المعاينة العشوائية الطبقية

وهي من أهم أساليب المعاينة المستخدمة في البحوث الاحصائية. و يتم اللجوء إليها إذا كان المجتمع يتكون من فئات (طبقات) يكون التجانس أو التقارب داخل كل واحدة منها أكبر من التجانس داخل المجتمع ككل (أي أن التشتت داخل المجتمع أكبر من التشتت داخل كل فئة من فئاته على حده).

في هذه الحالة يراعى في تمثيل الطبقة داخل العينة وزنها الحقيقي داخل المجتمع أو درجة التشتت داخل الطبقة.

بعد تحديد حجم كل طبقة، تسحب مفردات العينة عشوائياً من داخل الطبقة علما أن مجموع أحجام الطبقات يكون الحجم الاجمالي للعينة العشوائية الطبقية.

يتطلب تنفيذ العينة الطبقية العشوائية بعض الشروط الضرورية منها على الخصوص:

- وجود متغير له ارتباط كبير مع الصفة أو الخاصية المدروسة بواسطة البحث.
- توفر اطار أو قاعدة المعاينة على بيانات دقيقة من حيث مستوى ودرجة تشتت المتغير المستخدم في المجتمع وذلك من أجل تمكين مصمم العينة من تقسيم المجتمع إلى طبقات.

1. مصطلحات:

<u>المجتمع</u>

- N: حجم المجتمع
- y : المتغير المدروس
- $\bar{Y}$: المتوسط الحسابي للمجتمع

$$\bar{Y} = \sum_{1}^{H} W_h \bar{Y}_h$$

$$\bar{Y} = \frac{1}{N} \sum_{h}^{H} \sum_{i}^{N_h} y_{hi}$$

- P: نسبة الأفراد الذين تتوفر فيهم الخاصية المدروسة في المجتمع
- σ^2: التباين
- S^2: شبه التباين

الطبقات

- n هو حجم العينة
- n_h هو حجم العينة في الطبقة h
- H هو عدد الطبقات
- أحجام الطبقات: $h_1, h_2, h_3, \dots \dots \dots h_H$

هي على التوالي: $N_1, N_2, N_3, \dots \dots \dots N_H$

- $W_h = \frac{N_h}{N}$ هو وزن الطبقة h
- $f_h = \frac{n_h}{N_h}$ هي نسبة المعاينة في الطبقة h
- $\overline{Y}_h$ هو المتوسط الحسابي في الطبقة h

$$\bar{Y}_h = \frac{1}{N_h}\sum_{h}^{H} y_{hi}$$

- P_h نسبة الخاصية المدروسة في الطبقة h
- $\sigma^2{}_h$ التباين في الطبقة h
- $S^2{}_h$ هو شبه التباين في الطبقة h

2. المبدأ

في حالة المعاينة العشوائية البسيطة بدون إعادة، نتذكر صيغة تباين مقدر المتوسط الحسابي التي هي على الشكل التالي:

$$V\left(\hat{\bar{Y}}\right) = (1 - f)\frac{S^2}{n}$$

يستنتج من هذه المعادلة أن دقة المقدر ترتبط أساسا بمسألتين هامتين في مجال العينات و هما حجم العينة n ودرجة تشتت الظاهرة المدروسة داخل المجتمع S^2 .

من هذا المنطلق يمكن استخلاص أنه بالنسبة لحجم محدد فإن العينة تكون أكثر فعالية في مجتمع متجانس مقارنة مع مجتمع غير متجانس. وبعبارة أخرى كلما كان المجتمع غير متجانس كلما ارتفع تباين المقدر مما يجعل التقدير أقل دقة. فاللجوء إلى المعاينة في المجتمع غير متجانس جدا يقتضي سحب عينة ذات حجم أكبر.

ولتفادي الرفع من حجم العينة الذي عادة ما يكون مكلفا، يتعين تقسيم المجتمع إلى مجموعات أو فئات متجانسة.

وهكذا فإن مبدأ المعاينة الطبقية يرتكز على فكرة مفادها تكوين طبقات متجانسة داخل المجتمع أولا ثم سحب عينة عشوائية من كل طبقة بشكل يمكن من تقليص حجم العينة مقارنة مع ما يقتضيه الأمر في حالة اعتماد المعاينة العشوائية البسيطة.

3. الهدف من استعمال العينة العشوائية الطبقية

تهدف المعاينة الطبقية بالخصوص إلى:

- الرفع من دقة التقديرات بالنسبة للمجتمع المدروس.
- تقليص حجم العينة.
- ضمان الدقة الكافية بخصوص التقديرات المتعلقة ببعض الطبقات التي قد نعتبرها مجالات للدراسة على حدة.

تعتبر المعاينة العشوائية الطبقية أكثر نجاعة وفعالية من المعاينة العشوائية البسيطة.

4. تقدير المعلمات

المتوسط الحسابي

المقدر غير المتحيز للمتوسط $\overline{Y_s}$ في المعاينة العشوائية الطبقية هو نظيره $\overline{y_s}$ المحسوب انطلاقا من العينة ويعرف كما يلي:

$$\widehat{\overline{Y}} = \overline{y_s}$$

$$\overline{y_s} = \sum_{h}^{H} \frac{N_h}{N}\overline{y}_h$$

$$= \sum_{h}^{H} W_h \overline{y}_h$$

علما أن:

$$\overline{y}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} y_{hi}$$

التحيز

المتوسط الحسابي للعينة غير متحيز:

$$E(\overline{y_s}) = E\left(\sum_{h}^{H} W_h \overline{y}_h\right)$$

$$= \sum_{h}^{H} W_h\, E(\overline{y_h})$$

$$= \sum_{h}^{H} W_h \overline{Y}_h$$

$$= \sum_{h}^{H} \frac{N_h}{N} \frac{1}{N_h} \sum_{i=1}^{N_h} y_i$$

$$= \frac{1}{N} \sum_{h=1}^{H} \sum_{i=1}^{N_h} y_i$$

$$= \overline{Y}$$

♣ **تباين المتوسط الحسابي**

$$V(\overline{y_s}) = \sum_{h}^{H} W_h^2 \, V(\overline{y}_h)$$

$$= \sum_{h=1}^{H} W_h^2 \frac{N_h - n_h}{N_s} \frac{S_h^2}{n_h}$$

$$= \sum_{h=1}^{H} W_h^2 (1 - f_h) \, \frac{S_h^2}{n_h}$$

♣ **تقدير تباين المتوسط الحسابي**

علما أن شبه التباين S_h^2 غير معروف نقوم بتقديره ب s_h^2 الذي يتم حسابه انطلاقا من العينة المسحوبة من الطبقة h. إذن مقدر تباين المتوسط الحسابي هو:

$$\widehat{V(\overline{y_s})} = \sum_{h=1}^{H} W_h^2 (1 - f_h) \, \frac{s_h^2}{n_h}$$

♣ مجال الثقة

بالنسبة للمتوسط الحسابي للمجتمع، فان مجال الثقة بمستوى (1-α) هو:

$$\bar{y}_s - z_{\left(1-\frac{\alpha}{2}\right)}\sqrt{\sum_{h=1}^{H} W_h^2(1-f_h)\,\frac{s_h^2}{n_h}} < \bar{\boldsymbol{Y}} < \bar{y}_s + z_{\left(1-\frac{\alpha}{2}\right)}\sqrt{\sum_{h=1}^{H} W_h^2(1-f_h)\,\frac{s_h^2}{n_h}}$$

المجموع

في المعاينة الطبقية، يكتب المقدر غير المتحيز لمجموع متغير كمي في المجتمع:

$$\hat{Y} = \sum_{h}^{H} N_h \overline{y}_h$$

♣ التحيز

مجموع العينة غير متحيز:

$$E(\hat{Y}) = E\left(N\hat{\overline{Y}}\right)$$

$$NE\left(\hat{\overline{Y}}\right)$$

$$= \ Y$$

♣ تباين المجموع

$$V(\hat{Y}) \ = V(N\,\overline{y_s})$$

$$=N^2\, V(\,\overline{y_s})$$

$$= \sum_{h=1}^{H} N_h^2 (1 - f_h)\ \frac{S_h^2}{n_h}$$

♣ تقدير تباين المجموع

$$\widehat{V(\hat{Y})} \ = \sum_{h=1}^{H} N_h^2 (1 - f_h)\ \frac{s_h^2}{n_h}$$

♣ مجال الثقة

مجال الثقة للمجموع Y بمستوى (1-α) هو:

$$y_s - z_{(1-\frac{\alpha}{2})}\sqrt{\sum_{h=1}^{H} N_h^2(1-f_h)\ \frac{s_h^2}{n_h}} < \ \boldsymbol{Y}\ < y_s + z_{(1-\frac{\alpha}{2})}\sqrt{\sum_{h=1}^{H} N_h^2(1-f_h)\ \frac{s_h^2}{n_h}}$$

النسبة

إذا كان المتغير وصفي، فإن تقدير نسبة خاصية ما في المجتمع يتم بالكيفية الموضحة في الفصل السابق المتعلق بالمعاينة البسيطة عن طريق تقدير متغير y يساوي:

- 1 إذا كانت الوحدة تتوفر على الخاصية المدروسة
- 0 إذا كانت الوحدة لا تتوفر على الخاصية المدروسة

وهكذا فإن الصيغ المتعلقة بتقدير نسبة الخاصية في المجتمع وتباينها وتقدير التباين يكتب:

♣ نسبة الخاصية في المجتمع

$$\hat{P} = p_s$$

$$= \sum_{h}^{H} W_h\, p_h$$

♣ تباين النسبة

$$V(p_s) \ = \sum_{h=1}^{H} W_h^2(1-f_h)\frac{P_h Q_h}{n_h}$$

علما أن:

$$Q_h \ = \ 1 - P_h$$

♣ تقدير تباين النسبة

$$V(p_s) = \sum_{h=1}^{H} W_h^2 (1 - f_h) \frac{p_h q_h}{n_h}$$

♣ مجال الثقة

مجال الثقة للنسبة p_s بمستوى (1-α) هو:

$$p_s - z_{\left(1-\frac{\alpha}{2}\right)} \sqrt{\sum_{h=1}^{H} W_h^2 (1 - f_h) \frac{p_h q_h}{n_h}} < \mathrm{P} < p_s + z_{\left(1-\frac{\alpha}{2}\right)} \sqrt{\sum_{h=1}^{H} W_h^2 (1 - f_h) \frac{p_h q_h}{n_h}}$$

5. توزيع العينة حسب الطبقات

توجد ثلاثة مستويات لتوزيع العينات الطبقية وهي:

- التوزيع المتساوي.
- التوزيع المتناسب.
- التوزيع الأمثل.

5.1. التوزيع المتساوي

تستخدم هذه الطريقة في التوزيع عندما تكون فئات المجتمع متساوية وبالتالي يتم اعتماد نفس الحجم أو العدد من كل فئة كعينة تمثل هذا المجتمع. في هذه الحالة يكون توزيع العينة حسب الطبقات وفق الصيغة التالية:

$$n = \frac{n}{H}$$

5.2. التوزيع المتناسب

عندما نحدد نسبة معاينة $f_{h=}\ f$ متساوية في جميع الطبقات نتحدث عن معاينة طبقية متناسبة. للسحب الطبقي المتناسب خصائص هامة:

- احتمال السحب يكون متساويا بالنسبة لجميع عناصر قاعدة المعاينة و يماثل نسبة المعاينة الثابتة $f = \frac{\mathrm{n}}{\mathrm{N}}$.
- التوزيع المتناسب لا يرتبط بالمتغير $\mathrm{y_i}$ بل يتعلق فقط بحجم الطبقات.

$$n_h = n * W_h$$

5.3. التوزيع الأمثل

إذا كانت وحدات المجتمع تتسم أكثر بعدم التجانس من طبقة إلى أخرى وتكلفة البحث تتغير بشكل ملموس من طبقة الأخرى، نلجأ إلى التوزيع الأمثل. ويقتضي هذا النوع من التوزيع الرفع من نسبة المعاينة في الطبقات التي يكون فيها التباين مرتفعا وتقليل نسبة المعاينة في الطبقات التي تكون فيها التكلفة مرتفعة. هذا النوع من التوزيع يتطلب توفر معطيات دقيقة حول التباين في الطبقات وكلفة البحث.

$$C=\sum_h n_h\, C_h$$

في هذه الحالة التوزيع الأمثل للعينة على الطبقات يكتب.

$$n_h = n * \frac{\frac{N_h S_h}{\sqrt{C_h}}}{\frac{\sum_h N_h S_h}{\sqrt{C_h}}}$$

إذا كانت تكلفة البحث لا تتغير كثيرا من طبقة الى أخرى، $(C_{h=1})$ فإن هذا التوزيع يسمى توزيع Neyman. حيث نبحث عن h التي تقلص التباين داخل الطبقات. وتصبح الصيغة كالتالي:

$$n_h = n \frac{W_h S_h}{\sum_h^H W_h S_h}$$

أو:

$$\mathrm{n_h} = \frac{\mathrm{N_h S_h}}{\sum \mathrm{N_h S_h}}$$