أساليب المعاينة

Sampling Techniques

قسم الإحصاء

المرحلة الثانية

الفصل الدراسي الاول

2020-2019

فراس منذر

الفصل الأول

المقدمة

طرائق المعاينة هي الجزء الأساسي في علم الإحصاء والذي يهتم بعملية جمع البيانات لإجراء الدراسات العلمية والبحث العلمي، لتتم بعدها عملية تحليل البيانات بالطرائق الإحصائية المعروفة فإذا كان إسلوب جمع البيانات إسلوبا دقيقا فهذا يؤدي الى الحصول على نتائج وإستنتاجات منطقية ومقبولة .

بعض المفاهيم الأساسية:

قبل الشروع في تفاصيل أساليب المعاينة الإحصائية فإنه من الضروري معرفة بعض المفاهيم والمصطلحات التي تستعمل في مجال العينات:

الوحدة الإحصائية: Statistical Unit

الوحدة الإحصائية هي المفردة أو المشاهدة التي تؤلف جميعها المجتمع الإحصائي المدروس فقد تمثل الغنسان كما هو الحال التعداد السكاني، او قد تمثل منتجات حيوانية أو زراعية في دراسات الثروة الحيوانية والنباتية، او قد تكون جماذا كما هو الحال في دراسات وسائل النقل.

المجتمع: Population

يقصد بالمجتمع جميع الوحدات الإحصائية أو مفردات الظاهرة المدروسة، أو هو محجموعة الوحدات الإحصائية المعرفة بصورة واضحة بحيث يمكن تميزها لكي تدخل ضمن مشاهدات مجتمع ما، فسكان مدينة ما هو مجتمع مدروس، أو مجموعة منتجة من سلعة معينة تمثل مجتمع إنتاج معين.

العينة: Sample

هي مجموعة من الوحدات الإحصائية (جزء من المجتمع) يتم إختيارها وفق أسس وقواعد معينة لتمثل المجتمع المدروس تمثيلا صحيحا.

إسلوب المعاينة: Sampling Technique

ويمكن تقسيم الدراسات الإحصائية من حيث درجة الشمول الى نوعين من الدراسات النوع الأول دراسات بطريقة الحصر الشامل وهي الدراسات التي تشمل كافة مفردات المجتمع المدروس ويلجأ اليها الباحث عندما لاتكون لديه معرفة بطبيعة مفردات المجتمع المدروس، أما النوع الثاني من الدراسات فهو طريقة العينات فعندما يكون الباحث على دراية عن بعض المعلومات التي تخص المجتمع المدروس بحيث تمكنه من إختيار عينة تمثل ذلك المجتمع تمثيلا جيدا.

تجدر الإشارة الى إن إختيار العينات ليس مجرد إختيار مفردات تمثل المجتمع المدروس إنما هو عبارة عن طرائق وأساليب وقواعد تعتمد على نظرية الإحتمالات والرياضيات.

مميزات إسلوب المعاينة:

من اهم مميزات إسلوب المعاينة:

1. إختصار الوقت وتقليل الجهد وتقليص التكاليف.
2. سرعة وسهولة الحصول على النتائج.
3. توسيع وتعميم نتائج التحليل الإحصائي العينة على المجتمع المدروس.
4. البحث في دراسات لايمكن تحليلها إحصائيا من خلال إسلوب الحصر الشامل كما هو الحال في الدراسات الطبية، أو الصناعية.
5. من خلال إسلوب المعاينة يمكن معرفة دقة النتائج التي يتم الحصول عليها ومدى تمثيلها للمجتمع المدروس وهذه الميزة لاتتوفر في إسلوب الحصر الشامل.

بعض المقاييس الإحصائية:

من الضروري مراجعة لبعض المقاييس الإحصائية التي تمثل صفات معينة للمجتمع المدروس من خلال العينة وكما ياتي:

اذا كانت $x_1, x_2, \dots, x_n$ تمثل قياسات لمفردات عينة بحجم n فان الوسط الحسابي لتلك العينة يحسب كما ياتي:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad (1-1)$$

وان تباين العينة يحسب كما ياتي:

$$S_x^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} \qquad (1-2)$$

واذا كانت $X = (x_1, x_2, \dots, x_n)$ و $Y = (y_1, y_2, \dots, y)$ تمثل قياسات لمفردات عينتين بحجم n فان التباين المشترك بين المتغيرين X و Y يحسب كما ياتي:

$$Cov(X,Y) \; or \; S_{x,y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x}) \; (y_i - \bar{y})}{n-1} \qquad (1-3)$$

وان معامل الارتباط بين المتغيرين بين المتغيرين X و Y يمكن حسابه كما ياتي:

$$r_{x,y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x}) \; (y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} = \frac{S_{x,y}}{S_x \, S_y}$$

اذ ان $S_x \,, S_y$ هو الانحرافات المعيارية للمتغيرين X و Y ، وان $S_{x,y}$ هو التباين المشترك بين المتغيرين X و Y. وكما هو معروف فان معامل الارتباط هو مقياس لطبيعة وقوة العلاقة بين المتغيرات اذ تتراوح قيمته بين [-1 , 1] ويمكن تفسير طبيعة وقوة العلاقة بين متغيرين حسب قيمة معامل الارتباط كما هو مبين في الجدول التالي:

قيمة معامل الارتباط r	طبيعة وقوة العلاقة
$r = -1$	عكسية تامة
$-1 < r < -0.5$	عكسية قوية
$r = -0.5$	عكسية متوسطة
$-0.5 < r < 0$	عكسية ضعيفة
$0 < r < 0.5$	طردية ضعيفة
$r = 0.5$	طردية متوسطة
$0.5 < r < 1$	طردية قوية
$r = 1$	طردية تامة

مثال: للبيانات التالية:

x_i	20	40	60	25	35	30	50	5	10	25
y_i	30	60	80	100	200	175	150	50	125	100

احسب الوسط الحسابي والتباين والانحراف المعياري لكل عينة، ثم احسب التباين المشترك ومعامل الارتباط بين المتغيرين X و Y موضحا طبيعة وقوة العلاقة بين المتغيرين؟

الحل:

x_i	y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
20	30	20-30= -10	30-107= -77	100	5929	-10*-77=770
40	60	10	-47	100	2209	-470
60	80	30	-27	900	729	-810
25	100	-5	-7	25	49	35
35	200	5	93	25	8649	465
30	175	0	68	0	4624	0
50	150	20	43	400	1849	860
5	50	-25	-57	625	3249	1425
10	125	-20	18	400	324	-360
25	100	-5	-7	25	49	35
300	**1070**	**0**	**0**	**2600**	**27660**	**1950**

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{25 + 10 + \cdots + 20}{10} = \frac{300}{10} = 30$$

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{100 + 125 + \cdots + 30}{10} = \frac{1070}{10} = 107$$

$$S_x^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1} = \frac{2600}{9} = 288.88$$

$$S_y^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n-1} = \frac{27660}{9} = 2073.33$$

$$S_{x,y} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})\ (y_i - \bar{y})}{n-1} = \frac{1950}{9} = 216.67$$

$$r_{x,y} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})\ (y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}} = \frac{1950}{\sqrt{(288.88)(2073.33)}} = 0.28$$

واجب: اذا توفرت لديك البيانات والمعلومات التالية :

.1

x_i	0.5	0.4	0.1	1.5	2.5
y_i	8	5	2	10	10

2. $\sum_{i=1}^{20} x_i = 400 \, , \sum_{i=1}^{20} y_i = 1600 \, , S_x^2 = 64, S_y^2 = 100, S_{x,y} = -16$

احسب معامل الارتباط وبين طبيعة وقوة العلاقة بين المتغيرات؟

الفصل الثاني

المعاينة العشوائية البسيطة

Simple Random Sampling

تعرف العينة العشوائية البسيطة على انها عملية اختيار عدد معين من مفردات المجتمع بدون ارجاع للمفردة المسحوبة بحيث تكون لكل عينة من تلك العينات المسحوبة نفس فرصة الاختيار . المعاينة العشوائية البسيطة هي من اسهل اساليب المعاينة الاحتمالية لكنه ليس الاسلوب الاكثر شيوعا.

كيفية اختيار العينة العشوائية البسيطة:

ان افضل طريقة لسحب عينة عشوائية بسيطة بحجم (n) من مجتمع مؤلف من (N) من المفردات وبدون ارجاع لوحدة المعاينة المسحوبة هي عملية ترقيم وحدات المجتمع وسحبها عن طريق القرعة، وبذلك تكون عملية سحب العينة بشكل عشوائي تام. عليه فان:

عدد العينات التي يمكن سحبها يساوي $\binom{N}{n} = \frac{N!}{n!(N-n)!}$ ، وان فرصة او احتمال اختيار اي عينة يساوي $\frac{1}{N}$.

الرموز والمصطلحات:

يمكن تلخيص رموز ومصطلحات المعاينة العشوائية البسيطة كما مبين في الجدول التالي:

الخاصية	المجتمع	العينة
الوسط الحسابي	$\bar{Y} = \frac{\sum_{i=1}^{N} Y_i}{N}$	$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$
المجموع الكلي	$Y = \sum_{i=1}^{N} Y_i = Y_1 + Y_2 + .. + Y_N$	$y = \sum_{i=1}^{n} y_i = y_1 + y_2 + .. + y_N$
التباين	$\sigma_Y^2 = \frac{\sum_{i=1}^{N} (Y_i - \bar{Y})^2}{N}$	$S_y^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n-1}$

ان ايجاد خاصية معينة من خلال العلاقات الرياضية الخاصة بتلك الخاصية عن طريق اسلوب المعاينة يسمى بعملية التقدير (Estimation)، فمثلا القيمة التقديرية لمتوسط

المجتمع يرمز لها بالرمز $\hat{\bar{Y}}$ او القيمة التقديرية للمجموع الكلي لمفردات المجتمع $\hat{Y}$، كذلك فان القيمة التقديرية لتباين المجتمع يرمو لها بالرمز $\hat{\sigma}_Y^2$.

خصائص المعاينة العشوائية البسيطة:

سيتم التطرق الى بعض خصائص المعاينة العشوائية البسيطة من خلال النظريات والمبرهنات الموضوعة لهذا الغرض دون التطرق الى الاثباتات الرياضية لهذه النظريات والمبرهنات.

نظرية (2-1):

ان الوسط الحسابي للعينة $\bar{y}$ هو تقدير غير متحيز للوسط الحسابي للمجتمع $\bar{Y}$ اي ان

$$E(\bar{y}) = \bar{Y}$$

نظرية (2-2):

ان تباين الوسط الحسابي للعينة العشوائية البسيطة يعرف وفق الصيغة التالية:

$$\sigma_{\bar{y}}^2 \ or \ V(\bar{y}) = \frac{\sigma^2}{n}\ (1-f)$$

وان الخطا المعياري للوسط الحسابي للعينة يعرف وفق الصيغة التالية:

$$\sigma_{\bar{y}} = \frac{\sigma}{\sqrt{n}} \sqrt{(1-f)}$$

اذ ان f يمثل مايعرف بكسر المعاينة ، $f = \frac{n}{N}$.

نظرية (2-3):

ان تباين العينة S^2 هو تقدير غير متحيز لتباين المجتمع σ^2 ، اي ان :

$$E(s^2) = \sigma^2$$

نتيجة (1-2):

ان تباين المجموع الكلي للعينة العشوائية البسيطة يعطى وفق الصيغة التالية:

$$\sigma_{y}^{2} \ or \ V(y) = N^{2}\frac{\sigma^{2}}{n}\ (1-f)$$

وان الخطا المعياري للمجموع الكلي يعرف وفق الصيغة التالية:

$$\sigma_{y} = N\frac{\sigma}{\sqrt{n}}\ \sqrt{(1-f)}$$

نتيجة (2-2):

ان التقديرات الغير متحيزة لكل من تباين الوسط الحسابي وتباين المجموع الكلي للعينة العشوائية البسيطة هي

$$S_{\bar{y}}^{2} = \frac{S^{2}}{n}\ (1-f) \quad , \quad S_{y}^{2} = N^{2}\frac{S^{2}}{n}\ (1-f)$$

نتيجة (2-3):

ان التقديرات الغير متحيزة لكل من الانحراف المعياري للوسط الحسابي والانحراف المعياري المجموع الكلي للعينة العشوائية البسيطة هي

$$S_{\bar{y}} = \frac{S}{\sqrt{n}}\ \sqrt{(1-f)} \quad , \quad S_{y} = N\frac{S}{\sqrt{n}}\ \sqrt{(1-f)}$$

حدود الثقة: Confidence Limits

تعرف حدود الثقة على انها المدى الذي تقع فيه القيمة الحقيقية للمعلمة (الخاصية المقدرة، وان صيغة حدود الثقة للوسط الحسابي لعينة عشوائية بسيطة هي كما ياتي:

$$Pr\left[\bar{y} - \left\{t_{\alpha/2}\ S\sqrt{\frac{1-f}{n}}\right\} < \bar{Y} < \bar{y} + \left\{t_{\alpha/2}\ S\sqrt{\frac{1-f}{n}}\right\}\right]$$

مثال (2-1) : تم سحب عينة عشوائية بسيطة بحجم (3) مفردات من مجتمع مؤلف من اربعة مفردات هي (2،3،4،7).

1. احسب عدد العينات الممكن سحبها.
2. اثبت ان $E(\bar{y}) = \bar{Y}\ ,\ \ E(s^2) = \sigma^2$.

الحل:

$$m = \binom{N}{n} = \binom{4}{3}\frac{N!}{n!\,(N-n)!} = \frac{4!}{3!\,(4-3)!} = \frac{4\times 3!}{3!\,1!} = 4$$

$$\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i = \frac{1}{4}\ (2+3+4+7) = 4$$

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \bar{Y})^2 = \frac{\{(2-4)^2 + (3-4)^2 + (4-4)^2 + (7-4)^2\}}{4-1} = \frac{14}{3} = 4.66$$

$\bar{y}_1 = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{3}\ (2+3+4) = 3\ \ ,\bar{y}_2 = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{3}\ (2+3+7) = 4$

$\bar{y}_3 = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{3}\ (2+4+7) = 4.3, \bar{y}_3 = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{3}\ (3+4+7) = 4.7$

Samples	$\bar{y}_j$	$S_j^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \bar{y}_j\right)^2$
(2,3,4)	3	$S_1^2 = \frac{1}{3-1}(2-3)^2 + (3-3)^2 + (4-3)^2 = 1$
(2,3,7)	4	$S_2^2 = \frac{1}{3-1}(2-4)^2 + (3-4)^2 + (7-4)^2 = 7$
(2,4,7)	4.3	$S_3^2 = \frac{1}{3-1}(2-4.3)^2 + (4-4.3)^2 + (7-4.3)^2$ $= 5.29 + 0.09 + 7.29 = 6.335$
(3,4,7)	4.7	$S_4^2 = \frac{1}{3-1}(3-4.7)^2 + (4-4.7)^2 + (7-4.7)^2$ $= 2.89 + 0.49 + 5.29 = 4.33$

$$E(\bar{y}) = \frac{\sum_{j=1}^{m}\bar{y}_j}{m} = \frac{3+4+4.3+4.7}{4} = 4$$

$$E(\bar{y}) = \bar{Y} = 4$$

$$E(S^2) = \frac{\sum_{j=1}^{m} \bar{y}_j}{m} = \frac{1 + 7 + 6.335 + 4.335}{4} = 4.66$$

$$E(S^2) = \sigma^2 = 4.66$$

مثال (2-2) : مجتمع بحجم (1000) مفردة، سحبت منه عينة عشوائية بسيطة بحجم (100) مفردة بوسط حسابي يبلغ (50) بتباين يبلغ (16)، المطلوب تقدير:

1. تباين الوسط الحسابي.
2. حدود الثقة للوسط الحسابي عند مستوى معنوية (0.05).

الحل:

$$S_{\bar{y}}^2 = \frac{S^2}{n}(1-f) = \frac{16}{100}\left(1 - \frac{100}{1000}\right) = (0.16)(0.9) = 0.14$$

$$Pr\left[\bar{y} - \left\{t_{\alpha/2}\ S\sqrt{\frac{1-f}{n}}\right\} < \bar{Y} < \bar{y} + \left\{t_{\alpha/2}\ S\sqrt{\frac{1-f}{n}}\right\}\right] = (1-\alpha)\%$$

$$L = \bar{y} - \left\{t_{\alpha/2}\ S\sqrt{\frac{1-f}{n}}\right\} = 50 - (1.96)(4)\left(\sqrt{\frac{1-0.1}{100}}\right) = 0.95$$

$$L = 50 - (7.84)(0.09) = 50 - 0.7 = 49.3$$

$$U = \bar{y} + \left\{t_{\alpha/2}\ S\sqrt{\frac{1-f}{n}}\right\} = 50 + (1.96)(4)\left(\sqrt{\frac{1-0.1}{100}}\right)$$

$$U = 50 + (7.84)(0.09) = 50 + 0.7 = 50.7$$

$$Pr[49.3 < \bar{Y} < 50.7] = 0.95$$

مثال (2-3) : البيانات التالية تمثل عينة عشوائية بسيطة مسحوبة من مجتمع بحجم (500) مفردة:

y_i	50	30	60	20	55	75	25	35	100	10

المطلوب تقدير:

1. الوسط الحسابي والتباين والمجموع الكلي لمفردات المجتمع.

1. التقدير غير المتحيز لتباين الوسط الحسابي وتباين المجموع الكلي.

2. حدود الثقة للوسط الحسابي عند مستوى معنوية (0.1).