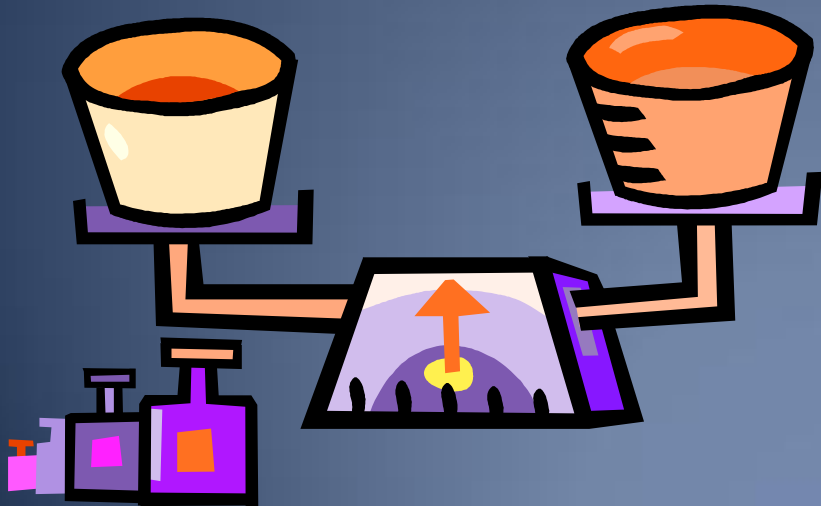


# Creating Weights to Improve Survey Population Estimates



**Pu-Shih Daniel Chen**

**Shimon Sarraf**

Indiana University  
Center for Postsecondary Research

# Case Study: NSSEville State University

- Student Population: 2000
- Survey Respondents: 1000

	Population		Respondents	
	Count	%	Count	%
FT/Male	1000	50%	200	20%
FT/Female	750	38%	600	60%
PT/Male	150	8%	125	13%
PT/Female	100	5%	75	8%
	2000		1000	

# Goals

- Introduce basic weighting concepts (simple random sample design)
  - Purpose
  - Types
  - Calculations
  - Limitations
- Enhance confidence to do it yourself
- Develop a more critical eye

# Purpose of Weights

- Corrects for one type of potential survey bias while estimating populations :

**Respondents do not represent  
population well**

- Weights cannot correct all types of survey error (questionnaire design, data collection, sampling, nonresponse effect)

# Purpose of Weights (cont.)

- Lack of representation comes from...
  - **Differential response rates**
    - 80% females **respond** but only 50% males (< 100% RR)

AND/OR

- **Unequal probability of selection**
  - 80% of females **selected** but only 50% of males (100% RR)

# Purpose of Weights (cont.)

- Using weights assumes:
  - Survey item results vary by subgroups
  - Negligible nonresponse effect
  - Aggregate analysis

# Is Weighting Necessary?

Check representation and identify sub-groups

	Population		Respondents
Male	58%	→	33%
Full-time	88%	→	80%

Determine how results vary by sub-groups

**Community Service Participation**

Gender		Enrollment Status	
Male	20%	Full-Time	80%
Female	60%	Part-Time	40%

# Types of Weights

## Scale & Proportional

- Representative sample created
- **Population** count

## Proportional

- Representative sample created
- **Respondent** count

- Both types of weights produce nearly identical item distributions, averages, and st. deviations
- 50% say “very often” using both weights



# Scale Weight Calculation

$$W_i = P_i / R_i$$

## Population Counts

	Full-time	Part-time
Male	P1	P3
Female	P2	P4

## Respondents Counts

	Full-time	Part-time
Male	R1	R3
Female	R2	R4

# Scale Weight Example

$$\mathbf{W}(\text{male/full-time}) = 1000/200 = 5$$

$$\mathbf{W}(\text{male/part-time}) = 150/125 = 1.2$$

$$\mathbf{W}(\text{female/full-time}) = 750/600 = 1.25$$

$$\mathbf{W}(\text{female/part-time}) = 100/75 = 1.33$$

## Population Counts

	Full-time	Part-time
Male	1000	150
	P1	P3
	P2	P4
Female	750	100

## Respondents Counts

	Full-time	Part-time
Male	200	125
	R1	R3
	R2	R4
Female	600	75

# Effects of Weighting

- When creating weights, we are trying to identify how many cases each respondent should represent
- Scale weights will make the number of respondents equal to the number of population

## Population Counts

	Full-time	Part-time
Male	1000	150
Female	750	100

## Respondents Counts

	Full-time	Part-time
Male	200	125
Female	600	75

# Adjusting Scale Weights

- The pros and cons of scale weights
- Scale Weights can be transformed into proportional weights so that the number of respondents is preserved

Degree of Freedom	Mean differences needed to achieve statistics significance at $p = .05$
1	12.71
10	2.23
100	1.98
$\infty$	1.96

# Proportional Weight Calculation

$$W_p = \frac{\text{Percent of Population}}{\text{Percent of Respondents}} = \frac{P_i / P_{total}}{R_i / R_{total}} = \frac{0.5}{0.2} = 2.5$$

## Population Counts

	Full-time	Part-time
Male	<b>1000</b> 50%	150 7.5%
Female	750 37.5%	100 5%

## Respondents Counts

	Full-time	Part-time
Male	<b>200</b> 20%	125 12.5%
Female	600 60%	75 7.5%

# Scale vs. Proportional Weights

Respondents	Unweighted Counts	Scale		Proportional	
		Weight	Weighted Counts	Weight	Weighted Counts
Male/FT	200	5	1000	2.50	500
Male/PT	125	1.2	150	0.60	75
Female/FT	600	1.25	750	0.63	375
Female/PT	75	1.33	100	0.67	50
<b>TOTAL</b>	<b>1000</b>		<b>2000</b>		<b>1000</b>

# Good & Bad News

- Bad News
  - SPSS cannot create weights for you
  - You create weighting variables on your own
- Good News
  - SPSS can determine cell counts
  - Just apply the formula provided to create your weight
  - **SPSS can apply weights for all your analyses**

# Determine Cell Counts with SPSS

## CROSSTABS

```

/TABLES=gender BY enrollment
/FORMAT= AVALUE TABLES
/CELLS= COUNT
/COUNT ROUND CELL .
    
```

gender \* enrollment Crosstabulation

Count		enrollment		
		Part-time	Full-time	Total
gender	Male	34	320	354
	Female	68	578	646
Total		102	898	1000

## Population Counts

	Full-time	Part-time
Male	1000	150
Female	750	100

## Respondents Counts

	Full-time	Part-time
Male		
Female		



# Weight Calculations Revisited

- Scale & Proportional Weight (NSSE Weight<sub>2</sub>)

$$W_2 = \frac{P_i}{R_i}$$

## Population Counts

	Full-time	Part-time
Male	P1	P3
Female	P2	P4

- ◆ Proportional Weight (NSSE Weight<sub>1</sub>)

$$W_1 = \frac{\frac{P_i}{P_{total}}}{\frac{R_i}{R_{total}}}$$

## Respondents Counts

	Full-time	Part-time
Male	R1	R3
Female	R2	R4

# Attaching Weights to the Data File

```
*****Attach Weights to the data file*****.  
IF (gender = 1 and enrollment = 1) WEIGHT = 4.41.  
IF (gender = 1 and enrollment = 2) WEIGHT = 3.13.  
IF (gender = 2 and enrollment = 1) WEIGHT = 1.47.  
IF (gender = 2 and enrollment = 2) WEIGHT = 1.3.  
  
FREQ WEIGHT  
  /statistics=sum.
```

# Attaching Weights to the Data File

The screenshot shows the SPSS Data Editor window with the 'Compute Variable' dialog box open. The 'Target Variable' is 'WEIGHT' and the 'Numeric Expression' is '4.15'. The 'If...' button is circled in red. The 'Compute Variable' dialog box also shows a list of variables on the left, including 'SurveyID', 'DST0601A', 'DST0601B', 'DST0601C', 'DST0601D', 'DST0602A', 'DST0602B', 'DST0603A', 'DST0603B', 'DST0603C', 'DST0603D', 'DST0603E', 'DST0603F', 'DST0603G', 'DST0603H', and 'IPEDS'. The 'Function group' is set to 'All'.

	SurveyID	DST0603
1	100296	
2	100301	
3	100302	
4	100304	
5	100313	
6	100314	
7	100316	
8	100375	
9	100384	4

# Check the Weights

## Statistics

WEIGHT1		
N	Valid	1000
	Missing	0
Sum		1000.36

## WEIGHT1

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.65	578	57.8	57.8	57.8
	.74	68	6.8	6.8	64.6
	1.56	320	32.0	32.0	96.6
	2.21	34	3.4	3.4	100.0
Total		1000	100.0	100.0	

# Applying Weights in SPSS

```
*****Apply Weights to Analyses*****.
```

```
Weight by WEIGHT1.
```

```
* Custom Tables.
```

```
CTABLES
```

```
  /VLABELS VARIABLES=enrollment ACa ACL SFI EEE SCE DISPLAY=DEFAULT
```

```
  /TABLE ACa [MEAN] + ACL [MEAN] + SFI [MEAN] + EEE [MEAN] + SCE [MEAN] BY  
  enrollment|
```

```
  /CATEGORIES VARIABLES=enrollment ORDER=A KEY=VALUE EMPTY=INCLUDE.
```

```
Weight off.
```

# Applying Weights in SPSS

The screenshot shows the SPSS Data Editor interface. The 'Data' menu is open, and 'Weight Cases...' is highlighted. The background shows a list of variables and a table of data.

Name	Label	Values
1 var1	Asked questio	{1, Never }...
2 var2	Coursework e	{1, Very little }.
3 var3	Participate in a	{1, Have not de
4 var4	Quality: Your r	{1, Unfriendly,
5 var5	Hours per 7-da	{1, 0}...
6 var6	Institutional e	{1, Very little }.
7 var7	Institutional co	{1, Very little }.
8 ACa	Academic Cha	None
9 ACL	Active and Coll	None
10 SFI		
11 EEE		
12 SCE		
13 gender		
14 enrollment		

The 'Weight Cases' dialog box is shown. It has two radio buttons: 'Do not weight cases' (unselected) and 'Weight cases by' (selected). Below the 'Weight cases by' option is a 'Frequency Variable' field with a list of variables: EEE, SCE, gender, enrollment, WEIGHT2, and WEIGHT1. The 'Current Status' is 'Do not weight cases'.

Weight Cases

Do not weight cases

Weight cases by

Frequency Variable:

Current Status: Do not weight cases

# Weighting Limitations

- Scale weights (count = population):
  - smaller standard errors → more significant results
- If nonresponders very different: inaccurate results
- Not appropriate for all survey items
- Precision may be lost in some situations
- Small cell sizes

# To Learn More about Weighting

## Basic Overview:

- Maletta, H. (2006). *Weighting*. Retrieved from Raynald Levesque's SPSS website: <http://www.spsstools.net/Tutorials/WEIGHTING.pdf>

## Advanced weighting techniques and implications:

- Thomas, S., Heck, R., & Bauer, K. (2005) Weighting and Adjusting for Design Effects in Secondary Data Analyses. In P. Umbach (Ed.), *Survey Research: Emerging Issues*. New Directions for Institutional Research, Number 127, 51-72.
- Dey, E. L. (1997). Working with low survey response rates: The efficacy of weighting adjustments. *Research in Higher Education*, 38, 215-227.



# Questions & Discussion

Contact: **Pu-Shih Daniel Chen, Ph.D.**

[pdchen@indiana.edu](mailto:pdchen@indiana.edu)

**Shimon Sarraf**

[ssarraf@indiana.edu](mailto:ssarraf@indiana.edu)

Presentation materials:

<http://nsse.iub.edu/conferences>