيه بعض الأسئلة أو حذف بعضها أو إضافة أسئلة أخرى أو وضع شروط أو قيود معينة عند إجراء التجربة أو جمع معلومة معينة.

٦- فى حالة المعاينة من أجل المسح فإنه يجب تنظيم العمل الميدانى وهو كل ما يختص بإدارة الأعمال من تدريب الأشخاص الذين سيقومون بجمع البيانات وكذلك الإشراف عليهم ووضع الخطط اللازمة فى حالة الفشل فى الحصول على معلومة معينة.

٧- استخدام الطرق الإحصائية المناسبة فى عرض وتلخيص وتحليل البيانات المتحصل عليها ومناقشتها وتفسيرها والاستدلال منها على معالم العشيرة المطلوب دراستها.

٨- يمكن استخدام ما سبق ذكره فى ٧- فى تقدير حجم العينة الأمثل فى الدراسات القادمة كما يمكن التعرف على المشاكل التى واجهت العدادين للتغلب عليها وإيجاد الحلول اللازمة لها وأيضاً الاستفادة منها فى تيسير العمل وبدقة وفى أقل فترة زمنية.

والإخلال بأى من الخطوات الرئيسية السابقة يؤدى إلى الحصول على بيانات غير دقيقة وبالتالى استنتاجات واستدلالات ليست سليمة بالضرورة.

صندوق ٥-١

المعاينة هى وسيلة عملية لدراسة العشائر لعدم إمكان قياس جميع أفراد العشيرة و/أو الاقتصاد فى التكلفة. فالعينة تؤخذ بطريقة محددة ولابد أن يتوافر فيها صفات معينة (سيتم شرح هذه الطرق والشروط فى الفصول التالية) حتى يمكن سحب التقديرات والقرارات المستمدة منها على العشيرة ككل فهي "منه وإليه"

٥-٣ طرق المعاينة Sampling methods

٥-٣-١ المعاينة الاحتمالية Probability sampling

يجرى السحب بتحديد احتمالات الاختيار لكل من الوحدات ثم سحب هذه الوحدات واحدة تلو الأخرى أو فى مجموعات حتى تستكمل العينة. قد تكون السحبات المتتالية فى العينة الاحتمالية مع إعادة الوحدات المختارة فى السحبات السابقة sampling with replacement أو بدون إرجاع sampling without replacement. وحيث إن المعاينة الاحتمالية تخضع لقوانين الاحتمالات وعليه فإنه يمكن حساب أخطاء المعاينة sampling errors وبالتالى مقدار الدقة المتوقعة فى التقديرات.

٥-٣-٢ المعاينة غير الاحتمالية

اختيار عينة بطريقة غير عشوائية. ومن أمثلتها: العينة التى تتكون مفرداتها من متطوعين لاختبار دواء معين فقد لا يوافق الأفراد الذين سيقع عليهم الاختيار، فى حالة

المعاينة الاحتمالية، لإجراء مثل ذلك الاختبار. وفى المعاينة غير الاحتمالية لا يستطيع الباحث قياس دقة النتائج المتحصل عليها حيث إنها لا تخضع لقوانين الاحتمالات كما أنها تؤدى فى الغالب إلى التحيز.

٥-٤ المعاينة العشوائية البسيطة Simple random sampling

هى إحدى صور المعاينة الاحتمالية. فإذا كان حجم العشيرة N ويراد سحب عينة عشوائية بسيطة حجمها n بدون إحلال فإنه يتم سحب وحدات تلك العينة واحدة تلو الأخرى، ويكون احتمال اختيار وحدة من الوحدات من السحب الأول مساوية n/N واحتمال اختيار الوحدة الثانية $(n-1)/(N-1)$ والثالثة $(n-2)/(N-2)$ ... وهكذا، والوحدة الأخيرة فى العينة $1/(N-n+1)$، مع مراعاة أن السحب يتم بدون إحلال without replacement أى أن وحدة المعاينة التى يتم سحبها لا تعاد مرة أخرى إلى العشيرة المسحوبة منها، وبالتالى فإن احتمال سحب وحدات تلك العينة والتى حجمها n من n سحبه بدون إحلال هو

$$(n/N)\cdot[(n-1)/(N-1)]\cdots[1/(N-n+1)] = 1/C_n^N \qquad (٥-١)$$

حيث: $C_n^N = \dfrac{N!}{(N-n)!n!}$ هى عدد التوافيق الممكنة لسحب n من وحدات المعاينة من عشيرة حجمها N.

وبالتالى فإن احتمال سحب أى عينة حجمها n من عشيرة حجمها N هو مقلوب عدد العينات المتساوية فى الحجم والتى لها أيضاً نفس الفرصة فى الاختيار. ويتم سحب مفردات العينة بأن ترقم العشيرة التى سيسحب منها العينة $1,2,\cdots,N$ ثم يتم سحب المفردات التى تكون تلك العينة إما بواسطة جداول الأعداد العشوائية أو عن طريق بعض البرامج الإحصائية باستخدام الحاسبات الآلية أو حتى تلك المتوفرة فى بعض الحاسبات.

ولتوضيح ذلك افترض وجود عشيرة مكونة من 6 وحدات معاينة يرمز لها a, b, c, d, e & f والمطلوب سحب عينة حجمها = 2 من هذه العشيرة فتكون كل العينات الممكن سحبها والتى لها نفس الحجم ونفس الاحتمال:

ab, ac, ad, ae, af, bc, bd, be, bf, cd, ce, cf, de, df, ef

أى عدد العينات الممكن سحبها هو $C_2^6 = \dfrac{6!}{2!\,4!} = 15$

وبالتالى فإن احتمال سحب أى عينة هو 1/15، وكل وحدة معاينة تتكرر عدداً متساوياً من المرات فى العينات وهو 5 مرات فى هذه الحالة ويعبر عنه رياضياً بما يلى:

$$C_{n-1}^{N-1} = \frac{(N-1)!}{[(N-n)!(n-1)!]} \qquad (٥-٢)$$

وفى المثال السابق فإن عدد المرات التى يتكرر فيها أى حرف $_{(6-1)}C_{(2-1)} = 5$.

وقد يتم سحب وحدات المعاينة مع الإحلال وفى هذه الحالة فإن احتمال سحب أى مفردة يكون متساويا ومقداره n / N، وفى هذه الحالة قد تتكرر نفس المفردة وهو أمر قد يكون غير منطقياً مع ملاحظة أن التقديرات وتبايناتها والتى تقدر من العينة تكون أبسط فى حالة المعاينة مع الإحلال عنها فى حالة المعاينة بدون إحلال.

٥-٤-١ اختيار عينة عشوائية بسيطة

يوجد العديد من جداول الأعداد العشوائية والمتوفرة فى كثير من كتب الإحصاء وهى عبارة عن الأرقام $0,1,2,\cdots,9$ موضوعة بطريقة عشوائية. ويمكن اختيار عينة عشوائية بسيطة باستخدام هذه الجداول بإحدى طريقتين وذلك بعد ترقيم وحدات المعاينة فى العشيرة هكذا: $1,2,\cdots,N$.

الطريقة الأولى:

إذا كان حجم العشيرة N وليكن 530 والمطلوب اختيار عينة عشوائية حجمها n ولتكن 10 وحدات. يتم اختيار إحدى الصفحات عشوائياً من جدول الأعداد العشوائية (جدول ١ ملحق أ) وتختار نقطة البداية فى الصفحة المختارة عشوائياً بأن يوضع الإصبع على أى نقطة للبدء منها. وحيث إن حجم العشيرة مكون من 3 أرقام فيتم اختيار 3 أعمدة ولتكن نقطة البداية السطر 00 والأعمدة 60,61,62 فتكون لدينا الأعداد من 000 إلى 999 وبالتالى تحذف كل الأعداد التى هى أكبر من 530 وأيضاً العدد 000. يتم استعراض أرقام الأعمدة الثلاثة من أعلى إلى أسفل ويتم اختيار الأعداد العشرة التى تمثل العينة وإن لم تستكمل الأعداد العشرة ممكن استخدام الأعمدة التالية للأعمدة السابقة وهى الأعمدة 65,66,67 وهكذا من أعلى إلى أسفل أيضاً. وإذا تكرر أى عدد فإنه يحذف.

وفي هذه الطريقة عند اختيار عينات مختلفة يجب اختيار نقط بداية جديدة عشوائيا. وبإتباع هذه الطريقة فإن مفردات العينة والتى حجمها 10 هى:

520, 514, 368, 359, 356, 512, 012, 104, 344 & 091

وتكون الأعداد الثلاثية المرفوضة 14 من جملة الأعداد الثلاثية 24 أى أن نسبة الرفض 58% تقريباً.

الطريقة الثانية:

تفضل عندما يكون حجم العشيرة أقل من 500 وليكن 125 مثلاً. تختار عشوائياً نقطة البداية وليكن الصف 00 وثلاثة أعمدة ولتكن 50,51,52. فى هذه الطريقة تحذف الأعداد 000 ويطرح 200 من جميع الأعداد ما بين 201 و 400، ويطرح 400 من جميع الأعداد والتى تقع بين 401 و 600، ويطرح 600 من جميع الأعداد بين 601 و 800، ويطرح 800 من جميع الأعداد من 801 و 999، ويتم رفض أى عدد أو باق أكبر من 125 وأى عدد يتكرر أكثر من مرة.

وبالتالى تكون مفردات العينة المسحوبة هى:

103, 068, 112, 024, 116, 116, 005, 025, 003, 002 & 059

ويلاحظ رفض العدد 116 لتكراره وأن نسبة الرفض فى هذه العينة 13/23 أى 57%.

وعند استخدام الطريقة الثانية فى حالة $N = 320$ (مثلا) يلاحظ طرح 400 من كل عدد بين 401 و 800 ورفض جميع الأعداد التى هى أكبر من 800 وأيضاً العدد 800 وكل البواقى التى هى أكبر من 320. وطرح 800 بين 801 و 999 حتى لا يعطى احتمال قبول أعلى للبواقى بين 001 و 199 مما يعطيه للبواقى بين 200 و320.

وإذا استخدمت الطريقة الأولى فى اختيار العينة والتى حجمها 10 من عشيرة حجمها 125 ولو فرضت نفس نقطة البداية السابقة مع حذف كل الأعداد التى هى أكبر من 125 وأيضاً 000 فإن مفردات العينة تكون:

103, 112, 116, 124, 056, 107, 026, 099, 065, 012

وتطلب السحب 90 رقما ثلاثياً لسحب 10 أرقام ثلاثية أى نسبة الرفض بلغت 80/90 أى حوالى 88% وهى أكبر بكثير مقارنة بالطريقة الثانية لسحب العينة من نفس العشيرة ويلاحظ هنا استخدام الأعمدة 50,51,52 والأعمدة 55,56,57 وأيضاً الأعمدة 60,61,62 وكانت نقطة البداية السطر 00.

٥-٤-٢ تقدير معالم العشيرة

يرمز للقيم التى تأخذها وحدات العشيرة بالرموز $Y_1, Y_2, \cdots, Y_N$ حيث حجم العشيرة N والوحدات التى تأخذها العينة بالرموز $Y_1, Y_2, \cdots, Y_n$ حيث n حجم العينة ويلاحظ أن Y_1 فى العينة ليست بالضرورة هى Y_1 فى العشيرة وقد تكون مجرد صدفة.

وفيما يلى أهم المعالم التى يتم تقديرها وسوف يرمز لمعالم العشيرة بالحروف الكبيرة capital letters بينما التقديرات من العينة بالحروف الكبيرة أيضاً مضافاً عليها علامة (^).

	من العينة	من العشيرة
المتوسط	$\hat{\bar{Y}} = \sum_{i=1}^{n} Y_i / n = Y./n$	$\bar{Y} = \sum_{i=1}^{N} Y_i / N = Y./N$
المجموع الكلى للعشيرة	$\hat{Y}. = N\hat{\bar{Y}}$ $= N(\sum Y_i / n)$	$Y. = Y_1 + Y_2 + \ldots + Y_N$ $= \sum Y_i$
التباين (متغير مستمر)	$\hat{S}^2 = \sum (Y_i - \hat{\bar{Y}})/(n-1)$ تقدير غير متحيز لـ S^2	$\sigma^2 = \sum (Y_i - \bar{Y})^2 / N$ $S^2 = \sum (Y_i - \bar{Y})^2 / (N-1)$

تباين المتوسط لعينة عشوائية بسيطة

التباين المقدر من العينة

$$S^2_{\hat{\bar{Y}}} = \frac{\hat{S}^2}{n}\left(\frac{N-n}{N}\right) = \frac{\hat{S}^2}{n}(1-f) \qquad (٥-٣)$$

التباين الحقيقى (الفعلى)

$$\sigma^2_{\bar{Y}} = \frac{S^2}{n}\left(\frac{N-n}{N}\right) = \frac{S^2}{n}(1-f) \qquad (٥-٤)$$

الإثبات:

حيث أن

$$n(\hat{\bar{Y}}-\bar{Y})=(Y_1-\bar{Y})+(Y_2-\bar{Y})+\cdots+(Y_n-\bar{Y}) \qquad (١)$$

$$E(Y_1+Y_2+\cdots+Y_n)=\frac{n}{N}(Y_1+Y_2+\cdots+Y_N) \qquad (٢)$$

إذاً

$$E[(Y_1-\bar{Y})^2+\cdots+(Y_n-\bar{Y})^2]=\frac{n}{N}[(Y_1-\bar{Y})^2+\cdots+(Y_N-\bar{Y})^2] \qquad (٣)$$

كما أن

$$\begin{aligned}E[(Y_1-\bar{Y})(Y_2-\bar{Y})+(Y_1-\bar{Y})(Y_3-\bar{Y})+\cdots+(Y_{n-1}-\bar{Y})(Y_n-\bar{Y})]\\ =\frac{n(n-1)}{N(N-1)}[(Y_1-\bar{Y})(Y_2-\bar{Y})+(Y_1-\bar{Y})(Y_3-\bar{Y})\\ +\cdots+(Y_{N-1}-\bar{Y})(Y_N-\bar{Y})] \qquad (٤)\end{aligned}$$

ويلاحظ أن المجموع يحتوى على عدد $\frac{n(n-1)}{2}$ من الحدود على اليسار وعلى $\frac{N(N-1)}{2}$ من الحدود على اليمين. وبتربيع المعادلة (١) واستخدام (٣) و (٤) فإن

$$(n^2)E(\hat{\bar{Y}}-\bar{Y})^2=\frac{n}{N}\left\{\begin{array}{l}(Y_1-\bar{Y})^2+\cdots+(Y_N-\bar{Y})^2\\ +\left[\begin{array}{l}\frac{2(n-1)}{N-1}[(Y_1-\bar{Y})(Y_2-\bar{Y})+\cdots\\ +(Y_{N-1}-\bar{Y})(Y_N-\bar{Y})]\end{array}\right]\end{array}\right\}$$

وبإتمام المربعات فإن:

$$(n^2)E(\hat{\bar{Y}}-\bar{Y})^2=\frac{n}{N}\left\{\begin{array}{l}\left(1-\frac{n-1}{N-1}\right)[(Y_1-\bar{Y})^2+\cdots+(Y_N-\bar{Y})^2]\\ +\left(\frac{n-1}{N-1}\right)[(Y_1-\bar{Y})+\cdots+(Y_N-\bar{Y})]^2\end{array}\right\}$$

ويلاحظ أن الحد الثانى والذى بداخل القوس المربع يساوى صفر، وبقسمة الطرفين على n^2 فإن

$$V(\hat{\bar{Y}}) = E(\hat{\bar{Y}} - \bar{Y})^2 = \left(\frac{N-n}{nN(N-1)}\right)\left(\sum_{i=1}^{N}(Y_i - \bar{Y})^2\right)$$

$$= \frac{S^2}{n}\left(\frac{N-n}{N}\right) = \frac{S^2}{n}(1-f)$$

وبالتالى فإن الخطأ المعيارى لـــ $\hat{\bar{Y}}$

$$\sigma_{\hat{\bar{Y}}} = \frac{S}{\sqrt{n}}\sqrt{(1-f)} \qquad (٥-٥)$$

وتباين $\hat{Y}.$

$$V(\hat{Y}.) = V(N\hat{\bar{Y}}) = \frac{N^2S^2}{n}(1-f) \qquad (٥-٦)$$

وبالتالى فإن الخطأ المعيارى لـــ $\hat{Y}.$

$$\sigma_{\hat{Y}.} = \frac{NS}{\sqrt{n}}\sqrt{(1-f)} \qquad (٥-٧)$$

ملحوظة: فى حالة تقدير التباين من العينة يستخدم $\hat{S}^2$ بدلاً من S^2

ويلاحظ أن الكمية $(N-n)/N$ تعرف بمعامل التصحيح للعشيرة المحدودة finite population correction factor والتى يمكن إعادة كتابتها على الصورة $1-(\frac{n}{N}) = 1-f$ حيث يطلق على $f = \frac{n}{N}$ كسر المعاينة sampling fraction ويمكن إهمال كسر المعاينة إذا كان أقل من 5% . ويشير بعض الإحصائيين إلى أنه يمكن إهمـــال كســـــر المعاينة حتى وإن وصل إلى 10% .

٥-٤-٣ خصائص التقديرات

كما ذكر من قبل فإن الهدف من المعاينة هو تقدير معالم العشيرة من العينة المسحوبة، وحتى توفى العينة بهذا الهدف يجب أن تتوافر فى هذه التقديرات خصائص معينة هى:

١- عدم التحيز unbiasdness
٢- الاتساق consistency
٣- أقل تباين minimum variance
٤- الكفاية sufficiency

وتعتمد دقة أى تقدير من العينة على الطريقة التى يحسب بها هذا التقدير من العينة وأيضاً على خطة المعاينة وهو ما يطلق عليه دقة المعاينة العشوائية البسيطة. وفيما يلى أهم تلك الخصائص للمتوسط الحسابى.

١ - عدم التحيز Unbiasedness

وهذا يعنى أن التقدير المحسوب من العينة $\hat{\bar{Y}}$ يكون غير متحيز لمعلم العشيرة $\bar{Y}$. ويمكن توضيح ذلك كالتالى:

حيث إن

$$E(\hat{\bar{Y}}) = \frac{\sum \hat{\bar{Y}}}{{}_N C_n} = \frac{\sum (Y_1 + Y_2 + \cdots + Y_n)}{(n) \dfrac{N!}{n!(N-n)!}}$$

وحيث إن كل Y_i تظهر فى عدد من العينات مقداره ${}_{N-1}C_{n-1}$ من معادلة (٥-٢) وبالتالى فإن:

$$\begin{aligned}\sum (Y_1 + Y_2 + \cdots + Y_n) &= {}_{N-1}C_{n-1} \, (Y_1 + Y_2 + \cdots + Y_N) \\ &= \frac{(N-1)!}{(n-1)!(N-n)!} (Y_1 + Y_2 + \cdots + Y_N)\end{aligned}$$

$$\begin{aligned}E(\hat{\bar{Y}}) &= \left(\frac{(N-1)!}{(n-1)!(N-n)!} \right) \left(\frac{n!(N-n)!}{n \, N!} \right) (Y_1 + Y_2 + \cdots + Y_N) \\ &= \frac{Y_1 + Y_2 + \cdots + Y_N}{N} = \bar{Y}\end{aligned}$$

أى أن $\hat{\bar{Y}}$ هو تقدير غير متحيز لمتوسط العشيرة. ومنه فإن $\hat{Y}. = N\hat{\bar{Y}}$ هو تقدير غير متحيز لمجموع العشيرة $Y.$

٢ - الاتساق Consistency

وهذا يعنى أن $\hat{\bar{Y}}$ و $N\hat{\bar{Y}}$ هما تقديرين متسقين لمتوسط ومجموع العشيرة على الترتيب. فالمتوسط $\hat{\bar{Y}}$ يقال إنه تقدير متسق consistent لمتوسط العشيرة $\bar{Y}$ إذا كان احتمال القيمة المطلقة للفرق بين القيمة الحقيقية والتقدير أكبر من قيمة معينة ولتكن ε

مهما تكن يؤول إلى الصفر كلما آلت n إلى مالا نهاية $n \to \infty$ ويمكن التعبير عن ذلك رياضياً بالصورة $\Pr(\hat{\bar{Y}} - \bar{Y}) > \varepsilon) \to 0 \text{ as } n \to \infty$. ويمكن إثبات ذلك باستخدام لا متساوية تشيبتشف Tchebychef 's inequality.

٣ - الكفاءة Minimum variance

تدل على أن المتوسط له أقل تباين وهو أيضاً يتحقق.

٤ - الكفاية Sufficiency

يقال أن المتوسط كاف sufficient statistic إذا كان يأخذ فى الاعتبار عند حسابه كل قيم وحدات المعاينة وهذا أيضاً يتحقق.

مثال ٥-١

فى عشرة من 5 وحدات كانت قيم Y_i هى 13,9,6,3,5 . احسب متوسط العينة فى كل من العينات العشوائية البسيطة الممكنة والتى حجم كل منها = 2 . ثم تحقق من أن $\hat{\bar{Y}}$ هو تقدير غير متحيز لمتوسط العشيرة. احسب التباين لكل من العينات وتحقق من أن $E(\hat{S}^2) = S^2$.

عدد العينات التى حجم كل منها 2 هو $_5C_2 = \frac{5!}{3!2!} = 10$

احتمال سحب أى عينة عشوائية $\frac{1}{_5C_2} = 1 \div \frac{5!}{3!2!} = \frac{1}{10}$

ويبين جدول ٥-١ العينات العشوائية البسيطة الممكنة وتباينها. وباستخدام بيانات هذا الجدول فإن:

المتوسط العام لمتوسطات العينات: $\frac{4 + 5.5 + \cdots + 11}{10} = 7.2$

وهو يساوى متوسط العشيرة 7.2

وبالتالى فإن $\hat{\bar{Y}}$ هو تقدير غير متحيز لمتوسط العشيرة.

تباين العشيرة:

$$S^2 = \sum_{c=1}^{5}(Y_i - \overline{Y})^2/(N-1)$$

$$= \frac{(5-7.2)^2 + (3-7.2)^2 + \cdots + (13-7.2)^2}{4} = 15.2$$

$$E(\hat{S}^2) = \frac{2 + 0.5 + \cdots + 8}{10} = 15.2$$

وبالتالى فإن $E(\hat{S}^2) = S^2$

جدول ٥-١ العينات العشوائية البسيطة الممكنة وتباينها لمثال ٥-١

العينة	المتوسط $\hat{\overline{Y}}$	التباين $\hat{S}^2 = \sum(Y_i - \hat{\overline{Y}})/(n-1)$
3, 5	4.0	2.0
6 , 5	5.5	0.5
9 , 5	7.0	8.0
13, 5	9.0	32.0
6, 3	4.5	4.5
9, 3	6.0	18.5
13, 3	8.0	50.0
9, 6	7.5	4.5
13, 6	9.5	24.5
13, 9	11.0	8.0

مثال ٥-٢

البيانات التالية تمثل الدخل الشهرى بالجنيه للعامل لعينة عشوائية بسيطة حجمها 30 مسحوبة من مصنع به 250 عامل

328	326	337	330	330	333	339	329	331	349
336	338	324	333	340	334	341	329	329	340
344	335	345	332	326	335	336	342	335	337

والمطلوب تقدير لمتوسط العشيرة وقيمته الكلية ثم حساب تباين المتوسط والخطأ المعيارى له وللقيمة الكلية.

إذا رمز للدخل الشهرى للعامل رقم i بـ Y_i

فإن $$\hat{\bar{Y}} = \sum_{i=1}^{30} Y_i / n = (328+339+\cdots+337)/30 = 10043/30 = 334.77 \text{ LE}$$

والتباين $$\hat{S}^2 = \frac{1}{n-1}\left[\sum Y_i^2 - \frac{(\sum Y_i)^2}{n}\right] = 36.87$$

وتباين المتوسط $$\hat{S}^2_{\bar{Y}} = \frac{\hat{S}^2}{n}\left(\frac{N-n}{N}\right) = \frac{36.87}{30}\left(\frac{250-30}{250}\right) = 1.08$$

الخطأ المعيارى للمتوسط $$\hat{S}_{\bar{Y}} = \sqrt{1.08} = 1.04 \text{ LE}$$

مجموع الدخل الشهرى للعمال فى المصنع $$N\hat{\bar{Y}} = 250 \text{ x } 334.77 = 83693 \text{ LE}$$

تباين القيمة الكلية:

$$\hat{S}^2_{\hat{Y}.} = \frac{N^2S^2}{n}\left(\frac{N-n}{N}\right) = \frac{(250)^2(36.87)^2(250-30)}{(30)(250)} = 67594.998$$

الخطأ المعيارى للقيمة الكلية $$\hat{S}_{\hat{Y}.} = \sqrt{67594.998} = 259.99 \text{ LE}$$

٥ – ٤ – ٤ حدود الثقة Confidence limits

لحساب حدى الثقة لمتغير ما يجب معرفة توزيع هذا المتغير. والكثير من المتغيرات التى نتعامل معها تتبع التوزيع الطبيعى أو على الأقل تميل أو تقترب من هذا التوزيع حيث إن التقريب الطبيعى ملائم لمعظم التطبيقات العملية من مصادر عدة. وقد وجد من الدراسات الخاصة بنظرية الاحتمالات أنه كلما زاد حجم العينة اقترب توزيع متوسطات العينات المسحوبة من هذه العشيرة إلى أن يكون طبيعياً وذلك يتحقق فى العشائر اللانهائية والتى لها انحراف معيارى منته. أما فى حالة العشائر المحددة فقد يحتاج الأمر إلى حجم أكبر للعينة يخضع لشروط لازمة وكافية والتى ينتهى تحتها توزيع متوسط عينة إلى التوزيع الطبيعى (Hajek, 1960). ويفشل التقريب الطبيعى عندما تحتوى العشيرة على وحدات متطرفة حيث تؤثر أيضاً على زيادة تباين العينة وانخفاض الدقة، وفى هذه الحالة قد يكون أخذ تعداد كامل أكثر حكمة أو عزل تلك الوحدات الشاذة إذا كان عددها ليس بالكثير، وقد يخفض إزاحة هذه

الوحدات المتطرفة من الالتواء ويقترب بالتوزيع من التوزيع الطبيعى. كما يمكن اللجوء إلى المعاينة الطبقية stratified sampling.

وعلى العموم فإن حدود الثقة لمتوسط العشيرة ومجموعها تكون كما يلى:

للمتوسط:

$$\hat{\bar{Y}} \pm \frac{t\,\hat{S}}{\sqrt{n}}\sqrt{1-f} \qquad (٥-٨)$$

وللمجموع:

$$N\hat{\bar{Y}} \pm \frac{t\,N\,\hat{S}}{\sqrt{n}}\sqrt{1-f} \qquad (٥-٩)$$

حيث t هى قيمة t الجدولية بمستوى معنوية معين وليكن α ودرجات حرية $n-1$ والتى حسب منها التباين.

مثال ٥-٣

فى المثال ٥-٢ احسب مدى الثقة لمتوسط دخل العامل فى المصنع عند مستوى معنوية $\alpha = 0.05$.

بتطبيق المعادلة (٥-٨) وحيث إن قيمة $t_{(.05,29)} = 2.045$ (جدول ٤ ملحق أ) فإن حدا الثقة هما:

$$334.77 \pm 2.045\,(1.04) = 334.77 \pm 2.13$$

وبالتالى فإن الحد الأعلى للثقة هو 336.9 جنيهاً و الحد الأدنى للثقة هو 332.6 جنيهاً

٥-٤-٥ تقدير نسبة (نسبة مجموعتين أو متوسطين)

قد تكون الكمية المراد تقديرها من العينة العشوائية البسيطة هى نسبة بين متغيرين يتغير كل منهما من وحدة معاينة إلى أخرى. ففى الإحصاءات المنزلية قد يكون من المفيد على سبيل المثال معرفة متوسط عدد أجهزة التكييف لكل شقة أو متوسط عدد الأطفال الذكور فى كل أسرة. ولكى تحصى هذه المفردات يمكن أن تسجل فى الشقة رقم i (حيث $i = 1, 2, \cdots, n$): عدد أجهزة التكييف (Y) وأيضاً عدد الحجرات (X) (مع اعتبار أن الصالة حجرة من الحجرات) وبالتالى يكون:

متوسط العشيرة:

$$R = \frac{\sum_{i=1}^{N} Y_i}{\sum_{i=1}^{N} X_i} \qquad (٥-١٠)$$

والتقدير من العينة:

$$\hat{R} = \frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} X_i} = \frac{\hat{\bar{Y}}}{\hat{\bar{X}}} \qquad (٥-١١)$$

وتظهر النسب أيضاً فى العديد من التطبيقات الأخرى فمثلا نسبة المساحة المزروعة بالذرة الشامية إلى مجموع المساحة فى المزرعة أو نسبة عدد الذكور إلى عدد الإناث فى الأقسام المختلفة فى إحدى الكليات الجامعية. ويلاحظ أن كلاً من البسط والمقام فى هذه الحالات متغير من عينة إلى أخرى وهذا يؤدى إلى أن يكون التوزيع الاحتمالى أكثر تعقيداً. وفي العينات ذات الحجم الصغير نسبياً يلاحظ أن $\hat{R}$ غير متناظر ويعد تقديراً منحازاً لـ R (عند تطبيق خصائص التقدير الجيد). لكن فى العينات كبيرة الحجم فإنه يمكن اعتبار أن R تتوزع طبيعياً.

تباين $\hat{R}$:

$$V(\hat{R}) \cong \left(\frac{1-f}{n\bar{X}^2}\right)\left(\frac{\sum_{i=1}^{N}(Y_i - RX_i)^2}{N-1}\right) \qquad (٥-١٢)$$

حيث $R = \bar{Y}/\bar{X}$ هى نسبة متوسطى العشيرة و $f = n/N$

وفى حالة استخدام العينة فإن:

$\frac{\sum_{i=1}^{n}(Y_i - \hat{R}X_i)^2}{n-1}$ يعتبر تقديراً للمقدار $\frac{\sum_{i=1}^{N}(Y_i - RX_i)^2}{N-1}$

والخطأ المعياري لـ $\hat{R}$

$$S(\hat{R}) = \left(\frac{\sqrt{1-f}}{(\sqrt{n})(\overline{X})}\right)\sqrt{\frac{\sum_{i=1}^{n}(Y_i - \hat{R}X_i)^2}{n-1}} \qquad (١٣-٥)$$

وإذا كان $\overline{X}$ غير معروف يعوض عنه بتقدير من العينة $\hat{\overline{X}}$

كما يمكن حساب $S(\hat{R})$ كما يلي:

$$S(\hat{R}) = \left(\frac{\sqrt{1-f}}{(\sqrt{n})(\overline{X})}\right)\sqrt{\frac{\sum Y_i^2 - 2\hat{R}\sum Y_i X_i + \hat{R}^2 \sum X_i^2}{n-1}} \qquad (١٤-٥)$$

مثال ٥-٤

الجدول التالى يبين عدد كل من الأبناء الذكور والإناث لعدد 20 أسرة من العاملين بإحدى الكليات الجامعية اختيروا بطريقة عشوائية من بين 500 أسرة.

رقم الأسرة	عدد أفراد الأسرة m	عدد الذكور X_1	عدد الإناث X_2	رقم الأسرة	عدد أفراد الأسرة m	عدد الذكور X_1	عدد الإناث X_2
١	2	1	1	١١	3	0	3
٢	2	2	0	١٢	3	1	2
٣	4	2	2	١٣	3	0	3
٤	4	1	3	١٤	4	2	2
٥	0	0	0	١٥	4	2	2
٦	3	3	0	١٦	3	2	1
٧	4	3	1	١٧	2	0	2
٨	3	2	1	١٨	5	3	2
٩	5	2	3	١٩	4	2	2
١٠	3	0	3	٢٠	3	1	2

والمطلوب تقدير:

١- متوسط عدد الذكور ومتوسط عدد الإناث للأسرة الواحدة والخطأ المعيارى لهما.

٢- متوسط عدد الأبناء والخطأ المعيارى له بغض النظر عن الجنس للأسرة الواحدة.

٣- النسبة المئوية والخطأ المعيارى للعدد الكلى من الذكور إلى العدد الكلى من الإناث.

من البيانات:

$$n = 20 \qquad \sum X_1 = 29 \qquad \sum X_2 = 35 \qquad \sum X_1 X_2 = 43$$

$$N = 500 \qquad \sum X_1^2 = 63 \qquad \sum X_2^2 = 81$$

١- متوسط عدد الذكور للأسرة الواحدة:

$$\overline{X}_1 = \frac{\sum X_1}{n} = \frac{29}{20} = 1.45$$

الخطأ المعيارى لمتوسط عدد الذكور للأسرة الواحدة:

$$\hat{S}_{\overline{X}_1} = \left(\frac{\sqrt{1-f}}{\sqrt{n}}\right)\sqrt{\frac{\sum X_1^2 - \frac{(\sum X_1)^2}{n}}{n-1}} = \left(\frac{\sqrt{1-\frac{20}{500}}}{\sqrt{20}}\right)\sqrt{\frac{63-\frac{(29)^2}{20}}{19}} = 0.23$$

وإذا أهمل معامل التصحيح للعشيرة المحدودة حيث إنه يمثل $(20/500)(100) = 4\%$ فقط فإن:

$$\hat{S}_{\overline{X}_1} = \frac{1}{\sqrt{n}}\sqrt{\frac{\sum X_1^2 - \frac{(\sum X_1)^2}{n}}{n-1}} = 0.235$$

الفرق بين التقديرين 0.005، أى أنه فى حالة إهمال معامل التصحيح للعشيرة المحدودة فإن الخطأ المعيارى يزيد بمقدار 2.17%.

متوسط عدد الإناث للأسرة الواحدة:

$$\overline{X}_2 = \frac{\sum X_2}{n} = \frac{35}{20} = 1.75$$

الخطأ المعيارى لمتوسط عدد الإناث للأسرة الواحدة:

$$\hat{S}_{\overline{X}_2} = \left(\frac{\sqrt{1-f}}{\sqrt{n}}\right)\sqrt{\frac{\sum X_2^2 - \frac{(\sum X_2)^2}{n}}{n-1}} = \left(\frac{\sqrt{1-\frac{20}{500}}}{\sqrt{20}}\right)\sqrt{\frac{81-\frac{(35)^2}{20}}{19}} = 0.22$$

٢- متوسط عدد الأبناء من الذكور والإناث معاً للأسرة الواحدة:

$$\overline{Y} = \frac{\sum m_i}{n} = \frac{2+2+\cdots+3}{20} = \frac{64}{20} = 3.2$$

تباين متوسط عدد الأبناء

$$\hat{S}^2_{\hat{\overline{Y}},} = [\sum m^2 - \frac{(\sum m)^2}{n}]/(n-1) = [(2)^2 + (2)^2 + \cdots + (3)^2 - \frac{(64)^2}{20}]/19$$
$$= 1.326$$

وعليه فإن الانحراف المعيارى لمتوسط عدد الأبناء

$$\hat{S}_{\hat{\overline{Y}}} = \sqrt{\frac{\hat{S}^2_{\hat{\overline{Y}}}}{n}(1-f)} = \sqrt{\left(\frac{1.326}{20}\right)\left(1-\frac{20}{50}\right)} = 0.25$$

٣- نسبة العدد الكلى من الذكور إلى العدد الكلى من الإناث:

$$\hat{R} = (\sum X_1 / \sum X_2) = 29/35 = 0.829$$

الانحراف المعيارى:

$$S(\hat{R}) = \frac{\sqrt{1-f}}{(\sqrt{n})\,(\overline{X}_2)}\sqrt{\frac{\sum X_1^2 - 2\hat{R}\sum X_1X_2 + \hat{R}^2\sum X_2^2}{n-1}}$$

$$= \frac{\sqrt{0.96}}{(\sqrt{20})(1.75)}\sqrt{\frac{63-2(0.829)(43)+(0.829)^2(81)}{19}} = 0.2$$

٥-٤-٦ معاينة النسب والنسب المئوية

قد يكون الغرض من المعاينة هو التعامل مع متغيرات لا تتبع التوزيعات المستمرة وإنما تنتمى إلى التوزيعات المتقطعة discrete distribution ومثال ذلك تلك المتغيرات التى تأخذ القيم فيها صفراً أو واحداً كتقسيم العشيرة إلى مدخنين وغير مدخنين، وقد تقسم العشيرة إلى عدة فئات A, B, C مثلا وقد يكون الاهتمام بالفئة A فيأخذ المتغير فى هذه الحالة القيمة واحد وفيما عدا ذلك من فئات يأخذ المتغير فيها القيمة صفراً ... وهكذا. وقد يكون المتغير أصلا مستمراً ولكن تم تقسيمه إلى فئات لا تداخل بينها وبالتالى يصبح متقطعاً كتقسيم أعمار العشيرة إلى فئات: أقل من 20 سنة ومن 20 إلى أقل من 30 ... وهكذا. والمطلوب فى مثل تلك الحالات تقدير العدد الكلى أو النسبة المئوية لوحدات فى العشيرة تتسم بخاصية مميزة أو تقع فى فئة معرفة. وفيما يلى مثال يوضح ذلك المفهوم.

افترض أن كل وحدة فى العشيرة تقع فى أحد الفئتين "فـ" أو مكملتها " فـ' " وبالتالى يكون:

عدد الوحدات من الفئة "فـ" فى		نسبة الوحدات من الفئة "فـ" فى	
العينة	العشيرة	العينة	العشيرة
a	A	$\hat{P} = a / n$	$P = A / N$

أى أن $\hat{P}$ تقدير لــ P وتقدير العينة لــ A هو $N\hat{P}$ أى $N(a / n)$، ومثل هذا المتغير يتبع توزيع "ذو الحدين" فى العشائر اللانهائية وقد يتبع التوزيع فوق الهندسى hypergeometric distribution فى العشائر المحدودة. وعلى العموم فإن توزيع "ذو الحدين" قد يؤخذ به تقريبياً.

تباين التقديرات:

إذا كان المتغير Y محل الدراسة يأخذ قيماً إما صفر أو واحد صحيح وبتطبيق ما سبق دراسته عند دراسة المعاينة للمتغيرات المستمرة فإن:

من العشيرة:

$$Y_{.} = \sum_{i=1}^{N} Y_i = A \quad ، \quad \overline{Y} = \frac{\sum_{i=1}^{N} Y_i}{N} = \frac{A}{N} = P$$

ومن العينة:

$$\hat{\bar{Y}} = \frac{\sum_{i=1}^{n} Y_i}{n} = \frac{a}{n} = \hat{P} \quad ، \quad \hat{Y}_. = \sum_{i=1}^{n} Y_i = a$$

التباين من العشيرة:

$$S^2 = \frac{\sum (Y_i - \bar{Y})^2}{N-1} = \frac{\sum_{i=1}^{N} Y_i^2 - N\bar{Y}^2}{N-1} \qquad (٥-١٥)$$

$$= \frac{1}{N-1}(NP - NP^2) = \frac{N}{N-1}P(1-P) = \frac{N}{N-1}PQ$$

التباين من العينة:

$$\hat{S}^2 = \frac{n}{n-1}\hat{P}\hat{Q} \qquad (٥-١٦)$$

تباين $\hat{P}$:

$$S_{\hat{P}}^2 = E(\hat{P} - P)^2 = \frac{S^2}{n}\left(\frac{N-n}{N}\right) = \frac{PQ}{n}\left(\frac{N-n}{N}\right) \qquad (٥-١٧)$$

ومن العينة:

$$\hat{S}_{\hat{P}}^2 = \hat{P}\hat{Q}\frac{N-n}{N(n-1)} \qquad (٥-١٨)$$

وإذا كانت N كبيرة جداً بالنسبة لـ n فإن $\hat{S}_{\hat{P}}^2 = \hat{P}\hat{Q}/(n-1)$ وتباين $\hat{A}$ حيث $\hat{A} = N\hat{P}$ وهو المجموع المقدر لوحدات الفئة "ف" وبالتالى:

$$S_{\hat{A}}^2 = \frac{N^2}{n}\left(\frac{N-n}{N-1}\right)PQ \qquad (٥-١٩)$$

ومن العينة:

$$\hat{S}_{\hat{A}}^2 = \frac{N(N-n)}{n-1}\hat{P}\hat{Q} \qquad (٥-٢٠)$$

ويعتبر $\hat{S}_{\hat{A}}^2$ تقديراً غير متحيز لـ $\hat{S}_{\hat{A}}$

مثال ٥-٥

ينتج مصنع 100 ألف جهاز تليفزيون فى الشهر. وفى عينة من 100 جهاز وجد أن 5 أجهزة منها بها عيوب فى الصناعة. فما هو تقدير العدد الكلى لأجهزة التليفزيون التى بها عيوب فى الصناعة فى الشهر ؟. وما هو الخطأ المعيارى لهذا التقدير؟.

$N = 100000 \qquad n = 100$

نسبة الأجهزة التى بها عيوب فى الصناعة: $\hat{P} = a/n = 5/100 = 0.05$

تقدير العدد الكلى للأجهزة التى بها عيوب فى الصناعة:

$$\hat{A} = N\hat{P} = (100000)(0.05) = 5000$$

الخطأ المعيارى للتقدير: $\hat{S}_{\hat{A}} = \sqrt{(100000)(99900)(.05)(.95)/(99)} = 2189.33$

وحيث إن $n/N = 100/100000 = 0.001$ وهى تمثل 0.1% وبالتالى يمكن وضع N بدلاً من $N-n$

وعليه فإن: $\hat{S}_{\hat{A}} = (N)\sqrt{\hat{P}\hat{Q}/(n-1)} = 2190.2$

وهذه القيمة لا تختلف كثيراً عن القيمة السابق حسابها. وإذا كان حجم العينة كبيراً فإنه يمكن أيضاً وضع n بدلاً من $n-1$

وبالتالى تكون المعادلة $\hat{S}_{\hat{A}} = (N)\sqrt{\hat{P}\hat{Q}/n}$

حدود الثقة للـ P

حيث إن المتغير يتبع فعلاً التوزيع فوق الهندسى وبالتالى فإن أحد الأشكال هو التقريب الطبيعى حيث تتوافر طرق عديدة لحساب حدود الثقة.

$$\hat{P} \pm \left[Z\sqrt{1-f}\sqrt{\hat{P}\hat{Q}/(n-1)} + \frac{1}{2n} \right] \qquad (٥-٢١)$$

حيث $f = n/N$ و Z قيمة المتغير الطبيعى الموافق لاحتمال الثقة. أما $1/2n$ وهو الحد الأخير فى الطرف الأيمن فهو تصحيح من أجل الاستمرارية continuity

مثال ٥-٦

فى المثال ٥-٥ احسب حدى الثقة لـ P عند مستوى معنوية 5%

بتطبيق المعادلة (٥-٢١) فإن حدى الثقة هما

$$0.05 \pm \left[1.96\sqrt{.999}\sqrt{(.05)(.95)/99} + \frac{1}{(2)(100)}\right] = 0.5 \pm .0479$$

وبالتالى فإن الحد الأعلى هو 0979. والحد الأدنى هو 0021.

ويلاحظ هنا تأثير $\hat{P}$ عند حساب الخطأ المعيارى، ولكى يقل الفرق بين الحدين الأدنى والأعلى فى مثل هذه الحالة فيجب زيادة حجم العينة المسحوبة من العشيرة.

٥-٤-٧ تقدير النسب فى المعاينة العنقودية

Estimation of proportions in cluster sampling

قد تمثل كل وحدة معاينة عنقوداً من العناصر عند تقدير نسبة العناصر الواقعة فى الفئة "ف". مثال ذلك إذا كانت وحدة المعاينة هى الأسرة وكانت العناصر هنا هو عدد كل من الأبناء الذكور والإناث فى كل أسرة أو أن تكون وحدة المعاينة صناديق البرتقال المصدرة إلى الخارج والعناصر هى أعداد البرتقال التالفة والسليمة فى كل صندوق ... الخ. وقد تكون أعداد العناصر متساوية وقد لا تكون فى كل وحدة من وحدات المعاينة.

أولاً: عندما تكون أعداد العناصر متساوية فى كل وحدات المعاينة ولتكن m:

إذا كانت نسبة العناصر التى تنتمى إلى الفئة "ف" هى $P_i = a_i / m$ فإن نسبة وحدات العينة الواقعة فى "ف" هى:

$$\hat{P} = \frac{\sum a_i}{n\, m} = \left(\frac{1}{n}\right)\sum_{i=1}^{n}\hat{P}_i \qquad (٥-٢٢)$$

وبالتالى فإن التباين الحقيقى

$$\sigma^2_{\hat{P}} = \left(\frac{1-f}{n}\right)\left(\frac{\sum_{i=1}^{N}(P_i - P)^2}{N-1}\right) \qquad (٥-٢٣)$$

وتقديره من العينة هو :

$$\hat{\sigma}_P^2 = \left(\frac{1-f}{n}\right)\left(\frac{\sum_{i=1}^{n}(P_i - \hat{P})^2}{n-1}\right) \qquad (٥-٢٤)$$

مثال ٥-٧

من بين 1000 صندوق معبأة بالبرتقال ومعدة للتصدير سحبت 5 صناديق بطريقة عشوائية بسيطة وقدر عدد البرتقالات التالفة فى كل صندوق علماً بأن كل صندوق يحتوى على مائة برتقالة وكان عدد الثمار التالفة هى 3,1,0,2,1 فى الصناديق المسحوبة. قدر العدد الكلى للثمار التالفة مع حساب الانحراف المعيارى.

$$\hat{P} = \frac{\sum a_i}{n\,m} = \frac{1+2+0+1+3}{(5)(100)} = 0.014$$

العدد الكلى للثمار التالفة فى كل الصناديق المعدة للتصدير:

$$(1000)(100)(0.014) = 14000$$

ويمكن تجاهل معامل التصحيح للعشيرة المحدودة حيث إن $f = 5/1000$ وقيمتها ضئيلة جداً، وبالتالى فإن:

$$\hat{\sigma}_P^2 = \frac{\sum(P-\hat{P})^2}{n(n-1)} = [\sum P_i^2 - (\sum P_i)^2/n]/[n(n-1)]$$

$$\hat{P}_1 = 1/100\ ,\ \hat{P}_2 = 2/100\ ,\ \hat{P}_3 = 0\ ,\ \hat{P}_4 = 1/100\ ,\ \hat{P}_5 = 3/100$$

$$\hat{\sigma}_P^2 = \frac{.0015 - .00098}{(5)(4)} = 0.000026$$

تباين العدد الكلى من الثمار التالفة:

$$(Nm)^2\hat{\sigma}_P^2 = [(1000)(100)]^2(0.000026) = 260000$$

الانحراف المعيارى للعدد الكلى من الثمار التالفة:

$$\sqrt{(Nm)^2\hat{\sigma}_P^2} = \sqrt{260000} = 509.9$$

ثانياً: عندما تكون أعداد العناصر غير متساوية فى وحدات المعاينة ويرمز لكل بالرمز m_i فإن:

$$\hat{P}_i = a_i / m_i \qquad (٥-٢٥)$$

ونسبة الوحدات فى العينة الواقعة فى الفئة "ف" هى:

$$\hat{P} = \sum a_i / \sum m_i \qquad (٥-٢٦)$$

وبالتالى فإن تباين $\hat{P}$:

$$\sigma_P^2 \cong \left(\frac{1-f}{n\overline{M}^2}\right)\left(\frac{\sum(a_i - Pm_i)^2}{N-1}\right) \qquad (٥-٢٧)$$

حيث P هى نسبة عناصر الفئة "ف" فى العشيرة و $\overline{M} = \sum^{N} m_i / N$ هو متوسط عدد العناصر فى العنقود الواحد والعبارة البديلة هى

$$\sigma_P^2 \cong \left(\frac{1-f}{n}\right)\sum^{N}\left(\frac{m_i}{\overline{M}}\right)^2\left(\frac{(P_i - P)^2}{N-1}\right) \qquad (٥-٢٨)$$

ويقدر هذا التباين من العينة كما فى المعادلة التالية

$$\hat{\sigma}_P^2 = \left(\frac{1-f}{(n)(\overline{m})^2}\right)\left(\frac{\sum a_i^2 - 2\hat{P}\sum a_i m_i + \hat{P}^2 \sum m_i^2}{n-1}\right) \qquad (٥-٢٩)$$

حيث $\overline{m} = (\sum m)/n$ هو متوسط عدد العناصر للعنقود الواحد فى العينة.

مثال ٥-٨

من بيانات المثال ٥-٤ قدر تباين نسبة الذكور وقارنه مع التقدير "ذو الحدين" للتباين.

فى حالة توزيع "ذو الحدين": $\hat{P} = 29/64 = 0.453$ ، $n = 64$

$$\hat{\sigma}_{\hat{P}}^2 = \hat{P}\hat{Q}/n = 0.00387$$

ومن أجل علاقة النسبة لاحظ وجود 20 عنقوداً أى أن: $n = 20$

العدد الكلي لوحدة المعاينة (الأسرة) i: = m_i

عدد الذكور فى الأسرة i: = a_i

وبالتالى $\overline{m} = (\sum m)/n = 64/20 = 3.2$

$$\sum m_i^2 = 320 \text{ ، } \sum a_i m_i = 106 \text{ ، } \sum a_i = 63$$

وبتطبيق المعادلة (٥-٢٩) مع تجاهل معامل التصحيح

$$\hat{\sigma}_{\hat{p}}^2 = \frac{63-(2)(0.453)(106)+(0.543)^2(230)}{(20)(3.2)^2(19)} = 0.00364$$

مما سبق يتضح أن التباين 0.00387 فى حالة توزيع "ذو الحدين" وهو أكبر بقليل عنه فى حالة النسبة وقد تكون أقل. وقد تكون هذه الفروق راجعة لنسبة الجنس والتى تقترب من 0.5 والتى قد لا تكون فى بعض الأسر.

صندوق ٥-٢

تنقسم طرق اختيار العينة إلى:

طرق احتمالية: وهى التى يختار فيها المجرب العينة بطريقة عشوائية تماماً مستخدماً بذلك جداول الأرقام العشوائية أو الحاسبات.

طرق غير احتمالية: وفيها يكون أفراد العينة متطوعين مثلاً لاختبار عقار معين.

والطرق الاحتمالية هى الأفضل وينصح بإتباعها كلما أمكن ذلك.

وإذا اختيرت العينة الاحتمالية بطريقة جيدة وبأعداد مناسبة فإن التقديرات المتحصل عليها تتميز بأنها غير متحيزة ومتسقة ولها أقل تباين وكافية. وهذه الصفات الأربع تم شرحها فى هذا الجزء من المؤلف.

٥-٤-٨ تقدير حجم العينة Estimation of sample size

تبين مما سبق أن اختيار عينة كبيرة جداً يتضمن هدراً للمصادر بينما العينة الصغيرة جداً تقلل من فائدة النتائج. وفى الحقيقة لا يمكن اتخاذ قرار سليم دائماً فى اختيار الحجم المناسب للعينة حيث إن المعلومات الكافية غير متوفرة دائماً.

أولاً: تحديد حجم العينة عند معاينة النسب

كما سبق فإن الوحدات تقسم إلى فئتين "ف" ، "فَ" فإذا كان الخطأ المسموح به (d) فى النسبة المقدرة لوحدات الفئة "ف" وهى $\hat{P}$ وإنه يمكن تجاوز الخطأ الفعلى للمقدار d ولكن باحتمال صغير α فإن المطلوب تحقيق $\Pr\left|(\hat{P}-P)\right| \geq d = \alpha$ وذلك فى معاينة عشوائية بسيطة مع افتراض أن $\hat{P}$ تتوزع طبيعياً وحيث إن:

$$\sigma_{\hat{p}} = \sqrt{\frac{PQ}{n}}\sqrt{\frac{N-n}{N-1}}$$

وعليه فإن العلاقة التى تربط n بالدرجة المطلوبة للدقة هى:

$$d = (Z)\left(\sqrt{\frac{PQ}{n}}\sqrt{\frac{N-n}{N-1}}\right)$$

حيث Z هى قيمة المتغير الطبيعى المعيارى التى تقطع مساحة قدرها $\alpha/2$ من كل من الذيلين (جدول ٣ ملحق أ) ومنها فإن:

$$n = \frac{Z^2\hat{P}\hat{Q}}{d^2} \Bigg/ \left[1 + \frac{1}{N}\left(\frac{Z^2PQ}{d^2} - 1\right)\right]$$

وإذا كانت N كبيرة واستخدمت $\hat{P}$ بدلاً من P فإن التقريب يعطى

$$n_\circ = \frac{Z^2\hat{P}\hat{Q}}{d^2} = \frac{\hat{P}\hat{Q}}{V} \qquad (٥-٣٠)$$

حيث التباين المرغوب لنسبة العينة

$$V = \frac{\hat{P}\hat{Q}}{n_\circ}$$

وعند التطبيق تحسب أولاً $n_\circ$ وإذا كانت $n_\circ / N$ صغيرة يمكن إهمالها ويمكن اعتبار أن $n_\circ$ تقريباً لا بأس به لتقدير حجم العينة n أما إذا كان غير ذلك فإنه يمكن حساب n من العلاقة

$$n = \frac{n_\circ}{1 + (n_\circ - 1)/N} \qquad (٥-٣١)$$

مثال ٥-٩

إذا كانت $\hat{P} = 0.4$، $d = 0.05$ ومستوى المعنوية $\alpha = 0.05$ وكان حجم العشيرة $N = 3000$. احسب حجم العينة فى حالة المعاينة العشوائية البسيطة.

بما إن $Z = 2$ (تقريباً، جدول ٣ ملحق أ) فإن

$$n_\circ = \frac{(2)^2 (0.4)(0.6)}{(0.05)^2} = 384$$

وحيث إن $n_\circ / N = 0.128$ أى 12.8% وبالتالى لا يمكن إهمالها فإن:

$$n = \frac{n_\circ}{1 + (n_\circ - 1)/N} = \frac{384}{1 + (384 - 1)/3000} = 340.5 \cong 341$$

ثانياً: تقدير حجم العينة العشوائية البسيطة فى حالة البيانات لمتغير مستمر

إذا كان متوسط عينة عشوائية بسيطة هو $\hat{\bar{Y}}$ والمطلوب هو تحقيق المعادلة الاحتمالية التالية

$$P\left(\left|\frac{\hat{\bar{Y}} - \bar{Y}}{\bar{Y}}\right| \geq r\right) = P\left(\left|\frac{N\hat{\bar{Y}} - N\bar{Y}}{N\bar{Y}}\right| \geq r\right) = P(|\hat{\bar{Y}} - \bar{Y}|) \geq r\bar{Y} = \alpha$$

حيث α قيمة احتمالية صغيرة وبافتراض أن $\hat{\bar{Y}}$ تتوزع طبيعياً و r الخطأ النسبى فى تقديرات مجموع أو متوسط العشيرة. وحيث أن:

$$\sigma_{\hat{\bar{Y}}} = \sqrt{\frac{N-n}{N}}\left(\frac{S}{\sqrt{n}}\right)$$

وبالتالى فإن

$$r\overline{Y} = Z\sigma_{\hat{\overline{Y}}} = Z\left(\sqrt{\frac{N-n}{N}}\right)\left(\frac{S}{\sqrt{n}}\right)$$

حيث Z هى قيمة المتغير الطبيعى المعيارى التى تقطع مساحة قدرها $\propto/2$ من كل من الذيلين. ويمكن إيضاح أن:

$$n = \left(\frac{ZS}{r\overline{Y}}\right)^2 \Bigg/ \left[1 + \frac{1}{N}\left(\frac{ZS}{r\overline{Y}}\right)^2\right]$$

ويلاحظ أن المعادلة تحتوى على $S/\overline{Y}$ وهو معامل اختلاف العشيرة وتخمينه أسهل من تخمين S نفسها.

وكتقدير مبدئى فإن

$$n_\circ = \left(\frac{ZS}{r\overline{Y}}\right)^2 \qquad (٥-٣٢)$$

وإذا كانت $n_\circ / N$ كبير فإن

$$n = \frac{n_\circ}{1 + (n_\circ / N)} \qquad (٥-٣٣)$$

وقد تقدر $n_\circ$ كما يلى:

$$n_\circ = \frac{Z^2 S^2}{d^2} \qquad (٥-٣٤)$$

حيث d هو الخطأ المطلق فى $\hat{\overline{Y}}$ بدلاً من الخطأ النسبى r أى أن $d = r\overline{Y}$.

مثال ٥-١٠

فى عشيرة حجمها 500 إذا كان متوسط متغير ما $\overline{Y} = 22$ والتباين $S^2 = 90$. فما هو حجم العينة العشوائية البسيطة لتقدير $\overline{Y}$ فى حدود خطأ مقدراه 10% وذلك باستثناء فرصة واحدة من عشرين.

باستثناء فرصة واحدة من عشرين تعنى أن مستوى المعنوية $\alpha = 0.05$ وبالتالى $Z \doteq 2$ تقريباً، $d = r\overline{Y} = (0.1)(22) = 2.2$ وبالتالى فإن:

$$n_{\circ} = \frac{(2)^2 (90)}{(2.2)^2} = 74.3$$

وحيث إن $n_{\circ} / N = 74.3/500$ أى 14.9% وبذلك لا يمكن إهمالها فإن

$$n = \frac{74.3}{1 + 74.3/500} = 64.7 \cong 65$$

وعلى القارئ أن يستنتج بنفسه أنه كلما زاد الخطأ المسموح به كلما قل حجم العينة والعكس صحيح بينما عكس ذلك فى حالة التباين حيث يزداد حجم العينة كلما زاد حجم التباين والعكس صحيح.

ملحوظة هامة:

إذا كان هناك بيانات لأكثر من متغير مستمر فى العشيرة وتم حساب n فى كل متغير فإن n الأكبر هى التى تؤخذ فى الاعتبار عند تحديد حجم العينة.

ومما سبق يتضح ضرورة إيجاد طريقة لتقدير تباين العشيرة حيث إنه أحد العوامل الأساسية فى تقدير حجم العينة وفيما يلى بعض هذه الطرق:

١- تؤخذ العينة على مرحلتين حيث تؤخذ عينة عشوائية بسيطة حجمها n_1 للحصول على $\hat{S}_1^2$ تقديراً للـ S^2 أو $\hat{P}_1$ تقديراً لـ P ومن ثم يمكن الحصول على n. وهذه الطريقة أجدر بالثقة ولكن يؤخذ عليها أنها تؤدى إلى طول فترة المعاينة.

٢- باستخدام نتائج مسح إحصائى استكشافى صغير غالباً ما يكون مقتصراً على جزء من العشيرة يسهل تناوله أو أنه يهدف للكشف عن حجم بعض المشاكل.

٣- استخدام نتائج مسوح إحصائية سابقة وهنا ترجع أهمية نشر أية معلومات حول الانحراف المعيارى تم الحصول عليها من مسوح سابقة أو أن تكون فى متناول يد الباحثين حتى يستطيعوا استخدامها عند الحاجة إليها.

٤- بعملية تخمين حول طبيعة العشيرة وذلك باستخدام بعض التوزيعات الرياضية البسيطة لتقدير التباين.

وهناك الكثير من الطرق التى تستخدم فى تقدير حجم العينة لا يتسع المجال لذكرها ويمكن الرجوع إلى أحد المراجع التى تتناول تقنية المعاينة الإحصائية.

ويوجد العديد من طرق المعاينة فبالإضافة إلى المعاينة العشوائية البسيطة التى تم تناولها فيما سبق توجد المعاينة العشوائية الطبقية stratified random sampling

والمعاينة النمطية systematic sampling والمعاينة العنقودية cluster sampling وغيرهم ولكن لن يتسع المجال لتناولهم فى هذا المؤلف.

٥-٥ مصادر الخطأ فى المعاينة

١- عدم التحديد الدقيق للعشيرة المراد معاينتها.

٢- الاختيار غير السليم لطريقة المعاينة.

٣- أخطاء نتيجة عدم الاستجابة non-response وهى تعنى الفشل فى مقياس بعض الوحدات فى العينة المختارة وذلك قد يرجع إلى الإخفاق فى تحديد موقع بعض وحدات العينة أو زيارتها أو أن يكون المنزل مغلقاً وليس به الأشخاص المعنيون أو أن يكون الأشخاص المعنيون بالبحث غير قادرين على الإجابة، فقد تكون المعلومات المطلوب الإجابة عليها غير متوفرة أو من غير الممكن الإدلاء بها، وقد ترجع عدم الاستجابة أيضاً إلى أن هناك أشخاصاً يرفضون إجراء مثل تلك المقابلات أو الإدلاء ببيانات عنهم وعن ممتلكاتهم ...الخ. وفى مثل تلك الأحوال تكون الزيارات المتكررة أحد الحلول الممكنة للتغلب على تلك الأخطاء، كما أن التوعية بضرورة وأهمية إجراء مثل تلك المسوح الإحصائية من الأهمية بمكان، كما تشكل الصياغة الجيدة للسؤال وطريقة تقديمه إمكانية الإجابة عليه بدقة وبدون تردد.

٤- أخطاء القياس نتيجة عدم دقة أداة القياس أو أن الأشخاص المعنيين لا يعطون إجابات دقيقة أو قد يكون بعضهم متحيزاً لإجابات معينة دون غيرها أو قد يكون هناك ارتباط بين أخطاء القياس بين وحدات العينة فى اتجاه ما قد يكون موجباً أو سالباً. وهنا يلزم إيجاد طريقة ما لتحديد القيمة الصحيحة وقد يكون إعادة القياس بطريقة مستقلة أكثر دقة وقد يكون تدريب القائمين بالحصر تدريباً دقيقاً وأن يكونوا من ذوى الخبرة حتى يمكنهم الحصول على البيانات المطلوبة بالدقة الكافية. وقد يكون إعادة قياس عينات جزئية أحد الحلول لتفادى أخطاء القياس. وقد تكون سرية البيانات المتحصل عليها ضرورة وخاصة عندما تكون الأسئلة عن واقع شخص كالسرقة مثلاً أو استخدام بعض العقاقير الطبية.

٥- أخطاء نتيجة عدم المعالجة الإحصائية السليمة للبيانات المتحصل عليها.

تمارين الباب الخامس

٥-١ فى عشيرة من 5 وحدات كانت قيمها 2 ,3 ,5 ,6 ,1 وسحبت عينة عشوائية بسيطة من وحدتين بدون إحلال، فما هو عدد العينات العشوائية البسيطة الممكن سحبها وما هو احتمال سحب كل منها ؟ احسب المتوسط والتباين وتحقق من أن المتوسط والتباين المحسوب من العينة هما تقديران غير متحيزين لمتوسط وتباين العشيرة.

٥-٢ إذا كان متوسط عدد كرمات العنب فى الأراضي المستصلحة حديثاً 900 فى الفدان. وكان متوسط إنتاج الكرمة الواحدة 13 كج بتباين مقداره 30 كج مربع. احسب كمية الإنتاج فى مساحة قدرها 100 فدان (فدان = $4200\ m^2$) والتى هى حجم العشيرة وانحرافه المعيارى إذا علمت أن المعاينة كانت عشوائية بسيطة وكان حجم العينة خمسة أفدنة. قدر حدود الثقة لكمية الإنتاج الكلية أيضاً بمستوى ثقة 95% .

٥-٣ سحبت عينة عشوائية بسيطة من 5 من العاملين في أحد المصانع وكانت دخولهم كما يلى: 350 ,400 ,200 ,330 ,500 جنيهاً مصرياً فى الشهر. فما هو متوسط دخل العامل فى المصنع إذا كان به 1000 عامل ثم احسب تباينه. ثم قدر مجموع الدخل المتحصل عليه لجميع العاملين فى المصنع فى الشهر الواحد وكذلك انحرافه المعيارى.

٥-٤ فى دراسة لاستهلاك الأصناف المختلفة للمنظفات الصناعية للملابس فى مدينة القاهرة سجلت المعلومات من السيارات المنتظرة فى إشارات المرور. حلل هذه الطريقة من المعاينة.

٥-٥ إذا كانت $\hat{P} = 0.25$ ، $d = 0.02$ ومستوى المعنوية 5% وحجم العشيرة هو 2000. احسب حجم العينة العشوائية البسيطة.

٥-٦ أعد حساب حجم العينة فى مثال ٥-١٠ فى حدود خطأ مقداره 5%، 15%، 20% وذلك بمستوى معنوية 5% . ماذا تستنتج؟

٥-٧ الجدول التالى يمثل الدخل الشهرى بالجنيه لــ 10 أسر مصرية أخذت بطريقة المعاينة العشوائية البسيطة من بين 200 أسرة وأيضاً تم رصد عدد أفراد الأسرة المدخنين وغير المدخنين:

الأسـرة	عدد الأفراد غير المدخنين	عدد الأفراد المدخنين	المنفق على الدروس الخصوصية	الدخــل
١	3	0	260	1300
٢	2	1	100	500
٣	5	1	400	2100
٤	2	0	250	1200
٥	3	0	180	900
٦	3	2	180	850
٧	4	1	270	1200
٨	3	2	200	900
٩	5	1	800	3000
١٠	2	3	500	2400

احسب النسبة المئوية للدخل المنفق على الدروس الخصوصية و خطأه المعيارى.

٥-٨ فى التمرين السابق المطلوب مقارنة علاقة النسبة بعلاقة التوزيع الثنائى وفقاً لحالة التدخين من عدمه. واحسب التباين فى كل.