



المعهد العربي للتدريب والبحوث الإحصائية

دورة متوسطة حول تصميم عينات المسوح الإحصائية

14

اعداد الاوزان في العينات

إعداد:

نايف عابد

nayif@pcbs.gov.ps nayifabed@yahoo.com

حزيران - 2022

مفهوم اوزان العينة

□ وزن وحدة المعاينة هو مقلوب احتمال اختيار وحدة المعاينة من إطار المعاينة. فإذا

كان احتمال الاختيار لوحدة المعاينة هو p_i فإن الوزن هو مقلوب الاحتمال $\frac{1}{p_i}$

□ يرتبط مفهوم الأوزان بالعينات الاحتمالية، ولذلك يجب حساب الأوزان بالاعتماد على

احتمالات الاختيار سواء كانت احتمالات متساوية او غير متساوية.

□ من الناحية التطبيقية فان وزن وحدة المعاينة يعبر عن عدد الوحدات التي تمثلها في

المجتمع بينما التعداد او الحصر الشامل تكون كل وحدة تعبر عن نفسها.

الهدف من الأوزان:

1. إيجاد تقدير المجموع للمجتمع، وذلك من خلال تقدير هارفتس تومسون

$$\begin{aligned}\hat{Y}_{HT} &= \sum_{i=1}^n \frac{y_i}{\pi_i} = \sum_{i \in S} \frac{y_i}{\pi_i} \\ &= \sum_{i=1}^n w_i y_i\end{aligned}$$

2. تقليل التحيز الناتج من عدم الاستجابة وزيادة الشمول

3. تمثيل العينة لمجتمع الدراسة، تقديرات العينة تعبر عن المجتمع

4. التعويض عن حالات نقص الشمول في الإطار

العينة الموزونة ذاتيا:

وهي العينة التي يتم اختيارها من مجتمع محدود بحيث كل وحدة لها نفس احتمال الاختيار وأمثلة عليها:

1- العينة العشوائية البسيطة أو المنتظمة: احتمال اختيار اي وحدة = حجم العينة/حجم

$$p_i = \frac{n}{N} \text{ المجتمع.}$$

فإذا كان مجتمع للدراسة فيه عدد الوحدات 100 ، وتم اختيار 10 وحدات، فإن احتمال

$$\text{الاختيار لكل وحدة متساوي ويساوي } 0.1 = \frac{10}{100} \text{ والوزن } = \frac{100}{10} = 10$$

2- العينة الطبقيّة ذات التوزيع المتناسب مع الحجم لكل طبقة:

نجد احتمال اختيار اي وحدة بقسمة حجم العينة في الطبقة على حجم الطبقة . $\frac{n_h}{N_h}$

ومثال على ذلك إذا كان هناك طبقتان أ، ب في كل منهما 200، 300 وتم اختيار

عينة طبقيّة من 100 وحدة بتوزيع متناسب مع الحجم فتكون عينة كل منهما 40، 60

ويكون احتمال الاختيار لكل وحدة متساوي.

احتمال اختيار الوحدة من أ $= 200/40 = 0.2$ والوزن $= 50$

احتمال اختيار الوحدة من ب $= 300/60 = 0.2$ والوزن $= 50$

3- العينة العنقودية ذات المرحلتين والتي فيها عينة المرحلة الأولى عناقيد يتم اختيارها بطريقة العينة الاحتمالية المتناسبة مع الحجم مع او بدون الارجاع، والمرحلة الثانية عينة عشوائية منتظمة/بسيطة حجمها ثابت من كل عنقود.

$$p_i = \frac{M_i}{\sum_i^N M_i} * \frac{m}{M_i} = \frac{m}{\sum_i^N M_i}$$

عينة ذات احتمال متناسب مع الحجم *pps* مع الارجاع

احتمال اختيار المرحلة
الاولى PSUs

احتمال اختيار المرحلة
الثانية SSUs

$$p_i = \frac{n * M_i}{\sum_i^N M_i} * \frac{m}{M_i} = \frac{nm}{\sum_i^N M_i}$$

عينة ذات احتمال متناسب مع الحجم *pps* بدون الارجاع

احتمال اختيار المرحلة
الاولى PSUs

احتمال اختيار المرحلة
الثانية SSUs

مراحل اعداد الاوزان

1- اعداد اوزان تصميم العينة Design Weight او Base Weight: من خلال ايجاد مقلوب احتمال الاختيار لكل وحدة معاينة، حيث تعكس الاوزان تصميم العينة، ونأخذ بالحسبان مراحل سحب العينة ونوعها. تعتبر اوزان التصميم اولية اذا كان هناك حالات عدم استجابة، وحالات زيادة شمول، او ان الاطار يعاني من نقص الشمول.

اذا كانت العينة من مرحلة واحدة فالوزن هو مقلوب احتمال الاختيار. $W_i = \frac{1}{p_i}$

اذا كانت العينة من عدة مراحل، وكانت عملية سحب العينة مستقلة في كل مرحلة،

فالوزن هو حاصل ضرب مقلوب الاحتمالات في كل المراحل. $W_i = \frac{1}{p_1} * \frac{1}{p_2} \dots \frac{1}{p_k}$

2- تعديل الأوزان الأولية بناء على عدم الاستجابة وزيادة الشمول:

□ يتم تعديل الأوزان بسبب عدم استجابة بعض الوحدات وبسبب زيادة الشمول في الإطار

□ المقصود بعدم الاستجابة هو عدم استجابة وحدة المعاينة بشكل كامل Unit Non-

response، وليس عدم استجابة جزئية (عدم استجابة البند Item Non-response).

□ يتم تعديل الأوزان بسبب عدم استجابة بعض الوحدات، للتعويض عن تجاوب هذه

الحالات، وذلك بزيادة وزن الوحدات المتجاوبة هي العينة.

□ تعديل الأوزان بسبب عدم الاستجابة يكون حلاً مقبولاً عندما يكون عدم الاستجابة

عشوائياً، بينما اذا كان عدم الاستجابة غير عشوائي فتعديل الاوزان لا يلغي التحيز

الناتج عن عدم عشوائية الاستجابة.

□ معالجة زيادة الشمول مختلفة عن معالجة عدم الاستجابة، بسبب اعتبار هذه الوحدات

لا تنتمي لمجتمع الهدف. وبالتالي يتم ازالة حالات زيادة الشمول من حجم العينة

وازالة اوزانهم من مجموع اوزان العينة.

□ لتعديل الأوزان حسب الاستجابة يتم حساب معامل تعديل للأوزان

□ الحالة الأولى: عندما تكون نسبة زيادة الشمول = 0

□ في البداية يتم تحديد فئة تعديل الأوزان (تكون فئات اجمالية، مثل: يتم تعديل الأوزان

حسب المناطق الجغرافية الرئيسية للأسر أو حسب الجنس ذكور وإناث للأفراد، أو

حسب النشاط الرئيسي للمنشآت.. الخ)

□ نجد معامل تعديل الأوزان كالتالي:

$$f_g = \frac{n_g}{r_g}$$

معامل تعديل الأوزان
للفئة g

عدد المفردات في العينة
من فئة التعديل g

عدد الوحدات المستجيبة
في فئة التعديل g

□ يتم ضرب وزن كل وحدة في العينة بمعامل تعديل الأوزان والذي هو عبارة عن

معكوس نسبة التجاوب.

□ الحالة الثانية: عندما تكون نسبة زيادة الشمول لا تساوي صفر، اي ان هناك حالات تم سحبها في العينة وبعد الوصول لها تبين انها لا تنتمي لمجتمع الهدف، مع أنها موجودة في الإطار.

□ يتم تحديد فئة تعديل الازان

□ نجد معامل تعديل الازان كالتالي:

$$f_g = \frac{n_g - n_{ocg}}{r_g}$$

عدد الوحدات في العينة
في فئة التعديل g

عدد حالات زيادة الشمول
في العينة من فئة التعديل g

معامل تعديل الازان
للفئة g

عدد الوحدات المستجيبة
في فئة التعديل g

□ يتم ضرب هذا المعامل في وزن الوحدات في العينة

3- تعديل الأوزان ليتوافق مع مصادر اخرى عن حجم مجتمع الهدف:

- يكون هذا التعديل حسب طبقات بعدية Post-strata
- عندما يتوفر بيانات عن مجموع وحدات مجتمع الهدف حسب مستويات معينة من مصادر اخرى تتصف بالدقة. مثل بيانات عن عدد الاسر والافراد من سجل السكان، او من تقديرات سكانية تم حسابها بشكل علمي دقيق.
- يتم مقارنة مجموع الأوزان للوحدات بعد التعديل لعدم الاستجابة وزيادة الشمول مع مجموع الوحدات من المصادر الاخرى التي تتصف بالدقة حسب فئات/مستويات معينة.
- في العادة يكون هناك اختلافات وعدم تطابق بين مجموع الاوزان من العينة وبين عدد الوحدات من المصادر الاخرى.

□ يتم ايجاد معامل تعديل للاوزان ليتطابق مجموعها مع اعداد المجتمع من المصادر الاخرى

$$f_{g_post-str} = \frac{N_g}{\sum_{i \in g} w_i}$$

معامل تعديل الازان
للطبقة البعدية g

مجموع الازان في
الطبقة البعدية g

عدد وحدات المجتمع من
مصدر اخر في الطبقة
البعدية g

□ يتم ضرب هذا المعامل في وزن الوحدات في العينة

4- حساب الوزن النسبي Relative Weight

□ يسمى احيانا اوزان معيارية standardized weights او Normalized weight

□ يتم حساب الوزن النسبي من خلال قسمة الوزن النهائي المعدل لكل وحدة معاينة على

متوسط الاوزان

□ يستخدم كبديل للوزن ، لكنه لا يصلح لتقدير مجموع المجتمع

$$w_{i_rw} = \frac{w_i}{\bar{w}}$$

الوزن النسبي للوحدة i

وزن وحدة المعاينة i

متوسط الاوزان

$$\bar{w} = \frac{\sum_{i=1}^{n_r} w_i}{n_r}$$

عدد الوحدات المستجيبة في العينة

ملاحظات حول حساب الازان:

- 1- مجموع الازان المعدلة النهائية يساوي حجم المجتمع
- 2- مجموع الازان النسبية يساوي حجم العينة ومتوسط الازان النسبية = 1
- 3- اذا كانت الازان متساوية فالتصميم للعينة موزون ذاتياً
- 4- كلما كانت الازان متقاربة في القيمة فهذا يعني ان التصميم قريب من التصميم الموزن ذاتياً وهذا يقلل من تباين التقدير
- 5- في حالة العينات المثبتة panel survey ، يتم اضافة معامل لتعديل الازان حسب تناقص العينة sample attrition
- 6- عادةً يتم حسب معامل تناقص العينة باستخدام الانحدار اللوجستي logistic regression

تمرین عملی