

Previous session:

Sampling in R

Using
Sampling Weights
In
R

أوزان التصميم Design Weights

- أوزان التصميم: الغرض الرئيسي من أوزان التصميم هو تصحيح حقيقة أنه في بعض الحالات ، يكون لدى المستجيبين احتمالات مختلفة ليكونوا جزءًا من العينة بسبب تصميم العينة المستخدم.
 - تصميم مقصود
 - تصميم ضعيف
- تُستخدم هذه الأوزان في مرحلة التحليل لتصحيح التحيز في الاختيار ، أي الخطأ الناتج عن الاحتمالات المختلفة بأن الأفراد من المجتمع المستهدف ليس لديهم نفس الاحتمالات ليتم تضمينهم في العينة.

Example: Employee Data

- لدينا مجتمع (مؤسسة) مكونة من 474 موظفا موزعين كما يلي:
- 54.4% ذكور و 45.6% إناث

Gender

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Female	216	45.6	45.6	45.6
Male	258	54.4	54.4	100.0
Total	474	100.0	100.0	

Example: Employee Data-Continue

- لدينا مجتمع (مؤسسة) مكونة من 474 موظفا موزعين كما يلي:
- 21.9% من الأقليات و 78.1% من غير الأقليات

Minority Classification

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid No	370	78.1	78.1	78.1
Yes	104	21.9	21.9	100.0
Total	474	100.0	100.0	

Example: Employee Data-Continue

- من خلال تحليل الانحدار لدراسة العلاقة الخطية بين سنوات التعليم والراتب (المعاش) السنوي، وجدنا أن:
- كل سنة تعليم إضافية تزيد الراتب السنوي للموظف حوالي 3909 دولار

Coefficients^a

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	
	B	Std. Error	Beta			
1	(Constant)	-18331.178-	2821.912		-6.496-	.000
	Educational Level (years)	3909.907	204.547	.661	19.115	.000

a. Dependent Variable: Current Salary

Example: Employee Data-Continue Selecting a stratified Sample

- نود سحب عينة من 100 موظف، بحيث تتكون من 50 موظف من طبقة الأقليات و 50 موظف من طبقة "غير الأقليات". لاحظ أن العينة غير ممثلة للمجتمع في متغير "الأقليات".
- حيث أن نسبة الأقليات في المجتمع = $104/474 = 0.22$ أي حوالي 22 %، في حين أن:
نسبة الأقليات في العينة = $50/100 = 0.5$ أي 50%

الوزن	نسبة العينة من الطبقة (احتمال الاختيار)	عدد الموظفين في العينة	عدد الموظفين في المجتمع	الطبقة
7.40	$50/370 = 0.135$	50	370	غير الأقليات
2.08	$50/104 = 0.481$	50	104	الأقليات
		100	474	الإجمالي

Effect of Weighting : Population parameters, Unweighted estimates and weighted estimates

- الجدول التالي يبين أثر التوزين على تقليل التحيز في التقديرات الناتج عن التحيز في العينة بسبب عدم إعطاء فرص متساوية لأفراد المجتمع

التقدير الموزون من بيانات العينة	التقدير غير الموزون من بيانات العينة	القيمة الحقيقية في المجتمع	التقدير
33,881	32,867	34,419	متوسط الراتب السنوي
% 52.4	58% .0	% 54.4	نسبة الذكور إلى المجموع
% 21.9	% 50	% 21.9	نسبة الأقليات إلى المجموع

Effect of Weighting : Regression Coefficients

أثر التوزين : معاملات معادلة الانحدار

- الجدول التالي يبين أثر التوزين على تقليل التحيز في تقديرات معاملات معادلة الانحدار، الناتج عن التحيز في العينة بسبب عدم إعطاء فرص متساوية لأفراد المجتمع

التقدير الموزون	التقدير غير الموزون	القيمة الحقيقية في المجتمع	التقدير
4353	3977	3929	معامل تأثير سنوات الدراسة على الراتب السنوي

النتائج

- التقديرات المرجحة (الموزونة) أقرب بكثير إلى قيم المجتمع (القيمة الحقيقية) من التقديرات غير المرجحة
- ومع ذلك ، قد لا تكون التقديرات الموزونة متطابقة مع قيم المجتمع (القيمة الحقيقية)،
- لذلك لجعل التقديرات الموزونة متطابقة قيم المجتمع (القيمة الحقيقية)، يلزم معايرة الأوزان

Result

- Weighted Estimates are much closer to the population (TRUE) values than Un-Weighted estimates
- However, Weighted estimates may not be identical to population (TRUE) values
- So to make the weighted estimates identical to the population (TRUE) values, weight calibration is needed (See SPSS-Example: Employee data)

Using weights in statistical functions

- Some R functions take a "*weights*=" argument, e.g.,
`lm(salary~educ, weights = dweights, data = empsam)`
- Shows that each year of education increases average yearly salary by 4005\$
- Without weights, coefficient would be 3155\$, because association differs by minority classification.

استخدام الأوزان في دوال برنامج R

- تأخذ بعض دوال R الوسيطة " " `weights =` ، على سبيل المثال ، عند تحليل العلاقة الخطية بين متغير "سنوات التعليم" ومتغير "الراتب الحالي" باستخدام نموذج تحليل الإنحدار الخطي `Linear Regression` نستخدم في R ما يلي:

```
lm(salary~educ, weights = wt_st, data = sample_st)
```

باستخدام بيانات العينة، الموزونة، سيكون المعامل 4354 دولارًا

- باستخدام بيانات إطار المعاينة (التعداد) يظهر أن كل سنة دراسية تزيد متوسط الراتب السنوي بمقدار 3929 دولار

```
lm(salary~educ, data = employee)
```

- باستخدام بيانات العينة، بدون أوزان ، سيكون المعامل 3977 دولارًا ، لأن الارتباط بين الراتب والتعليم يختلف باختلاف تصنيف "الأقلية".

```
lm(salary~educ, data = sample_st)
```

Using Survey Packages

- Other R functions, such as basic descriptive statistics, do not take weights
- Needs complex coding or additional packages:
 - 1. "survey" package to offer specific functions
 - 2. "srvyr" package" to integrate with "dplyr"
- They offer tools to deal with complex survey designs

Basic Logic of srvyr workflow

- **Create a survey object**

- `empsam_wt <- as_survey(empsam, weights = dweights)`
- Contains data + design information

- **Manipulate data**

- `empsam_wt %>% gender = "f") %>% summarise(survey_mean(salary))`

- **Calculate weighted statistics**

- `empsam_wt %>% summarise(survey_mean(salary, na.rm=T))`

What to remember?

- Survey **samples are often not balanced**, either on purpose or because of non-participation
- **Weighting needed** to deal with that, most large social surveys provide weights with data
- Some R functions take a `weight =` argument, for other statistical calculations, **srvyr** or **survey** packages are needed

Next session

Weights Calibration (weights Racking)