

# Weighting Survey Data

2nd edition



---

## Table of Contents

---

Overview and background .....	4
Commercial versus government and academic studies.....	5
Software.....	5
Weights, adjustment variables, and targets .....	6
Why and when to weight your survey data.....	7
Worked example.....	8
Adjustment variables and targets .....	10
Categorical adjustment variables.....	10
Targets involving multiple adjustment variables .....	11
Numeric adjustment variables.....	11
Avoid using variables with a high level of measurement error.....	11
Using market share as a numeric adjustment variable .....	12
Using dates (waves) in composite variables.....	12
Selecting adjustment variables.....	13
Inconsistent categories between targets and adjustment variables.....	14
The weight variable(s) .....	14
Creating sampling weights.....	16
Post-stratification weights .....	17
Design weights .....	17
Propensity weights (propensity score adjustment).....	18
Cell weights .....	19
Rim weights/Raking .....	19
Calibration weights .....	20
Calibrated weight.....	20
Volumetric weights .....	21
Weights for different purposes .....	21
Combining weights .....	22
More exotic type of weights.....	22
Frequency weights .....	22
Variance weights .....	23
Replicate weights .....	23
Replication weights .....	23
Expansion weights .....	23
Analytic weights .....	23
Importance weights .....	23
Self-weighting samples .....	24
Common mistakes .....	25
No sample for a target group.....	26
Targets for categorical adjustment variables that don't add up.....	26
Impossible numeric targets.....	26
Inconsistent categorical targets.....	27
Inconsistent numeric targets .....	27
Ignoring measurement error .....	28
Using weight factors instead of targets .....	28

Checking weights .....	29
Case study .....	30
Weighted tables of adjustment variables .....	31
Weighted table of key variables .....	31
The average of the weight variable .....	32
The range of the weight variable .....	32
The distribution of the weight variable .....	33
Effective sample size .....	34
Improving a weight .....	36
The bias versus error tradeoff .....	37
Changing the targets .....	37
Trimming (bounds on weights) .....	38
Analysis of weighted data .....	40
Bias and statistical inference .....	41
Approaches to incorporating weights in statistical analyses .....	41
Variance estimation .....	41
Resampling .....	41
Weight calibration .....	42
Using the observed sample size .....	42
Modifying the weight to have an average weight of 1 .....	42
Use the weight as is .....	42
Does choice of approach make a big difference? Yes .....	43
Approaches available by stat package .....	43
Q and Displayr .....	43
SPSS Statistics .....	44
R .....	44
When weighting does and does not work .....	44
Weighting in Displayr .....	46
Applying a weight to an analysis .....	47
Weighted versus unweighted data .....	47
Effective sample size .....	48
Using weights created in other software .....	48
Creating weights .....	48
Additional options .....	50
Weighting in Q .....	51
Applying a weight to an analysis .....	52
Weighted versus unweighted data .....	52
Effective sample size .....	52
Using weights created in other software .....	52
Cell weights .....	53
Rim weights .....	54
More complex weights .....	55
Weighting in R .....	56
Applying a weight to an analysis .....	57
Creating weights .....	58

# Overview and background

Weighting is a technique that adjusts some survey results to bring them into line with some known characteristics of the population. For example, if a sample contains 40% males and the population includes 49% males, weighting can adjust for this discrepancy.

This book describes all the key steps involved in weighting survey data (also known as *sample balancing*, *calibration*, *post-survey adjustment*, *raking*, *post-stratification*, and *non-response weighting*).

---

## Commercial versus government and academic studies

---

Commercial weighting tends to be different from that used in academic studies, which is, in turn, different from that used in official statistics and other government statistics. This book focuses on the commercial applications of weighting.<sup>1</sup>

A key aspect of commercial weighting is that one or a small number of weight variables are typically created for use in many subsequent analyses. Where there is only a single analysis of primary interest, such as a single regression model or an opinion poll, alternative approaches may be preferable, including:

- Performing a regression with the adjustment variables that would have been used to create the weights (see the next chapter) included as predictors.
- Using MrP (“Mister P”). See: <http://www.misterp.org/papers.html>.

---

## Software

---

This eBook provides instructions for creating weights in Q, Displayr, and R. For reasons discussed in the section on **SPSS Statistics**, the use of SPSS for analyzing surveys is only advisable if either:

- Performing everything in R using the R integration.
- Using the Complex Samples module.

---

<sup>1</sup> Government studies tend to involve multi-state samples with attempts at stratification and differential correction for design, coverage, and non-response factors. For more information about weighting government surveys, see G Kalton and I F-C (2003), “Weighting Methods”, *Journal of Official Statistics*, 19(2), 81-97. Academic studies tend to either not weight data at all, or, follow the approach of government studies.

# Weights, adjustment variables, and targets

This chapter:

- Explains the situations where it is appropriate weight.
- Provides a simple example of weighting
- Defines the various types of adjustment variables used to create weights
- Describes how to select adjustment variables.

---

## Why and when to weight your survey data

---

We should weight data when:

1. There are discrepancies between survey results and facts. For example:
  - Fact: 12% of US adults live in California
  - Result: 25% of respondents live in California
2. The discrepancy is believed to be caused by the survey having interviewed too few or too many people in one or more groups in the population. For example, if we have more people saying they are Californian because we interviewed too many. The main alternative to having interviewed too few or too many people is measurement error (e.g., people living in Texas but saying they live in California for some reason).
3. If the discrepancies are not addressed, key results of the study will be wrong. For example, suppose our study is focused on understanding attitudes to the environment, and we both have too many Californians and know that Californians differ from the average Americans in terms of their attitudes to the environment. In that case, we know we will have a problem unless we correct it via weighting.

Ideally, we should have all three of these conditions in place if weighting data. However, it is commonplace for people to weight when only the first two are in place, as:

- It can take time and skill to verify the third conclusion.
- If the third condition is not in place, the only consequence of weighting unnecessarily is that fewer results will be statistically significant than otherwise.
- Having weighted data is often seen by less technical clients as being a sign of hygiene.

---

## Understanding weighted calculations

---

Consider a tiny survey of five people, where we have asked them which brand of cola they last purchased. The first two said Coke. The following three said Pepsi. We can compute that 40% of people said Coke. This is an example of an *unweighted calculation*.

Now, consider the situation where we have also asked them how many they purchased, and the five respondents said 10, 5, 1, 2, and 2, respectively, as shown below. The total number of purchases is

20. If we sum up the Coke purchases, we get 15, which is 75%. This is an example of a *weighted calculation*, where:

- The number of purchases is the *weight variable*, and it contains in this example five *weights* (one for each person).
- If we were to replace the weight variable with five 1s instead, we would get the 40% obtained in the unweighted calculation. That is, unweighted calculations are just weighted calculations, where everybody has a weight of 1.

Last brand purchased	Number purchased	
Coke	10	← <b>Weight variable</b> , which contains five <b>weights</b> , one for each respondent
Coke	5	
Pepsi	1	
Pepsi	2	
Pepsi	2	

Weighted counts	%
Coke 15	75
Pepsi 5	25
Total 20	100

This example illustrates *volumetric weighting*. Our focus in this eBook is on *sampling weights*, which correct for over- and under-representation of critical groups in a study. With volumetric weights, we know the weights as they come from questions in the study. With sampling weights, we have to create weights such that when we perform weighted calculations, the survey results align with known facts about the market.

---

## Worked example

---

Consider the following data showing 10 people's favorite celebrity:

Brad Pitt  
 Brad Pitt  
 Brad Pitt  
 Brad Pitt  
 Brad Pitt  
 Brad Pitt  
 Brad Pitt



Brad Pitt  
Tiger Woods  
Tiger Woods

In this sample of ten, 80% of people have nominated Brad Pitt as their favorite celebrity. When we conduct surveys, we want to be able to draw conclusions about the entire population, rather than just our sample.

So instead of saying “80% of the sample nominate Brad Pitt as their favorite celebrity” we should aim to be conducting research that allows us to confidently state “Brad Pitt is the most popular celebrity with 80% of people nominating him as their favorite.”

Now consider the impact of some additional information about the gender of our ten respondents:

Brad Pitt	Female
Brad Pitt	Female
Brad Pitt	Female
Brad Pitt	Female
Brad Pitt	Female
Brad Pitt	Female
Brad Pitt	Female
Brad Pitt	Female
Tiger Woods	Male
Tiger Woods	Male

From this data we can see that the sample is unrepresentative in terms of gender, with eight of ten respondents being female (80%) and two being male (20%). A big difference when we consider that true representation in the world is about 50/50.

Furthermore, gender seems to be the sole determinant of preference. As the sample is not representative in terms of gender, and gender is correlated with our measure of favorite celebrity, it seems that any estimate of people’s favorite celebrity will only be valid if we consider the over-representation of women in the sample.

We can improve this estimate by weighting. A weight is computed for every respondent in a sample by dividing the correct proportion by the observed proportion. The correct proportion of males in our population is 50% and the observed proportion is 20%, so the weight for each male is  $50\%/20\%=2.5$  and the weight for each female is  $50\%/80\%=0.625$ . Thus, our data becomes:

<b>Favorite Celebrity</b>	<b>Weight</b>
Brad Pitt	0.625
Brad Pitt	0.625
Brad Pitt	0.625
Brad Pitt	0.625
Brad Pitt	0.625

Brad Pitt	0.625
Brad Pitt	0.625
Brad Pitt	0.625
Tiger Woods	2.5
Tiger Woods	2.5

Now we can compute our estimate of the proportion of people to list Brad Pitt as their favorite celebrity by summing up the weights of each of the respondents to prefer Brad Pitt and dividing this by the sum of all respondents' weights:

$$\frac{8 \times 0.625}{2 \times 2.5 + 8 \times 0.625} = \frac{5}{10} = 50\%$$

The approach described here for computing a weight is a relatively simple case, but the basic idea can be extended to deal with much more complicated cases.

Additional worked examples are available in our [recorded webinar on weighting](#).

---

## Adjustment variables and targets

---

When we weight data, we create a weight such that our weighted data matches some known facts about the world. These known facts are called *targets*. Examples of targets include:

- 12% of people living in California
- 12% of adults are aged 18 to 24
- 49% of adults are male
- 6% of adults who are aged 18 to 24 and are male
- 0.72% of adults who are aged 18 to 24 and are male and are living in California
- Coca-Cola's market share is 44%

### Categorical adjustment variables

In the earlier example about Brad Pitt and Tiger woods, gender was the *adjustment variable* and the adjustment targets were Male 50% and Female 50%. These targets are proportions. Targets can also be provided as population totals (e.g., Male 159,000,000 and Female 172,000,000).

As the variable contains categories, gender was a *categorical adjustment variable*.

## Targets involving multiple adjustment variables

The Excel screenshot below shows targets expressed in terms of age by gender by region. Note that these are expressed as percentages that add up to 100. It is possible in many programs to weight to actual population totals (e.g., the number of males aged 18 to 29 in DC, etc.). However, it tends to be safer to first weight to proportions and then, if necessary, multiply this weight by sample size. This is because when you enter population totals it becomes a lot harder to get an intuitive feel for the numbers, as they become too big to easily interpret.

	A	B	C	D	E	F	G
1			Northeast	Midwest	South	West	DC
2	18-29	Male	1.743	2.38	3.659	2.182	0.023
3		Female	1.793	2.379	3.718	2.13	0.029
4	30-44	Male	2.867	3.573	5.252	3.135	0.029
5		Female	3.035	3.644	5.517	3.169	0.031
6	45-64	Male	2.934	3.634	5.34	3.154	0.027
7		Female	3.157	3.761	5.775	3.271	0.034
8	65+	Male	1.552	1.865	2.746	1.542	0.014
9		Female	2.272	2.652	3.934	2.023	0.025

## Numeric adjustment variables

The examples so far have used *categorical adjustment variables*, where each adjustment variable consists of a set of mutually exclusive and exhaustive categories. It is also possible to use *numeric adjustment variables*, such as number of people in the household or number of products consumed.

When using numeric adjustment variables, the targets are the averages of these variables in the population.<sup>2</sup>

## Avoid using variables with a high level of measurement error

Consider the situation where a survey finds that the average household income is \$92,000, but the Census says the correct figure is \$62,000. It is possible that the survey over-represents high earners. But, it is more likely that there is a lot of measurement error, and people have told the government a different income than they indicates in the survey.

Weighting cannot fix measurement errors. Adjustment variables need to be variables with low levels or measurement error.

---

<sup>2</sup> Some software assumes that the targets are totals (i.e., the averages multiplied by the sample sizes).

## Using market share as a numeric adjustment variable

In order to use market share as a numeric adjustment variable it is necessary to create numeric variables that estimate the market share for each respondent. It is commonplace for people to compute this incorrectly, so a bit more description is provided here.

ID	Purchases		Incorrect calculations		Correct calculations	
	Coke	Pepsi	Coke	Pepsi	Coke	Pepsi
1	10	0	1	0	1.36363636	0
2	5	5	0.5	0.5	0.68181818	0.68181818
3	0	2	0	1	0	0.27272727
Average	5.0000000	2.3333333	0.5000000	0.5000000	0.6818182	0.3181818
Share	0.6818182	0.3181818	0.5000000	0.5000000	0.6818182	0.3181818

Consider the example above, where we have three respondents who have rated their frequency of consuming two brands, Coke and Pepsi. With this data, the market shares of the two brands are 68.2% and 31.8% respectively (e.g., for Coke it is  $(10 + 5 + 0) / (10 + 5 + 0 + 0 + 5 + 2)$ ).

A common mistake is to create adjustment variables by dividing one or more variables by the sum of the variables (at the respondent level). This is shown in the middle of the table. Why is it wrong? It fails to consider that respondent 3 consumes only one-fifth of the amount of colas as each of the other two respondents. Consequently, the averages become 0.5 for each brand, and so Coke is under-counted. It should be borne in mind that this calculation can be useful – it is technically known as average share of category requirements – but it is not market share.

The correct calculations are in the right column. The formula for each of the six cells is the value in the cell divided by the sum of the values in all six cells and multiplied by the sample size. For example, the 1.36 computed for Coke for respondent 1 is  $10 / (10 + 5 + 0 + 0 + 5 + 2) * 3$ .

A practical challenge when using market share as a numeric adjustment variable is that the totals used in calculating the market share can change due to the weighting. For example, the total purchases before the weighting was 22, as in the example above, but after weighting this total is different. The way to address this is to recalculate the market share and redo the weighting a few times until it stabilizes.

## Using dates (waves) in composite variables

It is often useful to weight within time periods. For example, if a survey is repeated every quarter months, it is usually appropriate to weight each quarter separately. This can be done either by constructing the weight within a sub-sample of the data, or by using the time periods as a variable. This is done as follows:

- If it does not already exist, create a categorical variable that represents the time periods.
- Use this categorical variable as an adjustment variable.

- Set the targets to  $1/k$ , where  $k$  is the number of time periods. So, if you have four waves, then each target is 0.25 (it is also possible to do other things, but your analysis will become complicated).
- Create a new set of all the other adjustment variables for each time period, recoding the observations not in that time period to 0. For example, if you are weighting by a composite age and gender variable with eight categories, and you have four time periods, then you need to create 32 variables.
- Divide each of the target values by  $k$ . For example, if the target for Females aged 18 to 24 is 0.1 in the total population, then the target set for each of the four time periods is 0.025.

---

## Selecting adjustment variables

---

The following considerations are relevant when selecting adjustment variables:

- Adjustment variables cannot include missing values.
- Inclusion of the adjustment variable in the weighting should change results. If the results are not affected, then there is no point in including the variable.
- Targets for the adjustment variables need to be known. For example, if you do not know market share, then a brand's market share cannot be used as an adjustment variable. For this reason, it is most common to weight surveys based on age, gender, and geography, as high-quality data is available on these in most countries.
- There's no need to manually create composite adjustment variables. Most weighting software will allow you to create these (as *cell weights*).
- The strength of the relationship of the adjustment variables to the key data collected in the study. The stronger the relationship, the more appropriate the adjustment variable. For example, if you are conducting a study looking at attitudes to social policies, voting history and education become appropriate adjustment variables. In contrast, they are likely less relevant in a study examining preferences for brands of tomato sauce. Two cases where this will commonly be the case are:
  - Non-proportional stratification. For example, if the data was collected using a *booster sample* of people in a group (e.g., buyers of the client's brands), then the variable used to perform this stratification needs to be used as an adjustment variable.
  - There is a relationship between the adjustment variables and known biases in the data collection process. For example, household size and structure are often related to how people are selected for studies (e.g., if conducting a study by phone, people in one-person households will typically be over-represented, as

the bigger a household, the less chance any one person in that household will respond).

---

## Inconsistent categories between targets and adjustment variables

---

Sometimes there is not a neat relationship between the categories used in a survey and those with reliable facts. For example:

- At the time of writing, the US Census collects sex as male and female, whereas Facebook has 58 gender categories (e.g., agender, androgyne, androgynous, bigender).
- It is common to ask people who they voted for and give them an option of *I'd prefer not to say*, whereas, in an election, such an option does not appear.

The simplest solution is to exclude such data from the weighting process. However, this is not a practical solution if the potential adjustment variable is correlated with key results. Solutions are:

- If the survey has extra categories:
  - Use judgment to group people into the most appropriate of target categories.
  - Treat them as missing values and impute the missing values.
  - Add the extra categories as targets, and assume that their size is the same as observed in the survey.
- If the target has extra categories, use judgment to apportion them in the survey question's categories.

---

## The weight variable(s)

---

The outcome of the process of creating weights is ideally a single variable called the *weight*, *weights*, *weight variable*, or *weighting variable*. Each observation in the data set should be assigned a value for this weight variable. An example of such a variable was provided in the second column in the worked example at the beginning of this chapter.

Typically, weight variables are created to have an average value of 1.0 and to have a minimum value of not less than 0, where a value of 0 indicates that the observations with this value are automatically excluded from any analyses performed using the weights.

Some weighting software can produce negative weights, but their use is not sensible in real-world applications.

# Creating sampling weights

In this book, “weights” is used as shorthand for *sampling weights*. This chapter reviews the main ways of creating sampling weights and discusses other (non-sampling) weights.



---

## Post-stratification weights

---

The types of weights described above are known in the academic literature as *post-stratification weights*. In practice, nearly all commercial applications of weighting involve post-stratification.

---

## Design weights

---

Design weights are used to compensate for non-proportional stratification. For example, in a population containing 100,000,000 men and 100,000,000 women, if you used quotas or stratification to achieve a sample of 80 men and 20 women, then a design weight is created to take this non-proportional stratification into account.

The probability of selecting men in this sample is  $80 / 100,000,000 = 0.0000008$  and the probability for women is  $0.0000002$ . So, each man in the sample is assigned a weight of 1,250,000 and each woman a weight of 5,000,000.

In jargon: design weights are computed as the inverse of the selection probability.<sup>3</sup>

Typically, these weights are then further adjusted to take an average value of 1.0 (i.e., the raw values are divided by their average).

Approaches to dealing with design weights include:

- Computing a design weight as described here and using it in analysis.
- Including the stratification variable as an adjustment variable when creating **Rim weights/Raking** or **Calibration**.

---

<sup>3</sup> D. G. Horvitz and D. J. Thompson (1952), "A Generalization of Sampling Without Replacement From a Finite Universe", Journal of the American Statistical Association, Vol. 47, No. 260 (Dec., 1952), pp. 663-685.

---

## Propensity weights (propensity score adjustment)

---

It is possible to use statistical and machine learning techniques to create weights and/or composite categorical adjustment variables. This approach is rare, so this section can be skipped by the less-interested reader.

This approach is useful in situations where there are many potential composite variables (which is rarely the case in practice). The basic process is:

- Obtain or create a data file which contains all the categorical adjustment variables and where the data is considered to be representative of the population of interest (e.g., the proportion of 18 to 24 year old men living in Hawaii in this data set should match the proportion in the population). This is called the *reference data set*.
- Merge the reference data set with the survey data set so that they are stacked on top of each other, with each of the adjustment variables merged together and a single extra variable denoting whether the observation in the combined data set is from the survey data set or the reference data set.
- Estimate a model using the merged data set, predicting which observations are in the survey data set using the adjustment variables as predictors. *Propensity weights* can then be computed in a number of ways, including:
  - As the inverse of the predicted probability of an observation being in the survey data set.
  - Computing Cell weights, where the categorical adjustment variable is created be either:
    - Grouping respondents according to quintiles of their predicted values in the data set of interest (i.e., the 20% of respondents with the highest predicted probability, next 20%, etc.). Logistic regression and random forest can be used for such a model. A particular benefit of this approach is that statistical tests can be used to work out which variables to include and to test for interactions.
    - Using the nodes created by tree-based methods, such as CART and CHAID.

There are further variants of this approach. For more information, please see these sources and the references within:

- Sunghee Lee and Richard Valliant (2009), Estimation for Volunteer Panel Web Surveys Using Propensity Score Adjustment and Calibration Adjustment, *Sociological Methods*, 37(3), 319-343.
- The Pew Research Center (2018), [For Weighting Online Opt-In Samples, What Matters Most?](#)

## Cell weights

When weighting is conducted using a single categorical adjustment variable, it is typically known as *cell weighting* and occasionally as weighting with *interlocking targets*. This is probably the most widespread approach to weighing.

Cell weighting involves targets that are created from combinations of multiple adjustment variables. For example, it is common to create a single composite variable of age, gender, and geography (e.g., Male 18-29 Midwest, Female 18-29 Midwest, Male 30-44 Midwest, etc.).

Total %		Northeast	Midwest	South	West	DC	NET
18-29	Male	1.23%	2.42%	2.98%	2.93%	.00%	9.56%
	Female	.72%	1.90%	2.16%	2.21%	.00%	6.99%
30-44	Male	2.06%	3.60%	4.32%	3.49%	.00%	13.46%
	Female	2.47%	3.55%	4.68%	2.93%	.00%	13.62%
45-64	Male	3.13%	4.11%	6.12%	3.91%	.00%	17.27%
	Female	3.24%	4.83%	7.45%	3.91%	.00%	19.42%
65+	Male	1.49%	2.47%	3.70%	1.64%	.00%	9.30%
	Female	1.90%	2.77%	3.60%	2.06%	.05%	10.38%
NET		16.24%	25.64%	34.99%	23.07%	.05%	100.00%

## Rim weights/Raking

Weighting that involves two or more categorical adjustment variables is typically known as *rim weighting* in commercial research and *raking* in government research. The term “rim” refers to the numbers being on the edge (rim) of the table, as opposed to the cells in the middle of the table (see the image in the previous section).

Rims can also be made up of multiple variables. For example:

	A	B	C	D	E	F
1		Male	Female		Northeast	19.353
2	18-29	9.987	10.049		Midwest	23.888
3	30-44	14.856	15.396		South	35.941
4	45-64	15.089	15.998		West	20.606
5	65+	7.719	10.906		DC	0.212

---

## Calibration weights

---

Weighting that involves a numeric adjustment variable is often referred to as *calibration*, although the term calibration can sometimes be used to refer to what has been defined as the process of scaling a weight (see [Weight calibration](#)) as well as to rim weighting.

Calibration is useful in the following circumstances:

- When numeric adjustment variables are required, as they cannot be used with rim and cell weighting.
- When there is a desire to incorporate a pre-existing design weight into the new weight, such that the new weight is as close to the existing design weight as possible.
- When there is a desire to *trim* the weight, by specifying the minimum and maximum possible values.

A variety of different algorithms have been developed for calibration. They differ in terms how they calculate the difference between the pre-existing design weight and the new weight.<sup>4</sup> The most common algorithms are *raking* (which is so-called because it gets the same result as traditional raking when applied to categorical data), *linear*, and *logistic*.<sup>5</sup>

There is no theory to suggest when to choose a particular algorithm nor when to trim. Further, experience suggests that they tend to have a marginal effect.

---

## Calibrated weight

---

The term *calibrated weight* is sometimes used to refer to a weight created using one of the techniques described above, but with each value divided by the [Effective sample size](#).

---

<sup>4</sup> Where there is no design weight, one is created where the same value is assigned to all observations. Most commonly this weight is 1.

<sup>5</sup> For more detail on algorithms, see Lumley, Thomas S. 2010. *Complex Surveys: A Guide to Analysis Using R*. Wiley Publishing.

---

## Volumetric weights

---

Volumetric weights reflect differences between observations in terms of their “size”. For example, in many studies of consumer goods, the weight is an estimate of the total amount of consumption that a respondent represents, which may be computed using something like household size, amount spent per week, or quantity consumed.

---

## Weights for different purposes

---

Sometimes multiple weights will be created for different purposes. For example:

- A weight may be created for the total sample including a boosted sub-sample and another which excludes the boosted sub-sample.
- Some analyses may have a different weight because they are known to be more biased by aspects of the data collection process.
- A volumetric weight may be created to consider differences between people in the amount they consume.
- One weight may be created for monthly analyses and another for quarterly analyses.
- Where analysis is conducted using rolling samples, a separate weight may be created for each rolling period (e.g., one weight for May to July, and another for June to August, etc.).

The last two of these should be avoided wherever possible. The reduction of bias achieved by such techniques will typically be negligible, while the increased difficulty of analysis and the resulting errors and time wasted will typically be non-negligible (e.g., there’s no rigorous way of comparing waves of a tracker that contain overlapping samples with different weights).

---

## Combining weights

---

In some situations, a sequence of weights will be created. There are a couple of general principles about how this should be done:

- Create design weights first, and then use these as an input into rim weighting or calibration (some algorithms permit the user to provide an existing weight as an input).
- Creating multiple weights and multiplying them together. In particular:
  - If you are confident that there will be no correlation between design weights and other weights, then you can multiply them together.
  - If wanting to perform volumetric weighting, first create a weight variable that ignores volume, and then create the weight to be used in analysis by multiplying this weight variable with the estimate of volume.

---

## More exotic type of weights

---

The weights described in this book are also known more generally as *sampling weights* or *probability weights*. These weights are created to reflect different probabilities of selection for inclusion in the sample. For example, if a sample contains 70% women but it is known that the population should only contain 51%, the data can be weighted to reflect this.

There are other types of weights.

### Frequency weights

Sometimes data files have frequency weights (also sometimes known as replication weights). For example, if two people gave the same answer in a survey, the data file may only contain the data for one of these people with a frequency weight being used to indicate that the data should be counted as two people (i.e., the weight would have a value of 2). Please note: if weighting data by profit, customer value, frequency or consumption (i.e., volumetric analysis), the weights are likely sampling weights, not frequency weights.

## Variance weights

Typically (but not always), this is just another name for frequency weights.

## Replicate weights

A replicate weight is a special type of sampling weight developed for protecting the privacy of individuals in surveys. Replicate weights are rarely used in practice and then virtually only by government statistical agencies (this rareness of use is due to the complexity of using replicate weights). Other than noting that special-purpose software is required to analyze data containing replicate weights (e.g., Westat, Stata), they are not discussed further.

## Replication weights

This is another name for frequency weights.

## Expansion weights

An expansion weight is used so that the weighted data reflects the size of the population of interest (e.g., if there are 300 million people in the population, but 300 people in the survey, a weight of 1,000,000 may be assigned to each respondent).

Expansion weights are also referred to as expansion factors, inflation factors and grossing-up factors. They are typically multiplied with sampling weights to create a single weight. This approach works when the analysis software can accommodate sampling weights but does not work if the analysis program treats weights as frequency weights (because the consequence of the expansion is to inflate the sample size and render all statistical tests invalid).

You can turn sampling weights into expansion weights by dividing the weight variable by the sample size and multiplying by the population size.

## Analytic weights

These are weights that reflect the different levels of precision of various observations. For example, if analyzing data where each observation is the average result from a geographic area, the analytic weight is proportional to the inverse of the estimated variance. Analytic weights are quite obscure in survey analysis and are not discussed further.

## Importance weights

These are weights that reflect the importance of observations in the sample. In commercial research, most commonly, importance weights are things like profitability, household size or typical consumption levels. Typically, importance weights can, and are, treated as being sampling weights.

## Self-weighting samples

A *self-weighting sample* is a sample that can be safely analyzed without any weight.



# Common mistakes

It is not unusual to get weird error messages from weighting software (e.g., “no convergence after 2500 iterations”). And, sometimes, the results just don’t make sense (e.g., too low effective sample sizes).

This chapter views the common mistakes that lead to such problems.

---

## No sample for a target group

---

A sub-group in the population can only be weighted to be greater than 0% if there are some sub-group members in the sample. For example, if the survey did not interview in Hawaii, then either:

- The target for Hawaii needs to be set to 0 (or removed from the adjustment variable and the targets).
- The Hawaii category needs to be merged with some other category.
- The state variable needs to be removed or replaced as an adjustment variable (e.g., by four regions).

---

## Targets for categorical adjustment variables that don't add up

---

For example:

- Targets expressed in percentages that do not add up to 100%.
- Population targets (i.e., which include expansion factors) that are different for different adjustment variables.

---

## Impossible numeric targets

---

The target for a numeric adjustment variable needs to be no greater than the highest observed value and no smaller than the lowest observed value for the variable. Otherwise, an error will occur.

---

## Inconsistent categorical targets

---

Consider the two sets of targets provided below. Provided that the underlying data in the survey is consistent, these targets are impossible, as the one on the left will attempt to create a weight where people with children account for 50% of the weighted sample, whereas the one on the right implies that 40% have children. No weight can satisfy these two sets of targets.

	A	B
1	No children	.5
2	All children under 12	.2
3	Some children over 12	.3

	A	B
1	Single, no children	.2
2	Single, children	.2
3	Couple, no children	.2
4	Couple, children	.2
5	Other, no children	.2

With large numbers of variables, this type of mistake is challenging to spot.

---

## Inconsistent numeric targets

---

Numeric targets also need to be consistent. Trivial cases are easy to spot. For example, if you have two share variables that sum up to more than 1.0, you will get an error (assuming that they only sum up to 1 in the data itself).

However, numeric variables can have difficult-to-spot problems. Suppose two numeric variables are highly correlated. In that case, you need to consider this correlation when specifying their targets (e.g., it may not be possible to increase the average body weight of a sample while reducing its average height).

---

## Ignoring measurement error

---

Weighting corrects for issues in the selection of the sample. It cannot fix measurement errors and, if the discrepancy between a result and a target is due to measurement error rather than sampling, weighting will further reduce the accuracy of the survey while at the same time making it appear more robust.

Consider height. Let's say the height used for your target was based on measurements taken in hospitals, and they reveal an average height for men of 175cm. But, in the survey, the average height is shown as 180cm. While the survey may have over-represented taller men, a much more likely explanation is that the men in the survey have exaggerated their heights. Consequently, weighting by this data would introduce a bias into the survey, whereby shorter people's data would be over-represented.

---

## Using weight factors instead of targets

---

Let's say that the sample contains 50% men, and the target is 70%. A weight factor of, say,  $.5/.7$  can be computed for men. A common mistake is to compute the weight factor and enter it as the target (the correct number to enter is 50%).

# Checking weights

This chapter describes how to check weights. It discusses:

- Weighted tables of adjustment variables
- Weighed table of key variables
- The average of the weight variable
- The range of the weight variable
- The distribution of the weight variable
- The effective sample size

## Case study

In this chapter we review a case study where data on attitudes to gay marriage is weighted by age gender and region. The data is from November 2004 and was collected by the Pew Research Center. After data cleaning, the sample size is 1,837.

The steps for reproducing this example in software are shown in the later chapters.

Initially, an attempt was made to create cell weights using the targets shown below based on age by gender by region. The raw data is [here](#) and the targets are below.

	A	B	C	D	E	F	G
1			Northeast	Midwest	South	West	DC
2	18-29	Male	1.743	2.38	3.659	2.182	0.023
3		Female	1.793	2.379	3.718	2.13	0.029
4	30-44	Male	2.867	3.573	5.252	3.135	0.029
5		Female	3.035	3.644	5.517	3.169	0.031
6	45-64	Male	2.934	3.634	5.34	3.154	0.027
7		Female	3.157	3.761	5.775	3.271	0.034
8	65+	Male	1.552	1.865	2.746	1.542	0.014
9		Female	2.272	2.652	3.934	2.023	0.025

Attempting to use these as weights led to an error, as the survey itself contained no respondents in DC except for one female aged 65+:

	Total %	Northeast	Midwest	South	West	DC	NET
18-29	Male	1.23%	2.42%	2.98%	2.93%	.00%	9.56%
	Female	.72%	1.90%	2.16%	2.21%	.00%	6.99%
30-44	Male	2.06%	3.60%	4.32%	3.49%	.00%	13.46%
	Female	2.47%	3.55%	4.68%	2.93%	.00%	13.62%
45-64	Male	3.13%	4.11%	6.12%	3.91%	.00%	17.27%
	Female	3.24%	4.83%	7.45%	3.91%	.00%	19.42%
65+	Male	1.49%	2.47%	3.70%	1.64%	.00%	9.30%
	Female	1.90%	2.77%	3.60%	2.06%	.05%	10.38%

This problem was rectified by merging DC with the South.

## Weighted tables of adjustment variables

The most fundamental check of a weighting is that the weighted analyses of the adjustment variables correspond to the targets used when creating the weights. Unless **Trimming** is used, the weighted values should be essentially identical to the targets.

The table below shows the survey after applying the weight variable.

	Total %	Northeast	Midwest	South	West	DC	NET
18-29	Male	1.74%	2.38%	3.68%	2.18%	.00%	9.99%
	Female	1.79%	2.38%	3.75%	2.13%	.00%	10.05%
30-44	Male	2.87%	3.57%	5.28%	3.14%	.00%	14.86%
	Female	3.04%	3.64%	5.55%	3.17%	.00%	15.40%
45-64	Male	2.93%	3.63%	5.37%	3.15%	.00%	15.09%
	Female	3.16%	3.76%	5.81%	3.27%	.00%	16.00%
65+	Male	1.55%	1.87%	2.76%	1.54%	.00%	7.72%
	Female	2.27%	2.65%	3.90%	2.02%	.06%	10.91%
	NET	19.35%	23.89%	36.10%	20.61%	.06%	100.00%

## Weighted table of key variables

The key variable of interest in the study measures whether people believe gay marriage should be legalized. Prior to weighting, the result was:

	%
Yes	31% ▼
No	63% ▲
Don't know	6% ▼
NET	100% ▲
Gay marriage should be legal SUMMARY	
sample size = 1946; 95% confidence level	

After weighting, the results only changed a little, suggesting that weighting probably is not required. However, as discussed above, often weighting is still applied in such situations.

	%
Yes	32%
No	61% ↑
Don't know	6% ↓
NET	100% ↑

Gay marriage should be legal SUMMARY

Weight: Weight: Age by Gender by Region; sample size = 1946; effective sample size = 1822 (94%); 95% confidence level

## The average of the weight variable

Typically, weights are created to have an average of exactly 1 (and not, for example, 1.01). It is always a good idea to check the average, as:

- If the average is not 1.0 it may signify that an error has occurred when creating the weight.
- If using non-specialist software for analyzing surveys, such as R and SPSS, analyses can be very wrong if the average is above this value.

## The range of the weight variable

The minimum and the maximum values of the weight variable are also indicators of the quality of the weight. Various rules of thumb exist regarding the maximum weight that should be used, such as that it should not exceed 3.5, 4, or 5. (Where the maximum is set to 5, it means that some observations will be weighted 5 times the average).

Similarly, rules of thumb also exist for the lowest value, such as that it must be greater than 0.2 or 0.33.



When checking the range, keep in mind that the minimum and maximum are related. That is, if the minimum is set to 0.2 and the maximum is set to 5, this means that some respondents will be given 25 times more weight in the analysis than others.

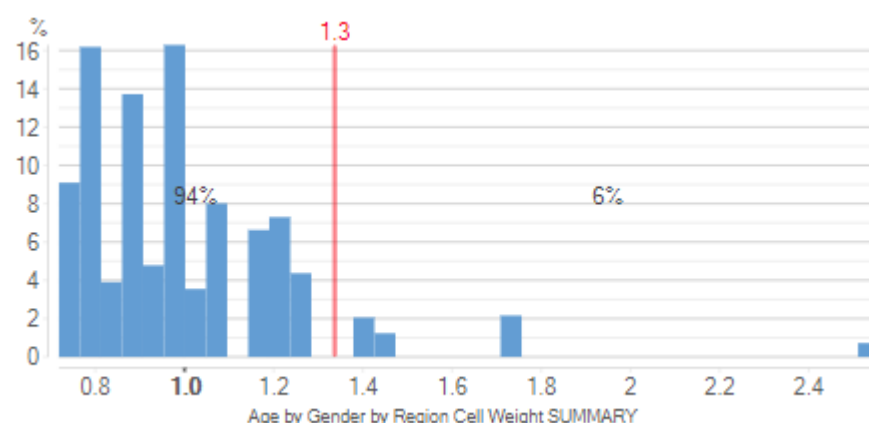
The minimum and maximum weights for the case study are shown below. This tells us that the maximum weight is more than three times the minimum weight. This falls within the ranges described above, which suggests it may not be a massive problem.

	Average	Minimum	Maximum
Age by Gender by Region Target	1.000 ↑	.742	2.534

A practical problem with such rules of thumb is that they are just made-up rules with no theoretical basis, as illustrated by the case study. Surveys often intentionally over-recruit subgroups (e.g., buyers of a particular brand), which means that large ranges of weights are appropriate.

## The distribution of the weight variable

The histogram of the weight variable is shown below. We can see that the maximum of 2.534 is an outlier, and only 6% of weights are greater than 1.3.



18-29 Female Midwest	1.3
18-29 Female Northeast + DC	2.5
18-29 Female South	1.7
18-29 Female West	1.0
18-29 Male Midwest	1.0
18-29 Male Northeast + DC	1.4
18-29 Male South	1.2
18-29 Male West	.7
30-44 Female Midwest	1.0
30-44 Female Northeast + DC	1.2
30-44 Female South	1.2
30-44 Female West	1.1
30-44 Male Midwest	1.0
30-44 Male Northeast + DC	1.4
30-44 Male South	1.2
30-44 Male West	.9
45-64 Female Midwest	.8
45-64 Female Northeast + DC	1.0
45-64 Female South	.8
45-64 Female West	.8
45-64 Male Midwest	.9
45-64 Male Northeast + DC	.9
45-64 Male South	.9
45-64 Male West	.8
65+ Female Midwest	1.0
65+ Female Northeast + DC	1.2
65+ Female South	1.1
65+ Female West	1.0
65+ Male Midwest	.8
65+ Male Northeast + DC	1.1
65+ Male South	.7
65+ Male West	.9

The distribution of the weights can further be explored by seeing how weights differ by the adjustment categories. The table to the right shows the average weight by age, gender, and region. We can see that the outlier is 18 to 29 year olds in the Northeast + DC.

---

## Effective sample size

---

*Effective sample size* is used as a way of quantifying the effect of weighting a survey. For example, if a survey of 1,000 people has an effective sample size for a statistic of 500, it means that the amount of sampling error is equivalent to that which would have been obtained by a study of 500 people that did not need to be weighted.<sup>6</sup>

The following formula, usually attributed to Leslie Kish, computes an *estimate* of the effective sample size:

$$\text{Effective sample size} = \frac{(\sum_{i=1}^n w_i)^2}{\sum_{i=1}^n w_i^2}$$

where  $w_i$  is the weight of the  $i$ th respondent (observation).

It is also useful to present the effective sample size as a proportion, representing the size of the effective sample size relative to the actual number of observations in the study.

In the case study, the effective sample size is 1,837, which is 94% of the original sample size, so the cost of the weighting in terms of reducing the effective sample size is near trivial.

A common question from people that are new to weighting is to ask “what is an appropriate effective sample size?”. There is no good answer to this question, other than to say it should have the same answer as “what is a good sample size?”. But, having said that, effective sample sizes of less than 70% tend to be a concern if not anticipated, and effective sample sizes below 20% are, if not anticipated, potentially highly problematic.

A common misunderstanding in survey analysis is that a survey has an effective sample size. This is rarely the case. Most statistics that are calculated have their own effective sample size. For example, in a table showing the proportion of the sample in different age categories, each of the categories will have a different effective sample size. Kish’s formula is used only because it has simple math, not because it is particularly accurate.

A better approach to computing effective sample size is to:

- Identify all the key analyses of interest.

---

<sup>6</sup> More formally, the *effective sample size* is an estimate of the sample size required to achieve the same level of precision as would be expected to be obtained by a simple random sample.

- Compute the effective sample size for each, using *Taylor Series Linearization* or some other appropriate method.
- If necessary, present the median or some other summary of the effective sample sizes.

The table below, on the left, shows the effective sample size computed for the key question of interest in the marriage study. The value of 1,837 shown for the NET corresponds to the initial Kish calculation. The more relevant numbers are those for Yes, No, and Don't Know, which are all roughly the same. The effective sample size shown in the footer is a summary of the values on the table.<sup>7</sup>

As the Don't Know response tends to be removed when computing polling results, the more accurate effective sample size in this example is the one shown on the table to the right. The 1,672 shown in each cell and in the footer is the more accurate number, as it takes into account the correlation between the weight and the data in the table, whereas the Kish number ignores this.

	%	Effective Sample Size
Yes	32%	1,781
No	61% ↑	1,804
Don't know	6% ↓	1,840
NET	100% ↑	1,837

Gay marriage should be legal SUMMARY  
Weight: Age by Gender by Region Cell Weight: sample size = 1946; effective sample size = 1820 (94%); 95% confidence level

	%	Effective Sample Size
Yes	34% ↓	1,672
No	66% ↑	1,672
NET	100% ↑	1,719

Gay marriage should be legal DK Removed SUMMARY  
Weight: Age by Gender by Region Cell Weight: sample size = 1822; total sample size = 1946; 124 missing; effective sample size = 1672 (92%); 95% confidence level

<sup>7</sup> It is a median. See:

[https://wiki.q-researchsoftware.com/wiki/Weights,\\_Effective\\_Sample\\_Size\\_and\\_Design\\_Effects](https://wiki.q-researchsoftware.com/wiki/Weights,_Effective_Sample_Size_and_Design_Effects).

# Improving a weight

This chapter explains how to improve a weight by changing the targets and trimming.

---

## The bias versus error tradeoff

---

Weights can be modified in a few ways. The two most common ways are by modifying the targets and by trimming, each of which is discussed in the next two sections. There are also more exotic ways of modifying weights, such as smoothing and using Bayesian estimates of weight factors; these are not discussed further.

The goal is to reduce the likely noise in the analyses (i.e., reduce the size of sampling error). This is most readily assessed using the effective sample size.

However, the flip side is that the initial weight will have led to less bias. This can be assessed by comparing the effect of different weights on key results of interest.

---

## Changing the targets

---

The most common strategies for changing the targets are:

- Merging categories of categorical adjustment variables. Often small effective sample sizes are caused by categories in the categorical adjustment variables having small sample sizes, which makes the resulting weights more affected by sampling error.
- Using rims versus cells.

Unfortunately, whatever has the biggest effect in terms of increasing the effective sample size of the weight may also have the biggest impact in terms of increasing bias, so care needs to be taken.

In the previous chapter, we examined a case study that had a single categorical adjustment variable with 32 categories, representing all the combinations of age, gender, and region.

Rim weighing can be conducted by instead using each of age, gender, and region as separate adjustment variables. This involves only have 10 targets. In this case, the resulting weight is grossly inferior to the original weight. In particular:

- The effective sample size from the resulting weight is 1,613 (83%), lower than the one from the cell weights, meaning that the sampling error will be higher.
- The effective sample size computed for the key analysis of interest (attitude to gay marriage) is 1,451, which is also lower.

- This weight makes no attempt to control for differences of age and gender within region, the resulting estimates involved from using this weight will be biased.

The effectiveness of increasing and decreasing the targets needs to be investigated on a case-by-case basis. In other studies, the effects of changing the number of categories can be the opposite of that observed in this case study.

---

## Trimming (bounds on weights)

---

Trimming a weight involves specifying the smallest (*lower*) and highest (*upper*) values that a weight may take.

There are three main approaches to trimming:

- *Simple trimming*: replacing values below some specified value with the specified lower bound and values above the upper value with the upper bound.
- *Constrained calibration*: algorithms that permit specification of targets and constraints regarding lower and upper bounds.<sup>8</sup>
- *Repeated raking and trimming*: computing weights using raking or calibration, trimming them, re-raking/calibrating, etc., until the weights stabilize.

Simple trimming is rarely the best solution. The resulting weight's average is typically changed from 1. And, the weighted tables of adjustment variables will not exactly match the targets. The other two approaches can be hit and miss.

A few comments about the effects of setting upper and lower bounds for weights via trimming are as follows:

- When setting lower and upper values when trimming it is important to remember that they are related. If the maximum is 5 then the minimum should be around 0.2 (i.e., lower = 1/upper).
- Sometimes reducing the upper bound can have a bigger impact than increasing the lower and vice versa.

---

<sup>8</sup> See Deville, J.-C., and Särndal, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87, 376-382 and Deville, J.-C., Särndal, C.-E., and Sautory, O. (1993). *Generalized raking procedures in survey sampling*. *Journal of the American Statistical Association*, 88, 1013-1020.

- The narrower the permitted range, the higher the effective sample size computed using Kish's formula, and typically the higher the effective sample size using other formulas. This is because the more trimming, the closer we come to the assumption of having no weight at all.
- When cell weighting is used, the effect of trimming is always to increase the magnitude of the difference between the targets and the weighted results. This is not always the case with calibration and raking.

# Analysis of weighted data

Traditional data analysis software is not designed for the use of sampling weights and will produce incorrect statistical inference ( $p$ -values, statistical tests, standard errors, and confidence intervals).



---

## Bias and statistical inference

---

Notwithstanding dumb mistakes, using weights when analyzing data always reduces bias.

The flipside of this is that weighting typically increases the estimated sampling error. The computed standard errors and confidence intervals are usually greater once weighting has been conducted,<sup>9</sup> *provided that the analysis is being done correctly* (more about this below).

---

## Approaches to incorporating weights in statistical analyses

---

There are at least six approaches in widespread use for performing statistical analyses on weighted data. Only one of these is valid.

### Variance estimation

Variance estimation refers to a class of techniques specifically designed for the analysis of weighted data. The most widely used of these is *Taylor Series Linearization*. Variance estimation is the “correct” approach. It is the approach typically used by government statisticians on very important studies. It is the approach taught in statistics courses. It is the only one approach that can be proved, with math, to be correct. All the others can be proved to be inferior.

### Resampling

Resampling creates a new synthetic data set by randomly selecting cases, with replacement, from an existing data set. Cases are selected with probability proportional to the weight. That is, a weighted bootstrap is used to create the data set.

This approach is always inferior to variance estimation due to the noise added from the random selection and by ignoring any correlations between the weight and the variables in the analysis.

---

<sup>9</sup> With non-proportionate stratification, the opposite can occur.

## Weight calibration

*Weight calibration* involves modifying the weight to have an average weight of  $ess/n$ , where  $ess$  is the effective sample size and  $n$  is the actual sample size, and using this weight in conventional statistical tests. The resulting weight is sometimes known as the **Calibrated weight**.

This approach is not “valid”.<sup>10</sup>

## Using the observed sample size

*Using the observed sample size* in any statistical test is less valid than any of the approaches described above, as it ignores the concept of effective sample size.

## Modifying the weight to have an average weight of 1

*Modifying the weight to have an average weight of 1* (if it doesn't already have this value) and using this weight in conventional statistical tests is the worst of the approaches discussed so far.

On average, half of the respondents will be given a weight of more than 1, so the certainty of these respondents' data is over-stated.

## Use the weight as is

The final approach is to use the weight “as is” in conventional statistical tests (i.e., as replication or frequency weights). If the weight already has an average of 1, this is identical to the previous approach. However, if the weight represents a probability of inclusion, or, is grossed up to the population (i.e., an expansion weight), the statistical tests will become even less accurate.

This last approach is the worst approach. It is the default in both SPSS and R.

---

<sup>10</sup> Robert E. Fay (1985), “A Jackknifed Chi-Squared Test for Complex Samples”, Journal of the American Statistical Association, March 1985 80(389):148-157.

---

## Does choice of approach make a big difference? Yes

---

The difference between these approaches is not small. For example, consider an extreme situation where a data set contains 1 man who said “Yes”, 50 women who said “No”, and another 50 women who said “Yes”. If you compare the men and women, we have 100% of men saying Yes ( $n = 1$ ) and 50% of women ( $n = 100$ ). It is hopefully obvious that the difference between the men and women will not be statistically significant.

If you use the average weight of 1 approach, the assumed comparison is 100% men ( $n = 50.5$ ) versus 50% women ( $n = 50.5$ ), and the difference will be concluded to be highly statistically significant. Obviously, most people would notice a mistake as big as this, but the problem with the approach is that all analyses have the same problem, just hopefully not to this extent.

Using the actual sample size suffers from the same problem. For example, if using a chi-square test or an ANOVA to test the difference between men and women and entering the total sample size of 101, you would still get the wrong answer. Yes, there are some analyses where you could enter the correct sample sizes (e.g., a  $t$ -test), but many, many, analyses only permit the total sample size to be estimated in statistical tests.

Weight calibration also suffers from the same problem, but in a less extreme fashion.

Unfortunately, the math of variance estimation is complicated, and for a surprisingly large number of problems (e.g., stat tests on multiple-response grid questions, and many multivariate statistical methods), the relevant math has yet to be derived.

---

## Approaches available by stat package

---

### Q and Displayr

Q and Displayr use Variance estimation (Taylor Series Linearization) for all crosstabs and all regression methods.

Weight calibration is used for all summary tables and other analyses. When weight calibration is used, the parameter estimates (percentages, means, coefficients, predictions) are correct, but the statistical inference is not.

## SPSS Statistics

By default, when a weight is applied in SPSS Statistics, it often defaults to assuming it is a frequency (replication weight), rather than a sampling weight. As a result, all statistical inference (statistical tests, standard errors, confidence intervals,  $p$ -values, and information criteria) give the wrong answer, but parameter estimates will often be correct.

“Often” is an important qualifier in the paragraph above. Some SPSS procedures either ignore weights entirely or round them to the nearest whole number, which renders the parameter estimates incorrect (see <https://www-01.ibm.com/support/docview.wss?uid=swg21477298>).

## R

By default, when a weight is applied in R functions, R defaults to assuming it is a frequency (replication weight), rather than a sampling weight. As a result, all statistical inference (statistical tests, standard errors, confidence intervals,  $p$ -values, and information criteria) give the wrong answer, but parameter estimates will often be correct.

A number of packages have been developed for R that can accommodate sampling weights, and which are designed specifically for survey analysis (e.g., the `survey` package). Example code is provided in the final chapter of this book.

---

## When weighting does and does not work

---

The theoretical assumption that justifies the use of weighting is that the probability that someone will choose to participate in a survey varies between people, and that differences in this probability can be explained with reference to the adjustment variables. Or more formally, that the data is *missing at random*.<sup>11</sup>

Consider the situation where a survey was well managed but due to some anticipated quirk of how the data was collected it over-represents males. If gender is known to relate to other variables of interest in the survey, it follows both that:

---

<sup>11</sup> E.g., see G Kalton and I F-C (2003), “Weighting Methods”, *Journal of Official Statistics*, 19(2), 81-97.

- The data needs to be weighted or otherwise any analyses will be misleading (as they will be biased by the misrepresentation of gender).
- Weighting will fix the problem.

However, what if the gender problem is one of many unknown problems in the data? Imagine that the survey is also biased towards low income people, towards people that like to do surveys and towards people that do not get out of doors much. Fixing the gender problem by weighting will not fix any of these other problems. Furthermore, in general, we will not know that we have these other problems, as we will have no data to check for representativeness.

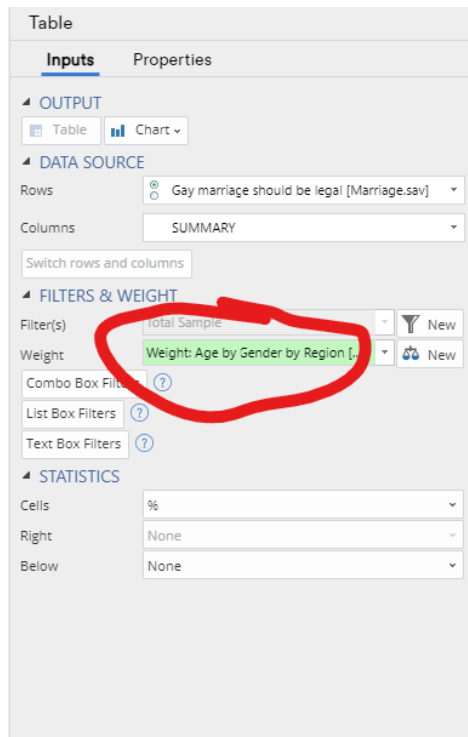
# Weighting in Displayr

---

## Applying a weight to an analysis

---

1. Create the analysis
2. Select the weight in **Inputs > FILTERS & WEIGHT > Weight** on the right side of the screen.



---

## Weighted versus unweighted data

---

Almost all statistics in Displayr are automatically weighted when a weight is applied. The only exceptions are those that refer explicitly to **Count** or **Sample size**; these are based on the unweighted sample size, excluding any observations with a weight of 0, less than 0 or a missing value.

When viewing a table, change statistics by clicking on the table and select the options from **Inputs > STATISTICS**.

---

## Effective sample size

---

All standard analyses in Displayr automatically show the effective sample size. Further, most crosstabs have it available as an optional statistic in **Inputs > STATISTICS**.

---

---

## Using weights created in other software

---

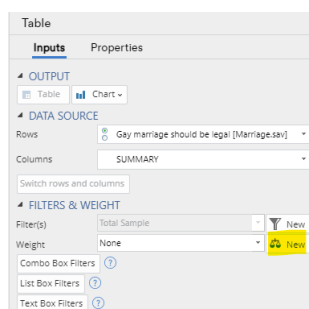
1. Find the variable in **Data Sets** tree (bottom-left of the screen)
  2. Click on it
  3. Check **Usable as a weight** (right-side of the screen)
- 

---

## Creating weights

---

1. Select a table that you wish to weight



2. Press the **New** button to create a weight.
3. Choose the first adjustment variable:



**New Weight**

Recompute weights for The entire dataset Make weights sum to sample size Select Design Weight Variable...

▼ Adjustment Variable(s):



4. To create cell weights, enter additional adjustment variables.
5. Enter targets. For example:

▼ Adjustment Variable(s): Age by Gender by Region

Age [Marriage.sav]  by  Region [Marriage.sav]

by  Gender [Marriage.sav] by  Select Adjustment variable(s)

by  Select Adjustment variable(s)

Targets are Percentages  

	Northeast	Midwest	DC, South	West
18-29, Male	1.743	2.38	3.682	2.182
18-29, Female	1.793	2.379	3.747	2.13
30-44, Male	2.867	3.573	5.281	3.135
30-44, Female	3.035	3.644	5.548	3.169
45-64, Male	2.934	3.634	5.367	3.154
45-64, Female	3.157	3.761	5.809	3.271
65+, Male	1.552	1.865	2.76	1.542

6. To create rim weights, press + Additional adjustment variable set and enter any targets. For example:



➤ Adjustment Variable(s): Age by Gender

---

▼ Adjustment Variable(s): Region

Region [Marriage.sav] by  Select Adjustment variable(s)

by  Select Adjustment variable(s)

Targets are Percentages  

Northeast	19.353
Midwest	23.888
South	35.941
West	20.606
DC	0.212
	100%

7. Press **+** **New Weight**.

---

## Additional options

---

- Apply any targets within date periods, by selecting date variables in **Recompute weights for**.
- Make weights as close to a *design weight* as possible by using **Select Design Weight Variable**.
- Trim using constrained calibration by using the **Minimum weight** and **Maximum weight** boxes.
- Choose **Algorithm** when applying calibration.

# Weighting in Q

---

## Applying a weight to an analysis

---

1. Create the analysis
2. Select the **Weight** in the dropdown at the bottom of the screen in the middle.

---

## Weighted versus unweighted data

---

Almost all statistics in Q are automatically weighted when a weight is applied. The only exceptions are those that refer explicitly to **n** or sample size; these are based on the unweighted sample size, excluding any observations with a weight of 0, less than 0 or a missing value.

When viewing a table, change statistics by clicking on the table and select the options from **Inputs > STATISTICS**.

---

## Effective sample size

---

All standard analyses in Q automatically show the effective sample size. Further, most crosstabs have it available as an optional statistic (right-click on a table and choose from the **Statistics** sub-menus).

---

## Using weights created in other software

---

1. Find the variable in the **Variables and Questions** tab
2. Click the yellow **W** button (in the **Tags** column).

---

## Cell weights

---

This section describes how to create cell weights when there is no existing weight variable to be incorporated. See below if you need to incorporate an existing weight, numeric adjustment variables, or trimming.

1. **Create > Variables and Questions > Variables > Weight**, which opens the **Construct Weight** dialog box, shown below.
2. Choose a categorical adjustment variable in the **Target question(s)** dropdown.
3. If desired, drag and drop the Targets categories onto each other to merge them.
4. If you wish to perform cell weighting with a composite variable, Q will automatically do this in the background. All that is required is that you select the other input variables. In the example below, for example, the composite variable involves age, gender, and region, and corresponds to the table of population targets shown in the chapter on [Adjustment variables and targets](#) (to use this table you would need to first convert its values to proportions, or, set **Target numbers** to **Population/Count**).
5. Press **OK**.

Construct Weight

Recompute for each: (do not recompute)

Q will compute the weight once for the total sample.

Weight set

Target question(s):

Age

Gender

(none)

Target column question(s):

Region

(none)

Target numbers

☒ Percentage %

☐ Population/Count

☐ JavaScript formula

Force sum of weights to equal sample size ☒

Targets:

	Northeast	Midwest	South	West	DC	
18-29, Male						0%
18-29, Female						0%
30-44, Male						0%
30-44, Female						0%
45-64, Male						0%
45-64, Female						0%
65+, Male						0%
65+, Female						0%
	0%	0%	0%	0%	0%	0%

Add Weight Set

Diagnostics Report

OK

Cancel

Help

## Rim weights

Rim weighting is performed in the same way as cell weighting (described in the previous section), except that you need to click **Add Weight Set** each time you wish to add an additional adjustment variable.

## More complex weights

Q has another approach for creating weights which is designed for more complicated types of weights (e.g., using numeric targets, trimming, and using existing weights as an input).

1. Press **Automate > Browse Online Library > Weight > Configure Weight from Variable(s)**.
2. If there is an initial weight that needs to be incorporated into the new weight (e.g., a weight created by whoever provided the data file, or a **design weight** or **propensity weight**), this is selected in **FILTERS & WEIGHT > Weight**.
3. If wanting to apply the filter to a subset of the population, select a filter in **Filter(s)**. The main use cases of this are when wanting to exclude some groups from the weight (e.g., booster samples, people with missing data), or when wanting to create different weights for different sub-samples (e.g., for different countries in a global study).
4. Create any required composite adjustment variables:
  - **Create > Variables and Questions > Variable(s) > R Variable**
  - Enter code in the following structure, replacing `Gender` and `Age` with whatever variable names are appropriate: `paste (Gender, Age)`
  - The resulting variable can be used with the more complicated weighting approach described below. However, if you wish to use it with cell weighting, change the **Question Type** to **Pick One**.
5. Add the targets for each variable. These are added by clicking the buttons below the target variables and typing or pasting the targets into the spreadsheets. The targets should be expressed as proportions. An example is to the right.
6. Select any numeric adjustment variables.
7. Set the targets for the numeric adjustment variables.
8. Press **CALCULATE**. If you get an error message, please read the error message shown on the top-right of the screen. See also [Error! Reference source not found.](#)
9. If you do not have an error message, click **Inputs > SAVE VARIABLE(S) > Save Weight Variable from Configuration** which will add the weight to the data file.

	A	B
1	18-29 Female	.125
2	18-29 Male	.125
3	30-44 Female	.125
4	30-44 Male	.125
5	45-64 Female	.125
6	45-64 Male	.125
7	65+ Female	.125
8	65+ Male	.125

# Weighting in R



---

## Applying a weight to an analysis

---

When conducting weighted analyses in R it is advisable to use purpose-built R packages designed for sampling weights, such as the `survey` package. It is not advisable to use traditional R functions that have a weight parameter, as typically these weight parameters are not assumed to be sampling weights. For example, below shows the code for applying a weight to a logistic regression in R via the standard `glm` function:

```
># Standard survey implementation
>marriage = foreign::read.spss('https://docs.displayr.com/images/8/89/Marriage.sav',
                             to.data.frame = TRUE)
>
># Creating a binary variable
>marriage$support.gay = as.integer(marriage$gay == "Yes")
>marriage$support.gay[is.na(marriage$support.gay)] = 0
>
># Standard binary logit
>summary(glm(support.gay ~ age.cat * female,
             data = marriage,
             family = binomial(logit),
             subset = race.wbh == "White",
             weight = weight))
non-integer #successes in a binomial glm!
Call:
glm(formula = support.gay ~ age.cat * female, family = binomial(logit),
    data = marriage, weights = weight, subset = race.wbh == "White")

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.4512  -1.2441  -0.8999   1.7128   3.2741

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    -0.2531     0.1174  -2.155   0.0311 *
age.cat30-44    -0.6201     0.1583  -3.916 8.99e-05 ***
age.cat45-64    -0.9055     0.1537  -5.890 3.85e-09 ***
age.cat65+     -1.7448     0.2286  -7.631 2.33e-14 ***
femaleFemale     0.3069     0.1648   1.862   0.0626 .
age.cat30-44:femaleFemale  0.1683     0.2156   0.780   0.4351
age.cat45-64:femaleFemale  0.4071     0.2114   1.926   0.0541 .
age.cat65+:femaleFemale   0.3663     0.2895   1.266   0.2057
```

Note that the computed  $p$ -value in the second to last row (representing the impact of females aged 45 to 64) shows a  $p$ -value of 0.0541. By contrast, the output below estimates logistic regression using a function specifically designed for the sampling weights. Note that the  $p$ -value is markedly different at 0.228069. Most of the time when the functions are used the results are less significant.

```
>summary(svyglm(support.gay ~ age.cat * female,
  subset = race.wbh == "White",
  design = svydesign(id = ~1,
    weights = ~weight,
    data = marriage),
  family = binomial(logit)))
non-integer #successes in a binomial glm!
Call:
svyglm(formula = support.gay ~ age.cat * female, design = svydesign(id = ~1,
  weights = ~weight, data = marriage), subset = race.wbh ==
  "White", family = binomial(logit))

Survey design:
svydesign(id = ~1, weights = ~weight, data = marriage)

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      -0.2531    0.1858  -1.362 0.173353
age.cat30-44      -0.6201    0.2427  -2.555 0.010718 *
age.cat45-64      -0.9055    0.2339  -3.871 0.000113 ***
age.cat65+        -1.7448    0.3124  -5.586 2.72e-08 ***
femaleFemale       0.3069    0.2803   1.095 0.273608
age.cat30-44:femaleFemale 0.1683    0.3521   0.478 0.632788
age.cat45-64:femaleFemale 0.4071    0.3376   1.206 0.228069
age.cat65+:femaleFemale 0.3663    0.4215   0.869 0.384868
```

Frequency tables and crosstabs are created using the `svytable` function. For example:

`svytable(~ age.cat + female, design)`, where the `design` is created as illustrated above.

## Creating weights

The code below computes the weight used in the cell weighting case study presented earlier:

```
# Standard survey implementation
marriage = foreign::read.spss('https://docs.displayr.com/images/8/89/Marriage.sav',
  to.data.frame = TRUE)

# Merging DC with North East
marriage$region.cat = as.character(marriage$region.cat)
marriage$region.cat[marriage$region.cat %in% c("DC", "Northeast")] = "Northeast + DC"
marriage$region.cat = factor(marriage$region.cat)

# Specifying the adjustment variables
formula = ~ age.cat + female + region.cat

# Loading the package
```

```

library(survey)

# The targets
nms = colnames(model.matrix(formula, data = marriage))
k = length(unique(marriage$age.by.gender.by.region))
targets = c(1, 0.302520, 0.310866, 0.186247, 0.523486, 0.195652, 0.359417,
0.206060)
names(targets) = nms

# Computing the weight and updating the design
design = calibrate(svydesign(id = ~1, weights = ~1, data = marriage),
                  formula,
                  population = targets * NROW(marriage),
                  calfun = "raking")

weights(design)

```

A few comments:

- The line starting with `formula` says to use each of `age.cat`, `female`, and `region.cat` as separate adjustment variables. If you wish to add numeric adjustment variables, they are added in the same way.
- The section titled `# The targets` sets up the targets in the expected format. Some specific things to be aware of:
  - It is a good idea to inspect `nms` in order to make sure that everything is lined up properly.
  - The first value determines the overall average weight; it is best left at 1.
  - When using categorical adjustment variables, the first category is always excluded (i.e., the 12 categories for the 3 adjustment variables are represented by only 9 targets).
  - When using a numeric adjustment variable, just add the target average to this vector.
  - If you wish to use an existing weight as an input, modify the `design` line, replacing the 1 in `weights = ~1` with the name of the weight variable (e.g., `weights = ~wgt`).
  - If you wish to apply trimming, replace `raking` with `logit` and add an argument called `bounds` (e.g., `calfun = "logit", bounds = c(0.5, 2)`).
- Create a composite variable in the data frame. E.g.,

```

marriage$age.by.gender.by.region = paste(age.cat, female, region.cat)
formula = ~ age.by.gender.by.region

```

# Want to cut your analysis and reporting time in half?

See Displayr in action →

Analysis and reporting software built to save you time

