

Post-stratification Weights

When conducting a survey, having a representative sample of the population is of paramount importance. But sometimes, despite your best efforts, you may accidentally (or sometimes intentionally) *oversample* some kinds of people and *undersample* others. In other words, the way a certain characteristic (such as age, education, race, sex, etc.) of your sample is distributed may differ from the way it is distributed in the population. For example, your sample may consist of 60 percent women, when women make up only 52 percent of the population. This introduces bias into any estimate you may obtain from your sample data because statistical procedures will give greater weight to those people you oversampled.. You can correct for these biases mathematically with a post-stratification survey weight.

In order to calculate a post-stratification weight, you need an auxiliary data set to which you can compare your sample data. For example, if you were conducting a survey of residents of Illinois, you would need census data or Current Population Survey data that show the demographic characteristics of the population of Illinois. You could then compare your sample to the auxiliary data file, to make sure that the distribution of demographic characteristics (such as age, education, race, sex, etc.) is similar to that of the auxiliary data. If the distributions are close enough, there is no need to calculate post-stratification weights. However, if they differ by more than a few percentage points, you will want to calculate the weights.

It is called a *post*-stratification weight because you can only compute it *after* you have collected all your data. The *stratification* part comes from the fact that you use various known *strata* (such as age group or sex distribution) of the population to adjust your sample data to conform more to the population's parameters.

Here is an example of how to compute a post-stratification weight. You may use a spreadsheet or a calculator to compute yours:

	Col. A	Col. B	Col. C	Col. D	
Age Group & Sex	Population N	Proportion of Population <u>Col. A</u> 8000	Sample n	Proportion of Sample <u>Col. C</u> 200	Weight <u>Col. B</u> <u>Col. D</u>
Age 18-34: Male	2,000	.25	30	.15	1.67
Age 18-34: Female	1,760	.22	40	.20	1.10
Age 35-64: Male	1,440	.18	50	.25	.72
Age 35-64: Female	1,600	.20	60	.30	.67
Age 65+: Male	400	.05	10	.05	1.00
Age 65+: Female	800	.10	10	.05	2.00
Total	8,000		200		

Creating a post-stratification weight in SPSS:

One way to create these weights in your data set, is simply to enter them by hand in one of the columns while you do data entry. Another way is to use SPSS commands in a syntax window after you have finished your data entry. Here are commands to construct a weight variable using the example above:

- Go to **File**→ **New**→ **Syntax**
- In the blank SPSS Syntax Editor that pops up, type:


```
if (AGEGROUP = 1 & SEX = 1) WEIGHT = 1.67.
if (AGEGROUP = 1 & SEX = 2) WEIGHT = 1.10.
if (AGEGROUP = 2 & SEX = 1) WEIGHT = .72.
if (AGEGROUP = 2 & SEX = 2) WEIGHT = .67.
if (AGEGROUP = 3 & SEX = 1) WEIGHT = 1.00.
if (AGEGROUP = 3 & SEX = 2) WEIGHT = 2.00.
execute.
```

(This assumes that AGEGROUP (coded 1, 2 & 3) and SEX (coded 1 & 2) are variables in your dataset, and WEIGHT is the name of the new weight variable you are creating.)

- Then choose **Run**→ **All**

If you are successful, you will see a new column on the far right side of your data which contains your weight variable.

Applying the post-stratification weight in SPSS:

Once you have created the post-stratification weight as a variable, you must tell SPSS to use it in your calculations. To do so:

- In the SPSS Data Editor window menus, go to **Data** → **Weight Cases...**
- Click in the circle next to **Weight Cases by**.
- Put the name of your weight variable in the **Frequency Variable:** box.
- Click the **Okay** button.

Once you have successfully applied the steps above, you should see the notice "Weight On" in the lower right hand corner of the SPSS Data Editor window. Always check your weights by running frequencies on the weighted data and comparing them to the auxiliary data. If you have calculated the weights correctly, the weighted sample data should have a very similar distributions (percentage-wise) of your strata variables.