الفصل السابع

التقدير بطريقة النسبة والانحدار والفرق

Ratio, Regression and Difference Estimation

1.7 مقدمة

قد تظهر في كثير من المسوحات معلومات مساعدة أو إضافية لمتغير له علاقة مع المتغير الذي نحن بصدد دراسته. هذه المعلومات يمكن الاستفادة منها لتحسين تصميم المعاينة. ومن ثم الوصول إلى تقديرات أكثر دقة لمعلمات المجتمع للمتغير الذي ندرسه. فمثلاً إذا كنا نقوم بمسح لتقدير مصروفات العائلة الشهرية لمدينة معينة.. فإن عدد أفراد العائلة له علاقة مباشرة وقوية مع مصروفات العائلة الشهرية. غالباً نستطيع الحصول على معلومات عن عدد أفراد العائلة في المجتمع من دوائر الأحوال المدنية أو من مسوحات شاملة جرى فيها جمع معلومات كاملة عن عدد أفراد العائلة وكذلك معلومات أخرى عن العائلة.

إن المعلومات الإضافية أو المساعدة التي يمكن الاستفادة منها تكون متوافرة على شكل مجاميع كلية أو معدلات، وليس عن كل وحدة من وحدات المجتمع؛ لذلك يمكن استخدام هذه المعلومات عند القيام بالمسح بالعينة، شريطة أن تكون هذه المعلومات الخاصة بالمتغير المساعد يمكن الحصول عليها بسهولة عند جمع معلومات حول المتغير الرئيس باستخدام العينة. ففي مثالنا الخاص بتقدير مصروفات العائلة الشهرية لا توجد صعوبة

عند سحب عينة من العائلات ومقابلتهم أن نسألهم بالإضافة إلى مقدار مصروفاتهم الشهرية عن عدد أفراد العائلة.

2.7 رموز ومصطلحات

إن هذه الرموز والمصطلحات التي سوف نعرفها هنا خاصة بالعينة العشوائية البسيطة. أما الرموز والمصطلحات الخاصة بالعينة العشوائية الطبقية فلا توجد صعوبة في الحصول عليها. وسوف نقوم بتعريف ما نحتاجه منها لاحقاً.

y_i : قيمة المتغير أو الصفة تحت الدراسة للوحدة i في المجتمع.

x_i : قيمة المتغير أو الصفة المساعدة في نفس الوحدة i في المجتمع.

Y : المجموع الكلي للمتغير y في المجتمع.

X : المجموع الكلي للمتغير x في المجتمع.

$R = \frac{Y}{X} = \frac{\bar{Y}}{\bar{X}}$: نسبة المجموع والوسط للمتغير y و x.

ρ : معامل الارتباط بين y و x في المجتمع.

لنفترض أننا نرغب في تقدير المجموع Y أو الوسط $\bar{Y}$ للمجتمع. وسحبنا عينة عشوائية بسيطة بحجم n من وحدات المجتمع. وحصلنا على n زوج من المشاهدات حول المتغيرين y و x. ولنفترض أن $\bar{y}$ و $\bar{x}$ الوسط الحسابي للعينة للمتغيرين y و x على التوالي، فعليه يكون $\hat{R}$ تقدير للنسبة R

$$\hat{R} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i} = \frac{\bar{y}}{\bar{x}}$$

و $\bar{y}_R$ تقدير الوسط الحسابي للمجتمع بطريقة النسبة

$$\bar{y}_R = \frac{\bar{y}}{\bar{x}} \bar{X} = \hat{R}\bar{X}$$

و Y_R تقدير المجموع الكلي للمجتمع بطريقة النسبة

$$Y_R = \frac{\bar{y}}{\bar{x}} X = R X$$

3.7 تقدير النسبة R في العينة العشوائية البسيطة

قد يحدث في كثير من الدراسات أن نجمع معلومات عن أكثر من متغير في الوقت نفسه. باستمرار توجد رغبة في دراسة العلاقة بين هذه المتغيرات، بالإضافة إلى دراسة هذه المتغيرات على حدة. إن النسبة بين متغيرين R من العلاقات التي نرغب في دراستها باستمرار. على سبيل المثال نرغب أن نعرف نسبة دخل الفرد أو العائلة المصروف على الطعام، أو نسبة الإنتاج التالف لأحد المصانع.

لنفترض أن المتغير الذي نرغب في دراسته y والمتغير الإضافي أو المساعد x فإن النسبة لهذين المتغيرين في المجتمع تسمى نسبة المجتمع للمتغيرين وتعرف

$$R = \frac{\bar{Y}}{\bar{X}} = \frac{Y}{X}$$

لتقدير هذه النسبة سحبنا عينة عشوائية بسيطة حجمها n وحدة من المجتمع، وجمعنا معلومات حول المتغيرين y وx، ولنفترض أن $\bar{y}$ و $\bar{x}$ هما الوسط الحسابي للعينة للمتغيرين y وx على التوالي. نقدر R باستخدام R والتي سبق تعريفها أعلاه. إن R هو تقدير متحيز إلى R أي إن $E(R) \neq R$. لمزيد من المعلومات يراجع حسين علوان (1993)و(1977) Cochran أو ترجمة أنيس كنجو (1995). ننتقل الآن لإيجاد تباين تقدير النسبة، إذا كان حجم العينة كبيراً فإن تباين R سيكون تقريباً

$$\mathrm{Var}(R) = \frac{1-f}{n\bar{X}^2}[S_y^2 + R^2 S_x^2 - 2RS_{xy}]$$

حيث إن

$$S_x^2=\frac{\sum_{i=1}^{N}(X_i-\overline{X})^2}{N-1}$$

و

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

و

$$f=\frac{n}{N}$$

و

$$S_{xy}=\frac{\sum_{i=1}^{N}(Y_i-\overline{Y})(X_i-\overline{X})}{N-1}$$

لا بد من تقدير $var(R)$ لأن R و S_x^2 و S_y^2 و S_{xy}^2 غير معلومات. لذلك فإن تقدير تباين $var(R)$ يمكن إيجاده باستخدام:

$$s^2(R)=\frac{1-f}{n\overline{x}^2}[s_y^2+R^2 s_x^2-2R\,s_{xy}]$$

حيث إن

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

و

$$s_y^2=\frac{\sum_{i=1}^{n}(y_i-\overline{y})^2}{n-1}$$

و

$$s_{xy}=\frac{\sum_{i=1}^{n}(y_i-\overline{y})(x_i-\overline{x})}{n-1}$$

إن $s^2(R)$ تقدير متحيز إلى $var(R)$ ولكن إذا كان حجم العينة كبيراً فإن قيمة التحيز تكون قليلة. بقي أن نكتب $s^2(R)$ بشكل آخر يسهل استخدامه في العمليات الحسابية:

$$s^2(R)=\frac{1-f}{n\bar{x}^2}\;\frac{\sum_{i=1}^{n}y_i^2+R^2\sum_{i=1}^{n}x_i^2-2R\sum_{i=1}^{n}y_i x_i}{n-1}$$

مثال (1): الجدول الآتي يمثل عدد الأسماك التي تم اصطيادها في بحيرة. وعدد الصيادين الذين استخدموا البحيرة مدة 20 يوماً. إذا كان N=200 يمثل عدد الأيام المسموح فيها الصيد خلال السنة.

اليوم	1	2	3	4	5	6	7	8	9	10
عدد الصيادين x_i	4	5	2	5	1	1	7	6	4	1
عدد الأسماك y_i	8	10	8	15	2	3	18	15	20	4
اليوم	11	12	13	41	15	16	17	81	19	20
عدد الصيادين x_i	2	4	1	2	3	4	4	2	0	2
عدد الأسماك y_i	6	14	3	7	8	10	12	4	0	7

استخدم هذه البيانات لتقدير نسبة عدد الأسماك إلى عدد الصيادين ثم أوجد تقدير معامل التغير إلى تقدير النسبة.

الحل:

$$\sum_{i=1}^{20}y_i=174 \text{ و } \sum_{i=1}^{20}x_i=60 \text{ و } \sum_{i=1}^{20}y_i^2=2094 \text{ و } \sum_{i=1}^{20}x_i^2=248 \text{ و } \sum_{i=1}^{20}x_i y_i=697$$

فعليه يكون تقدير النسبة

$$R = \frac{\bar{y}}{\bar{x}} = \frac{8.7}{3} = 2.9$$

و تقدير تباين النسبة

$$s^2(R) = \frac{N-n}{Nn\bar{x}^2}\,\frac{\sum_{i=1}^{n} y_i^2 + R^2\sum_{i=1}^{n} x_i^2 - 2R\sum_{i=1}^{n} y_i x_i}{n-1}$$

$$= \frac{(200-20)}{20(3)^2}\,\frac{2094 + (2.9)^2(248) - 2(2.9)(697)}{(200)(19)} = 0.0361$$

إن تقدير الخطأ المعياري لتقدير النسبة

$$s(R) = \sqrt{s^2(R)} = \sqrt{0361} = 0.19$$

كذلك فإن معامل التغير لتقدير النسبة

$$CV(R) = \frac{s(R)}{R} = \frac{0.19}{2.9} = 0.0655$$

إذا كان X المجموع الكلي للمتغير x معلوماً فإن تقدير النسبة يصبح

$$R = \frac{Y}{X}$$

وهذا تقدير غير متحيز إلى R وتباينه

$$Var(R) = \frac{1-f}{\bar{X}^2}\,\frac{S_y^2}{n}$$

لمزيد من المعلومات يراجع حسين علوان (1993) و(1977) Cochran أو ترجمة أنيس كنجو (1995).

4.7 تقدير الوسط الحسابي والمجموع باستخدام النسبة

إن الوسط الحسابي للمجتمع $\overline{Y}$ ومجموع المجتمع الكلي Y يمكن تقديرهما باستخدام تقدير النسبة على النحو الآتي:

تقدير الوسط الحسابي بطريقة النسبة

$$\overline{y}_R = R\overline{X}$$

تقدير المجموع بطريقة النسبة

$$Y_R = RX$$

نفترض معرفة الوسط الحسابي $\overline{X}$ والمجموع الكلي X للمجتمع للمتغير x واضح أن $\overline{y}_R$ و Y_R تقديران متحيزان إلى $\overline{Y}$ و Y على التوالي مع تباين تقريبي للتقديرين $\overline{y}_R$ و Y_R على التوالي

$$\mathrm{Var}(\overline{y}_R) = \overline{X}^2\,\mathrm{Var}(R) = \frac{1-f}{n}\,[S_y^2 + R^2S_x^2 - 2RS_{xy}]$$

$$\mathrm{Var}(Y_R) = N^2\,\mathrm{Var}(\overline{y}_R) = \frac{1-f}{n}\,N^2\,[S_y^2 + R^2S_x^2 - 2RS_{xy}]$$

وتقدير التباين للتقديرين $\overline{y}_R$ و Y_R هما على التوالي

$$s^2(\overline{y}_R) = \frac{(1-f)}{n}\;\frac{\sum_{i=1}^{n} y_i^2 + R^2\sum_{i=1}^{n} x_i^2 - 2R\sum_{i=1}^{n} y_i x_i}{n-1}$$

$$s^2(Y_R) = N^2\,s^2(\overline{y}_R)$$

باستطاعتنا المقارنة بين تباين الوسط الحسابي للعينة العشوائية البسيطة $\mathrm{var}(\overline{y})$ وتباين الوسط الحسابي للعينة العشوائية البسيطة باستخدام النسبة $\mathrm{var}(\overline{y}_R)$ لنجد أن

$$\mathrm{var}(\overline{y}_R) \le \mathrm{var}(\overline{y})$$

عندما تكون

$$2RS_{xy} \geq R^2S_x^2$$

أو

$$\rho \geq \frac{1}{2}\frac{CV(x)}{CV(y)}$$

تشير هذه النتائج إلى أن تقدير الوسط الحسابي بطريقة النسبة سيكون أكثر دقة من استخدام المتغير y بمفرده، إذا كان معامل الارتباط ρ أكبر من معامل تغير x على معامل تغير y .

مثال (2): في المثال (1) لنفترض أن X=650 عدد الصيادين في السنة. أوجد تقدير الوسط الحسابي بطريقة النسبة $\bar{y}_R$ وبطريقة y بمفرده وقارن بينهما.

الحل:

$$s^2(\bar{y}_R) = \frac{(N-n)}{Nn}\,\frac{\sum_{i=1}^{n} y_i^2 + R^2\sum_{i=1}^{n} x_i^2 - 2R\sum_{i=1}^{n} y_i x_i}{n-1}$$

$$= \frac{200-20}{20(200)}\,\frac{2094 + (2.9)^2(248) - 2(2.9)(697)}{19} = 0.3247$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{20}(174) = 8.7$$

$$s^2(\bar{y}) = \frac{N-n}{Nn}\,s_y^2 = \frac{200-20}{200}\,\frac{30.54}{20} = 1.3743$$

مما لا شك فيه لقد طرأ تحسن كبير على دقة التقدير للوسط الحسابي للمجتمع باستخدام النسبة.

5.7 اختيار حجم العينة

إن مقدار المعلومات المحتواة في العينة يعتمد على عاملين، الأول التشتت في البيانات، وهو الذي يمكن السيطرة عليه في بعض الأحيان وذلك من خلال تصميم المسح بالعينة. وأما العامل الثاني فهو حجم العينة n وبعد اختيار وتصميم العينة لا بد من تحديد حجم العينة. وسوف نناقش اختيار حجم العينة المطلوبة لتقدير معلمات المجتمع النسبة R والوسط الحسابي والمجموع الكلي للمجتمع.

1.5.7 اختيار حجم العينة لتقدير النسبة R

إن الطريقة التي سوف تتبع هنا لاختيار حجم العينة مماثله لما سبق في فصول سابقة. إن عدد الوحدات المطلوبة لتقدير نسبة المجتمع R ضمن حد معين لمقدار الخطأ في التقدير وليكن d ومع 95% ثقة إي

$$p(\left| R - R \right| < d) = 0.95$$

لذلك يجب حل المعادلة الآتية لإيجاد قيمة n

$$2\sqrt{var(R)} = d$$

سوف نستخدم $s^2(R)$ بدلاً من $var(R)$ لأن R و S_x^2 و S_y^2 و S_{xy}^2 غير معلومات.

إن حجم العينة المطلوب سيكون

$$n > \frac{Ns^2}{ND + s^2}$$

حيث إن $D = \frac{d^2\,\overline{X}^2}{4}$. ولكننا في الواقع لا نعرف قيمة σ^2. إذاً لا بد من تقديرها من معلومات سابقه لمسوحات مشابه أو بسحب عينة أولية حجمها n' ثم نستخدمها لتقدير σ^2 بما يلي

$$\sigma^2 = \frac{1}{n'-1}\sum_{i=1}^{n'}(y_i - R\,x_i)^2$$

ونعوض $\hat{\sigma}^2$ بدلاً عن σ^2 في المعادلة أعلاه، وكذلك إذا كانت قيمة $\overline{X}$ غير معلومة نقدرها باستخدام $\bar{x}$ الوسط الحسابي للعينة الأولية.

2.5.7 اختيار حجم العينة لتقدير الوسط الحسابي للمجتمع باستخدام النسبة

يمكن أن نحدد حجم العينة الذي نحتاجه لتقدير الوسط الحسابي للمجتمع باستخدام النسبة ضمن حد معين لمقدار الخطأ في التقدير وليكن d وبثقة 95% أي

$$p(\left|\overline{Y} - \bar{y}_R\right| < d) = 0.95$$

وذلك بحل إحدى المعادلتين

$$2\sqrt{Var(\bar{y}_R)} = d$$

أو

$$2\overline{X}\sqrt{Var(R)} = d$$

إذاً حجم العينة المطلوب لتقدير الوسط الحسابي للمجتمع مع خطأ في التقدير مقداره d وبثقة 95% فإن حجم العينة المطلوب يكون

$$n > \frac{Ns^2}{ND + s^2}$$

حيث إن $D = \dfrac{d^2}{4}$. نلاحظ هنا أننا لا نحتاج إلى تقدير $\overline{X}$ ولكن لا بد من تقدير σ^2 بالطريقة نفسها التي استخدمناها في تقدير حجم العينة للنسبة.

3.5.7 اختيار حجم العينة لتقدير المجموع الكلي للمجتمع باستخدام النسبة

لا نحتاج لإعادة الكلام نفسه الذي استخدمناه عند الكلام عن اختيار حجم العينة لتقدير R و $\overline{Y}$ أعلاه. ونقول إن حجم العينة المطلوب يكون

$$n > \frac{Ns^2}{ND + s^2}$$

حيث إن

$$. D = \frac{d^2}{4N^2}$$

مثال (3): لنفترض في مثالنا (1) أن البيانات المعطاة في الجدول تمثل عينه أولية أي إن $n' = 20$ وكذلك لنفترض أن $N = 200$ و $X = 650$. أوجد حجم العينة المطلوب لتقدير R و $\overline{Y}$ و Y إذا كانت قيمة d تساوي 0.3 و0.8 و100 على التوالي.

الحل: نقدر قيمة σ^2 باستخدام العينة الأولية

$$\sigma^2 = \frac{1}{n'-1}\sum_{i=1}^{n'}(y_i - R\,x_i)^2 = \frac{137.08}{19} = 7.2147$$

إذاً حجم العينة المطلوب لتقدير R

$$n = \frac{N\sigma^2}{ND+\sigma^2} = \frac{N\sigma^2}{N(d^2\overline{X}/4)+\sigma^2} = \frac{200(7.2147)}{200[(0.3)^2(3.25)^2/4]+7.2147} = 27$$

وكذلك فإن حجم العينة المطلوب لتقدير $\overline{Y}$

$$n = \frac{N\sigma^2}{ND+\sigma^2} = \frac{N\sigma^2}{N(d^2/4)+\sigma^2} = \frac{200(7.2147)}{200[(0.8)^2/4]+7.2147} = 37$$

وأخيراً فإن حجم العينة المطلوب لتقدير Y يكون

$$n = \frac{N\sigma^2}{ND+\sigma^2} = \frac{N\sigma^2}{N(d^2/N^2 4)+\sigma^2} = \frac{200(7.2147)}{200[(100)^2/4(200)^2]+7.2147} = 74$$

6.7 تقديرات النسبة في العينة العشوائية الطبقية

لقد لاحظنا من خلال دراستنا إلى الآن أن تقدير الوسط الحسابي والمجموع الكلي للمجتمع يمكن تحسينهما باستخدام العينة العشوائية

الطبقية. وأقصد بالتحسين تقليل التباين. وكذلك لاحظنا أن استخدام النسبة لتقدير الوسط الحسابي والمجموع للمجتمع يمكن أن يؤدي إلى تحسين التقدير لهذه المعلمات. والسؤال الذي يطرح نفسه ماذا لو استخدمنا النسبة عندما تكون العينة سحبت من مجتمع مقسم إلى طبقات؟ وهذا ما سنحاول دراسته هنا.

1.6.7 تقدير المجموع الكلي للمجتمع باستخدام النسبة

هناك طريقتان لتقدير المجموع الكلي باستخدام النسبة في العينة العشوائية الطبقية.

1- المجموع الطبقي النسبي المنفصل ويعرف

$$Y_{RS} = \sum_{i=1}^{K} \frac{\bar{y}_i}{\bar{x}_i} X_i = \sum_{i=1}^{K} R_i X_i$$

حيث إن R_i و X_i يمثلان تقدير النسبة والمجموع الكلي للطبقة i و $i = 1, 2, \ldots, K$ للمتغير x. ولا بد من معرفة X_i لجميع الطبقات. كذلك فإن مقدار التحيز للتقدير سيكون عبارة عن مجموع التحيزات لكل الطبقات وسيحمل التحيز على الأغلب الإشارة نفسها لجميع الطبقات. أما تباين Y_{RS} وتقدير تباينه فسيكونان على التوالي

$$Var(Y_{RS}) = \sum_{i=1}^{K} \frac{N_i^2}{n_i} (1 - f_i)[S_{yi}^2 + R_i^2 S_{xi}^2 - 2 R_i S_{xyi}]$$

$$s^2(Y_{RS}) = \sum_{i=1}^{K} \frac{N_i^2}{n_i} (1 - f_i)[s_{yi}^2 + R_i^2 s_{xi}^2 - 2R_i s_{xyi}]$$

2- المجموع الطبقي النسبي المشترك ويعرف

$$Y_{RC} = \frac{Y_{st}}{X_{st}} X = \frac{\sum_{i=1}^{K} N_i \bar{y}_i}{\sum_{i=1}^{K} N_i \bar{x}_i} X$$

حيث إن Y_{st} و X_{st} تقدير المجموع بالعينة العشوائية الطبقية للمتغيرينy وx وX يمثل المجموع الكلي للمتغيرx.

يلاحظ أن Y_{RC} فقط يحتاج المجموع الكلي X وليس المجموع الكلي لكل طبقه. كذلك لأن النسبة لا نجدها حتى الخطوة الأخيرة سيكون مقدار التحيز أقل مما هو عليه في Y_{RS} . أما التباين وتقدير التباين للتقدير فسيكونا على التوالي:

$$Var(Y_{RC}) = \sum_{i=1}^{K} \frac{N_i^2}{n_i}(1-f_i)[S_{yi}^2 + R^2 S_{xi}^2 - 2R\ S_{xyi}]$$

$$s^2(Y_{RC}) = \sum_{i=1}^{K} \frac{N_i^2}{n_i}(1-f_i)[s_{yi}^2 + R^2 s_{xi}^2 - 2R\ s_{xyi}]$$

وهذا يختلف عن تباين Y_{RS} ، هناك $R_i = \frac{Y_i}{X_i}$ نستخدم، أما هنا فنستخدم $R = \frac{Y}{X}$. وهذا هو الذي غير النتيجة لصالح التقدير النسبي المنفصل ليكون التقدير النسبي المشترك أقل دقة من التقدير النسبي المنفصل.

2.6.7 تقدير الوسط الحسابي للمجتمع باستخدام النسبة

كما هو الحال في تقدير المجموع يوجد لدينا تقديران لتقدير الوسط الحسابي في العينة العشوائية الطبقية وباستخدام النسبة.

1. **الوسط الحسابي الطبقي المنفصل بطريقة النسبة ويعرف**

$$\bar{y}_{RS}=\frac{1}{N}Y_{Rs}=\frac{1}{N}\sum_{I=1}^{K}\frac{\bar{y}_i}{\bar{x}_i}X_i=\sum_{I=1}^{K}W_i\frac{\bar{y}_i}{\bar{x}_i}\bar{X}_i=\sum_{i=1}^{K}W_i\,\bar{y}_{Ri}$$

أما التباين وتقدير التباين لهذا التقدير $\bar{y}_{RS}$ فسيكونا

$$\operatorname{var}(\bar{y}_{RS})=\frac{1}{N^2}\operatorname{Var}(Y_{RS})$$

$$s^2(\bar{y}_{RS})=\frac{1}{N^2}s^2(Y_{RS})$$

2. **الوسط الحسابي الطبقي المشترك بطريقة النسبة ويعرف**

$$\bar{y}_{RC}=\frac{1}{N}Y_{RC}=\frac{1}{N}\frac{\sum_{i=1}^{K}N_i\bar{y}_i}{\sum_{i=1}^{K}N_i\bar{x}_i}X=\frac{\sum_{i=1}^{K}W_i\bar{y}_i}{\sum_{i=1}^{K}W_i\bar{x}_i}\bar{X}$$

أما التباين وتقدير التباين للتقدير $\bar{y}_{RC}$ فيكونا

$$\operatorname{var}(\bar{y}_{RC})=\frac{1}{N^2}\operatorname{var}(Y_{RC})$$

$$s^2(\bar{y}_{RC})=\frac{1}{N^2}s^2(Y_{RC})$$

مثال (4): استخدم الجدول الآتي لتقدير $\bar{Y}$ وY باستخدام النسبة في العينة العشوائية الطبقية. أوجد تقدير الخطأ المعياري للتقديرات وقارنها مع النتائج التي نحصل عليها باستخدام العينة الطبقية بمفردها أي بدون استخدام النسبة.

الطبقة	N_i	W_i	n_i	$\bar{y}_i$	$\bar{x}_i$	X_i	s^2_{yi}	s^2_{xi}	s_{xyi}
1	700	0.2	40	50	32	21000	225	4	24
2	1750	0.5	50	30	18.5	31500	64	2.25	8.4
3	1050	0.3	40	20	7	8400	36	0.64	4.32

$$Y_{RS}=\sum_{i=1}^{K}\frac{\bar{y}_i}{\bar{x}_i}X_i=\sum_{i=1}^{K}R_i\,X_i=107893$$

$$Y_{Rc}=\frac{Y_{st}}{X_{st}}X=\frac{\sum_{i=1}^{K}N_i\,\bar{y}_i}{\sum_{i=1}^{K}N_i\,\bar{x}_i}X=\frac{108500}{62125}(60900)=106361$$

إذا أهملنا معامل التصحيح أي إن $(1-f_i)=1$ فإن

$$s^2(Y_{RS})=\sum_{i=1}^{K}\frac{N_i}{n_i}(1-f_i)[s_{yi}^2+R_i^2\,s_{xi}^2-2R_i\,s_{xyi}]=5026730$$

$$s\ (Y_{RS})=\sqrt{s^2(Y_{RS})}=\sqrt{5026730}=2242$$

$$s^2(\,Y_{RC})=\sum_{i=1}^{K}\frac{N_i}{n_i}(1-f_i)[s_{yi}^2+R^2\ s_{xi}^2-2R\ \ s_{xyi}]=5048091$$

$$s(Y_{RC})=\sqrt{s^2(\,Y_{RC})}=\sqrt{5048091}=2247$$

$$s^2(\bar{y}_{RS})=\frac{1}{N^2}s^2(Y_{RS})=0.641;\quad \bar{y}_{RS}=\sum_{i=1}^{3}W_i\,\bar{y}_{Ri}=30.83$$

$$;\quad s^2(\bar{y}_{RC})=\frac{1}{N^2}s^2(Y_{RC})=0.642\ \ \bar{y}_{RC}=\frac{\sum_{i=1}^{K}W_i\,\bar{y}_i}{\sum_{i=1}^{K}W_i\,\bar{x}_i}\bar{X}=30.39$$

وإذا استخدمنا المتغير y بمفرده في العينة العشوائية الطبقية

فإن تقدير المجموع

$$Y_{St}=108500$$

وتقدير التباين

$$s^2(Y_{st})=7361025$$

نلاحظ أن هناك زيادة في الدقة عندما استخدمنا النسبة في التقدير. وكذلك نلاحظ أن التقدير النسبي المنفصل أكثر دقة من التقدير النسبي المشترك.

7.7 التقديرات باستخدام الفرق في العينة العشوائية البسيطة

توجد طريقة أخرى للاستفادة أو لاستخدام المعلومات الإضافية أو المساعدة لتقدير الوسط الحسابي والمجموع الكلي للمجتمع، تسمى الفروق، وباستخدام هذه الطريقة يمكن أن نعرف تقدير الوسط الحسابي

$$\bar{y}_d = \bar{y} + (\bar{X} - \bar{x})$$

إذا كان الوسط الحسابي للمجتمع والعينة للمتغير X هما $\bar{x}$ و $\bar{X}$ على التوالي و $\bar{y}$ هو الوسط الحسابي للعينة للمتغير y. يمكن ملاحظة أن $\bar{y}_d$ هو تقدير غير متحيز إلى $\bar{Y}$ على شرط أن يكون $\bar{y}$ و $\bar{x}$ تقديرين غير متحيزين للوسطين الحسابيين للمجتمع $\bar{Y}$ و $\bar{X}$ على التوالي.

إن هذا التقدير لا يخلو من صعوبات ويجب استخدامه بحذر. بصورة عامة معلومات المتغير X يجب أن تكون من النوع نفسه والوحدات نفسها كما هي للمتغير y. فمثلاً يمكن أن يكون x يمثل معلومات سريعة عن إنتاج محصول معين وجد عن طريق تقدير سريع لإنتاج هذا المحصول لجميع الحقول، بينما y يمثل معلومات دقيقة وجدت عن إنتاج المحصول نفسه وذلك عن طريق سحب عينة عشوائية من الحقول لتقدير إنتاج هذا المحصول.

إن $\bar{y}_d$ هو تقدير غير متحيز إلى $\bar{Y}$ كما أشرنا إلى ذلك. أما تباين وتقدير التباين إلى $\bar{y}_d$ فهما على التوالي

$$\mathrm{var}(\bar{y}_d) = \frac{1-f}{n}\left(S_y^2 + S_x^2 - 2S_{xy}\right)$$

$$s^2(\bar{y}_d) = \frac{1-f}{n}\left(s_y^2 + s_x^2 - 2s_{xy}\right) = \frac{1-f}{n}\,\frac{\sum_{i=1}^{n}[(y_i - x_i) - (\bar{y} - \bar{x})]^2}{n-1}$$

مثال (5): لأجل القيام بتضمين مزارع الخضراوات يقوم أشخاص لهم خبرة في موضوع تخمين قيمة هذه المزارع. يوجد 180 مزرعة قدرت قيمتها بمبلغ 13320 دولاراً. لنفترض أن x_i ترمز إلى القيمة التخمينية للمزرعة و y_i ترمز إلى القيمة الحقيقية للمزرعة. سحبنا عينه عشوائية حجمها $n = 10$ مزارع أعطتنا النتائج المدونة في الجدول أدناه. استخدم هذه البيانات لتقدير المجموع الكلي لقيمة هذه المزارع باستخدام طريقة الفرق، ثم أوجد معامل التغير.

المزرعة	1	2	3	4	5	6	7	8	9	10
y_i	9	14	7	29	45	109	40	238	60	170
x_i	10	12	8	26	47	112	36	240	59	167

الحل:

$$\bar{y} = 72.1; \bar{x} = 71.7; \bar{X} = 74$$

$$\bar{y}_d = \bar{y} + (\bar{X} - \bar{x}) = 72.1 - (74 - 71.7) = 74.4$$

$$s^2(\bar{y}_d) = \frac{1-f}{n} \frac{\sum_{i=1}^{n}[(y_i - x_i) - (\bar{y} - \bar{x})]^2}{n-1} = 0.5928$$

معامل التغير لتقدير الوسط الحسابي

$$CV(\bar{y}_d) = \frac{s(\bar{y}_d)}{\bar{y}_d} = \frac{\sqrt{0.5928}}{74.4} = 0.013$$

إن تقدير المجموع الكلي بطريق الفرق يعرف

$$Y_d = N\bar{y}_d$$

أما التباين وتقدير التباين إلى Y_d فهما على التوالي

$$var(Y_d) = N^2 var(\bar{y}_d)$$

$$s^2(Y_d) = N^2 s^2(\bar{y}_d)$$

8.7 التقديرات باستخدام الانحدار في العينة العشوائية البسيطة

يستخدم النموذج الخطي أو ما يسمى بخط الانحدار البسيط كثيراً في التحليل الإحصائي. وسوف نستخدمه هنا لتقدير الوسط الحسابي والمجموع للمجتمع. ويمكن التعبير عن العلاقة بين المتغيرين y وx بمعادلة خط الانحدار

$$y_i = \alpha + bx_i + \varepsilon_i$$

حيث إن $(0, \sigma^2)$ ε_i مهما كانت قيم x. وباستخدام النموذج الخطي فإننا نستخدم $\bar{y}_{lr}$ كتقدير إلى $\bar{Y}$ ، عندما تكون β ميل المستقيم معلوم يعرف $\bar{y}_{lr}$ كما يأتي

$$\bar{y}_{lr} = \bar{y} + \beta(\bar{X} - \bar{x})$$

عندما $\beta = 1$ فإن $\bar{y}_d = \bar{y}_{lr}$. واضح أن $\bar{y}_{lr}$ تقدير غير متحيز إلى $\bar{Y}$ مع تباين

$$\operatorname{var}(\bar{y}_{lr}) = \frac{1-f}{n}(S_y^2 + \beta^2 S_x^2 - 2\beta S_{xy})$$

إن قيمة β التي سوف تعطينا أصغر تباين ممكن إلى $\bar{y}_{lr}$ هي

$$\beta_{opt} = \frac{S_{xy}}{S_x^2}$$

سيكون التباين الأمثل إلى $\bar{y}_{lr}$ كما يأتي

$$\operatorname{var}(\bar{y}_{lr}) = \frac{1-f}{n} S_y^2(1-\rho^2)$$

حيث إن

$$\rho^2 = \frac{(S_{xy})^2}{S_x^2 \, S_y^2}$$

و ρ^2 يسمى معامل التحديد.

إن قيمة β تكون غير معلومة في أغلب الأحيان لذلك لا بد من تقديرها باستخدام البيانات التي حصلنا عليها من المسح. نستطيع تقدير قيمة β باستخدام ما يسمى بطريقة المربعات الصغرى

$$\beta = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - [\sum_{i=1}^{n} x_i]^2}$$

الآن نستخدم β بدلاً من β ليصبح على الوجه الآتي

$$\overline{y}_{lr} = \overline{y} + \beta\ (\overline{X} - \overline{x})$$

إن الشروط التي تصاحب النموذج الخطي قد لا تتحقق إذا كنا نسحب عينه من مجتمع محدود. ومن هذه الشروط يجب أن تكون العلاقة بينy وx خطية، كذلك تباينات y_i لقيمة محددة من x تبقى نفسها مهما كانت قيمة x لا بد من الإشارة هنا أنه إذا كان المجتمع صغيراً، وكذلك حجم العينة صغيراً، فإن تجاوز بعض شروط النموذج قد يكون حرجاً. وقد يخلق صغر حجم العينة بعض الصعوبات. ولكن إذا كان النموذج صحيحاً للبيانات فإن $\overline{y}_{lr}$ سيكون أفضل تقدير إلى $\overline{Y}$. وسيكون أفضل حتى من $\overline{y}_R$.

نستطيع أن نقدر تباين $\overline{y}_{lr}$ وذلك باستخدام طريقة المربعات الصغرى فإن

$$s^2(\overline{y}_{lr}) = \frac{1-f}{n}\ \frac{n-1}{n-2}(s_y^2 - \beta\, s_{xy})$$

أما المجموع الكلي للمجتمع فيمكن تقديره

$$Y_{lr} = N\,\overline{y}_{lr} = Y + \beta\,(X - X)$$

وكذلك فإن تقدير تباين يكون

$$s^2(\,Y_{lr}) = N^2\ s^2(\overline{y}_{lr})$$

مثال (6): للمثال (5) أوجد $\bar{y}_{lr}$ و $s(\bar{y}_{lr})$ وقارنها مع النتائج التي حصلت عليها إلى $\bar{y}$ و $\bar{y}_d$ و $\bar{y}_R$.

الحل:

$$\bar{y}_{lr} = \bar{y} + \beta(\bar{X} - \bar{x}) = 72.1 + 0.9925(74 - 71.7) = 74.383$$

$$s^2(\bar{y}_{lr}) = \frac{1-f}{n}\frac{n-1}{n-2}(s_y^2 - \beta s_{xy})$$

$$= \frac{1-(10/180)}{10}\frac{9}{8}[5981.48 - (0.9925)(6020.59)] = .642$$

الجدول الآتي يلخص النتائج التي حصلنا عليها

المقدر	التقدير	تقدير الخطاء المعياري
y بمفرده	$\bar{y} = 72.10$	$s(\bar{y}) = 23.767$
النسبة	$\bar{y}_R = 74.41$	$s(\bar{y}_R) = 0.833$
الفرق	$\bar{y}_d = 74.40$	$s(\bar{y}_d) = 0.770$
الانحدار	$\bar{y}_{lr} = 74.38$	$s(\bar{y}_{lr}) = 0.801$

إن المفاضلة بين $\bar{y}_d$ و $\bar{y}$ كتقديرين للوسط الحسابي للمجتمع يعتمد على ميل المستقيم (أي خط انحدار y على x (فإن $\bar{y}_d$ يفضل على $\bar{y}$ إذا كان ميل المستقيم قريباً من القيمة المثلى أي

$$\beta_{opt} = \frac{S_{xy}}{S_x^2}$$

فإذا كانت

$$\beta_{opt} > \frac{1}{2}$$

يكون

$$\mathrm{var}(\bar{y}_d) < \mathrm{Var}(\bar{y})$$

إذاً $\bar{y}_d$ يفضل على $\bar{y}$. من خلال هذه النتائج، واضح أن $\bar{y}_d$ يفضل على $\bar{y}_R$ و $\bar{y}_{lr}$ وجميعهم يفضلون على $\bar{y}$.

بالأسلوب نفسه نستطيع أن نقارن بين $\bar{y}_d$ و $\bar{y}_R$ لنصل إلى أن $\bar{y}_d$ يفضل على $\bar{y}_R$ إذا كانت

$$\frac{1}{2} < \beta < \frac{R+1}{2}$$

حيث إن β تمثل ميل المستقيم وR نسبة المجتمع. إذا لم تكن β وR معلومتين، نقدرهما من البيانات لنحصل على

$$\beta = \frac{s_{xy}}{s_x^2} \text{ و } R = \frac{\bar{y}}{\bar{x}}$$

إذاً $\bar{y}_d$ يفضل على $\bar{y}_R$ إذا كانت

$$\frac{1}{2} < \beta < \frac{R+1}{2}$$

نلاحظ من النتائج التي حصلنا عليها في الجدول السابق أن $\bar{y}$ يُعد أسوأ تقدير إلى $\bar{Y}$ حيث أعطانا أكبر تقدير للخطأ المعياري بينما التقديرات الأخرى متقاربة من حيث قيمة الخطأ المعياري، ممكن أن يكون أفضلهم $\bar{y}_d$ لهذا المثال.

9.7 التقديرات باستخدام الانحدار والفرق في العينة الطبقية

هناك طريقتان لتقدير الوسط الحسابي في العينة العشوائية الطبقية باستخدام خط الانحدار.

1- الوسط الحسابي الطبقي المنفصل بطريقة الانحدار

يعرف الوسط الحسابي بطريقة الانحدار للطبقة i

$$\bar{y}_{lri} = \bar{y}_i + \beta_i(\bar{X}_i - \bar{x}_i)$$

حيث إن β_i تقدير لميل المستقيم في الطبقة i و $\overline{X}_i$ يمثل الوسط الحسابي للطبقة. إذاً الوسط الحسابي الطبقي المنفصل بطريقة الانحدار يعرف

$$\overline{y}_{lrs} = \sum_{i=1}^{K} W_i \, \overline{y}_{lri} = \sum_{i=1}^{K} W_i \, [\overline{y}_i + \beta_i \, (\overline{X}_i - \overline{x}_i)]$$

أما المجموع الكلي الطبقي المنفصل بطريقة الانحدار

$$Y_{lrs} = N \overline{y}_{lrs}$$

أما تباين $\overline{y}_{lrs}$ يمكن إيجاده

$$\text{var}(\overline{y}_{rs}) = \sum_{i=1}^{K} W_i^2 Var(\overline{y}_{lri}) = \sum_{i=1}^{K} W_i^2 \frac{1-f_i}{n_i} (S_{yi}^2 + \beta_i^2 S_{xi}^2 - 2\beta_i S_{xyi})$$

ويمكن تقدير تباين بما يأتي

$$s^2(\overline{y}_{rs}) = \sum_{i=1}^{K} W_i^2 \, \frac{1-f_i}{n_i} (s_{yi}^2 + \beta_i^2 \, s_{xi}^2 - 2\beta_i \, s_{xyi})$$

2- الوسط الحسابي الطبقي المشترك بطريقة الانحدار

يعرف على الوجه الآتي

$$\overline{y}_{lrc} = \overline{y}_{st} + \beta (\overline{X} - \overline{x}_{st}) = \sum_{i=1}^{K} W_i \, [\overline{y}_i + \beta (\overline{X}_i - \overline{x}_i)]$$

أما المجموع الكلي الطبقي المشترك بطريقة الانحدار فيعرف

$$Y_{lrc} = N \overline{y}_{lrc}$$

أما تباين $\overline{y}_{lrc}$ فيمكن إيجاده

$$Var(\overline{y}_{lrc}) = \sum_{i=1}^{K} W_i^2 \, \frac{1-f_i}{n_i} (S_{yi}^2 + \beta^2 S_{xi}^2 - 2\beta S_{xyi})$$

حيث إن

$$\beta = \frac{\sum_{i=1}^{K} W_i^2 \frac{1-f_i}{n_i} S_{xyi}}{\sum_{i=1}^{K} W_i^2 \frac{1-f_i}{n_i} S_{xi}^2}$$

ويمكن تقدير التباين وذلك باستبدال β و S_{yi}^2 و S_{xi}^2 و S_{xyi} بتقديراتها من العينة وهي على التوالي β و s_{yi}^2 و s_{xi}^2 و s_{xyi} .

النقطة الأخيرة في هذا الفصل إذا عوضنا عن $\beta = 1$ في الوسط الحسابي الطبقي المشترك بطريقة الانحدار نحصل على

$$\bar{y}_{dst} = \bar{y}_{st} + (\bar{X} - \bar{x}_{st})$$

وهذا هو الوسط الحسابي الطبقي المشترك بطريقة الفرق مع تباين

$$Var(\bar{y}_{dst}) = \sum_{i=1}^{K} W_i^2 \frac{1-f_i}{n_i}(S_{yi}^2 + S_{xi}^2 - 2S_{xyi})$$

تمارين

1. لتحديد نسبة المصروف من دخل العائلة الشهري على الطعام، سحبت عينة عشوائية بسيطة حجمها $n=12$ عائلة من منطقة سكنية صغيرة، عدد العائلات فيها $N=200$ والنتائج كما هي مبينة في الجدول الآتي:

العائلة	1	2	3	4	5	6	7	8	9	10	11	12
الدخل	200	170	450	300	600	550	1000	1200	150	220	250	320
المصروف	125	110	200	170	250	240	350	500	105	130	140	160

أ- قدر النسبة R ثم أوجد تقدير الخطأ المعياري.

ب- أوجد فترة 95% ثقة للنسبة R.

2. استخدم البيانات في السؤال السابق علماً بأن معدل دخل العائلة الشهري في القرية $\overline{X} = 455$ أوجد:

أ. تقدير معدل مصروفات العائلة الشهرية على المواد الغذائية بالاعتماد على المعلومات المتوافرة حول مصروفات العائلة على المواد الغذائية وحدها.

ب. استخدم النسبة لتقدير معدل مصروفات العائلة الشهرية على المواد الغذائية.

ت. استخدم خط انحدار y على x لتقدير مصروفات العائلة الشهرية على المواد الغذائية.

ث. أوجد تقدير الخطأ المعياري للحالات الثلاثة أعلاه، ثم علق على النتائج التي حصلت عليها.

3. يقوم أحد الخبراء بتخمين وزن التفاح الموجود على كل شجره، إذا كان لدينا حديقة فيها 100 شجرة، خُمن وزن التفاح لهذه الأشجار x=12500 كيلو غرام، سحبت عينة عشوائية بسيطة حجمها 10 شجرات، وجرى قطف التفاح وجرى وزنه لإيجاد الوزن الحقيقي للتفاح في كل شجرة، الجدول التالي يبين الوزن التخميني والحقيقي لعشر شجرات

الشجرة	1	2	3	4	5	6	7	8	9	10
الوزن الحقيقي y_i	55	42	46	39	71	61	58	57	58	67
الوزن التخميني x_i	56	47	48	40	78	59	52	58	55	67

أ. قدر وزن التفاح الموجود في الحديقة باستخدام طريقة الفرق، ثم أوجد الانحراف المعياري.

ب. قدر وزن التفاح الموجود في الحديقة باستخدام خط انحدارy علىx ثم أوجد الانحراف المعياري للتقدير.

ت. قارن بين الطريقتين، أيُّها في اعتقادك أفضل لتقدير وزن التفاح الموجود في الحديقة؟

4. استخدم البيانات السؤال الأول لتحديد حجم العينة اللازم لتقدير النسبةR إذا كان حد الخطأ لتقدير النسبة d=0.15.

5. استخدم البيانات المعطاة في السؤالين الأول والثاني لتحديد حجم العينة المطلوب، لتقدير الوسط الحسابي للطرق الثلاثة إذا كان مقدار حد الخطأ لتقدير الوسط الحسابيd=0.15.

6. .سحبت عينة عشوائية بسيطة بحجم 5 أشخاص من الذين أنهوا برنامجاً للحمية، الجدول الآتي يبين الأوزان قبل وبعد الالتحاق بالبرنامج.

الشخص	1	2	3	4	5
الوزن قبل البرنامج	120	140	160	130	180
الوزن بعد البرنامج	100	105	110	115	120

أ. قدر R وأوجد فترة 95% ثقة للقيمة الحقيقية للنسبةR.

ب. قدر معدل أوزان الأشخاص الملتحقين بالبرنامج بعد إتمامه ومن ثم أوجد فترة 90% ثقة باستخدام النسبة.

7. لنفترض أن عينة عشوائية بسيطة حجمها 10 طلاب اختيرت من طلبة أحد الأقسام العلمية في الجامعة، وسئلوا عن مقدار دخلهم الشهري (x) ومصروفاتهم الشهرية لشراء الكتب العلمية (y). الجدول الآتي يعطينا إجابات الطلاب. علماً بأن عدد طلاب القسم يبلغ 500 طالب.

الطالب	1	2	3	4	5	6	7	8	9	10
(x)	20	15	22	30	40	35	25	32	42	24
(y)	1100	1200	1300	1400	1600	1300	1250	1300	1650	1200

أ. قدر معدل مصروفات الطلاب في القسم على مشتريات الكتب باستخدام النسبة R ومقدار الخطأ المعياري للتقدير.

ب. قدر معدل مصروفات الطلاب الشهرية لشراء الكتب العلمية باستخدام خط انحدارy على x ومن ثم أوجد مقدار الخطأ المعياري للتقدير.

ت. استخدم y بمفرده لتقدير معدل مصروفات الطلاب الشهرية ولشراء الكتب العلمية ومن ثم أوجد مقدار الخطأ المعياري للتقدير.

ث. قارن بين ما حصلت عليه باستخدام الطرق الثلاث أعلاه ومن ثم أوصِ بأفضلهم لكي تستخدم لاحقاً في مثل هذه التقديرات.

8. ترغب الغرفة الصناعية في تقدير دخل الشركات الصناعية لعام 2005 العاملة ضمن المنطقة الموجودة فيها، يبلغ عدد الشركات الصناعية25 شركة. الدخل لجميع الشركات الصناعية لعام 2001 بلغ 670 مليون دولار. قامت الغرفة بسحب عينة عشوائية حجمها 9 شركات وحصلت على معلومات حول دخل هذه الشركات لعامي 2001 و2005، الجدول الآتي يعطينا الدخل بملايين الدولارات

الشركة	1	2	3	4	5	6	7	8	9
2001	20	19	30	26	25	33	14	39	27
2005	23	24	28	27	30	38	16	44	29

أ. استخدم النسبة لتقدير مجموع دخل الشركات الصناعية لعام 2005 ومن ثم أوجد فترة 95% ثقة للدخل الحقيقي للشركات.

ب. استخدم الفرق لتقدير مجموع دخل الشركات الصناعية لعام 2005 ومن ثم أوجد فترة 95% ثقة للدخل الحقيقي للشركات.

ت. استخدم طريقة الانحدار لتقدير مجموع دخل الشركات الصناعية لعام 2005 ومن ثم أوجد فترة 95% ثقة للدخل الحقيقي للشركات.

ث. أي الطرق الثلاثة الواردة أعلاه أفضل لتقدير مجموع دخل الشركات الصناعية؟ ولماذا؟

9. يوجد في إحدى المدن 500 شركة يعمل في هذه الشركات 500.000 موظف. جرى سحب عينة عشوائية بحجم 5 شركات وحصلنا على البيانات الآتية:

عدد الموظفين(x)	30	12	15	25	35
المصروف على الإنترنت بآلاف الدولارات (y)	10	12	15	25	35

استخدم طريق الانحدار لتقدير المصروفات الشهرية على الإنترنت ومن ثم أوجد فترة 90% ثقة

للقيمة الحقيقية للمصروفات الشهرية على الإنترنت.

10. يقوم أحد المصانع بإنتاج نوعين من السلع لغرض عرضهما في الأسواق. يمكن أن تعد النوعين من السلع، طبقتين لتقدير إجمالي حجم المبيعات من السلعتين. سحبت عينة عشوائية بسيطة من زبائن السلعتين، وسئلوا لإعطاء عدد الوحدات التي يمكن أن يشتروها من السلعتين للسنة القادمة y علماً بأن عدد الوحدات التي قام بشرائها نفس الزبائن من السلعتين للسنة الماضية x متوافرة لدينا، الجدول الآتي يعطينا البيانات للسلعتين. علماً بأن $X_1 = 24500$ مجموع الوحدات التي بيعت السنة الماضية من السلعة الأولى، وعدد الزبائن للسلعة الأولى $N_1 = 120$. كذلك فإن عدد الزبائن للسلعة الثانية للسنة الماضية ومجموع الوحدات التي بيعت من السلعة الثانية هي على التوالي $N_2 = 180$ و $X_2 = 21200$.

السلعة	x_i	204	143	82	256	198				
الأولى	y_i	210	160	75	280	190				
السلعة	x_i	137	189	119	63	103	107	159	63	87
الثانية	y_i	150	200	125	60	110	100	180	25	90

أ. قدر إجمالي المبيعات للسلعتين للسنة القادمة باستخدام طريقة النسبة وخط انحدار y على x والفرق.

ب. أوجد تقدير الخطأ المعياري للطرق الثلاثة أعلاه، ثم وازن بين هذه الطرق.

ت. ما هي الطريقة التي تفضل استخدامها لتقدير إجمالي المبيعات، ولماذا؟

المراجع العربية

1. سليمان محمد طشطوش. (2001) أساسيات المعاينة الإحصائية، دار الشروق للنشر والتوزيع، عمان.

2. حسين علوان (1994). طرق المعاينة، دار الفرقان، عمان.

3. هلال عبود البياتي وصبري رديف العاني (1981). العينات، جامعة بغداد، بغداد.

4. وليد عبدالحميد نوري وعبدالمجيد حمزة الناصر (1981). العينات، وزارة التعليم العالي والبحث العلمي، بغداد.

5. وليم كوكران – ترجمة أنيس كنجو (1995). تقنية المعاينة الإحصائية، جامعة الملك سعود، الرياض.

References

1. Barnett, V. (1991). Sampling Survey Principles and Methods, Edward Arnold, London.
2. Benedetto, J. J. and Ferreira, P. J. (2001). Modern Sampling Theory, Birkhauser Boston.
3. Brewer, K. W. R. (1963). Ratio Estimation in Finite Populations: Some Results Deducible from the Assumption of an Underlying Stochastic Process, *Aust. J. Statistics.* 5, 93-105.
4. Cassel, C. M., Sarndal, C. E. and Wretman, J. H. (1976). Some Results on Generalized Difference and Generalized Regression Estimation for Finite Population, *Biometrika* 63, 615-620.
5. Chaudhuri, A. and Stenger, H. (1992). Survey Sampling: Theory and Methods, Marcel Dekker, New York.
6. Cochran, W. G. (1977). Sampling Technique, 3rd Ed. Wiley, New York.
7. David, I. P. and Sukhatme, B. V. (1974). On the Bias and Mean Square Error of the Ratio Estimator, *J. Amer. Statist. Asso.* 69, 464-466.
8. Deng, L. Y. and Wu, C. F. J. (1987). Estimation of the Variance of the Regression Estimator, *J. Amer. Statist. Asso.* 82, 568-576.
9. Durbin, J. (1959). A Note on the Application of Quenouille's Methods of Bias Reduction to Estimation of Ratios, *Biometrika* 46, 477-480.
10. Foreman, E. K. (1991). Survey Sampling Principles, Marcel Dekker, New York.
11. Goodman, L. A. and Hartley, H. O. (1958). The Precision of Unbiased Ratio-type Estimators. *J. Amer. Statist. Asso.* 53, 491-508.
12. Govindarajulu, Z. (1999). Elements of Sampling Theory and Methods, Prentice Hall.
13. Hajek, J (1981). Sampling from Finite Population, Marcel Dekker, New York.
14. Hansen, M. H., and Hurwitz, W. N. (1943). On the Theory of Sampling from Finite Populations. *Ann. Math. Statist.* 14,333-362.
15. Hedayat, A. S. and Sinha, B. K. (1991). Design and Inference in Finite Population Sampling, Wiley, New York.

16. Hartley, H. O. and Ross, A. (1954). Unbiased Ratio Estimates, *Nature* 174, 270-271.
17. Kish, L., Namboodiri, N. K. and Pillai, R. K. (1962). The Ration bias in Surveys. *J. Amer. Stat. Assoc.* 57, 863-876.
18. Kish, L (1995). Survey Sampling, Wiley, New York.
19. Lahiri, D. B. (1951). A Method for Sample Selection Providing Unbiased Ratio, Estimate. *Bull. Int. Stat. Inst.* 33, 133-140.
20. Levy, P. S. and Lemeshow S. (1991). Sampling of Population: Methods and Applications, Wiley, New York.
21. Lohr, S. L. (1999). Sampling: Design and Analysis, Duxbury, New York.
22. Mickey, M. R. (1959). Some Finite Population Unbiased Ratio and Regression Estimators, *J. Amer. Statist. Asso.* 54, 594-612.
23. Murthy, M. N. (1957). Ordered and Unordered Estimators in Sampling without Replacement. *Sankhya* 18, 379-390.
24. Nanjamma, N. S., Murthy, M. N. and Sethi, V. K. (1959). Some Sampling Systems Providing Unbiased Ratio Estimators, *Sankhya* 21, 299-314.
25. Olkin, I. (1958). Multivariate Ratio Estimation for Finite Populations, *Biometrika* 45, 154-165.
26. Quenouille, M. H. (1956). Notes on Bias in Estimation *Biometrika* 43, 353-360.
27. Raj, D. (1968). Sampling Theory, McGraw Hill, New York.
28. Rao, J. N. K. (1968). Some Small Sample Results in Ratio and Regression Estimation. *J. Ind. Stat. Assoc.*, 6, 160-168.
29. Rao, J. N. K. (1969). Ratio and Regression Estimators. In: New Developments in Survey Sampling (Johnson, N. L. and Smith, H. Jr. Eds.) 213-234, Wiley, New York.
30. Rao, J. N. K. and Vijayan, K. (1977). On Estimating the Variance in Sampling with Probability Proportion to Aggregate Size, *J. Amer. Statist. Asso.* 72, 579-584.
31. Rao, J. N. K. and Webster, J. T. (1966). On Two Methods of Biased Reduction in the Estimation of Ratio, *Biometrika* 53, 315-321.
32. Rao, J. N. K. (1988). Ratio and Regression Estimators. In: Handbook of Statistics, Vol. 6, Sampling (Krishnaiah, P. R. and Rao, C. R, Eds.), 449-468, North Holland, Amsterdam.
33. Rao, P. S. R. S. and Rao, J. N. K. (1971). Small Sample Results for Ratio Estimation, *Biometrika* 58, 625-630.
34. Rao, T. J. (1981a). A Note on Unbiasedness in Ratio Estimators, *J. Statist. Plann. Infère.* , 5, 335-340.
35. Rao, T. J. (1981b). On a Class of Almost Unbiased Ratio Estimators, *Ann. Inst. Statist. Math*, 33, 225-231.
36. Ray, S. K. and Sahai, A. (1980). Efficient Families of Ratio and Product Type Estimators, *Biometrika* 67, 211-215.
37. Royall, R. M. (1970). On Finite Population Sampling Theory under Certain Linear Regression Models, *Biometrika* 57, 377-387.
38. Royall, R. M. and Cumberland, W. G. (1981a). An Empirical Study of the Ratio Estimator and Estimators of its Variance, *J. Amer. Statist. Asso.* 76, 66-77.
39. Royall, R. M. and Cumberland, W. G. (1981b). The Finite Population Linear Regression Estimator and Estimators of its Variance: An Empirical Study, *J. Amer. Statist. Asso.* 76, 924-930.

40. Royall, R. M. and Eberhardt, K. R. (1975).Variance Estimation of the Ratio Estimator, *Sankhya* C 37, 43-52.
41. Sampath, S. (2000). Sampling Theory and Methods, CRC Press.
42. Singh, P. and Srivastava, A. K. (1980). Sampling Schemes Providing Unbiased Regression Estimators, *Biometrika* 67, 205-209.
43. Singh, R. and Sukhatme, B. V. (1973). Optimum Stratification with Ratio and Regression Methods of Estimation, *Ann. Inst. Statist. Math.* 25, 627-633.
44. Scheaffer, R.L., Mendenhall, W. and Ott L. (1996). Elementary Sampling Survey, 5th ed., Duxbury, New York.
45. Smith, T.M.F. (1976). The Foundation of Survey Sampling: A review, *Journal of Royal Statistics Society* A139, 183-204.
46. Srivenkataramana T. (1980). A Dual to Ratio Estimators in Sample Surveys, *Biometrika* 67, 199-204.
47. Sukhatme, P. V., Sukhatme, B. V., Sukhatme, S., and Ashok, C. (1984). Sampling Theory of Surveys with Applications, 3rd ed., Ames (Iowa): Iowa State University Press.
48. Thompson, S. K. (2002). Sampling, 2end Wiley, New York.
49. Tin, M. (1965). Comparison of some Ratio Estimators, *J. Amer. Statist. Asso.* 60, 294-307.
50. Tryfos, P. (1996). Sampling Methods for Applied Research, Wiley, New York.
51. Williams, W. H. (1963). The Precision of some Unbiased Regression Estimators, *Biometrics* 19, 352-361.
52. Yates, F. (1981). Sampling Methods for Censuses and Surveys 4th Ed., Grittin, London

الفصل الثامن

العينة العشوائية المنتظمة

Systematic Random Sample

1.8 مقدمة

لنفرض أننا نرغب في سحب عينة حجمها n من مجتمع يحتوي على N من الوحدات، ويوجد لدينا كشف كامل بأسماء هذه الوحدات، يمكن أن نحدد أعضاء العينة في الكشف ومن ثم نقوم بمشاهدة هذه الوحدات في المجتمع، لجمع المعلومات المطلوبة منها. فمثلاً يمكن أن نختار مجموعة من طلاب الجامعة، وذلك باستخدام الكشوفات التي تنشرها الجامعة بأسماء طلابها، أو يمكن أن نختار مجموعة من الكتب، وذلك باستخدام الكشوفات الموجودة في المكتبة بأسماء الكتب المحفوظة فيها. إن اختيار عينة عشوائية بسيطة يبدو صعباً جداً في مثل هذه الحالات، ويحتاج إلى وقت وجهد كبيرين. لنتصور أننا نرغب في اختيار 500 طالب من بين 15000 طالب مسجلين في إحدى الجامعات، لذلك هناك رغبة كبيرة للبحث عن طريقة بديلة.

إن الطريقة الشائعة الاستخدام في مثل هذه الحالات هي ما تسمى بالعينة العشوائية المنتظمة وهذه تسمى أحيانا العينة المنتظمة. ونستطيع أن نعرف العينة العشوائية المنتظمة على أنها العينة التي يتم اختيارها بواسطة اختيار وحدة واحدة بطريقة عشوائية من بين أول k من العناصر الموجودة في الكشف

أو الإطار، ثم بعد ذلك تؤخذ بقية الوحدات بشكل تكون فيه الوحدات المتتالية ذات أبعاد متساوية فيما بينها والعدد منها مساوٍ إلى k.

لنَعد الآن إلى مثالنا الذي أردنا فيه اختيار 500 طالب من بين 15000 من طلبة الجامعة. لا بد من تحديد قيمة k الذي يمكن أن نجد قيمتها باستخدام $k = N/n$ ففي مثالنا هذا $k = 30$ بطبيعة الحال فإن الكشف بأسماء الطلاب مرقم من 1 إلى 15000 وإذا لم يكن مرقماً بهذه الطريقة فإنه من السهولة القيام بترقيمه. نقوم الآن باختيار رقم وبطريقة عشوائية من بين أول ثلاثين رقماً أي بين 1 إلى30. ولنفرض أننا اخترنا الرقم 7 بطريقة عشوائية فسيكون الطالب رقم 7 في العينة، ومن ثم نقوم بأخذ الطالب من المجموعة الثانية أي من بين الطلاب الذين أرقامهم بين 31 إلى 60 والذي يكون تسلسله $7 + k = 37$ ومن المجموعة الثالثة الطالب الذي تسلسله $7 + 2k = 67$ وهكذا إلى نهاية الكشف.

2.8 طرق اختيار وحدات العينة المنتظمة

نستخدم العينة العشوائية المنتظمة باستمرار، إذا كان هناك إطار كامل وحديث متوافر. سوف نتكلم عن طريقتين لاختيار وحدات العينة المنتظمة من المجتمع.

1.2.8 الطريقة الخطية المنتظمة

إن هذه الطريقة هي الطريقة الشائعة الاستخدام. نفرض أن لدينا مجتمعاً مكوناً من N عنصر، وهذه العناصر مرتبة بطريقة خطية، تعطى هذه العناصر أرقاماً من 1 إلى N ثم نجد قيمة $k = N/n$ ، ونقوم باختيار عنصر من بين 1 إلى k بطريقة عشوائية ولنفرض أننا اخترنا العنصر رقم i حيث إن i أقل أو يساوي k، تسمى k بفترة الاختيار. وسوف تحتوي العينة التي تسحب بهذه

الطريقة على العناصر $i, i+k, i+2k, \ldots, i+(n-1)k$ باستخدام هذه الطريقة سوف نحصل على k من العينات الممكنة السحب. إن الجدول الآتي يعطينا العينات الممكنة السحب حيث إن y_i ترمز إلى قيمة العنصر i.

رقم العينة

1	2	…	i	…	k
y_1	y_2	…	y_i	…	y_k
y_{1+k}	y_{2+2k}	…	y_{i+2k}	…	y_{2k}
⋮	⋮	⋮	⋮	⋮	⋮
$y_{1+(n-1)k}$	$y_{2+(n-1)k}$	…	$y_{i+(n-1)k}$	…	y_{nk}

هناك مشكلة تواجهنا عندما لا تكون $k=N/n$ مساوية إلى عدد صحيح، نقوم بتقريب k إلى أقرب عدد صحيح. إن هذه الطريقة يمكن أن تؤدي لأن تكون العينات الممكنة السحب من المجتمع غير متساوية الحجم.

مثال (1): نفرض أن لدينا مجتمعاً مكوناً $N=11$ وحدة وأردنا أن نسحب عينة منتظمة حجمها $n = 4$ في هذه الحالة $k=11/4=3$.سيكون لدينا العينات الممكنة الظهور هي (1,4,7,10) أو(2,5.8,11) أو(3,6,9). واضح أن العينة الأخيرة عدد وحداتها 3 وهو لا يساوي عدد وحدات العينات الأولى.

2.2.8 طريقة الاختيار الدائرية المنتظمة

لحل مشكلة عدم تساوي عدد الوحدات في العينات الممكنة الظهور عندما تكون $N \neq nk$ يمكن استخدام طريقة الاختيار الدائرية. إن الطريقة الجديدة تتضمن اختيار وحدة بطريقة عشوائية من بين العناصر 1 إلى N ولنفرض أن هذا العنصر رقمه i و $i \le N$ ثم نقوم باختيار العناصر $i+jk$ إذا كانت $i+jk \le N$ لكل $j = 0, 1, 2, \ldots, (n-1)$ وكذلك نختار العناصر

$i+jk-N$ إذا كانت $i+jk>N$ إن قيمة k تكون عدداً صحيحاً أو مقرباً إلى أقرب عدد صحيح.

مثال (2): لنَعُد إلى مثالنا السابق الذي فيه $N=11$ و$n=4$ و$k=3$ لنستخدم الطريقة الدائرية لاختيار جميع العينات الممكنة.

وهي كما يأتي: (4,7,10,2) و(3,6,9,1) و(2,5,8,11) و(1,4,7,10) و(9,1,4,7) و(8,11,3,6) و(7,10,2,5) و(6,9,1,4) و(5.8.11.3) و(10.2.5.8) و(11,3,6,9).

يمكن بسهولة أن نثبت أن كل وحدة من وحدات المجتمع لها نفس الفرصة في الظهور. وكذلك واضح أن كل عينة من العينات الممكنة الظهور وبحجم n أيضاً لها الفرصة نفسها في الظهور.

3.8 مزايا وعيوب العينة المنتظمة

إن من أهم مزايا العينة المنتظمة

1. أنها سهلة الاستخدام أو التنفيذ في الميدان، ولذلك تُعد أقل عرضة لأن يخطئ العاملون في اختيار وحداتها بالمقارنة إلى العينة العشوائية البسيطة أو العينة العشوائية الطبقية خصوصاً إذا لم يكن هناك إطار جيد متوافر.
2. أنها توفر معلومات أكثر لكل وحدة تكلفة بالقياس إلى العينة العشوائية البسيطة.
3. أنها تغطي المجتمع بصورة أفضل من العينة العشوائية البسيطة أو الطبقية. لذلك نراها تستخدم كثيراً في الغابات لتقدير كمية الخشب الموجودة وفي البحيرات لتقدير عدد الأسماك.
4. يمكن استخدام العينة العشوائية المنتظمة حتى ولو كنا لا نعرف قيمةN.

5. أحيانا تعطي تقديراً أكثر دقة من العينة العشوائية البسيطة، خصوصاً إذا كان المجتمع لا يحتوي على دورات معينة أو مرتباً ترتيباً تصاعدياً أو تنازلياً.

أما أهم عيوبها فهي

1. إذا كان المجتمع يحتوي على دورات متكررة فيجب استخدام العينة المنتظمة بحذر شديد.
2. لا يمكن الحصول على تقدير غير متحيز لتباين تقدير الوسط الحسابي أو المجموع الكلي باستخدام عينة واحدة، لأننا استخدمنا العشوائية في الاختيار مرة واحدة، أي باختيار عنصر واحد فقط من عناصر العينة.
3. إذا كان المجتمع يحتوي على دورات غير معروفة لنا فإن استخدام العينة المنتظمة يؤدي إلى تحيز في تقدير الوسط الحسابي أو المجموع.

4.8 تقدير الوسط الحسابي والمجموع الكلي للمجتمع

إن الهدف الرئيس لأكثر المسوحات بالعينة هو تقدير مَعْلَم أو أكثر من مَعَالم المجتمع. لذلك سوف نتكلم عن تقدير الوسط الحسابي والمجموع للمجتمع.

1.4.8 تقدير الوسط الحسابي للمجتمع

يمكننا تقدير الوسط الحسابي للمجتمع باستخدام العينة العشوائية المنتظمة باستخدام الوسط الحسابي للعينة

$$\bar{y}_{sy} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

كذلك يمكن تقدير تباين الوسط الحسابي للعينة $\bar{y}_{sy}$ بما يأتي:

$$s^2(\bar{y}_{sy}) = \frac{N-n}{N}\frac{s^2}{n};\ s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y}_{sy})^2$$

إذا كانت N غير معلومة فإننا نهمل معامل التصحيح للمجتمع المحدود $\frac{N-n}{N}$. إن تقدير الوسط الحسابي للمجتمع باستخدام $\bar{y}_{sy}$ الوسط الحسابي للعينة المنتظمة، وكذلك تقدير التباين للوسط الحسابي للعينة $\bar{y}_{sy}$ مماثل إلى الوسط الحسابي وتقدير تباين الوسط الحسابي للعينة العشوائية البسيطة $\bar{y}$ ولكن هذا لا يعني أن تباين المجتمع للتقديرين متساويان حيث إن تباين $\bar{y}$

$$Var(\bar{y}) = \frac{\sigma^2}{n}\frac{N-n}{N-1} = \frac{S^2}{n}(\frac{N-n}{N})$$

وتباين $\bar{y}_{sy}$

$$Var(\bar{y}_{sy}) = \frac{\sigma^2}{n}[1+(n-1)\rho]$$

حيث إن ρ يمثل معامل الارتباط بين أزواج من وحدات العينة نفسها. واضح أن $Var(\bar{y})$ و $Var(\bar{y}_{sy})$ غير متساويين.

إذا كانت قيمة ρ قريبة من واحد هذا يعني أن العناصر داخل العينة المنتظمة متماثلة إلى حد كبير بالنسبة إلى المتغير الذي نحن بصدد دراسته، وهذا سوف يؤدي إلى أن يكون تباين الوسط الحسابي للعينة المنتظمة أكبر من تباين الوسط الحسابي للعينة العشوائية البسيطة. أما إذا كانت قيمة ρ سالبة فربما تكون العينة المنتظمة أفضل من العينة العشوائية البسيطة. نلاحظ أن قيمة ρ لا يمكن أن تكون كبيرة في السالب وإلا سيؤدي ذلك إلى أن تكون قيمة $Var(\bar{y}_{sy})$ سالبة.

لا نستطيع الحصول على تقدير غير متحيز إلى $Var(\bar{y}_{sy})$ وذلك باستخدام عينة منتظمة واحدة، هذا لا يعني أنه لا يمكن الحصول على تقدير تباين $Var(\bar{y}_{sy})$. عندما تكون العينة المنتظمة قريبة أو مماثلة إلى العينة البسيطة يمكن أن يكون تقدير تباين الوسط الحسابي للعينة المنتظمة $s^2(\bar{y}_{sy})$ قريباً جداً من تقدير تباين الوسط الحسابي للعينة العشوائية البسيطة $s^2(\bar{y})$.

2.4.8 تقدير المجموع الكلي للمجتمع

إن تقدير المجموع الكلي للمجتمع يتطلب معرفة حجم المجتمع N .ولكن في كثير من الحالات تكون N غير معروفة. لذلك كان أحد أسباب استخدام العينة العشوائية المنتظمة هو عدم معرفةN . ولكن إذا كانت N معروفة نستطيع أن نقدر المجموع الكلي للمجتمع باستخدام

$$Y_{sy} = N\bar{y}_{sy}$$

مع تقدير التباين

$$s^2(Y_{sy}) = N^2 s^2(\bar{y}_{sy}) = N^2 \frac{N-n}{N}\frac{s^2}{n}$$

نلاحظ أن هذه النتائج مماثلة لتلك التي حصلنا عليها باستخدام العينة العشوائية البسيطة لتقدير المجموع الكلي للمجتمع.

5.8 مقارنة العينة المنتظمة بالعينة العشوائية البسيطة

إن مقارنة العينة العشوائية المنتظمة بالعينة العشوائية البسيطة يعتمد إلى حد كبير على طبيعة المجتمع تحت الدراسة؛ لذلك سوف نقارن بين طريقتي المعاينة لأنواع محددة من المجتمعات.

1.5.8 المجتمع عشوائي (Random Population)

يمكن أن نقول إن المجتمع عشوائي إذا كانت عناصره لا تتبع ترتيباً معيناً أي مرتبة عشوائياً. إن وحدات العينة العشوائية المنتظمة التي تُسحب من مجتمع عشوائي يتوقع أن تكون مختلفة فيما بينها، أي إن $\rho \approx 0$؛ لذلك إذا كانت N كبيرة فإن $Var(\bar{y}_{sy}) \approx Var(\bar{y})$ أي تباين الوسط الحسابي للعينة المنتظمة مساوٍ تقريباً إلى تباين الوسط الحسابي للعينة العشوائية البسيطة. فمثلاً إذا أردنا أن نقدر المعدل التراكمي لطلبة الجامعة، فإن الإطار عبارة عن كشف بأسماء جميع طلبة الجامعة مرتبين حسب الأحرف الهجائية للاسم الأول للطالب أو الرقم الجامعي. واضح أنه لا توجد علاقة بين ترتيب أسماء الطلاب حسب الأحرف الهجائية أو أرقامهم الجامعية ومعدلاتهم التراكمية، لذلك نستطيع أن نَعُدَّ المجتمع عشوائيًا بالنسبة إلى علامات الطلاب، وفي هذه الحالة تكون العينة العشوائية المنتظمة مماثلة للعينة العشوائية البسيطة.

2.5.8 المجتمع مرتب (Ordered Population)

إن المجتمع يعتبر مرتباً إذا كانت عناصره مرتبة بناءً على مقادير بحسب مخطط معين. فمثلاً في حالات المسح لتقييم تدريس أحد المدرسين في مساق عدد الطلبة المسجلين فيه كبير، في العادة يطلب من الطلبة تقييم مدرسهم وفق نظام عددي من 1 إلى 10 حيث يكون واحد أقل علامة وعشرة أكبر علامة، ومن ثم يقومون بترتيب استمارات الإجابة وفق نظام معين، كأن نبدأ بأعلى العلامات ثم نبدأ في النزول حتى الوصول إلى أقل العلامات، ومن ثم نسحب عينة منتظمة من بين هذه الاستمارات، في هذه الحالة يكون المجتمع مرتبًا ترتيباً تنازلياً.

إن العينة المنتظمة التي تسحب من مجتمع مرتب بصورة عامة تكون عناصرها مختلفة، ومن ثم ستكون قيمة $\rho \leq 0$. ويمكن إثبات أنه إذا كانت N كبيرة وقيمة ρ أقل أو تساوي صفراً فإن $Var(\bar{y}_{sy}) \leq Var(\bar{y})$. هذا يعني أن العينة العشوائية المنتظمة المسحوبة من مجتمع مرتب تعطينا معلومات أكثر لكل وحدة تكلفة من العينة العشوائية البسيطة. نستطيع تقدير $Var(\bar{y}_{sy})$ باستخدام

$$s^2(\bar{y}_{sy}) = \frac{N-n}{N}\frac{s^2}{n}$$

في هذه الحالة يُعد التقدير محافظاً؛ لأنه سوف يكون أكبر مما يفترض أن يكون.

3.5.8 المجتمع دوري (Periodic Population)

يُعد المجتمع دورياً إذا كانت تغيرات عناصره تأخذ شكلاً دورياً. مثلاً إذا أردنا تقدير عدد المصلين الذين يرتادون مسجداً معيناً خلال شهر. بالتأكيد إذا أخذنا عدد المصلين الذين يأتون إلى صلاة الفجر يومياً مدة شهر سوف يكون تقديرنا لعدد المصلين أقل مما يجب، لأن عدد المصلين الذين يحضرون صلاة الفجر أقل من عددهم في الصلوات الأخرى. وبالمقابل لو أننا أخذنا عدد المصلين الذي يحضرون صلاة الظهر سوف يكون تقديرنا لعدد المصلين أكثر مما يجب، حيث إن عدد المصلين الذين يحضرون صلاة الظهر في المسجد يكون أكبر من بقية الصلوات وخصوصاً الذين يحضرون صلاة الجمعة. لذلك يُعد هذا النوع من المجتمعات مجتمعاً دورياً.

إن عناصر العينة المنتظمة المسحوبة من مجتمع دوري يمكن أن تكون متماثلة أو قريبة من بعضها البعض، ومن ثم سوف تكون قيمة ρ أكبر من صفر $(\rho > 0)$. على سبيل المثال إن عدد المصلين الذين يصلون صلاة الفجر في

المسجد ستكون قريبة من بعضها البعض خلال أيام الشهر. يمكن أن نثبت إذا كانت N كبيرة و $\rho > 0$ فإن $\mathrm{Var}(\bar{y}_{sy}) > \mathrm{Var}(\bar{y})$. لذلك فإن العينة المنتظمة تعطينا معلومات أقل لكل وحدة كلفة بالمقارنة بالعينة العشوائية البسيطة إذا كان المجتمع دورياً.

للتخلص من مشكلة المجتمع الدوري يمكن أن نغير نقطة البداية العشوائية عدة مرات. إن هذه الطريقة سوف تقلل فرصة اختيار وحدات من المجتمع عندها نفس شكل التغير في المجتمع. فمثلاً في مثالنا أعلاه نستطيع أن نختار نقطة البداية بطريقة عشوائية، أي نختار صلاة من بين الصلوات الخمس خلال اليوم الأول من الأسبوع الأول، فمثلاً ظهرت صلاة المغرب نستمر مدة أسبوع بجمع بيانات حول عدد المصلين الذين يحضرون صلاة المغرب في المسجد حتى نهاية الأسبوع الأول، ومن ثم نختار وبطريقة عشوائية مرة أخرى إحدى الصلوات من بداية الأسبوع الثاني، ونجمع بيانات حول عدد المصلين الذين يصلون هذه الصلاة في المسجد مدة أسبوع آخر، وهكذا إلى نهاية الشهر.

يمكننا أن نفترض أن العينة المنتظمة التي تسحب بهذه الطريقة تكون مماثلة للعينة المنتظمة التي تسحب من مجتمع عشوائي، ومن ثم نستطيع أن نقدر تباين $\bar{y}_{sy}$ باستخدام

$$s^2(\bar{y}_{sy}) = \frac{N-n}{N}\frac{s^2}{n}$$

6.8 تقدير النسبة

غالباً ما نرغب في استخدام البيانات التي حصلنا عليها باستخدام العينة العشوائية المنتظمة لتقدير النسبة في المجتمع P ويمكن الحصول عليها إذا عرفنا y_i على الوجه الآتي

$y_i = 1$ إذا كانت الوحدة تحمل صفة مميزة

$y_i = 0$ خلاف ذلك أي إن الوحدة لا تحمل الصفة المميزة

فيكون $\hat{p}_{sy}$ تقدير النسبة في العينة العشوائية المنتظمة

$$\hat{p}_{sy} = \bar{y}_{sy} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{a}{n}$$

حيث إن a عدد الوحدات في العينة الذين يحملون الصفة المميزة. أما تقدير التباين فيكون

$$s^2(\hat{p}_{sy}) = \frac{N-n}{N}\frac{\hat{p}_{sy}\hat{q}_{sy}}{n-1} = (1-f)\frac{\hat{p}_{sy}\hat{q}_{sy}}{n-1}$$

يمكن أن نهمل $(1-f)$ معامل التصحيح للمجتمعات المحدودة إذا كانت N كبيرة أو غير معلومة. نلاحظ هنا أن تقدير النسبة في العينة العشوائية المنتظمة مماثل لتقدير النسبة في العينة العشوائية البسيطة التي سبق أن تكلمنا عنها. لكن هذا لا يعني أن تباين المجتمع $Var(\hat{p}_{sy})$ للعينة المنتظمة مساوٍ $Var(\hat{p})$ للعينة العشوائية البسيطة. ومع ذلك إذا كانت N كبيرة وكذلك البيانات في العينة المنتظمة غير مرتبطة، أي $\rho = 0$ سيكون تباين المجتمعات للنسبتين $\hat{p}$ و $\hat{p}_{sy}$ متقاربين.

7.8 اختيار حجم العينة

إن تحديد حجم العينة من الأمور المهمة جداً في المسح بالعينة، وذلك لارتباط حجم العينة بالدقة ودرجة الثقة المطلوبتين للتقدير. ناهيك عن علاقة حجم العينة بتكاليف المسح بالعينة وسرعة تنفيذه.

1.7.8 اختيار حجم العينة لتقدير الوسط الحسابي

لتحديد حجم العينة المطلوب لتقدير الوسط الحسابي للمجتمع $\overline{Y}$ سوف نتبع الطريقة نفسها التي سبق أن اتبعناها في الفصول السابقة. لنفرض أننا حددنا مقدار الخطأ في التقدير بما يساوي d ومقدار الثقة بما يساوي 95% أي

$$P(\left|\overline{Y} - \overline{y}_{sy}\right| < d) = 0.95$$

فإن حجم العينة المطلوب يمكن إيجاده بحل المعادلة الآتي لإيجاد قيم n

$$2\sqrt{Var(\overline{y}_{sy})} = d$$

هذه المعادلة سوف تعتمد على قيمة S^2 و ρ اللذين يجب أن يكونا معلومين، على الأقل، القيمة التقريبية لهما، ويمكن إيجاد القيمة التقريبية باستخدام معلومات سابقه متوافرة من مسوحات مماثلة أو بسحب عينة أولية عشوائية من المجتمع. سوف لا نستخدم هذه المعلومات لإيجاد قيمة S^2 و ρ ولكن سوف نستخدم حجم العينة التي يمكن إيجادها في حالة العينة العشوائية البسيطة. وهذه الطريقة سوف تعطينا حجم عينة أكبر إذا كان المجتمع مرتباً وحجم عينة أقل إذا كان المجتمع دورياً؛ لاحظنا أن تباين الوسط الحسابي للعينة العشوائية البسيطة والوسط الحسابي للعينة العشوائية المنتظمة متقاربان إذا كان المجتمع عشوائياً. لذلك فإن حجم العينة المطلوب سيكون

$$n \geq \frac{NS^2}{ND + S^2}$$

حيث إن $D = \frac{d^2}{4}$ و S^2 على الأرجح تكون غير معلومة نحددها باستخدام معلومات سابقة أو سحب عينة عشوائية أولية.

2.7.8 اختيار حجم العينة لتقدير المجموع الكلي

سوف نتبع الطريقة نفسها التي اتبعناها في العينة العشوائية البسيطة لتحديد حجم العينة لتقدير المجموع الكلي للمجتمع Y وذلك باختيار d ليكون حد الخطأ لتقدير المجموع الكلي ودرجة الثقة لتكون95%؛ لذلك فإننا نجد قيمة n بحل المعادلة

$$2\sqrt{N^2 Var(\overline{y}_{sy})} = d$$

لذلك فإن

$$n \geq \frac{NS^2}{ND + S^2}$$

حيث إن $D = \frac{d^2}{4N^2}$ ، ونقدر S^2 إذا كانت غير معلومة باستخدام معلومات سابقة أو سحب عينة عشوائية أولية.

مثال (3): إذا كانت شركة توزيع الكهرباء على المستهلكين في إحدى المدن ترغب في تقدير معدل وقت تأخير دفع فواتير الكهرباء من قبل المستهلكين. إن العينة العشوائية المنتظمة تبدو ملائمة لسحب عينة من بين 2500 فاتورة متأخرة الدفع. لقد أجرت البحث نفسه في السنة الماضية فوجدت أن $s^2 = 100$. ما هو حجم العينة الذي يلزمنا لتقدير الوسط الحسابي للمجتمع $\overline{Y}$ الذي يمثل معدل وقت تأخر دفع الفواتير. علماً بأنه تم تحديد حد الخطأ في التقدير $d = 2$ يوم.

الحل:

لابد من إيجاد قيمة

$$D = \frac{d^2}{4} = \frac{(2)^2}{4} = 1$$

و

$$n \geq \frac{NS^2}{ND+S^2} \approx \frac{Ns^2}{ND+s^2} = \frac{(2500)(100)}{2500+100} = 96.19$$

لذلك لا بد لإدارة الشركة أن تسحب عينة حجمها $n = 97$ لتقدير معدل وقت التأخير في دفع الفواتير المتأخرة الدفع.

3.7.8 اختيار حجم العينة لتقدير النسبة

كذلك فإن الطريقة التي سوف تستخدم هنا لاختيار حجم العينة لتقدير النسبة في العينة العشوائية المنتظمة هي نفسها المستخدمة في العينة العشوائية البسيطة؛ ولذلك إذا حددنا حد الخطأ في التقدير بمقدار d ومقدار الثقة بمقدار 95% أي

$$P(|P-\hat{p}_{sy} < d|) = 0.95$$

نقوم بحل المعادلة

$$2\sqrt{Var(\hat{p}_{sy})} = d$$

لإيجاد قيمة n

$$n \geq \frac{NPQ}{ND+PQ}$$

حيث إن $D = \frac{d^2}{4}$ و $Q = 1-P$. ولا بد من الإشارة هنا إلى أن قيمة P تكون غير معلومة ولا بد من تقديرها باستخدام معلومات سابقة أو سحب عينة عشوائية أولية. ويمكن استخدام $P=0.5$ في كل الأحوال، حيث إنه إذا استخدمنا $P=0.5$ فسوف يعطينا أكبر حجم عينة ممكناً وهذا يجعلنا

مطمئنين إلى أن حجم العينة الذي حصلنا عليه هو أكبر حجم يمكن الحصول عليه.

مثال (4): لمعرفة رأي سكان مدينة معينة في إنتاج إحدى الشركات المحلية فيما إذا كان مقبولاً أم لا. قرر الباحث سحب عينة منتظمة من سجلات الأحوال المدنية في المدينة، حيث هنالك $P=5000$ اسم لشخص عمره فوق 16 سنة مسجل في هذه السجلات. لقد تم تحديد مقدار الخطأ في تقدير النسبة للمجتمع P بما يساوي 3% أي $d=0.03$. يبدو أنه لا توجد هنالك معلومات عن قيمة P لذلك قرر الباحث استخدام $P=0.5$.

الحل:

لابد من إيجاد قيمة

$$D=\frac{d^2}{4}=\frac{(0.03)^2}{4}=0.000225$$

و

$$n\ ^{3}\ \frac{NPQ}{ND+PQ}=\frac{(5000)(0.5)(0.5)}{5000(0.000225)+(0.5)(0.5)}=909.01$$

لا بدمن الإشارة هنا إلى أن قيمة $n=910$ هو أكبر قيمة ممكنة لحجم العينة حيث إن $P=0.5$ تعطينا هذه النتيجة كما أشرنا إلى ذلك سابقاً.

8.8 تكرار العينة المنتظمة

لقد سبق أن قلنا: إننا لا نستطيع أن نقدر تباين الوسط الحسابي للعينة المنتظمة $\bar{y}_{sy}$ باستخدام عينة واحدة. لذلك استخدمنا العينة العشوائية البسيطة لتقدير تباين $\bar{y}_{sy}$. ولكن في كثير من الحالات العملية، العينة المنتظمة لا تماثل العينة العشوائية البسيطة. إذاً لا بد من استخدام طريقة بديلة لتقدير تباين $\bar{y}_{sy}$. إن تكرار العينة العشوائية المنتظمة هو أحد الطرق التي يمكن

استخدامها. وكما هو واضح من التسمية لا بد من تكرار العينة المنتظمة للحصول على أكثر من عينة واحدة. فعلى سبيل المثال في مجتمع يحتوي على $N=250$ وحدة، نستطيع أن نحصل على 10 عينات حجم كل عينة 5 وحدات بسحب وحدة واحدة من بين كل 50 وحدة من وحدات المجتمع، ونرمز لها $50 - \text{in} - 1 \quad 10$ وهذا يساوي عدد الوحدات لعينة واحدة إذا سحبنا وحدة واحدة من بين كل 5 وحدات، ونرمز لها $5 - \text{in} - 1 \quad 1$.

نلاحظ أن الطريقتين سوف تعطينا عدد الوحدات نفسها التي سوف نسحبها من المجتمع لتقدير الوسط الحسابي للمجتمع $\overline{Y}$. ولكن الطريقة الثانية سوف تعطينا $n_s = 10$ عينات نستطيع أن نستخدمها لتقدير $\text{Var}(\overline{y}_{sy})$.

لتوضيح فكرة اختيار العينات العشوائية المنتظمة المكررة من المجتمع لنفرض أن لدينا مجتمعاً حجمه $N = 960$ وحدة والتي يمكن أن نعطيها أرقاماً من 1 إلى 960. إذا أردنا اختيار عينة منتظمة حجمها $n=60$ نختار $k=N/n=16$ ومن ثم نختار عدداً عشوائياً بين 1 إلى 16 ليكون نقطة البداية العشوائية للعينة، ونختار بقية الوحدات بالأسلوب الذي تكلمنا عنه سابقاً. الآن ما هي الطريقة التي يمكن استخدامها لاختيار $n_s = 10$ عينات منتظمة بدلاً من العينة العشوائية أعلاه. نبدأ بتحديد قيمة $k' = n_s k$، لقد سبق أن حددنا $k = 16$ إذاً $k' = 10(16) = 160$. إن الخطوة الثانية التي نقوم بها هي اختيار 10 أرقام وبطريقة عشوائية من بين الأرقام الواقعة بين 1 إلى 160 ليكون كل رقم من هذه الأرقام نقطة البداية العشوائية للعينات المنتظمة $n_s = 10$. ومن ثم نكمل كل عينة بطريقة منتظمة كما مر معنا سابقاً، ولكن $k' = 160$ لنحصل على 10 عينات حجم كل عينة 6 وحدات.

لنفرض أننا سحبنا 10 أرقام بطريقة عشوائية بين 1 إلى 160 وحصلنا على الأرقام التالية $66, 18, 6, 22, 50, 70, 140, 120, 90, 155$ نكمل كل العينات لنحصل على الجدول 1.

جدول (1) اختيار 10 عينات منتظمة مكررة

العينة	نقطة البداية العشوائية	العنصر الثاني	العنصر الثالث	العنصر الرابع	العنصر الخامس	العنصر السادس
1	6	166	326	486	646	806
2	18	178	338	498	658	818
3	22	182	342	502	662	822
4	50	210	370	530	690	850
5	66	226	386	546	706	866
6	70	230	390	550	710	870
7	90	250	410	570	730	890
8	120	280	440	600	760	920
9	140	300	460	620	780	940
10	155	315	475	635	795	955

لقد استخدمنا $n_s = 10$ وذلك للحصول على عدد كافٍ من الأوساط الحسابية للعينات وهذا هو القدر الذي نحتاجه لتقدير $Var(\bar{y}_{sy})$. لقد اخترنا $k' = 160$ لتعطينا العدد نفسه من العناصر الذي حصلنا عليه باختيار عينة منتظمة واحدة $1 - in - k$ و $k = 16$ وأخيراً فإن $k' = kn_s = 160$.

لتقدير الوسط الحسابي للمجتمع $\bar{Y}$ باستخدام $(1 - in - k)$ n_s عينة منتظمة سوف نستخدم

$$\bar{y}_{syw} = \frac{1}{n_s} \sum_{i=1}^{n_s} \bar{y}_i$$

حيث إن $\bar{y}_i$ يمثل الوسط الحسابي للعينة المنتظمة i و $i = 1, 2, ..., n_s$ وتقدير تباين $\bar{y}_{syw}$

$$s^2(\bar{y}_{syw}) = \frac{N - n}{N} \frac{\sum_{i=1}^{n_s} (\bar{y}_i - \bar{y}_{syw})^2}{n_s(n_s - 1)}$$

وكذلك يمكن استخدام العينات العشوائية المنتظمة المكررة لتقدير المجموع الكلي للمجتمع Y

$$Y_{syw} = N\bar{y}_{syw} = N\sum_{i=1}^{n_s}\frac{\bar{y}_i}{n_s}$$

مع تقدير لتباين Y_{syw}

$$s^2(Y_{syw}) = N^2 s^2(\bar{y}_{syw})$$

و لمزيد من المعلومات مع بعض الأمثلة راجع (1996) Scheaffer, Retal.

تمارين

1. لنفترض أن أحد البنوك لديه كشف بأسماء الزبائن الذين حصلوا على قرض من البنك، مسلسل بحسب تاريخ منح القرض من قبل البنك للعشرين سنة الماضية. هناك اتجاه لدى الزبائن بعدم دفع ما تبقى بذمتهم، وذلك لارتفاع تكاليف المعيشة للسنوات الأخيرة. يرغب البنك في تقدير مجموع المبالغ المستحقة وغير المدفوعة. ماذا تقترح له؟ هل يستخدم العينة العشوائية البسيطة أم العينة العشوائية المنتظمة ولماذا؟
2. تقوم إحدى الشركات بوضع أسماء العاملين فيها مرتبين حسب الأحرف الأبجدية ضمن كل فئة معينة لمقدار الدخل، وهذه الفئات تبدأ من أعلى الدخول إلى أقل الدخول، إذا كان الهدف تقدير معدل دخل العامل، هل نستخدم العينة العشوائية المنتظمة أم العينة العشوائية الطبقية، أم العينة العشوائية البسيطة لسحب عينة من المجتمع؟ وضح إجابتك.
3. قام قسم مراقبة الجودة في أحد المصانع باستخدام العينة المنتظمة، لتقدير معدل وزن المادة المستخدمة لصناعة علبة مشروبات غازية محمولة على حزام الإنتاج للمصنع. إن البيانات في الجدول الآتي تمثل عينة عشوائية منتظمة 1-in-50 لإنتاج المصنع ليوم واحد. قدر، ثم أوجد فترة 90% ثقة للوسط الحسابي . لنفرض أن N=1800.

وزن العلبة بالغرامات

11.80	12.01	12.03	12.01	11.97	12.00
11.83	12.00	11.98	12.03	11.98	11.91
11.88	11.90	11.87	11.98	12.01	11.87
11.89	11.94	11.93	11.91	11.87	12.05
12.05	11.93	11.97	11.95	11.93	11.72
12.04	12.02	12.05	11.87	11.98	11.85

4. استخدام البيانات في السؤال الثالث لتحديد حجم العينة المطلوب لتقدير الوسط الحسابي إذا كان حد الخطأ لتقدير الوسط الحسابي هو d=0.7 ودرجة ثقة 95%.
5. إذا كان قسم مراقبة الجودة في السؤال الثالث يهمه معرفة ما إذا كانت العلبة مطابقة للمواصفات الموضوعة لها أم لا، فقام بفحص نفس العينة فوجد أن 6 علب في العينة لا تنطبق عليها المواصفات. قدر نسبة العلب الصالحة للاستعمال، ثم أوجد معامل التغير لتقدير النسبة.
6. يعطي الجدول الآتي أعداد العاملين في 11 شركة صناعية في إحدى المدن جرى سحبها باستخدام العينة المنتظمة من بين 300 شركة عاملة في المدينة.

الشركة	1	2	3	4	5	6	7	8	9	10	11
عدد العاملين	110	50	202	60	90	304	240	92	160	220	20

أ. قدر مجموع العاملين في القطاع الصناعي في هذه المدينة.
ب. أوجد تقدير التباين للمجموع الكلي ومن ثم أوجد 95% فترة ثقة للقيمة الحقيقية لعدد العاملين في القطاع الصناعي في هذه المدينة.
ت. قارن بين تقدير التباين في الفرع ب أعلاه وتقدير التباين للمجموع الكلي باستخدام العينة العشوائية البسيطة.

7. ترغب إحدى كليات المجتمع في تطوير علاقتها مع أهل المنطقة المحيطة بالكلية. قامت الكلية بسحب عينة منتظمة 1-in-300 من كشف بأسماء طلبة الكلية يحتوي على4500 طالب وسألتهم عن مقدار ما صرفوا على الملابس في الفصل الماضي. البيانات الآتية تمثل مقدار ما صرفه الطلاب الذين تم سحبهم بالعينة:

30, 60, 20, 15, 28, 15, 24, 34, 10, 50, 20, 25, 33, 48, 23

أ. قدر معدل مصروف طلبة الكلية على الملابس وأوجد فترة 95% ثقة للقيمة الحقيقية لمعدل المصروفات الشهرية على الملابس.

ب. قدر مجموع المصروف على الملابس من قبل طلبة الكلية، ثم أوجد فترة 95% ثقة للقيمة الحقيقية لمجموع المصروف على الملابس من قبل طلبة الكلية.

8. استخدم البيانات في السؤال السابق لتحديد حجم العينة لتقدير مجموع المصروف خلال الفصل ليكون ضمن 1000± وبدرجة ثقة 95%.

9. تتوقع دائرة السير أن تتجاوز إحدى نقاط التفتيش على الأقل N=4000 سيارة. حدد حجم العينة لتقدير النسبة P إذا كان حد الخطأ في التقدير d= 0.015 ودرجة الثقة 95%.

المراجع العربية

1. سليمان محمد طشطوش (2001)، أساسيات المعاينة الإحصائية، دار الشروق للنشر والتوزيع، عمان.
2. حسين علوان (1994)، طرق المعاينة، دار الفرقان، عمان.
3. هلال عبود البياتي وصبري رديف العاني (1981)، العينات، جامعة بغداد، بغداد.
4. وليد عبد الحميد نوري وعبد المجيد حمزة الناصر (1981) العينات، وزارة التعليم العالي والبحث العلمي، بغداد.
5. وليم كوكران –ترجمة أنيس كنجو– تقنية المعاينة الإحصائية (1995)، جامعة الملك سعود، الرياض.

References

1. Barnett, V. (1991). Sampling Survey Principles and Methods, Edward Arnold, London.
2. Benedetto, J. J. and Ferreira, P. J. (2001). Modern Sampling Theory, Birkhauser Boston.
3. Bellhouse, L. D. (1988). Systematic Sampling. In: Handbook of Statistics, Vol. 6, Sampling (Krishnaiah, P. R. and Rao, C. R., Eds.), 125-145, North Holland, Amsterdam.
4. Bellhouse, L. D. and Rao, J. N. K. (1975). Systematic Sampling in the Presence of a Trend. *Biometrika* 62, 694- 697.
5. Brewer, K. W. R. (1963). A Model of Systematic sampling with Unequal Probabilities. *Australian Jour. Stat.* 5, 5-13.
6. Chaudhuri, A. and Stenger, H. (1992). Survey Sampling: Theory and Methods, Marcel Dekker, New York.
7. Cochran, W. G. (1963). Relative Accuracy of Systematic and Stratified Random Samples for a Certain Class of Populations. *Ann. Math. Statist.* 17, 164-177.
8. Cochran, W. G. (1977). Sampling Technique, 3rd Ed. Wiley, New York.
9. Das, A. C. (1950). Two-dimensional Systematic Sampling and the Associated Stratified Random Sampling, *Sankhya*, 10, 95-108.
10. Deming, W. E. (1960). Sampling Design in Business Research. Wiley, New York.
11. Finney, D. J. (1948). Random and Systematic Sampling in Timber Surveys. *Forestry*, 22, 1-36.
12. Foreman, E. K. (1991). Survey Sampling Principles, Marcel Dekker, New York.
13. Gautschi, W. (1957). Some Remarks on Systematic Sampling. *Ann. Math. Statist.* , 28, 385-394.
14. Govindarajulu, Z. (1999). Elements of Sampling Theory and Methods, Prentice Hall.
15. Goodman, R. and Kish, L. (1950). Controlled Selection-a Technique in Probability Sampling. *J. Amer. Statist. Assoc.*, 45, 350-372.
16. Hajek, J (1981). Sampling from Finite Population, Marcel Dekker, New York.

17. Hansen, M. H., and Hurwitz, W. N. (1943). On the Theory of Sampling from Finite Populations. *Ann. Math. Statist.* 14, 333-362.
18. Iachan, R. (1982). Systematic Sampling a Critical Review. *Internat. Statist. Rev.* 50, 293-303.
19. Hedayat, A. S. and Sinha, B. K. (1991). Design and Inference in Finite Population Sampling, Wiley, New York.
20. Kish, L (1995). Survey Sampling, Wiley, New York.
21. Koop, J. C. (1963). On Splitting a Systematic Sample for Variance Estimation. *Ann. Math. Statist.* 42, 1084-1087.
22. Koop, J. C. (1988). The Technique of Replicated or Interpenetrating Samples, In: Handbook of Statistics, Vol. 6, Sampling (Krishnaiah, P. R. and Rao, C. R., Eds.), 125-145, North Holland, Amsterdam.
23. Kunte, S. (1978). A Note on Circular Systematic Sampling Desig, *Sankhya*, C40, 72-73.
24. Levy, P. S. and Lemeshow S. (1991). Sampling of Population: Methods and Applications, Wiley, New York.
25. Lohr, S. L. (1999). Sampling: Design and Analysis, Duxbury, New York.
26. Madow, W. G. and Madow, L. H. (1944). On the Theory of Systematic Sampling. *Ann. Math. Statist.* 15, 1-24.
27. Madow, L. H. (1946). Systematic Sampling and its Relation to other Sampling Designs. *J. Amer. Statist. Assoc.*, 41, 207-214.
28. Madow, L. H. (1949). In the Theory of Systematic Sampling II. *Ann. Math. Statist.* 20, 333-354.
29. Madow, W. G. (1953). On the Theory of Systematic Sampling III: Comparison of Centered and Random Start Systematic Sampling. *Ann. Math. Statist.* 24, 101-106.
30. Mahalanobis, P. C. (1946). Recent Experiments in Statistical Sampling in the Indian Statistical Institute. *J. Roy. Statist. Soc.* A109, 325-378, reprinted in Sankhyii.
31. Milne, A. (1959). The Centric Systematic Area Sample Treated as A Simple Random Sample. *Biometrics,* 15, 270-297.
32. Murthy, M. N. and Rao, T. J. (1988). Systematic Sampling with Illustrative Examples, In: Handbook of Statistics, Vol. 6, Sampling (Krishnaiah, P. R. and Rao, C. R., Eds.), 125-145, North Holland, Amsterdam.
33. Raj, D. (1968). Sampling Theory, McGraw Hill, New York.
34. Rao, J. N. K. (1985). Conditional Inference in Survey Sampling. Survey Methodology, D. I. *Inst. Statist. Math.* No.1, 15-31.
35. Sampath, S. (2000). Sampling Theory and Methods, CRC Press.
36. Scheaffer, R.L., Mendenhall, W. and Ott L. (1996). Elementary Sampling Survey, 5th ed., Duxbury, New York.
37. Singh, D., Jindal, K. and Grag, J. N. (1968).On Modified Systematic Sampling, *Biometrika,* 55, 541-546.
38. Singh, D. and Singh, P. (1977). New Systematic Sampling. *J. Statist. Plann. Infer.* 1, 163-178.
39. Sukhatme, P. V., Sukhatme, B. V., Sukhatme, S., and Ashok, C. (1984). Sampling Theory of Surveys with Applications, 3rd ed., Ames (Iowa): Iowa State University Press.
40. Thompson, S. K (2002). Sampling, 2end Wiley, New York.

41. Tornqvist, L. (1963). The Theory of Replicated Systematic Cluster Sampling with Random Start. *Rev. Internat. Statist. Inst.*, 31, 11-23.
42. Wu, C. F. (1984). Estimation in Systematic Sampling with Supplementary Observations. Sankhya, B46, 306-315.
43. Yates, F. (1981). Sampling Methods for Censuses and Surveys 4th Ed., Grittin, London.
44. Yates, F. (1948). Systematic Sampling, *Phil. Trans. Roy. Soc. London*, A 241, 345-371.
45. Zinger, A. (1964). Systematic Sampling in Forestry. *Biometrics*, 20, 553-565.

الفصل التاسع

العينة العنقودية بمرحلة واحدة

Single – Stage Cluster Sample

1.9 مقدمة

إن الهدف للمسح بالعينة مصمم للحصول على معلومات محددة عن معالم المجتمع بأقل كلفة ممكنة، وكما رأينا أن العينة العشوائية الطبقية أفضل من العينة العشوائية البسيطة للأسباب التي ذكرناها في حينها، سوف نتناول في هذا الفصل العينة العنقودية في مرحلة واحدة والتي في بعض الأحيان تعطي معلومات أكثر دقة لكل وحدة تكاليف من طرق المعاينة التي تكلمنا عنها سابقاً.

يمكن أن نعرف العينة العنقودية بمرحلة واحدة على أنها عينة عشوائية بسيطة، تكون فيها وحدات المعاينة عبارة عن مجموعة من العناصر أو عنقود منها، لقد رأينا أن العينة العشوائية البسيطة والعينة العشوائية الطبقية تحتاج إلى إطار المعاينة، ولكن في الغالب تكون عملية الحصول أو تكوين إطار للمعاينة شاقة ومكلفة كثيراً. وأحياناً لا يوجد إطار ولا يمكن تكوينه. لذا فإن السبب الأول لاستخدام العينة العنقودية هو عدم توافر إطار للمعاينة، أو لكون الحصول عليه يكلف كثيراً. أما السبب الثاني فهو أن تكاليف الحصول على المعلومات تزداد كلما تباعدت عناصر المجتمع عن بعضها الآخر.

توجد أمثله كثيرة لاستخدامات العينة العنقودية، سوف أكتفي بذكر مثال واحد سأشرح من خلاله عملية سحب العينة العنقودية في مرحلة واحدة. لنفرض أننا نريد أن نقدر معدل دخل العائلة في إحدى المدن، إن عملية إعداد إطار لجميع وحدات المجتمع التي تمثل هنا العائلات الساكنة في المدينة قد تكون صعبة أو مكلفة خصوصاً إذا كان عدد سكان المدينة كبيراً، غالباً نستطيع الحصول على كشف بقطاعات المدينة السكنية. لذا سوف نَعُدَّ أن كل قطاع يمثل وحدة من وحدات المجتمع، وطبعاً كل قطاع يحتوي على عدد من العائلات. يمكن أن نعرف عدد القطاعات الموجودة في المدينة، ولتقدير معدل دخل العائلة سوف نقوم بسحب عينة عشوائية بسيطة من القطاعات، ومن ثم نقوم بجمع معلومات عن جميع العائلات الموجودة في هذه القطاعات التي سحبت في العينة. ونستخدم هذه المعلومات لتقدير معدل دخل العائلة في المدينة. مما لا شك فيه أن العائلات الموجودة في كل قطاع ستكون متجاورة ومتجانسة تقريباً مع بعضها؛ لذا سيؤدي هذا إلى خفض تكاليف المعاينة.

2.9 تقدير الوسط الحسابي وا لمجموع الكلي للمجتمع

إن العينة العنقودية ما هي إلا عينة عشوائية بسيطة في كل وحدة من وحدات المعاينة توجد مجموعة من العناصر، لذلك سوف تكون عملية تقدير مَعَالم المجتمع مشابهة إلى حد ما لتلك التي حصلنا عليها في العينة العشوائية البسيطة، وخصوصاً تقدير الوسط الحسابي للمجتمع.

سوف نقوم باستخدام الرموز والمصطلحات الآتية في هذا الفصل

N : عدد العناقيد أو المجموعات في المجتمع.

n: عدد العناقيد أو المجموعات التي اختيرت باستخدام العينة العشوائية البسيطة.

m_i : عدد العناصر في العنقود i و $i = 1, 2, \ldots, N$

$M=\sum_{i=1}^{N} m_i$: عدد العناصر في المجتمع.

$\overline{M}=M/N$: معدل حجم العنقود في المجتمع.

$\overline{m}=\frac{1}{n}\sum_{i=1}^{n} m_i$: معدل حجم العنقود في العينة.

y_i : مجموع المشاهدات للمتغير y في العنقود i.

1.2.9 تقدير الوسط الحسابي للمجتمع

سوف نستخدم الوسط الحسابي للعينة للمتغير y، والذي سنرمز له $\overline{y}_{cs}$ كتقدير للوسط الحسابي للمجتمع ويعرف

$$\overline{y}_{cs}=\frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} m_i}$$

أما تقدير تباين $\overline{y}_{cs}$ فيكون

$$s^2(\overline{y}_{cs})=\frac{N-n}{N\,n\,\overline{M}^2}\,\frac{\sum_{i=1}^{n}(y_i-m_i\,\overline{y}_{cs})}{n-1}$$

نلاحظ أن $\overline{y}_{cs}$ يأخذ شكل تقدير النسبة R والتي سبق أن تكلمنا عنها، كلٌّ ما في الأمر استبدلنا m_i بدلاً عن x_i. لذلك من السهولة ان نتوصل إلى أن $\overline{y}_{cs}$ تقدير متحيز إلى $\overline{Y}$ الوسط الحسابي للمجتمع. وتباينه $Var(\overline{y}_{cs})$ بأخذ شكل تباين النسبة R. إن تقدير التباين $s^2(\overline{y}_{cs})$ يعتبر تقديراً متحيزاً. لكن تقل كمية التحيز كلما كبر حجم العينة. هنا إذا كانت $n>20$ تعتبر

كبيرة. ولكن إذا كان حجم العناقيد متساوياً أي $m_1 = m_2 = ... = m_n = m$ يصبح تقديراً غير متحيز.

مثال (1): لتقدير المدخولات السنوية للعائلة في إحدى المدن، سحبت عينة عشوائية مكونة من 25 قطاعاً من بين 415 قطاعاً موجوداً في المدينة. استخدم البيانات الموجودة في الجدول الآتي لتقدير معدل دخل العائلة، وكذلك أوجد 95% فترة ثقة لمعدل لمدخولات العائلة السنوية.

مجموع الدخول (y_i)	عدد العائلات (m_i)	العنقود (i)	مجموع الدخول (y_i)	عدد العائلات (m_i)	العنقود (i)
49,000	10	14	96,000	8	1
53,000	9	15	121,00	12	2
50,000	3	16	42,000	4	3
32,000	6	17	65,000	5	4
22,000	5	18	52,000	6	5
45,000	5	19	40,000	6	6
37,000	4	20	75,000	7	7
51,000	6	21	65,000	5	8
30,000	8	22	45,000	8	9
39,000	7	23	50,000	3	10
47,000	3	24	85,000	2	11
41,000	8	25	43,000	6	13
			54,000	5	13

الحل:

$$\bar{y}_{cs} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} m_i} = \frac{1,329,000}{151} = 8,801$$

كذلك

$$\sum_{i=1}^{n}(y_i - m_i \bar{y}cs)^2 = \sum_{i=1}^{25} y_i^2 - 2\bar{y}_{cs}\sum_{i=1}^{25} y_i m_i + \bar{y}_{cs}^2 \sum_{i=1}^{25} m_i^2$$

$$= 82,039,000,000 - 2(8,801)(8,403,000) + (8,801)^2(1,047)$$

$$= 15,227,502,247$$

بما أن $\bar{M}$ غير معروفة نقدرها باستخدام

$$\bar{m} = \frac{1}{n}\sum_{i=1}^{25} m_i = \frac{151}{25} = 6.04$$

$$s^2(\bar{y}_{cs}) = \frac{N-n}{N\,n\,\bar{M}^2}\,\frac{\sum_{i=1}^{n}(y_i - m_i\,\bar{y}_{cs})}{n-1} = \frac{(415-25)}{(415)(25)(6.04)^2}\,\frac{15,227,502,247}{24} = 653,785$$

لذا فإن تقدير الخطأ المعياري

$$s(\bar{y}_{cs}) = \sqrt{653,785} = 808.57$$

وكذلك فترة 95% ثقة تقريبية للوسط الحسابي $\bar{Y}$

$$\bar{y}_{cs} \pm z_{\alpha/2} s(\bar{y}_{cs})$$

$$8,801 \pm 1.96(808.57); \quad (7216.2,\ 10385.79)$$

نلاحظ أن فترة الثقة كبيرة.بالإمكان تصغيرها؛ وذلك بزيادة حجم العينة.

2.2.9 تقدير المجموع الكلي للمجتمع

سوف نستخدم طريقتين لتقدير المجموع الكلي للمجتمع

1. **تقدير المجموع الكلي بالاعتماد على M**

سنرمز إلى تقدير المجموع الكلي هنا Y_{csm} ويعرف

$$Y_{csm} = M\bar{y}_{cs}$$

أما تقدير تباينه فيمكن إيجاده كما يأتي

$$s^2(Y_{csm}) = M^2 s^2(\overline{y}_{cs})$$

ملاحظة: إن هذا التقدير مفيدٌ إذا كان مجموع عناصر المجتمع M معلوم.

مثال (2): استخدم البيانات في مثال (1) لإيجاد تقدير المجموع الكلي للمجتمع. إذا علم أن هناك 2.500 عائلة في المدينة. وكذلك أوجد 90% فتره ثقة للمجموع الكليY.

$$Y_{csm} = M\overline{y}_{cs} = 2500(80801) = 22002,500$$

كذلك فإن 90% فترة ثقة للمجموع Y

$$(2,2002,500 \mp 3,325,243) \;;$$
$$22,002,500 \pm 1.645\sqrt{(2,500)^2(653,785)}$$

2- تقدير المجموع الكلي دون الاعتماد على M

غالباً ما تكون عدد عناصر المجتمع غير معلومة، لذلك لا نستطيع استخدام Y_{csm} ، ولكن نستطيع إيجاد تقديرٍ آخر لا يعتمد علىM. لنفرض أن $\overline{y}_c$ على أنه معدل مجموع العناقيد في العينة ويعرف

$$\overline{y}_c = \frac{1}{n}\sum_{i=1}^{n} y_i$$

يعتبر $\overline{y}_c$ تقديراً غير متحيز إلى معدل مجموع العناقيد في المجتمع. كذلك يمكن أن نقدر المجموع الكلي Y بما يأتي

$$Y_{cs} = N\overline{y}_c = \frac{N}{n}\sum_{i=1}^{N} y_i$$

نلاحظ أن Y_{cs} تقدير غير متحيز إلى Y مع تقدير للتباين:

إذا

$$s^2(Y_{cs}) = N^2 s^2(\overline{y}_{cs}) = N^2 \frac{N-n}{Nn} \frac{\sum_{i=1}^{n}(y_i - \overline{y}_c)^2}{n-1}$$

كان هناك تغيرٌ كبيرٌ بين حجم العناقيد، وكان حجم العناقيد له علاقة قوية مع مجموع العناقيد سيكون تباين Y_{cs} بصورة عامة أكبر من تباين Y_{csm}. كذلك نلاحظ أن Y_{cs} لم يستخدم المعلومات المتوافرة حول حجم العناقيد $m_1, m_2, \ldots, m_n$ لذلك يعتبر أقل دقة من Y_{csm} الذي استخدم هذه المعلومات.

مثال (3): استخدم البيانات في مثال (1) لتقدير المجموع الكلي وكذلك أوجد فترة 90% ثقة للمجموع الكلي

الحل:

$$Y_{cs} = \frac{N}{n}\sum_{i=1}^{n} y_i = \frac{415}{25}(1,328,000) = 22,061,400$$

$$\sum_{i=1}^{N}(y_i - \bar{y}_c)^2 = \sum_{i=1}^{n} y_i^2 - \frac{1}{n}(\sum_{i=1}^{n} y_i)^2 = 82,039,000,000 - \frac{1}{25}(1,329,000)^2$$
$$= 11,389,360,000$$

كذلك فإن فترة 90% ثقة للمجموع Y

$$Y_{cs} \pm z_{\alpha/2}\sqrt{s^2(Y_{cs})}$$

$$22,061,400 \pm 1.645\sqrt{(415)^2\ \frac{415-25}{415(25)}\ \frac{11,389,360,000}{24}}$$

$$22,061,400 \pm 2,883,619$$

3.9 تقدير الوسط الحسابي والمجموع إذا كان حجم العناقيد متساوياً

توجد حالات كثيرة يكون فيها حجم العناقيد متساوياً أي $m_1 = m_2 = \ldots = m_n = m$ على سبيل المثال: يقوم أحد المصانع للمعلبات بوضع كل 64 علبة من إنتاجه في صندوق واحد، فإذا كان إنتاج المصنع يومياً 1000

صندوق، وأرادت إدارة مراقبة الإنتاج أن تسحب عينة عشوائية مقدارها 30 صندوقاً يومياً وفحصها. نلاحظ هنا أن N=1000 وn=30 و $m_1 = m_2 = ... = m_{30} = 64$ وM=mN=64000.

سيكون تقدير الوسط الحسابي للمجتمع مع تقدير للتباين في هذه الحالة كما يأتي

$$\bar{y}_{cs} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} m_i} = \frac{1}{nm}\sum_{i=1}^{n} y_i$$

$$s^2(\bar{y}_{cs}) = \frac{N-n}{nNm^2}\,\frac{\sum_{i=1}^{n}(y_i - m\bar{y}_{cs})^2}{n-1}$$

أما تقدير المجموع الكلي ففي هذه الحالة M=mN ستكون معلومة في الغالب. لذلك فإن تقدير المجموع الكلي سيكون

$$Y_{cs} = M\bar{y}_{cs} = M\frac{1}{nm}\sum_{i=1}^{n} y_i$$

أما تقدير التباين فسيكون

$$s^2(Y_{cs}) = M^2 s^2(\bar{y}_{cs})$$

مثال (4): أراد بائع صحف في إحدى المناطق السكنية المؤلفة من 4000 منزل أن يقدر عدد الصحف التي تشترى من قبل سكان كل منزل. إذا علمنا أن المنطقة مؤلفة من 400 قطاع، كل قطاع يحتوي على 10 منازل. قام بسحب عينة عشوائية مؤلفة من 4 قطاعات وحصل على النتائج الآتية. أوجد فترة 95% ثقة للوسط الحسابي $\bar{Y}$.

المجموع	عدد الصحف المشتراة من قبل المنازل المختلفة										القطاع
19	1	1	2	1	3	3	2	1	4	1	1
20	1	3	2	2	3	1	4	1	1	2	2
16	2	1	1	1	1	3	2	1	3	1	3
20	1	1	3	2	1	5	1	2	3	1	4

الحل:

$$\bar{y}_{cs} = \frac{1}{nm}\sum_{i=1}^{n} y_i \frac{19+20+16+20}{4(10)} = 1.875$$

$$\sum_{i=1}^{4}(y_i - m\bar{y}_{cs})^2 = \sum_{i=1}^{4} y_i^2 - nm^2 \bar{y}_{cs}^2$$

$$= (19)^2 + (20)^2 + (16)^2 + (20)^2 - 4(10)^2(1.875)^2 = 10.75$$

$$s^2(\bar{y}_{cs}) = \frac{N-n}{nNm^2} \frac{\sum_{i=1}^{n}(y_i - m\bar{y}_{cs})^2}{n-1} = \frac{(400-4)}{400(4)(10)^2} \frac{10.75}{3} = 0.0089$$

لذا فإن فترة 95% ثقة للوسط الحسابي $\bar{Y}$

$$\bar{y}_{cs} \pm z_{\alpha/2}\sqrt{s^2(\bar{y}_{cs})}$$

$$1.875 \mp 1.96\sqrt{0.0089}; \quad (1.875 \mp 0.1849)$$

4.9 اختيار حجم العينة لتقدير الوسط الحسابي وا لمجموع الكلي

تتأثر كمية المعلومات في العينة العنقودية بعدد العناقيد وحجم العنقود النسبي. وكما لاحظنا فإن تقدير الخطأ المعياري يعتمد على مقدار التشتت بين مجاميع العناقيد، لذلك فإن أي محاوله لتقليل الخطأ المعياري لا بد فيها من اختيار عناقيد يكون التشتت بين مجاميعها قليلاً.

1.4.9 اختيار حجم العينة لتقدير الوسط الحسابي

سوف نفترض أن حجم العنقود (وحدات المعاينة) تم اختيارها وبقي أن نختار عدد العناقيد n يمكن كتابة تباين المجتمع إلى $\bar{y}_{cs}$

$$Var(\bar{y}_{cs}) = \frac{N-n}{nN\bar{M}^2} S_c^2$$

بحيث إن S_c^2 تباين المجتمع إلى

$$s_c^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - m_i\,\bar{y}_{cs})^2$$

بما أن S_c^2 غير معلومة سيتم تقديرها باستخدام s_c^2. حيث إن S_c^2 و $\bar{M}$ غير معلومات لذا فإن اختيار حجم العينة، أي عدد العناقيد تعتريه صعوبات، وللتغلب على هذه الصعوبات فإننا نستخدم الأسلوب نفسه الذي اتبعناه في تقدير النسبة R ، وذلك بأن نقوم بتقدير S_c^2 و $\bar{M}$ إما باستخدام بيانات لتجارب سابقه مماثلة أو باستخدام عينة عشوائية أولية.

سوف نستخدم الأسلوب نفسه الذي سبق أن استخدمناه في اختيار حجم العينة حيث نقوم بحل المعادلة

$$2\sqrt{Var(\bar{y}_{cs})} = d$$

لإيجاد قيمة n. هنا d ترمز إلى حد خطأ التقدير. نستطيع إيجاد حجم العينة التقريبي لتقدير $\bar{Y}$ مع حد لخطأ التقدير مقداره d ودرجة ثقة 95% بما يأتي

$$n \geq \frac{NS_c^2}{ND + S_c^2}$$

حيث إننا نقدر S_c^2 و $\bar{M}$ باستخدام s_c^2 و $\bar{m}$ على التوالي، وكذلك فإن

$$D = \frac{d^2\,\bar{M}}{4}$$

مثال (5): استخدم الجدول في مثال (1) لإيجاد حجم العينة المطلوب بحيث يكون معدل دخل العائلة ضمن 500.

لا بد من تقدير S_c^2 باستخدام البيانات المتوافرة حول 25 عائلة باعتبارها عينة عشوائية أولية

$$s_c^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - m_i \bar{y}_{cs})^2 = 634,479,260$$

كذلك نقدر $\bar{M}$ باستخدام $\bar{m} = 6.04$. إذاً سيكون حجم D التقريبي

$$D = \frac{d^2 \bar{m}^2}{4} = \frac{(500)^2(6.04)^2}{4} = 2,280,100$$

$$n \geq \frac{Ns_c^2}{ND + s_c^2} = \frac{(415)(634,479,260)}{(415)(2,280,100) + 634,479,260} = 167$$

2.4.9 اختيار حجم العينة لتقدير المجموع الكلي

كما مر معنا في حالة تقدير المجموع كان هناك طريقتان لتقدير المجموع الكلي. كذلك توجد طريقتان لاختيار حجم العينة لتقدير المجموع.

1. **عندما تكون M معلومة**

إذا كان حجم المجتمع M معلوماً فإن حجم العينة المطلوب لتقدير المجموع يكون

$$n \geq \frac{NS_c^2}{ND + S_c^2}$$

حيث إن S_c^2 تقدر باستخدام s_c^2 وذلك باستخدام معلومات سابقة أو سحب عينة عشوائية أولية وكذلك فإن

$$D = \frac{d^2}{4N^2}$$

مثال (6): استخدم الجدول في مثال (1) لإيجاد حجم العينة المطلوب لتقدير Y إذا كان M=2500 وd=1.000.000 ودرجة ثقة 95%.

الحل: أوجدنا في المثال السابق $s_c^2 = 634,479,260$ ، فإن

$$D = \frac{d^2}{4N^2} = \frac{(1,000,000)^2}{4(415)^2} = 1,415,589.5$$

$$n \;^{3} \frac{Ns_c^2}{ND + s_c^2} = \frac{(415)(634,479,260)}{(1,451,589.5)(415) + 634,479,260} = 213$$

2. **عندما تكون M معلومة**

لقد سبق أن قدرنا تباين Y_{cs} باستخدام

$$s^2(Y_{cs}) = N^2 \frac{N-n}{nN} s_t^2$$

حيث إن

$$s_t^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y}_c)^2$$

لذا فإن تباين المجتمع إلى Y_{cs} سيكون

$$Var(Y_{cs}) = N^2 \frac{N-n}{nN} S_t^2$$

حيث تم تقدير S_t^2 باستخدام s_t^2. إذاً نقدرY ضمن حد خطأ للتقدير مقداره d يمكن إيجاد حجم العينة الذي يحقق لنا هذه الشروط بحل المعادلة

$$2\sqrt{Var(Y_{cs})} = d$$

لنحصل على

$$n \geq \frac{NS_t^2}{ND + S_t^2}$$

وبما أن S_t^2 غير معلومة نقدرها باستخدام s_t^2 وذلك باستخدام معلومات سابقة أو سحب عينة عشوائية أولية. كذلك فإن

$$D = \frac{d^2}{4N^2}$$

مثال (7):لنفرض أن M غير معلوم في مثال (1) كم حجم العينة الذي نحتاجه لتقدير المجموع إذا كان مقدار حد الخطأ المسموح به d=1,000,000

الحل: باستخدام المعلومات في الأمثلة السابقة نستطيع بسهولة إيجاد قيمة n كما يأتي

$$s_t^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y}_c)^2 = \frac{1}{24}(11,389,360,000) = 474,556,667$$

$$D = \frac{d^2}{4N^2} = \frac{(1,000,000)^2}{4(415)^2} = 2,903,179$$

$$n \geq \frac{NS_t^2}{ND + S_t^2} = \frac{(415)(474,556,667)}{(415)(2,903,179) + 474,556,667} = 183$$

5.9 تقدير النسبة

لنفرض أننا نرغب في تقدير نسبة المجتمع في مثالنا الخاص بتقدير معدل دخل العائلة، يمكن أن يرغب الباحث في تقدير نسبة العائلات التي تعتمد في دخولها على فرد واحد من أعضاء العائلة. إن أفضل تقدير لنسبة المجتمع P هو نسبة العينة $\hat{p}_{(cs)}$ حيث تعرف

$$\hat{p}_{(cs)} = \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} m_i}$$

a_i : ترمـز إلى مجمـوع العناصـر في العنقـود i ممّـن يحملـون صـفه معينـة و m_i يمثل عدد العناصر في العنقود.

ملاحظ: $\hat{p}_{(cs)}$ مشابه لـ $\bar{y}_{cs}$ الذي مر علينا سابقاً ما عدا أننا استخدمنا a_i بدلاً من y_i. إن تقدير تباين النسبة $\hat{p}_{(cs)}$ يكون

$$s^2(\hat{p}_{cs}) = \frac{N-n}{nN\overline{M}^2}\,\frac{\sum_{i=1}^{n}(a_i - m_i\,\hat{p}_{cs})^2}{n-1}$$

مثال (8): عـودة إلى الجـدول الخـاص بالمثـال (1) إذا كانـت a_i تمثـل عـدد العـائلات في العنقـود i والـذين يعتمـدون في دخلـهم علـى فـرد واحـد مـن أعضـاء العائلة هي كما يأتي على التوالي:

1,7,4,3,0,4,3,3,1,3,2,4,1,4,5,2,3,1,2,3,2,4,4,3,3. أوجـد تقـدير النسـبة P، كذلك أوجد فترة 95% ثقة للنسبة لتقدير النسبةP.

الحل:

$$\hat{p}_{(cs)} = \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} m_i} = \frac{72}{151} = 0.477$$

لتقدير التباين لا بد من حساب

$$\sum_{i=1}^{n}(a_i - m_i\,\hat{p}_{cs})^2 = \sum_{i=1}^{n} a_i^2 - 2\hat{p}_{cs}\sum_{i=1}^{n} a_i m_i + \hat{p}_{cs}^2\sum_{i=1}^{n} m_i^2$$

$$= 262 - 2(0.477)(511) + (0.477)^2(1047) = 12.729$$

نقدر $\overline{M}$ باستخدام

$$\overline{m} = \frac{1}{n}\sum_{i=1}^{n} m_i = \frac{151}{25} = 6.04$$

$$s^2(\hat{p}_{cs}) = \frac{415-25}{(415)(25)(6.04)^2}\,\frac{12.729}{24} = 0.00055$$

إن فترة 95% ثقة للنسبة لتقدير P

$$\hat{p}_{cs} \pm z_{\alpha/2}\sqrt{s^2(\hat{p}_{cs})}$$

$$0.477 \mp 1.96\sqrt{0.0055}\,; \quad 0.477 \mp 0.046$$

6.9 اختيار حجم العينة لتقدير النسبة

إن إيجاد حجم العينة لتقدير النسبة P يمكن إيجاده باستخدام الطريقة نفسها التي أوجدنا فيها حجم العينة لتقدير $\overline{Y}$ ، لذا كان حد خطأ التقدير d ودرجة الثقة 95%، وذلك بحل المعادلة

$$2\sqrt{Var(\hat{p}_{cs})} = d$$

لإيجاد قيمة n

$$n \geq \frac{NS_c^2}{ND + S_c^2}$$

حيث إن

$$D = \frac{d^2\,\overline{M}^2}{4}$$

و S_c^2 يجري تقديرها باستخدام

$$s_c^2 = \frac{1}{n-1}\sum_{i=1}^{n}(a_i - \hat{p}_{cs}\,m_i)^2$$

وذلك باستخدام معلومات سابقة أو سحب عينة عشوائية أولية.

مثال (9): عودة إلى المثال (8) أوجد حجم العينة المطلوبة لتقدير P ضمن حد لخطأ للتقدير مقداره 0.04 ودرجة ثقة 95%.

الحل: إن أفضل طريقة لتقدير s_c^2 باستخدام

$$s_c^2 = \frac{1}{n-1}\sum_{i=1}^{n}(a_i - \hat{p}_{cs}\,m_i)^2 = \frac{12.729}{24} = 0.53$$

كذلك نقدر $\overline{M}$ باستخدام $\overline{m} = 6.04$ إذن

$$D = \frac{d^2 \overline{m}^2}{4} = \frac{(0.04)^2 (6.04)^2}{4} = 0.0146$$

$$n \geq \frac{N s_c^2}{ND + s_c^2} = \frac{(415)(0.53)}{(415)(0.0146) + 0.53} = 34$$

تمارين

1. ترغب شركة لإنتاج ماكينات الخياطة أن تقدر معدل تكاليف تصليح الماكينات التي تبيعها إلى مصانع مختلفة شهرياً، لا تستطيع أن تحصل على تكاليف التصليح لكل ماكينة، ولكن تستطيع الحصول على مجموع ما أنفقته على التصليح وعدد الماكينات الموجودة في كل مصنع يستخدم ماكيناتها. لذلك قررت أن عينة عنقودية تكون ملائمة، حيث إن كل مصنع يمثل عنقوداً. سحبت عينة عشوائية بسيطة حجمها 10 مصانع من 90 مصنعاً. يعطي الجدول الآتي عدد الماكينات ومجموع تكاليف التصليح لكل مصنع. قدر معدل تكاليف التصليح للماكينة وأوجد فترة 95% ثقة للوسط الحسابي للمجتمع $\overline{Y}$.

المصنع	1	2	3	4	5	6	7	8	9	10
عدد الماكينات	3	7	10	9	2	12	15	3	5	8
تكاليف التصليح	50	110	220	140	60	280	240	45	60	230

2. قدر مجموع ما تنفقه الشركة على التصليح ثم أوجد فترة 95% ثقة للمجموع الكلي لتكاليف التصليح في السؤال الأول.
3. بعد مراجعة سجلات البيع للشركة توصلوا إلى أن مجموع ما باعت الشركة للمصانع المختلفة في السؤال الأول هو 800 ماكينة خياطة. استخدم هذه المعلومات الجديدة لتقدير مجموع ما أنفقته الشركة على التصليح، ثم أوجد فترة 95% ثقة لمجموع تكاليف التصليح.
4. نفس الشركة في السؤال الأول ترغب بتحديد حجم العينة n لتقدير معدل تكاليف التصليح للشهر القادم. كم عدد العناقيد التي يجب أن

يختاروها لتقدير معدل تكاليف التصليح إذا كان حد الخطأ في التقدير d=2.

5. ترغب إدارة شؤون الطلاب تقدير نسبة الطلاب من أهل المدينة التي تتواجد فيها الجامعة والساكنين في سكن الجامعة. يتكون سكن الجامعة من بناية واحدة من أحد عشر طابقاً، كل طابق يحتوي على 70 غرفة. لنفرض أننا قمنا بسحب أربعة طوابق بطريقة عشوائية وحصلنا على النتائج الآتية:

الطابق	2	5	8	10
عدد الطلبة	10	12	5	17

قدر نسبة الطلاب من أهل المدينة والساكنين في سكن الجامعة ومن ثم أوجد فترة 95% ثقة للنسبة الحقيقية لعدد الطلاب الساكنين في سكن الجامعة من أهل المدينة التي توجد فيها الجامعة.

6. حدد أحد الباحثين الاقتصاديين بحثاً لتقدير معدل مصروفات العائلة على المنافع في إحدى المدن، ولكن لا يوجد كشف بجميع العائلات الساكنة في المدينة، لذلك قرر سحب عينة عشوائية بسيطة بحجم 11 قطاعاً من مجموع قطاعات المدينة البالغة 60 قطاعاً. قام العادون بجمع معلومات عن مصروفات العائلات على المنافع لكل القطاعات التي سحبت منها العينة. الجدول الآتي يعطينا عدد العائلات في كل قطاع ومجموع ما أنفق على المنافع. قدر معدل ما تنفقه العائلة على المنافع في المدينة، ثم أوجد معامل التغير لهذا التقدير.

القطاع	1	2	3	4	5	6	7	8	9	10	11
عدد العائلات	55	60	63	58	71	78	69	57	52	71	73
المجموع	2210	2390	2430	2380	2761	3000	2780	2370	2000	2800	2900

7. قدر مجموع ما تنفقه المدينة على المنافع، علماً بأن عدد العائلات القاطنة في المدينة غير معروف، ثم أوجد معامل التغير لهذا التقدير للبيانات في السؤال الخامس.
8. قرر الباحثون القيام بتقدير مجموع ما تنفقه مدينة مجاورة على المنافع مع حد خطأ في التقدير d=4000. استخدم البيانات الموجودة في السؤال الخامس لتحديد حجم العينة المطلوب للقيام بالبحث في المدينة المجاورة.
9. ترغب إحدى شركات سيارات الأجرة في إحدى المدن تقدير نسبة الإطارات الموجودة في سياراتها، والتي تُعد تالفة أو غير صالحة للاستعمال، إذا كان لدى الشركة 180 سيارة. قامت بسحب عينة عشوائية بسيطة بحجم n=25 باعتبار أن كل سيارة تمثل عنقوداً بأربعة إطارات. فوجدت عدد الإطارات غير الصالحة كما يأتي

2, 4, 0, 1, 0, 2, 4, 1, 3, 1, 2, 0, 1, 1, 2, 2, 4, 1, 0, 0, 3, 1, 2, 2, 1

قدر نسبة الإطارات غير الصالحة ثم أوجد فترة 95% ثقة للنسبة الحقيقية.

المراجع العربية

1. سليمان محمد طشطوش (2001)، أساسيات المعاينة الإحصائية، دار الشروق للنشر والتوزيع، عمان.
2. حسين علوان (1994)، طرق المعاينة، دار الفرقان، عمان.
3. هلال عبود البياتي وصبري رديف العاني (1981)، العينات، جامعة بغداد، بغداد.
4. وليد عبدالحميد نوري وعبدالمجيد حمزة الناصر (1981) العينات، وزارة التعليم العالي والبحث العلمي، بغداد.
5. وليم كوكران –ترجمة أنيس كنجو– تقنية المعاينة الإحصائية (1995)، جامعة الملك سعود، الرياض.

References

1. Barnett, V. (1991). Sampling Survey Principles and Methods, Edward Arnold, London.
2. Chaudhuri, A. and Stenger, H. (1992). Survey Sampling: Theory and Methods, Marcel Dekker, New York.
3. Cochran, W. G. (1977). Sampling Technique, 3rd Ed. Wiley, New York.
4. Cornfield, J. (1951). The Determination of Sample Size. *Amer. J. Publ. Healt*h, 41, 654-661.
5. Deming, W. E. (1960). Sampling Design in Business Research. Wiley, New York.
6. Durbin, J. and Sturat, A. (1954). Callbacks and Cluster Sampling Survey: An Experimental Study. *J. Roy. Stat. Soc.* A117, 387-428.
7. Foreman, E. K. (1991). Survey Sampling Principles, Marcel Dekker, New York.
8. Govindarajulu, Z. (1999). Elements of Sampling Theory and Methods, Prentice Hall.
9. Hajek, J (1981). Sampling from Finite Population, Marcel Dekker, New York.
10. Hansen, M. H., and Hurwitz, W. N. (1943). On the Theory of Sampling from Finite Populations. *Ann. Math. Statist.* 14, 333-362.
11. Hedayat, A. S. and Sinha, B. K. (1991). Design and Inference in Finite Population Sampling, Wiley, New York.
12. Kish, L (1957). Confidence Limits for Clustered Samples. *Amer. Soc. Rev.,* 22, 154-165.
13. Kish, L (1995). Survey Sampling, Wiley, New York.
14. Lahiri, D. B. (1951). A Method for Sample Selection Providing Unbiased Ratio Estimate. *Bull, Intern. Statist. Inst.* 31, 333-362.
15. Levy, P. S. and Lemeshow S. (1991). Sampling of Population: Methods and Applications, Wiley, New York.
16. Lohr, S. L. (1999). Sampling: Design and Analysis, Duxbury, New York.
17. Raj, D. (1968). Sampling Theory, McGraw Hill, New York.

18. Sampath, S. (2000). Sampling Theory and Methods, CRC Press.
19. Scheaffer, R.L., Mendenhall, W. and Ott, L. (1996). Elementary Sampling Survey, 5th ed., Duxbury, New York.
20. Sedransk, J. and Smith, P. J. (1988). Inference for Finite Population Quantiles. In: Handbook of Statistics, Vol. 6, Sampling (Krishnaiah, P. R. and Rao, C. R., Eds.), 267-289, North Holland, Amsterdam.
21. Sukhatme, P. V., Sukhatme, B. V., Sukhatme, S., and Ashok, C. (1984). Sampling Theory of Surveys with Applications, 3rd ed., Ames (Iowa): Iowa State University Press.
22. Thompson, S. K. (2002). Sampling, 2end Wiley, New York.
Yates, F. (1981). Sampling Methods for Censuses and Surveys 4th Ed., Grittin, London.

الفصل العاشر

العينة العنقودية بمرحلتين

Two – Stage Cluster Sample

1.10 مقدمة

إن العينة العنقودية بمرحلتين ما هي إلا امتداد للعينة العنقودية بمرحلة واحدة، والتي تكلمنا عنها في الفصل السابق. وكما رأينا سابقاً فإن العنقود في بعض الأحيان عبارة عن تجمع طبيعي لمجموعة من العناصر، مثل قطاع يحتوي على مجموعة من المنازل أو صندوق يحتوي على مجموعة وحدات من إنتاج مصنع معين. يحتوي العنقود في أغلب الأحيان على مجموعة كبيرة من العناصر، يصعب جمع معلومات عن جميع هذه الوحدات، أو قد يحتوي العنقود على مجموعة من العناصر المتشابهة، لذلك فإن جمع معلومات عن مجموعة صغيرة من هذه العناصر يكون كافياً لإعطائنا معلومات عن العناصر داخل العنقود الواحد.

يمكن أن تُعرف العينة العنقودية بمرحلتين على أنها اختيار عينة عشوائية بسيطة من العناقيد، ومن ثم اختيار عينة عشوائية بسيطة من العناصر من جميع العناقيد التي اختيرت في المرحلة الأولى، فعلى سبيل المثال في مثالنا الذي تحدثنا عنه في الفصل السابق حول تقدير مصروفات العائلة في مدينة ما، حيث قسمنا المدينة إلى مجموعة من القطاعات (عناقيد) كل قطاع يحتوي على مجموعة من المنازل (العائلات)، سوف نقوم أولاً باختيار عينة

عشوائية بسيطة حجمه n من القطاعات (عناقيد)، ثم من كل قطاع تم اختياره نقوم بسحب عينة عشوائية بسيطة من المنازل، ونقوم بجمع معلومات من هذه المنازل.

هناك تشابه كبير بين العينة العنقودية بمرحلتين والعينة الطبقية. يمكن أن نعتبر المجتمع مقسماً إلى مجموعات من العناصر غير المتداخلة، إذا اعتبرنا هذه المجموعات طبقات، ومن ثم نختار عينة عشوائية بسيطة من كل مجموعة، ولكن إذا اعتبرنا هذه المجموعات عناقيد نختار عينة عشوائية بسيطة من بين هذه العناقيد، ومن ثم نختار عينة عشوائية بسيطة من بين عناصر كل عنقود جرى اختياره. إن العينة الطبقية تعطينا تقديرات إلى مَعَالم المجتمع بكون تباينها قليلاً، عندما يكون التغير بين عناصر المجموعة الواحدة قليلاً. أما العينة العنقودية بمرحلتين فتكون جيدة عندما يكون تغير العناصر داخل كل مجموعة كبيراً، والمجموعات (العناقيد) مشابه بعضها بعضًا.

أما مزايا العينة العنقودية بمرحلتين فهي ذاتها التي تكلمنا عنها في العينة العنقودية بمرحلة واحدة. ومن أهم هذه المزايا أنها لا تحتاج إلى إطار للمعاينة لتنفيذها، وكما نعرف فإن عملية إعداد إطار المعاينة مكلفة وشاقة كثيراً. وبالمقابل فإن عملية الحصول على كشف لجميع العناقيد الموجودة في المجتمع سهلة في الغالب. أما الميزة الأخرى فإن العينة العنقودية بمرحلتين تقلل تكاليف الحصول على المعلومات المطلوبة، وذلك لتقارب وحدات المعاينة في العينة العنقودية مما يقلل تكاليف السفر والوقت الذي يصرفه العادون في الحصول على المعلومات.

2.10 تقدير الوسط الحسابي والمجموع الكلي للمجتمع

سوف نستخدم الرموز والمصطلحات الآتية في هذا الفصل

N : عدد العناقيد في المجتمع.

n : عدد العناقيد في العينة التي اختيرت باستخدام العينة العشوائية البسيطة.

M_i : عدد العناصر في العنقود i.

m_i : عدد العناصر التي اختيرت من العنقود i باستخدام العينة العشوائية البسيطة.

$M = \sum_{i=1}^{N} M_i$: عدد العناصر في المجتمع.

$\overline{M}$: متوسط حجم العنقود في المجتمع.

y_{ij} : قيمة العنصر j المسحوب من العنقود i للمتغير y.

$\overline{y}_i = \frac{1}{m_i}\sum_{j=1}^{m_i} y_{ij}$: متوسط العينة للعنقود i.

1.2.10 تقدير الوسط الحسابي للمجتمع

لإيجاد تقدير الوسط الحسابي للمجتمع $\overline{Y}$ يمكن أن نستفيد من الطريقة التي استخدمناها في تقدير المجموع الكلي للمجتمع والتي لا تعتمد على M. نستطيع أن نحصل على تقدير غير متحيز للوسط الحسابي للمجتمع $\overline{Y}$ يمكن أن نعرفه

$$\overline{y}_{cs2} = \frac{N}{Mn}\sum_{i=1}^{n} M_i \overline{y}_i$$

أما تقدير تباين $\overline{y}_{cs2}$ فيكون

$$s^2(\overline{y}_{cs2}) = \frac{N-n}{N}\frac{s_b^2}{n\overline{M}^2} + \frac{1}{nN\overline{M}^2}\sum_{i=1}^{n} M_i^2 \frac{M_i - m_i}{M_i}\frac{s_i^2}{m_i}$$

حيث إن

$$s_b^2=\frac{1}{n-1}\sum_{i=1}^{n}(M_i\,\bar{y}_i-\bar{M}\,\bar{y}_{cs2})^2;\ s_i^2=\frac{1}{m_i-1}\sum_{i=1}^{m_i}(y_{ij}-\bar{y})^2;\ i=1,2,\ldots,n$$

نلاحظ أن $\bar{y}_{cs2}$ يعتمد على M عدد العناصر في المجتمع، وكذلك فإن s_i^2 عبارة عن تباين العينة العشوائية البسيطة المسحوبة من العنقود i.

مثال (1): توجد لدى إحدى شركات خياطة الملابس الجاهزة 20 مخيطة موزعة على أجزاء مختلفة من البلد. ترغب إدارة الشركة في تقدير معدل عدد الساعات التي تتوقف فيها ماكينات الخياطة عن العمل بسبب أعطال مختلفة شهرياً. بما أن المخيطات موزعة بصورة واسعة على البلد، وكل مخيطة تحتوي على عدد كبير من الماكينات، لذلك فإن عملية مشاهدة وفحص جميع سجلات الصيانة لكل الماكينات العاملة في الشركة تُعد عملية صعبة وتأخذ وقتاً طويلاً. لذلك قررت الشركة استخدام أسلوب المعاينة العنقودية بمرحلتين، في المرحلة الأولى جرى سحب 4 مخيطات بصورة عشوائية من بين العشرين مخيطة التابعة للشركة. أما المرحلة الثانية فقد جرى سحب عدد معين من الماكينات وبصورة عشوائية من كل مخيطة من المخيطات الأربع التي سحبت أولاً، ومن ثم جرت عملية مشاهدة سجلات الصيانة للماكينات التي سحبت من كل مخيطة، وحصلنا على النتائج المبينة في الجدول الآتي:

المخيطة	M_i	m_i	عدد ساعات التوقف شهرياً y_{ij}				$\bar{y}_i$	s_i^2
1	15	3	12	8	3		7.67	20.34
2	15	3	15	10	1		8.67	50.34
3	7	2	6	5			5.5	0.5
4	20	4	20	7	4	1	8	70

إذا كان عدد ماكينات الخياطة في جميع فروع الشركة M=300

الحل:

$$\bar{y}_{cs2}=\frac{N}{nM}\sum_{i=1}^{n}M_i\,\bar{y}_i=\frac{20}{4(400)}(443.5)=7.39$$

$$s_b^2=\frac{1}{n-1}\sum_{i=1}^{n}(M_i\,\bar{y}_i-\bar{M}\,\bar{y}_{cs2})^2=2678.84$$

$$\bar{M}=\frac{300}{20}=15$$

$$s^2(\bar{y}_{cs2})=\frac{N-n}{N}\,\frac{s_b^2}{n\bar{M}^2}\,\frac{1}{nN\bar{M}^2}\sum_{i=1}^{n}M_i^2\,\frac{M_i-m_i}{M_i}\,\frac{s_i^2}{m_i}$$

$$=\frac{20-4}{20}\,\frac{2678.84}{4(15)^2}+\frac{9849.55}{4(20)\,(15)^2}=2.928$$

2.2.10 تقدير المجموع الكلي للمجتمع

يمكن أن نحصل على تقدير غير متحيز إلى المجموع الكلي للمجتمع وذلك باستخدام $\bar{y}_{cs2}$ تقدير الوسط الحسابي للمجتمع، لذلك فإن تقدير المجموع الكلي وتقدير تباينه هما على التوالي

$$Y_{cs2}=M\,\bar{y}_{cs2}=\frac{N}{n}\sum_{i=1}^{n}M_i\,\bar{y}_i$$

$$s^2(Y_{cs2})=\frac{N-n}{N}\,\frac{N^2}{n}\,s_b^2+\frac{N}{n}\sum_{i=1}^{n}M_i^2\,\frac{M_i-m_i}{M_i}\,\frac{s_i^2}{m_i}$$

نلاحظ أننا لا نحتاج لمعرفة M عدد عناصر المجتمع للحصول على تقدير المجموع الكلي للمجتمع.

مثال (2): أوجد المجموع الكلي وتقدير الخطأ المعياري لتقدير المجموع الكلي للبيانات في المثال السابق.

الحل :

$$Y_{cs2}=M\,\bar{y}_{cs2}=300\,(7.39)=2217$$

$$s^2(Y_{cs2})=M\sqrt{s^2(\bar{y}_{cs2})}=300\sqrt{2.928}=513.376$$

3.10 تقدير النسبة للوسط الحسابي

لقد لاحظنا أن تقدير الوسط الحسابي للمجتمع قبل قليل يعتمد على M عدد عناصر المجتمع، ولكن في كثير من الأحيان قد لا تكون عدد عناصر المجتمع معلومة، لذلك لا بد من تقدير الوسط الحسابي $\overline{Y}$ باستخدام المعلومات المتوافرة لدينا من العينة، وذلك عن طريق تقدير M، حيث يمكن أن نقدره باستخدام

$$M = N\sum_{i=1}^{n}\frac{M_i}{n}$$

إذا عوضنا عن قيمةM باستخدام M في $\overline{y}_{cs2}$ نحصل على ما يسمى بتقدير النسبة

$$\overline{y}_{csR} = \frac{\sum_{i=1}^{n} M_i \overline{y}_i}{\sum_{i=1}^{n} M_i}$$

كذلك فإن تقدير التباين إلى $\overline{y}_{csR}$

$$s^2(\overline{y}_{csR}) = \frac{N-n}{N}\frac{s_t^2}{n\,\overline{M}^2} + \frac{1}{nN\,\overline{M}^2}\sum_{i=1}^{n} M_i^2\,\frac{M_i - m_i}{M_i}\,\frac{s_i^2}{m_i}$$

حيث إن

$$s_t^2 = \frac{1}{n-1}\sum_{i=1}^{n} M_i^2(\overline{y}_i - \overline{y}_{csR})^2 \;;\; s_i^2 = \frac{1}{m_i - 1}\sum_{j=1}^{m_i}(y_{ij} - \overline{y}_i)^2$$

نلاحظ أن $\overline{y}_{csR}$ هو تقدير متحيز إلى $\overline{Y}$، كذلك فإن تقدير تباينه يعتمد على $\overline{M}$ حيث يمكن تقديره باستخدام $\overline{m}$. لا بد من الإشارة هنا إلى أن قيمة التحيز تصبح صغيرة وغير مهمة كلما كبر حجم العينة.

مثال (3): إذا كانت M غير معلومة في مثال **(1)** قدر معدل عدد ساعات التوقف لماكينات الخياطة في الشركة، كذلك أوجد معامل التغير.

الحل:

$$\bar{y}_{csR} = \frac{\sum_{i=1}^{n} M_i \bar{y}_i}{\sum_{i=1}^{n} M_i} = \frac{443.6}{57} = 7.78$$

$$\bar{m} = \frac{1}{n}\sum_{i=1}^{n} M_i = \frac{57}{4} = 14.25$$

$$s_t^2 = \frac{1}{n-1}\sum_{i=1}^{n} M_i^2 (\bar{y}_i - \bar{y}_{csR})^2 = \frac{455.027}{3} = 151.676$$

$$s^2(\bar{y}_{csR}) = \frac{(20-4)}{20(4)(14.25)^2}(151.676) + \frac{9849.55}{4(20)(14.25)^2} = 0.0755$$

$$CV(\bar{y}_{csR}) = \frac{\sqrt{s^2(\bar{y}_{csR})}}{\bar{y}_{csR}} = \frac{\sqrt{0.0755}}{7.78} = 0.035$$

4.10 تقدير نسبة المجتمع

إن تقدير النسبة في المجتمع يستخدم كثيراً في الحياة العملية، ففي مثالنا في بداية هذا الفصل حول تقدير مصروفات العائلات يمكننا تقدير نسبة العائلات التي تعتمد في مصروفاتها الشهرية على فرد واحد من أفراد العائلة.

لذلك لا بد من إعادة تعريف y_{ij} على الوجه الآتي

$y_{ij} = 1$ إذا كان العنصر j في العنقود i يحمل صفة مميزة

$y_{ij} = 0$ خلاف ذلك أي أن العنصر لا يحمل الصفة المميزة

بما أنه في العادة M غير معلومة سوف نقدر النسبة P دون الاعتماد على قيمة M. إذا كان $\hat{p}_i$ يمثل نسبة الذين يحملون صفة معينة أو ينتمون إلى فئة معينة في العنقود i فإن تقدير النسبة P للمجتمع يمكن إيجاده باستخدام

$$\hat{p}_{csR} = \frac{\sum_{i=1}^{n} M_i \hat{p}_i}{\sum_{i=1}^{n} M_i}$$

وكذلك فإن تقدير تباين النسبة

$$s^2(\hat{p}_{csR}) = \frac{N-n}{N}\frac{s_t^2}{n\,\overline{M}^2} + \frac{1}{nN\,\overline{M}^2}\sum_{i=1}^{n} M_i^2 \frac{M_i - m_i}{M_i}\frac{\hat{p}_i q_i}{m_i - 1}$$

حيث إن

$$s_t^2 = \frac{1}{n-1}\sum_{i=1}^{n} M_i^2(\hat{p}_i - \hat{p}_{csR})^2$$

مثال (4): عودة إلى مثال **(1)** إذا كانت نسبة الماكينات التي تحتاج إلى صيانة كبيرة، وهي على التوالي: $\hat{p}_1 = 0.4, \hat{p}_2 = 0.3, \hat{p}_3 = 0.25, \hat{p}_4 = 0.15$. حيث إن $\hat{p}_i$ و i=1,2,3,4 يمثل النسبة في العنقود i .أوجد $\hat{p}_{csR}$ وتقدير خطئه المعياري إذا كانت M غير معلومة.

الحل:

$$\hat{p}_{csR} = \frac{\sum_{i=1}^{n} M_i \hat{p}_i}{\sum_{i=1}^{n} M_i} = \frac{125.25}{57} = 0.267$$

$$\overline{m} = 14.25$$

$$s_t^2 = \frac{1}{n-1}\sum_{i=1}^{n} M_i^2(\hat{p}_i - \hat{p}_{csR})^2 = \frac{9.7846}{3} = 3.26$$

$$s^2(\hat{p}_{csR}) = \frac{20-4}{20}\frac{3.26}{4(14.25)^2} + \frac{60.66}{4(20)\,(14.25)^2} = 0.0069$$

$$s(\hat{p}_{csR}) = \sqrt{s^2(\hat{p}_{csR})} = \sqrt{0.0069} = 0.083$$

5.10 اختيار حجم العينة

إن اختيار حجم العينة في العينة العنقودية بمرحلتين يُعد أكثر تعقيداً منه في العينة العنقودية بمرحلة واحدة. إذ لا بد من اختيار عدد العناقيد n، وكذلك لا بد من اختيار حجم العينة من كل عنقود m_i و$i=1,2,\dots,n$. إن اختيار n وm_i يعتمد على مصدرين من مصادر التغير الأول بين العناقيد والثاني داخل كل عنقود. بصورة عامة نسحب حجم عينة أكبر من العناقيد ذات التغير الكبير، إذا كانت العناقيد متجانسة داخلياً، أي عناصر كل عنقود متقاربة من بعضها بعضا، ولكن إذا كان هنالك اختلافات كبيرة بين أوساط العناقيد المختلفة، نقوم بسحب عناصر قليلة من كل عنقود، وبالمقابل نرفع عدد العناقيد المسحوبة (أي نزيد حجم n ونقلل حجم m_i). وإذا كانت هنالك اختلافات كبيرة داخل كل عنقود، ولكن أوساط العناقيد متقاربة، نقوم بزيادة العناصر المسحوبة من كل عنقود ونقلل عدد العناقيد (أي نزيد حجم m_i ونقلل حجمn).

لنفرض أن أحجام العناقيد متساوية أي $M_1 = M_2 = L = M_N = \overline{M}$ كذلك $m_1 = m_2 = \dots = m_n = m$. إذا عوضنا عن قيمة M_i و m_i كذلك أهملنا معامل التصحيح فإننا سنحصل على تقدير إلى الوسط الحسابي $\overline{y}_{cs2}$ للمجتمع وتباينه على التوالي

$$\overline{y}_{cs2} = \frac{1}{n}\sum_{i=1}^{n}\overline{y}_i$$

$$Var(\overline{y}_{cs2}) = \frac{S_b^2}{n} + \frac{S_w^2}{nm}$$

حيث إن S_b^2 التباين بين أوساط العناقيد و S_w^2 التباين بين العناصر داخل العناقيد. لإيجاد حجم العينةn وm إما أن نصغر التباين $Var(\overline{y}_{cs2})$ عندما تكون التكاليف ثابتة، أو نقلل التكاليف عندما يكون التباين مساوياً إلى

كمية ثابتة. لا بد من معرفة التكاليف، لذلك نفرض أن تكاليف معاينة العنقود الواحد C_1 وتكاليف معاينة كل عنصر داخل العناقيد C_2 ومجموع التكاليف $C = nC_1 + nmC_2$. إن قيمة m التي تصغر $Var(\bar{y}_{cs2})$ إذا كانت التكاليف C ثابتة أو نصغر C إذا كان التباين ثابتاً هي

$$m = \sqrt{\frac{C_1 S_w^2}{C_2 S_b^2}}$$

بعد أن نحدد قيمة m (عدد العناصر التي سوف نختارها من كل عنقود) نقوم بتحديد عدد العناقيدn إذا كان التباين ثابتاً باستخدام

$$Var(\bar{y}_{cs2}) = \frac{S_b^2}{n} + \frac{S_w^2}{nm}$$

أو إذا كانت التكاليف ثابتة نستخدم

$$C = nC_1 + nmC_2$$

نلاحظ أن S_w^2 وS_b^2 غير معروفتين لذلك لا بد من تقديرهما إما باستخدام مسوحات مماثلة سابقة، أو بسحب عينة عشوائية أولية، واستخدام البيانات في العينة الأولية لإيجاد تقدير غير متحيز إلى S_w^2 هو

$$s_w^2 = \frac{1}{n}\sum_{i=1}^{n} s_i^2$$

ومن ثم نجد تقديرًا إلى S_b^2 هو

$$s_b^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\bar{y}_i - \bar{y}_{cs2})^2 - \frac{s_w^2}{m}$$

مثال (5): لمثالنا **(1)** لنفرض أن M=15 وm=4 كذلك $\bar{y}_1 = 13.4$ و $\bar{y}_2 = 8.67$ و $\bar{y}_3 = 4$ و $\bar{y}_4 = 10.34$ ، وكذلك فإن $s_1^2 = 20.34$ و $s_2^2 = 50.38$ و $s_3^2 = 7$ و $s_4^2 = 77.34$. أوجد قيمة n وm إذا ك أن $Var(\bar{y}_{cs2}) = 2$ و $C_1 = 9$ و $C_2 = 1$ وC=250.

الحل :

$$s_w^2 = \frac{1}{4}\sum_{i=1}^{4} s_i^2 = 37.51; \bar{y}_{cs2} = \frac{1}{4}\sum_{i=1}^{4}\bar{y}_i = 7.67$$

$$s_b^2 = \frac{1}{3}\sum_{i=1}^{4}(\bar{y}_i - \bar{y}_{cs2})^2 - \frac{s_w^2}{4} = 15.41 - \frac{37.51}{4} = 6.03$$

$$m = \sqrt{\frac{C_1 S_w^2}{C_2 S_b^2}} = \sqrt{\frac{(9)(7.67)}{(1)(6.03)}} = 4$$

إذا كان التباين إلى $\bar{y}_{cs2}$ ثابتاً ومساوياً إلى 2 أي $Var(\bar{y}_{cs2}) = 2$ نجد قيمة n باستخدام

$$Var(\bar{y}_{cs2}) = \frac{S_b^2}{n} + \frac{S_w^2}{4n} = 2 \text{ or } 6.03 + \frac{37.51}{4} = 2n$$

إذاً $n = 8$

إذا كانت C=250، نجد قيمة n باستخدام

$$C = nC_1 + nmC_2 \text{ or } 250 = n + 36n$$

إذاً $n = 7$

تمارين

1. يرغب صاحب مشتل في تقدير معدل أطوال الشتلات الموجودة، في حقل مقسم إلى 50 قطعة تختلف قليلاً في الحجم. لذلك قرر أن يسحب عينة بحجم 10% من كل قطعة باستخدام العينة العنقودية بمرحلتين، الجدول الآتي يعطينا البيانات اللازمة.

القطعة	عدد الشتلات	عدد الشتلات في العينة	أطوال الشتلات في العينة
1	52	5	12,11,12,10,13
2	56	6	10,9,7,9,10,8
3	60	6	6,5,7,5,4,6
4	46	5	7,8,7,7,8
5	49	5	10,11,15,12,11
6	51	5	14,15,13,12,13

قدر معدل أطوال الشتلات في الحقل ثم أوجد فترة 95% ثقة للوسط الحسابي للمجتمع.

2. في السؤال الأول لنفرض عدد الشتلات الموجودة في الحقل 2500، استخدم هذه المعلومات الإضافية لتقدير معدل أطوال الشتلات، ثم أوجد معامل التغير للتقدير.

3. لدى إحدى الشركات مجموعة كبيرة من المطاعم موزعة على 20 مدينة في إحدى البلدان، ترغب الشركة معرفة نسبة المطاعم التي لا تلتزم بشروط النظافة المطلوبة من قبل الشركة. المطاعم الموجودة في المدينة الواحدة يبدو أنها تُظهر نفس المستوى من النظافة. لذلك قررت إدارة الشركة سحب عينة عنقودية بمرحلتين، قامت بسحب 5 مدن بطريق عشوائية، ثم سحبت نصف المطاعم الموجودة في هذه المدن. يوضح الجدول الآتي البيانات المطلوبة. قدر نسبة المطاعم التي تمتلكها الشركة والتي لا

تلتزم بشروط النظافة المطلوبة من قبل الشركة، ثم أوجد فترة 90% ثقة للنسبة الحقيقية.

المدينة	عدد المطاعم في العينة	عدد المطاعم في المدينة	عدد المطاعم التي لا تلتزم بشروط النظافة
1	4	9	1
2	5	10	2
3	7	15	3
4	6	12	2
5	3	6	1

4. إذا علمنا أن عدد المطاعم التي تمتلكها الشركة في السؤال الثالث 400 مطعم. قدر نسبة المطاعم التي لا تلتزم بشروط النظافة، ثم أوجد معامل التغير لهذا التقدير.
5. يرغب أحد الباحثين الاجتماعيين في تقدير عدد المتقاعدين في إحدى المدن، لذلك قرر أن يستخدم العينة العنقودية بمرحلتين، وذلك بسحب عينة من قطاعات المدينة أولاً، ومن كل قطاع يسحب عينة من العائلات، علماً بأنه يوجد في المدينة 300 قطاع. قدر مجموع المتقاعدين القاطنين في هذه المدينة، ثم أوجد فترة 95% ثقة للمجموع الحقيقي.

قطاع	عدد العائلات	عدد العائلات في العينة	عدد المتقاعدين
1	18	3	1, 0, 2
2	4	3	0, 3, 0
3	9	3	1, 1, 2
4	12	3	0, 1, 1

6. استخدم البيانات في السؤال الخامس لتقدير معدل عدد المتقاعدين في العائلة، ثم أوجد معامل التغير لهذا التقدير.
7. يحتوي أحد المجتمعات على 10.000 عنصر. قسم المجتمع إلى 50 مجموعة جزئية حسب إحدى المواصفات. سحبت عينة عنقودية بمرحلتين: في المرحلة الأولى جرى سحب عينة من أربع مجموعات،

ومن كل مجموعة جرى سحب 40 عنصراً. الجدول الآتي يلخص النتائج التي حصلنا عليها.

$\hat{p}_i$	s_i^2	$\bar{y}_i$	n_i	N_i	المجموعة
0.42	540	142	50	150	1
0.38	482	154	50	250	2
0.36	623	130	50	200	3
0.32	508	165	50	180	4

احسب تقدير معدل المجتمع، وتقدير النسبة، ومن ثم أوجد فترة 95% ثقة لكلا التقديرين.

المراجع العربية

1. سليمان محمد طشطوش (2001)، أساسيات المعاينة الإحصائية، دار الشروق للنشر والتوزيع، عمان.
2. حسين علوان (1994)، طرق المعاينة، دار الفرقان، عمان.
3. هلال عبود البياتي وصبري رديف العاني (1981)، العينات، جامعة بغداد، بغداد.
4. وليد عبدالحميد نوري وعبدالمجيد حمزة الناصر (1981) العينات، وزارة التعليم العالي والبحث العلمي، بغداد.
5. وليم كوكوران –ترجمة أنيس كنجو– تقنية المعاينة الإحصائية (1995)، جامعة الملك سعود، الرياض.

References

1. Barnett, V. (1991). Sampling Survey Principles and Methods, Edward Arnold, London.
2. Brewer, K. W. R. and Hanif, M. (1970). Durbin's new Multistage Variance Estimator. *J. Roy. Stat. Soc.* B32, 302-311.
3. Chaudhuri, A. and Stenger, H. (1992). Survey Sampling: Theory and Methods, Marcel Dekker, New York.
4. Cochran, W. G. (1977). Sampling Technique, 3rd Ed. Wiley, New York.
5. Cornfield, J. (1951). The Determination of Sample Size. *Amer. J. Publ. Health*, 41, 654-661.
6. Deming, W. E. (1960). Sampling Design in Business Research. Wiley, New York
7. Foreman, E. K. (1991). Survey Sampling Principles, Marcel Dekker, New York.
8. Govindarajulu, Z. (1999). Elements of Sampling Theory and Methods, Prentice Hall.
9. Hajek, J (1981). Sampling from Finite Population, Marcel Dekker, New York.
10. Hansen, M. H., and Hurwitz, W. N. (1943). On the Theory of Sampling from Finite Populations. *Ann. Math. Statist.* 14, 333-362.
11. Hedayat, A. S. and Sinha, B. K. (1991). Design and Inference in Finite Population Sampling, Wiley, New York.
12. Kish, L and Hess, I. (1959). On Variance of Rations and their Differences in multistage Samples, *J. Amer. Stat. Assoc.* 54, 416-446.
13. Kish, L (1995). Survey Sampling, Wiley, New York.
14. Lahiri, D. B. (1951). A Method for Sample Selection Providing Unbiased Ratio Estimate, *Bull, Intern. Statist. Inst.* 31, 333-362.
15. Levy, P. S. and Lemeshow S. (1991). Sampling of Population: Methods and Applications, Wiley, New York.
16. Lohr, S. L. (1999). Sampling: Design and Analysis, Duxbury, New York.

17. Murthy, M. N. (1967). Sampling Theory and Methods. Calcutta, India: Statistical Publication Society.
18. Raj, D. (1966). Some Remarks on a Simple Procedure for Sampling without Replacement, *J. Amer. Statist. Assoc.* 61, 391-397.
19. Raj, D. (1968). Sampling Theory, McGraw Hill, New York.
20. Sampath, S. (2000). Sampling Theory and Methods, CRC Press.
21. Scheaffer, R.L., Mendenhall, W. and Ott L. (1996). Elementary Sampling Survey, 5th ed., Duxbury, New York.
22. Sedransk, J. and Smith, P. J. (1988). Inference for Finite Population Quantiles. In: Handbook of Statistics, Vol. 6, Sampling (Krishnaiah, P. R. and Rao, C. R., Eds.), 267-289, North Holland, Amsterdam.
23. Sukhatme, P. V., Sukhatme, B. V., Sukhatme, S., and Ashok, C. (1984). Sampling Theory of Surveys with Applications, 3rd ed., Ames (Iowa): Iowa State University Press.
24. Thompson, S. K. (2002). Sampling, 2end Wiley, New York.
25. Yates, F. (1981). Sampling Methods for Censuses and Surveys 4th Ed., Grittin, London.

الفصل الحادي عشر

المعاينة مختلفة الاحتمالات

Varying Probability Sampling

1.11 مقدمة

لقد تكلمنا في الفصول العشر الأولى من هذا الكتاب عن اختيار وحدات العينة من المجتمع، بحيث إن جميع وحدات المجتمع لها الفرصة نفسها في الظهور في العينة، كما هي الحالة في العينة العشوائية البسيطة أو الطبقية أو المنتظمة ...إلخ. ولكن عندما تكون وحدات المعاينة مختلفة كثيراً في الحجم فإن العينة العشوائية البسيطة أو المنتظمة لا تعطينا تقديراً جيداً، وذلك بسبب التغير الكبير في الوحدات بالنسبة إلى الصفة y التي تجري مشاهدتها أو دراستها، فعلى سبيل المثال إذا كانت الوحدة تمثل حقولاً زراعية أو مدناً واضح أن مساحات الحقول لا تكون متساوية وكذلك عدد سكان المدن المختلفة لا يكون متساوياً. يمكن في مثل هذه الحالات أن نحصل على تقدير أفضل لو أعطينا احتمالاً أكبر لاختيار الوحدات الكبيرة. يمكن اختيار العينة مع الإرجاع أو دون إرجاع، هذا النوع من طرق المعاينة يسمى

Sampling with Probability proportional to size

أو(PPS Sampling) سوف نستخدم الرمز PPS في هذا الفصل للدلالة على أن وحدات العينة سحبت باحتمال يتناسب مع حجمها. كذلك سوف يقتصر كلامنا في هذا الفصل فقط على ما يسمى بالمعاينة مع احتمال يتناسب مع الحجم (PPS) سوف لا نتطرق إلى طرق أخرى يكون احتمال سحب الوحدات

من المجتمع غير متساوٍ. لتوضيح الفكرة؛ لنفرض أننا نرغب في تقدير عدد فرص العمل المتوافرة في الشركات العاملة في البلد، واضح أن الشركات ذات الحجم الكبير يكون توافر فرص العمل فيها أكبر من الشركات الصغيرة، ومن ثم لو أننا سحبنا عينة عشوائية بسيطة من الشركات دون التمييز بينها حسب حجمها فإن تقديرنا لعدد فرص العمل لا يكون دقيقاً، وذلك لأن التباين سوف يكون كبيراً، لأن التغير كبير بين حجم الوحدات والتي تمثل هنا شركات عاملة في البلد.

2.11 العينة العشوائية البسيطة مع الإرجاع

لنفرض أن لدينا مجتمعاً عدد وحداته N ،ولنفرض أننا نرغب في دراسة الصفة y للوحدة i و $i= 1, 2, \ldots, N$ ولنفرض كذلك أنَّ احتمال سحب الوحدات من المجتمع يتناسب مع حجمها X_i أي $p_i = X_i/X$ ، بحيث إن $X = \sum_{i=1}^{N} X_i$.

وكذلك نرفض أننا سوف نختار وحدات العينة n مع الإرجاع، وكذلك لنفرض أن (y_i, p_i) يمثل القيمة والاحتمال للوحدة i و $i= 1, 2, \ldots, n$ من العينة. ويمكن ملاحظة أن y_i/p_i و $i= 1, 2, \ldots, n$ مستقلة ولها نفس التوزيع. إذا كان $p_i = 1/N$ تكون العينة عشوائية بسيطة. ولهذا نلاحظ أن العينة العشوائية البسيطة ما هي إلا حالة خاصة من العينة المختلفة الاحتمالات.

يمكن الحصول على تقدير غير متحيز لتقدير المجموع الكلي Y للمجتمع وباحتمالات متناسبة مع حجم الوحدة (PPS) كما يأتي:

$$Y_{PPS} = \frac{1}{n}\sum_{i=1}^{n} y_i/p_i$$

مع تباين

$$Var(Y_{PPS}) = \frac{1}{n}\sum_{i=1}^{n} p_i (y_i/p_i - Y)^2$$

وتقدير تباين

$$s^2(Y_{PPS}) = \frac{1}{n(n-1)}\sum_{i=1}^{n} p_i(y_i/p_i - Y_{PPS})^2$$

ويمكن مقارنة هذه المعاينة بالعينة العشوائية البسيطة لنحصل على $Var(Y_{PPS}) \leq Var(Y)$. لبرهنة هذه النتيجة ولمزيد من التفصيلات حول هذا النوع من المعاينة يراجع Singh and Chaudhary (1986).

3.11 العينة العشوائية البسيطة دون الإرجاع

لقد سبق وتكلمنا عن سحب عينة عشوائية بسيطة ودون إرجاع من المجتمع، وهذه الطريقة تعطي جميع وحدات المجتمع نفس الفرصة في الظهور في العينة. ولكن في بعض الحالات يكون إعطاء احتمالٍ متساوٍ لجميع وحدات المجتمع بالظهور في العينة غير ملائم، وذلك لكون وحدات المجتمع غير متساوية أو مختلفة في الحجم، والحجم له تأثير مباشر على قيمة المشاهدة في الوحدة.

سوف نحاول توضيح فكرة سحب الوحدة من المجتمع مع احتمال تناسب مع حجم الوحدة. لنفرض أننا نرغب في تقدير مجموع محصول القمح للحقول المزروعة بالقمح في إحدى المحافظات، وقمنا بإعداد كشف بجميع الحقول المزروعة بالقمح في هذا الفصل في المحافظة، نستطيع أن نسحب عينة عشوائية بسيطة. وباحتمال سحب متساوٍ لجميع الحقول كما مر معنا في الفصل الخامس. ولكن واضح أن الوحدة هنا تمثل حقلاً والحقول غير متساوية في مقدار المساحة المزروعة، ومن ثم فإن إعطاء احتمال متساوٍ لجميع الحقول بالظهور في العينة لا يكون ملائماً في مثل هذه الحالة، لكون مساحة الحقل لها علاقة مباشرة بمقدار القمح الذي سنحصل عليه من الحقل. لنفرض أن

لدينا خارطة عن الحقول المزروعة بالقمح في المحافظة، نستطيع استخدام هذه الخارطة لسحب عينة عشوائية من جميع الحقول وباحتمال يتناسب مع حجم الحقل (المساحة) X_j و $j=1, 2, \ldots, N$. واضح أنه توجد علاقة قوية بين إنتاج الحقل من القمح Y_j و $j= 1, 2, \ldots, N$ وحجم الحقل X_j، ولنفرض أنه $Y_j = kX_j$ هذا يعني أن ناتج القمح يتناسب مع حجم الحقل. ولنفرض أننا سحبنا عينة حجمها n من الحقول ودون إرجاع وحصلنا على المشاهدات $y_1, y_2, \ldots, y_n$ وكذلك مساحة هذه الحقول $x_1, x_2, \ldots, x_n$، وكذلك نفرض أننا نعرف مجموع المساحة المزروعة بالقمح لهذا الموسم X .فإن تقدير مجموع إنتاج القمح (المجموع الكلي للمجتمع)

$$Y_{PPS} = \frac{X}{n}\sum_{j=1}^{n}\frac{y_j}{x_j}$$

واضح أن Y_{PPS} تقدير غير متحيز إلى $Y = \sum_{j=1}^{N} Y_j$ المجموع الكلي للمجتمع مع تباين

$$Var(Y_{PPS}) = \frac{X}{n}\sum_{j=1}^{N}\left(Y_j/X_j\right)^2 - \left(Y^2/n\right)$$

في حالة كون $Y_j = kX_j$ و $Y= kX$ نحصل على تقدير مثالي أي إن $Var(Y_{PPS}) = 0$ وهذا غير ممكن في الحياة العملية؛ لأنه إذا كان $Y=kx$، ونعرف X نستطيع معرفة Y دون الحاجة إلى التقدير.

إن استخدام العينة العشوائية ذات الاحتمالات المتناسبة مع حجم الوحدة، يصاحبه مشكلات عدة من أهمها كيف نختار وحدات العينة مع احتمال يتناسب مع أحجامها ودون إرجاع، وكذلك كيف نحصل على عينات s_i واحتمال سحب كل عينة π_i وأخيراً كيف يمكن اشتقاق الخصائص

الإحصائية للتقديرات والتي تعتمد على (s_i, π_i) ، يمكن التغلب على بعض هذه المشكلات أحياناً وذلك بسحب وحدات العينة مع الإرجاع. كما مر معنا أعلاه. يراجع Singh and Chaudhary (1986) لمزيد من المعلومات حول بعض طرق اختيار وحدات العينة ودون إرجاع من المجتمع وكذلك حول بعض التفصيلات الأخرى.

4.11 تقدير النسبة R

لنفرض أننا سحبنا عينة $(y_1, x_1), (y_2, x_2), ..., (y_n, x_n)$ دون إرجاع حيث إن احتمال الحصول على Y_j بتناسب مع X_j هذا يعني

$$P(y_j) = kX_j = X_j / \sum_{i=1}^{N} x_i$$

كذلك نفرض أننا نعرف $X = \sum_{i=1}^{N} x_i$ وأن $Y_i = \beta X_i$. نلاحظ المعاينة مع احتمالات متناسبة مع الحجم للمتغير x هي نفسها احتمالات متناسبة مع الحجم للمتغير y لأن

$$P(y_j) = kX_j = X_j / \sum_{i=1}^{N} x_i = Y_j / \sum_{i=1}^{N} y_i$$

لنأخذ التقدير

$$Y_{PPS} = \frac{1}{n} \sum_{j=1}^{n} y_j / p_j = \frac{X}{n} \sum_{i=1}^{n} y_i / x_i$$

إن $E(Y_{PPS}) = Y$ أي أنه تقدير غير متحيز مع تباين يساوي صفراً. نلاحظ أن Y_{PPS} مشابه إلى تقدير النسبة

$$R_1 = \frac{1}{n} \sum_{i=1}^{n} y_i / x_i$$

ولكن مع الفرق لإيجاد R_1 استخدمنا المعاينة العشوائية البسيطة متساوية الاحتمالات، أما Y_{PPS} استخدمنا المعاينة المختلفة الاحتمالات أو بالتحديد احتمال سحب الوحدات يتناسب مع حجمها. ولابد من الإشارة هنا أننا في الحياة الواقعية لا يمكن أن نحصل على $Y_i = \beta X_i$ ولكن قد نحصل على حالات قريبة من هذه العلاقة ولذلك يبقى مفيداً في الحياة العملية.

لنفرض أننا لدينا الأمثلة الآتية Y يمثل إنتاج القمح وX مجموع الدونمات المزروعة بالقمح، Y يمثل إنتاج شركات من بضاعة معينة وX مجموع العاملين في هذه الشركات، Y يمثل مصروف العائلات وX حجم العائلات. والسؤال الآن: كيف نختار وحدات عينة باحتمال يتناسب مع حجم الوحدات؟ إذا كان لدينا $X_1, X_2, ..., X_N$. سوف نتناول طريقتين .

1. **الطريقة الأولى استخدام المجموع الجزئي**

يمكن أن نجد المجموع الجزئي وهو

$$X_1,\ X_1 + X_2,\ X_1 + X_2 + X_3,\ ...,\ \sum_{i=1}^{N} X_i$$

نقوم بسحب رقم عشوائي Z يقع بين 1 إلى $X = \sum_{i=1}^{N} x_i$ ونختار x_j عندما

$$\sum_{i=1}^{j-1} X_i < Z \le \sum_{i=1}^{j} X_i$$

ولكن هذه الطريقة تستغرق وقتاً طويلاً ولا بد من توافر كشف بجميع $X_1, X_2, ..., X_N$.

2. **الطريقة الثانية تسمى (Lahiri's Method)**

لنفرض أننا عندنا فكرة عن أكبر قيمة من بين X_j و $j= 1, 2, \ldots, N$ ولنعطها اسماً X_{max} ، نقوم الآن بسحب رقمين عشوائيين مستقلين عن بعضهما بعضا ، الأول عدد صحيح يقع بين $(1, N)$ والثاني يقع بين $(0, X_{max})$. لنفرض أنهما j و Z على التوالي، إذا كان $Z < X_j$ نأخذ X_j لتكون هي المشاهدة المطلوبة ، خلاف ذلك نرفض الرقمين ونسحب رقمين جديدين. وهكذا. ولكن هذه الطريقة ستكون حساسة جداً للخطأ في قيمة X_{max} ؛ لأننا في الغالب لا نعرفه ونحاول تحديده بطريقة أو بأخرى، لمزيد من المعلومات يراجع Lahiri (1951).

5.11 العينة العشوائية المنتظمة

سوف نتناول هنا العينة المنتظمة مع كون احتمال سحب وحداتها يتناسب مع حجم الوحدة، والتي اقترحها Madow (1949)، يمكن تلخيص هذه الطريقة، بأن نقوم بترتيب وحدات المجتمع بصورة عشوائية، ومن ثم نقوم ببناء ما يسمى بالمجاميع التراكمية $T_i = \sum_{j=1}^{i} X_j$ و $i= 1, 2, \ldots, N$ حيث إن X_j يمثل حجم الوحدة i. لاختيار عينة بحجم n وباحتمال متناسب مع حجم الوحدة، نقوم بسحب رقم عشوائي وليكن R يقع بين 1 إلى k و $k = \frac{T_N}{n}$ تكون الوحدة المعادلة للرقم $R + jk$ و $j=0, 1, \ldots, (n-1)$ في العينة.

مثال (1): لنفرض أن لدينا قرية تحتوي على 10 مناطق زراعية تحتوي على العدد الآتي من الحقول: 27, 28, 35, 24, 26, 40, 25, 45, 30, 50 ونريد أن

نختار عينة عشوائية تتكون من 3 مناطق وبطريقة منتظمة مع احتمال يتناسب مع عدد الحقول في هذه المناطق.

الحل:

$$T_N = 330;\ k = \frac{T_N}{n} = \frac{330}{3} = 110$$

لنفرض أن الرقم الذي اختير بطريقة عشوائية يقع بين 1 إلى 110 هو51. إذن الوحدات المقابلة للمجاميع التراكمية 271 ,161 ,51 تكون في العينة المنتظمة مع احتمال يتناسب مع حجم الوحدات (المناطق). يمكن استخدام هذه الطريقة إذا كان k عدداً غير صحيح وذلك بأخذ أقرب عدد صحيح إلى T_N/n .

لقد اشتق Hartley and Rao (1962) تقديراً غير متحيز لتقدير المجموع الكلي وهو

$$Y_{HR} = \sum_{i=1}^{n} \frac{y_i}{\pi_i} = \sum_{i=1}^{n} \frac{y_i}{np_i}$$

ولكن المشكلة التي واجهتنا في تقدير التباين في العينة العشوائية المنتظمة والتي تكلمنا عنها في الفصل الثامن، تواجهنا هنا في تقدير تباين Y_{HR} ، ولقد اقترح Hartley and Rao تبايناً إلى Y_{HR} يمكن استخدامه إذا كان حجم العينة صغيراً قياساً إلى حجم المجتمع وهو:

$$Var(Y_{HR}) \approx \frac{1}{n} \sum_{i=1}^{N} (y_i/p_i - Y)^2\, p_i\, [1 - (n-1)p_i]$$

ويمكن تقدير هذا التباين باستخدام

$$s^2(Y_{HR}) = \frac{1}{n^2(n-1)} \sum_{i=1}^{n} \sum_{j>i}^{n} [1 - n(p_i + p_j) + \sum_{j'}^{N} p_{j'}^2] (\frac{y_i}{p_i} - \frac{y_j}{p_j})^2$$

لتفاصيل ومعلومات أكثر عن الموضوع يراجع Hartley and Rao (1962).

6.11 العينة العنقودية بمرحلة واحدة

لقد مر معنا في الفصل التاسع في بعض الأحيان نستطيع أن نقلل التباين للتقدير عندما نقوم باختيار وحدات العينة باحتمال يتناسب مع حجم الوحدة، في الحقيقة أن العينة العنقودية غالباً ما تعطينا حالات مثالية لفكرة اختيار وحدات العينة باحتمال يتناسب مع حجم الوحدة، لأن عدد الوحدات في العنقود الواحد m_i في الغالب غير متساوية.

إن استخدام العينة العنقودية مع احتمال يتناسب مع حجم العنقود m_i يعطينا فائدة كبيرة في تقليل التباين خصوصاً إذا كان مجموع العناقيد y_i له علاقة قوية أو عالية مع عدد العناصر في العنقود. لنفرض $p_i = \frac{m_i}{M}$ احتمال أن الوحدة i تظهر في العينة، حيث إن $M = \sum_{i=1}^{N} m_i$ وN يمثل عدد العناقيد في المجتمع، لذلك فإن تقدير المجموع الكلي للمجتمع يكون

$$Y_{pps} = \frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{p_i} = \frac{M}{n} \sum_{i=1}^{n} \frac{y_i}{m_i}$$

أما تقدير الوسط الحسابي للمجتمع فيكون

$$\bar{y}_{pps} = \frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{m_i}$$

إن تقدير التباين إلى Y_{pps} و $\bar{y}_{pps}$ فهما على التوالي

$$s^2(Y_{pps}) = \frac{M^2}{n(n-1)} \sum_{i=1}^{n} (y_i/m_i - \bar{y}_{pps})^2$$

$$s^2(\bar{y}_{pps}) = \frac{1}{n(n-1)} \sum_{i=1}^{n} (y_i/m_i - \bar{y}_{pps})^2$$

7.11 العينة العنقودية بمرحلتين

إن عدد الوحدات للعناقيد مختلفٌ من عنقود إلى آخر أحياناً بشكل كبير، لذلك فإن سحب العناقيد من المجتمع باحتمالات متناسبة مع حجم هذه العناقيد غالباً ما يفضل على سحبها باحتمالات متساويةً، والذي تكلمنا عنه في الفصل العاشر، بصورة عامة نستخدم احتمالات متناسبة مع حجم العنقود في المرحلة الأولى فقط، لأن العناصر داخل كل عنقود غالباً ما تكون متساوية الحجم. لذلك سوف نتكلم عن تقدير للوسط الحسابي، والمجموع الكلي للمجتمع يعتمد على احتمالات متناسبة مع الحجم في المرحلة الأولى فقط. لقد حصلنا على تقدير للوسط الحسابي في الفقرة أعلاه وهو عبارة عن

$$\bar{y}_{pps} = \frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{m_i} = \frac{1}{n} \sum_{i=1}^{n} \bar{y}_i$$

و $\bar{y}_i$ وجدناه باستخدام جميع عناصر العنقود i. هنا نحسب $\bar{y}_i$ بالاعتماد على عينة عشوائية بسيطة من عناصر العنقود i، لذلك $\bar{y}_i$ هنا ما هو إلا تقدير إلى الوسط الحسابي للعنقود i. إن تقدير الوسط الحسابي في العينة العنقودية

بمرحلتين مع استخدام احتمالات متناسبة مع حجم العنقود في المرحلة الأولى سيكون

$$\bar{y}_{pps} = \frac{1}{n}\sum_{i=1}^{n}\bar{y}_i$$

مع تقدير التباين

$$s^2(\bar{y}_{pps}) = \frac{1}{n(n-1)}\sum_{i=1}^{n}(\bar{y}_i - \bar{y}_{pps})^2$$

وكذلك فإن تقدير المجموع الكلي سيكون

$$Y_{pps} = \frac{M}{n}\sum_{i=1}^{n}\bar{y}_i$$

مع تقدير للتباين

$$s^2(Y_{pps}) = \frac{M^2}{n(n-1)}\sum_{i=1}^{n}(\bar{y}_i - \bar{y}_{pps})^2$$

References

1. Hartley, H. O. and Rao, J. N. K. (1962). Sampling without Replacement, J. Amer., Statist. Assoc., 61, 739-748.
2. Horvitz, D. G. and Thompson, D. J. (1952). A Generalization of Sampling without Replacement, *J. Amer. Statist. Assoc.* 47, 663-685.
3. Lahiri, D. B. (1951). A Method of Sample Selection Providing. Unbiased Ratio Estimates Bull. Int. Statist., 33, 133, 140.
4. Lahiri, D. B. (1951). A Method of Sample Selection Providing. Unbiased Ratio Estimates Bull. Int. Statist., 33, 133, 140.
5. Madow, N. G. (1949). On Theory of Systematic Sampling Ann. Math. Statist., 20, 333-354.
6. Ray, D. (1956). Some Estimates in Sampling with Varying Probabilities without Replacement. *J. Amer. Statist. Assoc.* 51, 269-284.
7. Samford, M. R. (1967). On Sampling without Replacement with Unequal Probabilities of Selection, *Biometrika* 54, 499-513.
8. Scheaffer, R.L., Mendenhall, W. and Ott, L. (1996). Elementary Sampling Survey, 5th ed., Duxbury, New York.
9. Singh, D. and F. S. Chaudhary (1986). Theory and Analysis of Sample Survey Designs, John Wiley & Sons, New York.
10. Thompson, S. K (2002). Sampling, 2end Wiley, New York.

الفصل الثاني عشر

المعاينة المزدوجة

Double Sampling

1.12 مقدمة

لقد سبق أن تكلمنا في الفصل السابع عن عدة طرق للمعاينة تعتمد على المعلومات المساعدة أو الإضافية أو ما يسمى بالمتغير المساعد x ،ولقد لاحظنا أن التقدير باستخدام النسبة R أو خط انحدار y على x أو الفرق يتطلب معرفة الوسط الحسابي للمجتمع للمتغير x ،وإذا كنا نرغب في تقسيم المجتمع إلى طبقات بالاعتماد على المتغير x فإن ذلك يتطلب معرفة التوزيع التكراري للمتغير المساعد x.

عندما لا تكون المعلومات الخاصة بالمتغير المساعد x متوافرة، ولكن في بعض الأحيان غير مكلفة أن نسحب عينة عشوائية كبيرة من المجتمع للمتغير X وحده ونجمع معلومات حوله. إن الهدف من هذه العينة الحصول على تقدير جيد إلى الوسط الحسابي للمجتمع أو التوزيع التكراري للمتغير المساعد X في كثير من المسوحات التي يكون الهدف منها تقدير معلمات المجتمع للمتغير y وحده، يمكن أن يكون مجديًا ونافعًا أن يخصص بعض الموارد المخصصة للمسح لسحب عينة أولية للحصول على معلومات حول المتغير المساعد x على الرغم من أن ذلك يعني أن حجم العينة لتقدير المتغير y سوف يتقلص، إن هذه الطريقة تعرف باسم المعاينة المزدوجة (Double Sampling).

إن هذه الطريقة للمعاينة مفضلة، إذا كانت الدقة التي نحصل عليها من خلال استخدام السبة، أو خط الانحدار، أو استخدام الطبقية تكون أكبر فيما لو استخدمنا جميع الموارد لسحب عينة بحجم معين للمتغير الأساس y وحده.

2.12 تقدير النسبة باستخدام العينة المزدوجة

الرموز والمصطلحات الآتية سوف نستخدمها في هذا الفصل

y_i : قيمة المتغير الأساسي، أو الصفة تحت الدراسة للوحدة i في المجتمع.

x_i : قيمة المتغير المساعد، أو الصفة المساعدة في نفس الوحدة i في المجتمع.

n' : عدد الوحدات التي جرى سحبها في العينة الأولى.

n : عدد الوحدات التي جرى سحبها في العينة الثانية.

ستجري مشاهدة كلا المتغيرين y وx لكل وحدة من وحدات العينة الثانية، أما في العينة الأولى فسوف تتم مشاهدة المتغير x فقط.

إذا كان المتغيرين y وx مرتبطين بعلاقة خطية قوية بحيث إذا كانت $x_i = 0$ فإن قيمة $y_i = 0$ ، لذا فإن تقدير النسبة مع العينة المزدوجة سيؤدي إلى تقدير أفضل لمعدل أو مجموع المجتمع للمتغير الأساسي y.

لنفترض أنه من مجتمع حجمه N جرى سحب عينة عشوائية مع عدم الإرجاع وبحجم n' من الوحدات وتمت مشاهدة المتغير x، كذلك تم سحب عينة عشوائية مع عدم الإرجاع وبحجم n من بين وحدات العينة الأولى n' وجرى مشاهدة المتغير y بالإضافة إلى المتغير x.

نستطيع أن نجد من العينة الصغيرة نسبة العينة وهو

$$r = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i}$$

حيث إن المجموع أخذ لجميع عناصر العينة الثانية.

نستطيع من العينة الكاملة التي تم مشاهدة المتغير x فيها أن نقدر مجموع المجتمع لهذا المتغير وهو

$$\hat{t}_x = \frac{N}{n'} \sum_{i=1}^{n'} x_i$$

ويمكننا أن نجد القيمة التقريبية لتباين $\hat{t}_x$

$$Var(\hat{t}_x) \approx N(N-n')\frac{s^2}{n'} + N^2(\frac{n'-n}{n'})\frac{s_r^2}{n}$$

حيث أن σ^2 عبارة عن تباين المجتمع للمتغير y و σ_r^2 يمثل تباين المجتمع حول خط النسبة ويمكن أن نعرفه بما يأتي :

$$\sigma_r^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - Rx_i)^2$$

بحيث أن $R = \tau / \tau_x$ وهو نسبة المجتمع. ويمكننا أن نحصل على تقدير للتباين $var(\hat{\tau}_x)$ وهو

$$s^2(\hat{\tau}_x) \approx N(N-n')\frac{s^2}{n'} + N^2(\frac{n'-n}{n'n(n-1)}) \sum_{i=1}^{n}(y_i - rx_i)^2$$

حيث أن s^2 هو تباين العينة للمتغير y .

مثال 1: تمت مشاهدة 240 غزال في مسح من الجو لعينة تتكون من 20 وحدة، علماً أن المجتمع يتكون من 100 وحدة متساوية المساحة تقريباً، ولقد

تم إرسال مجموعة من العادين؛ لكي يقوموا بمشاهدة 5 وحدة من بين الوحدات العشرين التي تم مسحها من الجو؛ لمشاهدتها بصورة تامة على الأرض، ولقد تمت مشاهدة 56 غزال في هذه الوحدات الخمس من الجو، ولكن عندما تمت مشاهدتها على الأرض تم رصد 70 غزال، وهذا يمثل الرقم الحقيقي لعدد الغزلان في هذه القطع الخمس. قدّر عدد الغزلان في منطقة الدراسة.

الحل:

$$r = \frac{\sum_{i=1}^{5} y_i}{\sum_{i=1}^{5} x_i} = \frac{70}{56} = 1.25$$

باستخدام العينة الكاملة $n' = 20$ نستطيع أن نقدر المجموع الكلي للمجتمع

$$\hat{\tau}_x = \frac{N}{n'}\sum_{i=1}^{n'} x_i = \frac{100}{20}\sum_{i=1}^{20} x_i = \frac{100}{20}(240) = 1200$$

الآن نستطيع أن نقدر عدد الغزلان باستخدام تقدير النسبة والمعتمد على العينة المزدوجة وهو

$$\hat{\tau}_r = r\,\hat{\tau}_x = 1.25(1200) = 1500$$

أي أننا نقدر مجموع الغزلان في منطقة الدراسة بنحو 1500 غزال.

3.12 توزيع العينة في المعاينة المزدوجة لتقدير النسبة

سيكون تقدير النسبة باستخدام العينة المزدوجة أكثر فاعلية إذا كان المتغيران y وx مرتبطين بعلاقة خطية قوية، بحيث إن قيمة $y=0$ إذا كانت $x=0$، وع كون تكلفة قياس المتغير x أرخص من y. إن القيمة المثلى لنسبة

حجم العينة الجزئية n إلى حجم العينة الكلية n' يعتمد على تكاليف مشاهدة أو قياس المتغيرين y وx وكذلك على قوة العلاقة بينهم.

لنفرض أن تكاليف مشاهدة وحدة واحدة للمتغير x هي c' وتكاليف مشاهدة المتغير y هي c، لذا فإن مجموع التكاليف C ستكون

$$C = c'n' + cn$$

إذا كانت التكاليف C ثابتة فإن الحد الأدنى لقيمة التباين للمتغير $\hat{\tau}_r$ يمكن إيجادها باستخدام العلاقة الآتية:

$$\frac{n}{n'} = \sqrt{\frac{c'}{c}\left(\frac{\sigma_r^2}{\sigma^2 - \sigma_r^2}\right)}$$

4.12 المعاينة المزدوجة للتقسيم المجتمع إلى طبقات

لا يمكن تصنيف الوحدات إلى طبقات في بعض الحالات إلا بعد أن نقوم بعملية سحب الوحدة، على سبيل المثال عندما نقوم بسحب عينة من الأشخاص ومن ثم الاتصال بهم هاتفياً لجمع معلومات منهم، لا يمكن تصنيفهم حسب الجنس أو المستوى التعليمي أو العمر أو نوع العمل إلا بعد الاتصال بهم هاتفياً. إن طريقة تقسيم المجتمع إلى طبقات بعد سحب العينة والتي سبق وأن تحدثنا عنها في الفصل السادس يمكن أن تكون مفيدة في مثل هذه الحالات إذا كان وزن الطبقة $W_i = N_i/N$ معلوماً لكل طبقة. إذا كان وزن الطبقة غير معلوم فيمكن استخدام المعاينة المزدوجة مع عينة أولية كبيرة يمكن استخدامها لتصنيف الوحدات إلى طبقات، ومن ثم نقوم بسحب عينة طبقية من العينة الأولية. يمكن أن تكون العينة المزدوجة مفيدة إذا كانت الوحدات في كل عينة متماثلة وإذا كان مشاهدة المتغير x أسهل وأقل تكلفة من المتغير y.

نقوم بسحب عينة عشوائية أولية بحجم n' وحدة من المجتمع ذي الحجم N، ومن ثم نقوم بتصنيف هذه الوحدات إلى طبقات في كل طبقة n'_i وحدة جرى مشاهدتها في الطبقة i و$i= 1, 2, \ldots, K$. نقوم بتقدير وزن الطبقة في المجتمع W_i باستخدام وزن الطبقة في العينة $w_i = \dfrac{n'_i}{n'}$ ، ومن ثم نقوم بسحب عينة طبقية من العينة الأولية بحيث نسحب n_i وحدة من الوحدات n'_i في الطبقة i، وأخيراً نقوم بمشاهدة وتسجيل جميع قيم المتغير y لكل وحدة في العينة الثانية. يمكن أن نعرف الوسط الحسابي للعينة الثانية للطبقة i

$$\bar{y}_i = \frac{1}{n_i}\sum_{i=1}^{n_i} y_i$$

وتقدير الوسط الحسابي للمجتمع يكون

$$\bar{y}_d = \sum_{i=1}^{K} w_i \bar{y}_i$$

التقدير $\bar{y}_d$ سيكون تقدير غير متحيز إلى الوسط الحسابي للمجتمع مع تباين

$$\operatorname{var}(\bar{y}_d) = \frac{N-n'}{N}\frac{\sigma^2}{n'} + E\sum_{i=1}^{K}[(\frac{n'_i}{n'})^2\ (\frac{n'_i-n_i}{n'_i})\frac{\sigma^2_{i(s_1)}}{n_i}]$$

حيث أن σ^2 يمثل التباين الكلي للمجتمع و $\sigma^2_{i(s_1)}$ يمثل تباين المجتمع في الطبقة i للوجه الأول للعينة s_1 . يمكن أن نحصل على تقدير غير متحيز لتباين $\bar{y}_d$ وهو

$$s^2(\bar{y}_d) = \frac{N-1}{N}\sum_{i=1}^{K}(\frac{n'_i-1}{n'-1} - \frac{n_i-1}{N-1})\frac{w_i s_i^2}{n_i} + \frac{N-n'}{N(n'-1)}\sum_{i=1}^{K} w_i(\bar{y}_i - \bar{y}_d)^2$$

إذا جرى تحديد النسبة n_i/n'_i وجرى استبعاد العينات بحجم $n'_i = 0$ فإن تباين $\bar{y}_d$ يكون تقريباً

$$\operatorname{var}(\bar{y}_d) = \frac{N-n'}{Nn'}\sigma^2 + \sum_{i=1}^{K}\frac{W_i\sigma_i^2}{n'}(\frac{n'_i}{n_i}-1)$$

ويبقي تقدير التباين كما هو لهذه الحالة. يراجع Thompson (2002) لمزيد من المعلومات.

5.12 التقديرات باستخدام الانحدار في العينة المزدوجة

من بعض تطبيقات المعاينة المزدوجة يمكن استخدام المتغير المساعد تقدير الوسط الحسابي $\overline{Y}$ للمجتمع باستخدام الانحدار. لنفرض أنه جرى سحب عينة عشوائية بحجم n' من الوحدات وتمت مشاهدة المتغير المساعد x، كذلك تم سحب عينة عشوائية بحجم n وجرى مشاهدة المتغير الأساسيy بالإضافة إلى المتغير المساعد x. لذا فإن تقدير $\overline{Y}$ باستخدام انحدار y على x يمكن تعريفه

$$\overline{y}_{lr} = \overline{y} + b(\overline{x}' - \overline{x})$$

حيث $\overline{x}'$ و $\overline{x}$ هما متوسطا المتغير المساعدx في العينتين الأولى والثانية وb معامل انحدار المربعات الصغرى للمتغير y على x محسوباً من العينة الثانية. وإذا لم نفترض بوجد علاقة خطية بين المتغيرين في المجتمع فإن $\overline{y}_{lr}$ سيكون منحازاً. وإذا افترضنا يمكننا إهمال $1/n$ و$1/n'$ ويمكننا إعطاء تقدير تقريبي لتباين $\overline{y}_{lr}$

$$var(\overline{y}_{lr}) \approx \frac{S_y^2(1-\rho^2)}{n} + \frac{\rho^2 S_y^2}{n'} - \frac{S_y^2}{N}$$

حيث S_y^2 يمثل تباين المجتمع للمتغير y و ρ معامل الارتباط للمجتمع بين المتغيرين y وx.

أما تقدير التباين $1/n$ و $1/n'$ فسيكون

$$s^2(\overline{y}_{lr}) = \frac{s_{yx}^2}{n} + \frac{s_y^2 - s_{yx}^2}{n'} - \frac{s_y^2}{N}$$

حيث

$$s_{yx}^2 = \frac{1}{n-2}[\sum_{i=1}^{n}(y_i - \overline{y})^2 - b^2\sum_{i=1}^{n}(x_i - \overline{x})^2]$$

و

$$s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2$$

أما إذا لم يكن بالإمكان إهمال 1/n لصغر حجم العينة، فإن التقدير المقترح لتقدير التباين سيكون

$$s^2(\overline{y}_{lr}) = s_{yx}^2\left[\frac{1}{n} + \frac{(\overline{x}' - \overline{x})^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}\right] + \frac{s_y^2 - s_{yx}^2}{n'} - \frac{s_y^2}{N}$$

لمزيد من المعلومات يراجع (1977) Cochran، أو ترجمة أنيس كنجو(1995).

6.12 تكرار المعانية لنفس المجتمع

لقد أصبح الاعتماد على العينة لجمع المعلومات حول مجتمعات الدراسة المختلفة وخصوصاً الإنسانية منها والتي هي لمدد زمنية متقاربة، من الأمور المتعارف عليها بل أحياناً أصبحت ضرورةً ملحة؛ وذلك لكون المجتمعات متحركة ومتبدلة بصورة كبيرة، وأصبحت المسوحات التي تجرى كل عشر سنوات مثل المسوحات السكانية قليلة الفائدة بعد سنة أو سنتين؛ وذلك للتغيرات السريعة والكبيرة التي تحصل بسبب تبدل آراء الناس حول مختلف الموضوعات؛ وذلك للتطور العلمي والتكنولوجي الذي يشهده العالم هذه السنوات وخصوصاً العشرة سنوات الأخيرة.

عندما تجرى المسوحات بواسطة العينة ولمُددٍ متقاربة يكون الباحث في موضع مثالي ليقوم بتقديرات واقعية للتكاليف والتباينات ويمكنه من تطبيق الطريقة التي تمكنه من الوصل إلى فاعلية مثلى للمعاينة. من الأسئلة المهمة هي ما هي الطريقة التي تتم بها تغير العينة مع مرور الزمن؟ وما هي المدة الزمنية اللازمة للتغير؟ هنالك اعتبارات كثيرة قد تؤثر على القرار. ربما لأنَّ

بعض الناس لا يرغبون بإعطاء المعلومات نفسها مرة بعد أخرى، وربما يكون بعض الناس متأثرين ببعض المعلومات المسبقة حول المسح قد تمنعهم من إعطاء معلومات دقيقة في المقابلة الأولى، في بعض الأحيان التعاون يكون أفضل في المقابلة الثانية منه في الأولى والمعلومات التي تعطى تكون أكثر دقة خصوصاً إذا كانت المعلومات المطلوبة تقنية أو سرية أو شخصية.

إذا حصلنا على بيانات من عينات متتابعة، فهناك ثلاثة أنواع من التقديرات ربما يفكر الباحث القيام بها وهي:

1. التغير في الوسط الحسابي للمجتمع $\overline{Y}$ من مناسبة إلى المناسبة التي تليها.
2. معدل الوسط الحسابي للمجتمع $\overline{Y}$ لجميع المناسبات.
3. المعدل للوسط الحسابي للمجتمع $\overline{Y}$ في المناسبات الأحدث.

في كثير من المسوحات بالعينة يكون التركيز على المعدل للمناسبات الأحدث، خصوصاً إذا كان التغير في المجتمع يكون سريعاً وكبيراً من مدة زمنية لأخرى، ولكن إذا كان التغير بطيئاً فإنَّ المعدل لجميع المناسبات هو الأفضل، أما التغير في الوسط الحسابي للمجتمع من مناسبة لأخرى فيستعمل في الغالب لمعرفة التغيرات التي قد تحصل على الوسط الحسابي إذا حدثت ظروف نتوقع أنها تؤثر على الوسط الحسابي للمجتمع بشكل ملحوظ. على سبيل المثال رفع رواتب العاملين بالدولة وتأثيره على ارتفاع الأسعار أو زيادة الاستثمار.

لنفرض أننا لدينا الحرية في تغير تركيب العينة أو الاحتفاظ به، وأن حجم العينة الكلي سيبقي نفسه في جميع المناسبات، إذا رغبنا الحصول على أعلى دقة فيمكننا وضع العبارات الآتية حول سياسة الاستبدال:

1. من الأفضل أن نحافظ على نفس العينة خلال جميع المناسبات إذا رغبنا في تقدير التغير.
2. إذا رغبنا في تقدير المتوسط لجميع المناسبات من الأفضل سحب عينة جديدة في كل مناسبة.
3. إذا كنا نركز على التقديرات الراهنة، فإننا نحصل على الدقة نفسها سواء احتفظنا بالعينة نفسها أو بدَّلناها في كل مناسبة، وقد يكون من الأفضل تغير جزء من العينة في كل مناسبة.

تمارين

1. في مسح جوي لأربع قطع جرى سحبها من بين عشر قطع في منطقة الدراسة باستخدام العينة العشوائية البسيطة، لوحظ أنَّ عدد طيور البط التي جرى مشاهدتها في هذه القطع الأربعة هي على النحو الآتي: 44, 55, 16 ,4. ولكن عند فحص الصور الفوتوغرافية للقطعة الأولى والثالثة (التي جرى سحبهما باستخدام العينة العشوائية البسيطة بصورة دقيقة) تبين أن عدد طيور البط في هاتين القطعتين هو 56 و6. قدر مجموع البط في منطقة الدراسة باستخدام تقدير النسبة، ومن ثم أوجد تقدير التباين لهذا التقدير.

2. لتقدير المصروفات الشهرية للمصروفات على العلاج الطبي في مدينة تحتوي على 5000 عائلة، جرى سحب عينة عشوائية تحتوي على500 عائلة، فوجد أن هنالك 336 عائلة لديها أطفال و164 دون أطفال، ومن ثم جرى سحب عينة جزئية طبقية تحتوى على 112 عائلة لديها أطفال و41 لا يوجد لديها أطفال، ولقد جرى جمع معلومات حول المصروفات الشهرية لأفراد العينة الجزئية وحصلنا على المعلومات الآتية:

	الوسط الحسابي	الانحراف المعياري	حجم العينة
دون أطفال	$280	$160	41
مع أطفال	$110	$60	112

قدر معدل المصروفات الطبية الشهرية للعائلة، ومن ثم قدر تباين التقدير.

3. في أحد تطبيقات المعاينة المزدوجة مع الانحدار كان حجم العينة الصغيرة 87 وحجم العينة الكبيرة 300. النتائج الآتية تتعلق بالعينة الصغيرة

$$\sum_{i=1}^{87}(y_i-\bar{y})(x_i-\bar{x})=5114 \text{ و } \sum_{i=1}^{87}(x_i-\bar{x})^2=3248 \text{ و } \sum_{i=1}^{87}(y_i-\bar{y})^2=17283$$

أوجد تقدير الخطأ المعياري لتقدير الانحدار.

المراجع العربية

1. وليم كوكوران – ترجمة أنيس كنجو .(1995) تقنية المعاينة الإحصائية ، جامعة الملك سعود ، الرياض.

References

1. Cochran, W. G. (1961). Comparison of Methods for Determined Stratum Boundaries. *Bull. Inst. Stat. Inst.*, 38, 345-358.
2. Cochran, W. G. (1977). Sampling Technique, 3rd Ed. Wiley, New York.
3. Rao, J. N. K. (1973). On Double Sampling for Stratification and Analytical Surveys. *Biometrika*, 60, 125-133.
4. Sӓmdal, C. E. and Swensson, B. (1987). A General View of Estimation of Two Phases of Selection with Applications to Two-phase Sampling and Nonresponse. *International Statistical Review,* **55**,279-294.
5. Schafer, J. L. (1997). Analysis of Incomplete Multivariate Data, London Chapman & Hall.
6. Singh, D. and Chaudhary, F. S. (1986). Theory and Analysis of Sample Survey Designs, New Delhi, Wiley Eastern.
7. Thompson, S. K. (2002). Sampling, 2end Wiley, New York.