TOPIC 7
ANALYSING WEIGHTED DATA

You don't have to eat the whole ox to know that the meat is tough.

Samuel Johnson

Introduction

different analysis for sample data

Up until now, all of the analysis techniques have only dealt with producing the mean and variance from the raw data. But what do we do if we have weighted data? This chapter will discuss firstly what a weight is and then look at how we can calculate summary statistics from weighted data. Finally, we will look at the Standard Error and discuss how it differs from the Standard Deviation.

What are weights?

the sample represents the population

We have already seen in Topic 1 that often it is better to take a sample of the population rather than going to the entire population. Remember that a sample is a small number of statistical units selected to represent the entire population. After we have completed the survey, we are often required to make inferences about the population using the sample.

For example, let us say that we have just completed a sample survey asking about how many shells of Kava you drank last week. We could easily apply the techniques in the first 6 topics of this book to produce the mean, median and variance **of this sample**. However, how could we estimate the total amount of money spent on Kava in Vanuatu or the total number of people that drank Kava?

estimates for the population use weights

To produce estimates about the entire population we allocate each person in the sample a weight.

A weight indicates how many people in the population this person represents.

For example, if our Kava Survey went to 1,500 people and there are 150,000 people in the population then each person in the survey would receive a weight of $^{150,000}/_{1,500}$ =100. So each person in the sample would be representing 100 people in the population.

weights depend on sample selection

In the Kava Survey, each person received the same weight, however this is not always the case. The weight of each statistical unit will depend on the way that the units were selected and how many units did not respond. However, the following things should always be true:

3 'rules' for weights

1. if you add up all of the weights they will equal the number of units in the population;
2. the weight of a unit indicates how many units in the population it is representing; and
3. weights should always be larger than 1. If you are in a survey then you must be representing at least one person.....yourself!

weights do not have to be whole numbers

Although it sounds strange, weights do not need to be a whole number. That is, you can have a weight of 4.53 or 88.763. Although it may seem strange that a person can be representing 88.763 people in the population, it is just our "best guess" of how many people with similar characteristics are in the population.

Using weights in analysis

sample values estimate the population

One important difference when we are analysing weighted data is that we don't know the exact value for the population. For example, when we are working out the mean value for a census, we know every value in the population, so the result will be exactly right. However, when we take a sample we are not going to everyone in the population, just a few people or statistical units who we think represent the population.

Sometimes the sample selected will provide a good or accurate representation of the population, and at other times it might not be so good. It is important to select the sample in such a way as to best represent the population.

TIP

When we calculate the mean, median and variance using weighted data, we say that we are **estimating** the population value, that is, we are estimating the population mean, median or variance.

at best an estimate

So if we are working on census data, then we can say that we are calculating the mean, but if we are using a sample to estimate the mean of a population then we are estimating the mean.

sample error is the measure of the sample's representation of the population

These sample estimates of the population values are subject to **sampling error**. This is error because we have chosen a sample, and not the whole population. If we have chosen the sample correctly, our estimates of the population values will be very close to, or even the same as, the population values. In these situations the sampling error is very small. Remember that population estimates are also affected by non-sampling errors which are very difficult to calcuate.

assume here that the weights have been defined for you

When you are working with weighted data, the weights will be provided to you with the data. Weights are determined by a number of factors, such as how the sample was selected, response rates and other criteria usually specified by a mathematical statistician. Do not worry about how to calculate the

weights, it is often complex and out of the scope of this course.

Estimating the mean

formula

For weighted data the formulae for estimating the mean is:

Estimate of the population mean from the sample = $\frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$

where w_i is the weight for each unit x_I is the value for the variable

example

For example, if we had the following results from a survey:

Table 7.1 Estimating the population mean from sample data

Value (x_i)	Weight (w_i)	Weight × Value ($w_i x_i$)
5	10	50
3	20	60
6	15	90
7	10	70
9	5	45
Total	**60**	**315**

Source: Illustrative data only.

mean estimate

Then our estimate of the population mean would be:

$$= \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} \qquad = \frac{315}{60}$$

$$= 5.25$$

Estimating the variance of the population

estimate variance

For weighted data the formula for estimating the variance of the population looks complicated, because the weights are integrated into the calculation:

$$\text{Variance estimate} = \frac{\sum_{i=1}^{n} w_i x_i^2 - \frac{(\sum_{i=1}^{n} w_i x_i)^2}{(\sum_{i=1}^{n} w_i)}}{(\sum_{i=1}^{n} w_i) - 1}$$

example

We can use the same sample survey data as before to estimate the variance for the population. Of course, it is much easier to make the estimates if you use a computer to perform the calculations using formulas

Table 7.2 Estimating the population variance from sample data

Value (x_i)	Weight (w_i)	Weight × Value ($w_i x_i$)	Weight × Value squared ($w_i x_i^2$)
5	10	50	250
3	20	60	180
6	15	90	540
7	10	70	490
9	5	45	405
Total	60	315	1,865

Source: Illustrative data only.

variance estimate

Then our estimate of the population variance would be:

$$\text{Variance estimate} = \frac{\sum_{i=1}^{n} w_i x_i^2 - \frac{(\sum_{i=1}^{n} w_i x_i)^2}{(\sum_{i=1}^{n} w_i)}}{(\sum_{i=1}^{n} w_i) - 1} = \frac{1{,}865 - \frac{(315)^2}{60}}{60 - 1} = \frac{211.25}{59} = 3.58$$

Creating frequency distributions

sum the weights

We can also create frequency distributions of the population using weights. It is very similar to creating the frequency distributions that we created in Topic 3. The main difference is that our frequency is now the sum of the weights in the class interval.

example

For example, let us say that we have the following data:

Table 7.3 Creating frequency distributions from sample data

Variable (x_i)	Weight (w_i)
2	5
3	6
3	6

4	5
4	5
4	5
5	4
5	7
7	8
8	5
9	4

Source: Illustrative data only.

class intervals

Let us use the following classes: 2–3, 4–5, 6–7, 8–9

frequency

Then our frequencies for our population would be:

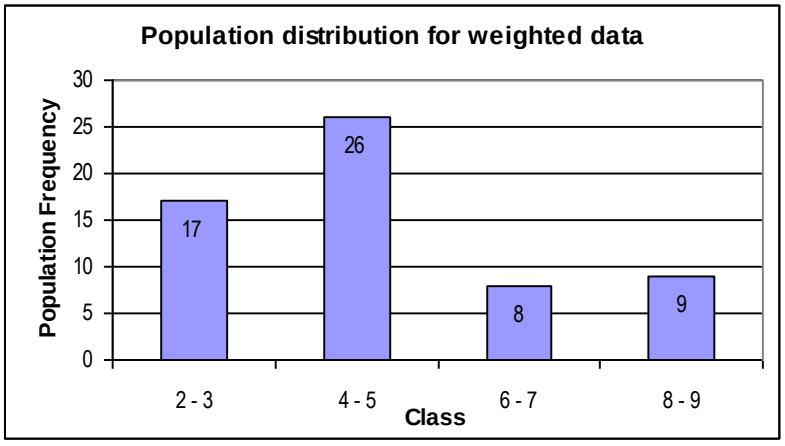
Table 7.4 Creating frequency distributions from sample data

Class	Weights in Class (w_i)	Frequency (Σw_i)
2–3	5 + 6 + 6	17
4–5	5 + 5 + 5 + 4 + 7	26
6–7	8	8
8–9	5 + 4	9

Source: Illustrative data only.

Charts

Once you have the frequency distribution, you chart the data in the same way you chart unweighted data, using the same guidelines for quantitative and qualitative data.



Population proportions

estimate population proportions

Often we want to know the number of people in the population with a certain characteristic. As for many of the measures demonstrated in this chapter, you use the sum of the weights rather than the frequency of the value or variable.

For example, we may want to estimate the number of people in the population who drink Kava. The following formula will estimate the population proportion from the sample:

formula

$$\textbf{Population proportion estimate } (p) = \frac{\sum w_{\text{i (with characteristic)}}}{\sum w_{i\text{ (all units)}}}$$

So you take the sum of the weights of the variables with the characteristic and divide it by the sum of all the weights in the sample.

example

So let us say that we had the following sample of people, and we asked them if they drank Kava last night:

Table 7.5 Sample of people and if they drank Kava last night

Person	Weight	Drank Kava
John	10	Yes
Simon	9	Yes
Howard	7	No
Jack	8	Yes
Theto	10	No
Jean Mark	9	No
Meryline	4	Yes

Source: Illustrative data only.

working

Our estimate of the proportion of people in the population who drank Kava last night would be:

$$\text{Population proportion estimate of kava drinkers } (p) = \frac{\sum w_{\text{i (with characteristic)}}}{\sum w_{i\text{ (all units)}}}$$

$$= \frac{(10+9+8+4)}{(10+9+7+8+10+9+4)} = \frac{31}{57} = 0.5438$$

result

So we can say that we estimate that 54.38% of people in the population drank kava last night.

Standard errors

sample data estimates the population

Whenever a sample survey is conducted, an additional dimension of error is introduced due to sampling only a subset of the population. The theory relating to this type of error is well developed and the ability to calculate the error introduced by this collection methodology is one of the attractive features of sample surveys.

The usual quantity calculated to measure the accuracy of a sample estimate is the “standard error” of the estimate. The standard error of an estimate enables us to make certain probability statements about the estimate, if certain conditions are met (these conditions are not constrictive as long as the sample

sizes are not very small). We can say that we are 95% certain that the true value of what we are trying to measure will be within two standard errors of our sample estimate.

how accurate is the sample data?

The Standard Error should not be confused with the Standard Deviation.

- ☆ **Standard Error:** measure of how accurate an estimate is from a sample survey.
- ☆ **Standard Deviation:** measure of the spread of the values in the population.

The Standard Deviation is a component of the Standard Error, such that the greater the Standard Deviation, the greater the Standard Error.

The **larger** our Standard Error is, the **worse** the estimate is.

sampling techniques available

There are many different sample selection techniques available, and quite often the preferred option is a combination of numerous techniques.

Some common techniques applied by survey statisticians include:

i) Simple Random Sampling: is the most simple form of sampling, and involves assigning a random number to the units in the population of interest and using this random number to select the sample. Usually results in a well spread out sample.

ii) Systematic Sampling: is also a simple form of sampling and involves listing the units in the population of interest (often ordered by a particular variable/s), and then running a skip through the list to select the sample. Also results in a well spread out sample.

iii) Stratified Sampling: involves splitting the units in the population of interest into sub-populations (strata), and selecting separate samples within each stratum. Has the benefit of being able to control sample sizes for sub-populations.

iv) Probability Proportional to Size (PPS) Sampling: for this technique, the size of a unit determines the likelihood of that unit being selected. That is, if villages are being selected for the survey, and the size measure being adopted is the number of households in each village, then a village with twice as many households as another village, will have twice as much chance of being selected.

v) Cluster Sampling: involves selecting clusters of units instead of units spread out randomly. Has the benefit of cost saving due to the reduced travel cost. Unfortunately, as the sample is no longer well spread, the standard errors will increase.

vi) Multi Stage Sampling: involves selecting the sample in more than one stage. For example, rather than go to every village, simply select a sample of villages, and from those selected villages, select a sample of households. Once again this technique has cost savings.

As mentioned, it is rare for just one of these techniques to be applied to a sample survey. Often 2, 3 or even 4 of these techniques are adopted for the one survey. For example, for a HIES, you may first split the population of interest into two strata (urban and rural). From there you may wish to list all the villages in each of the stratum and apply PPS sampling to select the villages. For each selected village, it may then be desirable to run a skip through the village in order to select 10 households. This approach uses Stratified Sampling, PPS Sampling, Systematic Sampling, and Two-Stage Sampling.

formula

The formula used to calculate the Standard Error for a sample survey is dependent on the sample selection methodology applied. When a combination of sampling techniques is applied, then the

formula can become extremely complex. Often approximation techniques (eg, Jack-knife variance estimation) are adopted to overcome this problem.

To give an example of what the formulae look like however, to calculate the Standard Error of the estimated population mean, population total and population proportion from a simple random sample we use the following formula:

***SE* (estimated population mean)** = $\sqrt{\{(1-(\frac{n}{N}))*\frac{s^2}{n}\}}$

***SE* (estimated population total)** = $\sqrt{N^2*\{(1-(\frac{n}{N}))*\frac{s^2}{n}\}}$

***SE* (estimated population proportion)** = $\sqrt{\{(1-(\frac{n}{N}))*\frac{(p(1-p))}{n}\}}$

Where s^2 is the estimate of the population variance

n is the sample size

N is the number of units in the population

p is the estimate of the proportion in the population

kava example

So in the Kava example above, our proportion was 0.5438, our sample is size 7 and our population is 57, so the Standard Error is:

SE (estimate of the proportion in the population)

$$= \sqrt{\{(1-(\frac{7}{57}))*\frac{(0.5438(1-0.5438))}{7}\}} \qquad = 0.1763$$

So we say that the standard error of the sample proportion is 0.1763.

interpreting standard errors

The first thing we need to ask ourselves when we produce a standard error is "what does it mean?"

If you take 2 standard errors either side of an estimate, then you are 95% confident that the true value will lie within this range. That is, for the kava example above, we can say that we are 95% confident that the true value for the proportion of kava drinkers is between:

$$0.5438 \pm (2*0.1763)$$

$$= (0.1912, 0.8964)$$

relative standard error (RSE)

Another way of showing the accuracy of an estimate is to compute the relative standard error (RSE), which is simply the standard error as a percentage of the estimate. For the example above, the relative standard error for the estimate would be calculated as follows:

$$RSE = 0.1763/0.5438*100 = 32.4\%$$

It differs from person to person as to what constitutes a good estimate in terms of RSE, but most statisticians would agree that any estimate with an RSE below 5% is a good one. When running a

sample survey, it is desirable to produce key estimates from the survey with RSEs of 5% or below. For other estimates of significance, the RSEs should not exceed 20%. It would therefore be fair to conclude that the estimated proportion of Kava drinkers above is not a very good one. This is largely due to the sample size of 7 being too small. If the sample size was increased to 30, and an estimate of 0.5438 was still generated, then an RSE of 11.5% would have been achieved.

SE (estimate of the proportion in the population)

$$= \sqrt{\{(1-(\frac{30}{57}))*\frac{(0.5438(1-0.5438))}{30}\}} \qquad = 0.0626$$

$$RSE = 0.0626/0.5438*100 = 11.5\%$$

zero error in a Census

One additional note is that if we run a census then the sample size equals the population size ($n = N$). So, the estimate of the population value should be exactly correct because we have every one in the population in the sample.

We can also see that in the formula, that when n = N then:

$$(1-\frac{n}{N}) \quad = \quad (1-\frac{N}{N}) \quad = \quad 1-1 \quad = \quad 0$$

So there is no error in our estimate.

non-sampling errors

Remember that sampling error is only one component of the total survey error. There are many ways that error can be introduced to surveys other than the use of sampling methods. Field enumeration error, respondent error, questionnaire design flaws and processing errors can all introduce errors to the overall survey (note that these errors are independent of the sampling process and would have occurred even if a census had been conducted).

(NB: This component of the total survey error, can often be significantly higher than the sampling error, so every care should be taken to undertake correct survey procedures, to minimise this impact.)

TIP

The important thing is to minimise these 'non-sampling' errors and ensure that errors that introduce a systematic bias to the survey results are avoided.

... ExerciseS ...

Using the information below, calculate the following:

(a) Standard error for the mean annual household income
(b) Standard error for the total annual household expenditure
(c) Standard error for the proportion of households with annual household income > $10,000
(d) The equivalent RSE for each of these 3 estimates

(nb: assume that the sample was selected using Simple Random Sampling)

- Number of households in the population (N): 123,653
- Number of households in the sample (n): 3,425
- Estimated average household income = $8,932
- Estimated total household expenditure = $970,552,397
- Sample variance (S^2) for annual household income: 645,657,234
- Sample variance (S^2) for annual household expenditure: 553,265,867
- Proportion of households in the sample with annual household income > $10,000: 0.32

Excel – standard errors

When calculating standard errors you have to take into account the sample design used. The worksheet method here is applicable for a **simple random sample** (that is, that the sample households were randomly selected from a list of households).

If your sample was selected by 'clusters', with blocks of households selected and all households within these blocks being selected, this 'clusters' the sample selection and increases the sample error of the results. It is more complicated to calculate the standard error from a 'cluster' sample and, unfortunately, the effect of such clustering is not well known.

Even in developed countries such as Australia the effects of clusters are not known for certain surveys (like household income and expenditure). However, it is suspected that the effect on the standard error is an increase between 10% and 30%. If you are calculating standard errors from a cluster sample contact the Statistics Programme at the SPC for technical assistance.

Remember that sampling error is only one component of the total survey error. There are many ways that error can be introduced to surveys other than the use of sampling methods. Field enumeration error, respondent error, questionnaire design flaws and processing errors can all introduce errors to the overall survey (note that these errors are independent of the sampling process and would have occurred even if a census had been conducted). The important thing is to minimise these 'non-sampling' errors and ensure that errors that introduce a systematic bias to the survey results are avoided. In the analysis of the results there was no evidence of such systematic bias being introduced to the survey results, but this does not mean that this did not occur, and was hidden from the analysis phase of the project.

Setting up a worksheet to calculate standard errors

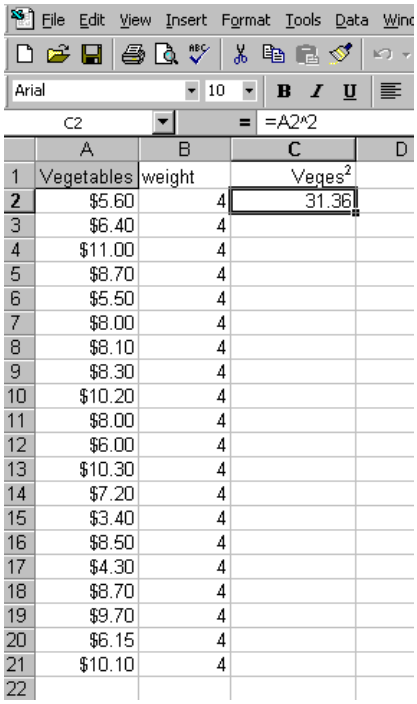
Basically you set up a worksheet which calculates the different parts of the Standard Error formula – so from your data you find the values for x_i^2 , Σx_i, Σx_i^2 , $\bar{x}$, etc. It can be complicated to set up the worksheet, so for any advice contact the Statistics Programme at the SPC. As well as the variable you are calculating the standard error estimates for, you will also need the sample weights (to make the estimate of *N*).

WARNING!

The example data used here is limited to one variable with 20 records and equal weights. You would work through the same steps if you had 200 variables and 60,000 records with the same weight. If, for example, you have different weights for different regions, you would calculate the SE for each region SEPARATELY.

1. The first thing is to **square the observations**. In this example, the variable is called Veges, and is expenditure on fruit and vegetables over a two-week period.

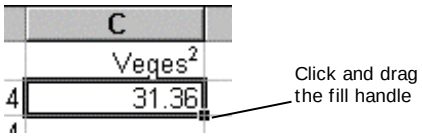
   The formula to square a value is: **=cell reference ^ 2 (enter)**. If you were calculating the standard error for more than one variable, you could use another worksheet in the same workbook to square the values, but here it is beside the 'test' data.



	A	B	C	D
1	Vegetables	weight	Veges^2	
2	$5.60	4	31.36	
3	$6.40	4		
4	$11.00	4		
5	$8.70	4		
6	$5.50	4		
7	$8.00	4		
8	$8.10	4		
9	$8.30	4		
10	$10.20	4		
11	$8.00	4		
12	$6.00	4		
13	$10.30	4		
14	$7.20	4		
15	$3.40	4		
16	$8.50	4		
17	$4.30	4		
18	$8.70	4		
19	$9.70	4		
20	$6.15	4		
21	$10.10	4		
22				

2. You copy the formula down the rest of the page by clicking and dragging on the 'fill handle' at the bottom right corner of the cell with the formula in it.

   Click and drag on the fill handle to copy the formula down the column:



3. **Calculate Σx_i^2:** type in a label for the calculation to sum the squared values in a cell below your data. The formula to sum values is: **=sum(cell ref) (enter).** Here the cell reference is C2:C21. Your screen should look like this:

C23 = =SUM(C2:C21)

	A	B	C	D
1	Vegetables	weight	Veges^2	
4	$11.00	4	121	
5	$8.70	4	75.69	
6	$5.50	4	30.25	
7	$8.00	4	64	
8	$8.10	4	65.61	
9	$8.30	4	68.89	
10	$10.20	4	104.04	
11	$8.00	4	64	
12	$6.00	4	36	
13	$10.30	4	106.09	
14	$7.20	4	51.84	
15	$3.40	4	11.56	
16	$8.50	4	72.25	
17	$4.30	4	18.49	
18	$8.70	4	75.69	
19	$9.70	4	94.09	
20	$6.15	4	37.8225	
21	$10.10	4	102.01	
22				
23	Sum Xi squared		1271.64	
24	Sum Xi			

4. **Calculate Σx_i:** type in a label for the calculation to sum the values in a cell below your data. The formula to sum values is: **=sum(cell ref) (enter).** Here the cell reference is A2:A21. Your screen should look like this:

C24 = =SUM(A2:A21)

	A	B	C	D
1	Vegetables	weight	Veges^2	
5	$8.70	4	75.69	
6	$5.50	4	30.25	
7	$8.00	4	64	
8	$8.10	4	65.61	
9	$8.30	4	68.89	
10	$10.20	4	104.04	
11	$8.00	4	64	
12	$6.00	4	36	
13	$10.30	4	106.09	
14	$7.20	4	51.84	
15	$3.40	4	11.56	
16	$8.50	4	72.25	
17	$4.30	4	18.49	
18	$8.70	4	75.69	
19	$9.70	4	94.09	
20	$6.15	4	37.8225	
21	$10.10	4	102.01	
22				
23	Sum Xi squared		1271.64	
24	Sum Xi		154.15	
25	n (sample size)			

5. **Find *n* and *N* (the estimated population size):** find *n* by counting the number of observations in the sample using the formula **= count (A2:A21)**. Find *N* by summing the weights using the formula = sum(B2:B21). Your worksheet should now look like this:

C25 | =COUNT(A2:A21)

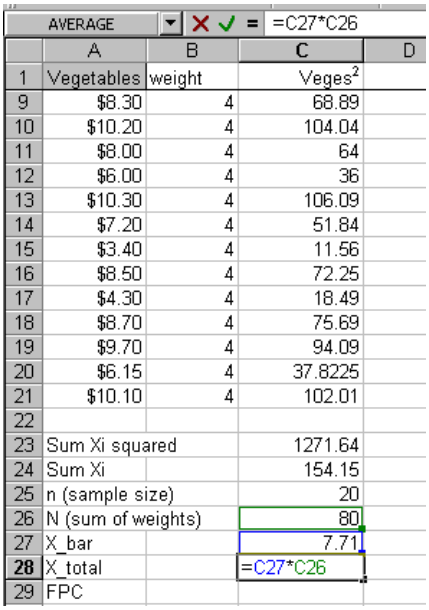
	A	B	C	D
1	Vegetables	weight	Veges^2	
6	$5.50	4	30.25	
7	$8.00	4	64	
8	$8.10	4	65.61	
9	$8.30	4	68.89	
10	$10.20	4	104.04	
11	$8.00	4	64	
12	$6.00	4	36	
13	$10.30	4	106.09	
14	$7.20	4	51.84	
15	$3.40	4	11.56	
16	$8.50	4	72.25	
17	$4.30	4	18.49	
18	$8.70	4	75.69	
19	$9.70	4	94.09	
20	$6.15	4	37.8225	
21	$10.10	4	102.01	
22				
23	Sum Xi squared		1271.64	
24	Sum Xi		154.15	
25	n (sample size)		20	
26	N (sum of weights)			

6. **Calculate the sample mean $\bar{x}$:** find the arithmetic mean of the sample data using the Excel formula: **=C24/C25**. You divide the sum of the observations by the count of the observations to give you $\bar{x}$ - the mean of the sample. Your worksheet should now look like this:

AVERAGE | =C24/C25

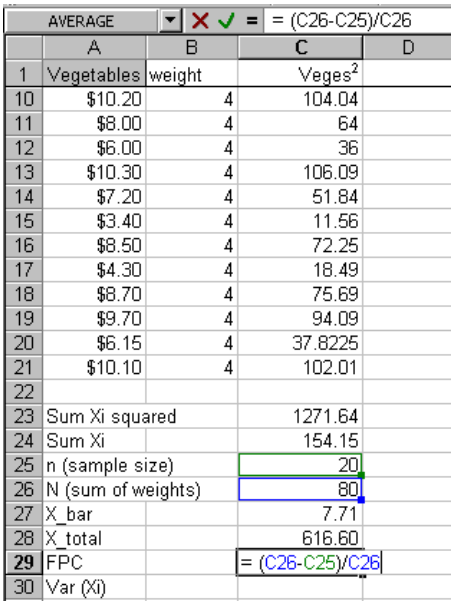
	A	B	C	D
1	Vegetables	weight	Veges^2	
8	$8.10	4	65.61	
9	$8.30	4	68.89	
10	$10.20	4	104.04	
11	$8.00	4	64	
12	$6.00	4	36	
13	$10.30	4	106.09	
14	$7.20	4	51.84	
15	$3.40	4	11.56	
16	$8.50	4	72.25	
17	$4.30	4	18.49	
18	$8.70	4	75.69	
19	$9.70	4	94.09	
20	$6.15	4	37.8225	
21	$10.10	4	102.01	
22				
23	Sum Xi squared		1271.64	
24	Sum Xi		154.15	
25	n (sample size)		20	
26	N (sum of weights)		80	
27	X_bar		=C24/C25	
28	X_total			

7. **Calculate the estimate of the total for the population:** find the estimated population total by multiplying the sample mean by the sum of the weights. Enter the Excel formula **= (C26*C27)**. Your worksheet should now look like this:



AVERAGE	=C27*C26		
	A	B	C	D
1	Vegetables	weight	Veges^2	
9	$8.30	4	68.89	
10	$10.20	4	104.04	
11	$8.00	4	64	
12	$6.00	4	36	
13	$10.30	4	106.09	
14	$7.20	4	51.84	
15	$3.40	4	11.56	
16	$8.50	4	72.25	
17	$4.30	4	18.49	
18	$8.70	4	75.69	
19	$9.70	4	94.09	
20	$6.15	4	37.8225	
21	$10.10	4	102.01	
22				
23	Sum Xi squared		1271.64	
24	Sum Xi		154.15	
25	n (sample size)		20	
26	N (sum of weights)		80	
27	X_bar		7.71	
28	X_total		=C27*C26	
29	FPC			

8. **Calculate the Finite Population Correction (FPC) factor:** the variance calculation is multiplied by this factor to compensate for the sample data. You do not have to multiply by the FPC. The formula for the FPC = $N - n/N$. Enter the Excel formula **= (C26 – C25)/C26**. Your worksheet should now look like this:

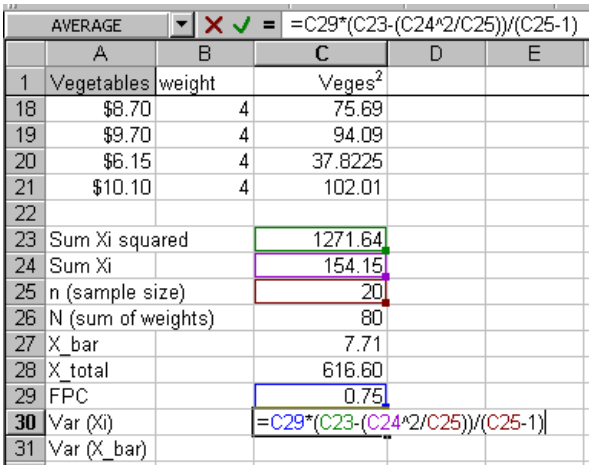


AVERAGE	= (C26-C25)/C26		
	A	B	C	D
1	Vegetables	weight	Veges^2	
10	$10.20	4	104.04	
11	$8.00	4	64	
12	$6.00	4	36	
13	$10.30	4	106.09	
14	$7.20	4	51.84	
15	$3.40	4	11.56	
16	$8.50	4	72.25	
17	$4.30	4	18.49	
18	$8.70	4	75.69	
19	$9.70	4	94.09	
20	$6.15	4	37.8225	
21	$10.10	4	102.01	
22				
23	Sum Xi squared		1271.64	
24	Sum Xi		154.15	
25	n (sample size)		20	
26	N (sum of weights)		80	
27	X_bar		7.71	
28	X_total		616.60	
29	FPC		= (C26-C25)/C26	
30	Var (Xi)			

9. Calculate the sample variance: you now have all the values required to calculate the sample variance

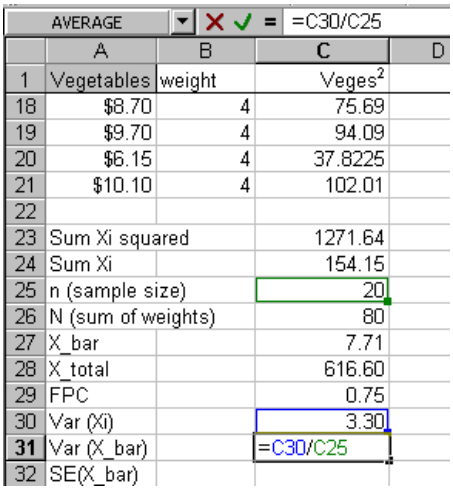
using the formula $FPC * \frac{(\sum x_i^2) - \frac{(\sum x_i)^2}{n}}{n-1}$.

Enter the formula in Excel **=C29*(C23-(C24^2/C25))/(C25-1)**. Your worksheet should now look like this:



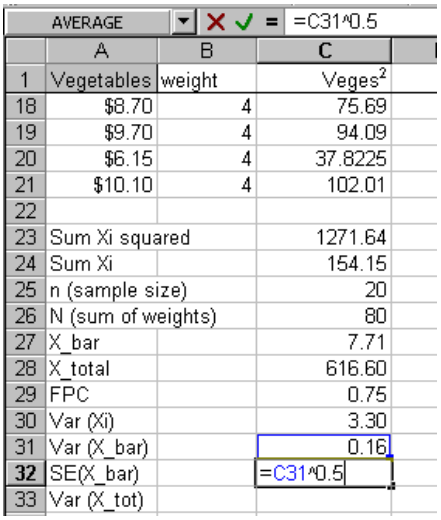
	A	B	C	D	E
1	Vegetables	weight	Veges^2		
18	$8.70	4	75.69		
19	$9.70	4	94.09		
20	$6.15	4	37.8225		
21	$10.10	4	102.01		
22					
23	Sum Xi squared		1271.64		
24	Sum Xi		154.15		
25	n (sample size)		20		
26	N (sum of weights)		80		
27	X_bar		7.71		
28	X_total		616.60		
29	FPC		0.75		
30	Var (Xi)		=C29*(C23-(C24^2/C25))/(C25-1)		
31	Var (X_bar)				

10. Calculate the variance of the sample mean: this is the result of the dividing the sample variance by *n* – the sample size. Enter the formula in Excel **= C30/C25**. Your worksheet should now look like this:



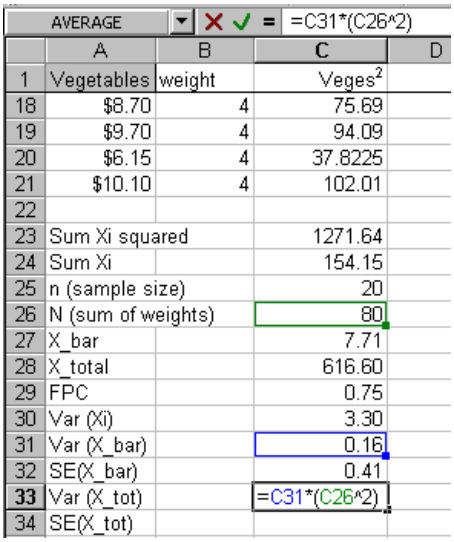
	A	B	C	D
1	Vegetables	weight	Veges^2	
18	$8.70	4	75.69	
19	$9.70	4	94.09	
20	$6.15	4	37.8225	
21	$10.10	4	102.01	
22				
23	Sum Xi squared		1271.64	
24	Sum Xi		154.15	
25	n (sample size)		20	
26	N (sum of weights)		80	
27	X_bar		7.71	
28	X_total		616.60	
29	FPC		0.75	
30	Var (Xi)		3.30	
31	Var (X_bar)		=C30/C25	
32	SE(X_bar)			

11. Calculate the standard error (SE) of the sample mean: the standard error is the square root of the sample variance. Enter the formula in Excel **=C31^0.5**. This is the same as typing =sqrt(C31). Your worksheet should now look like this:



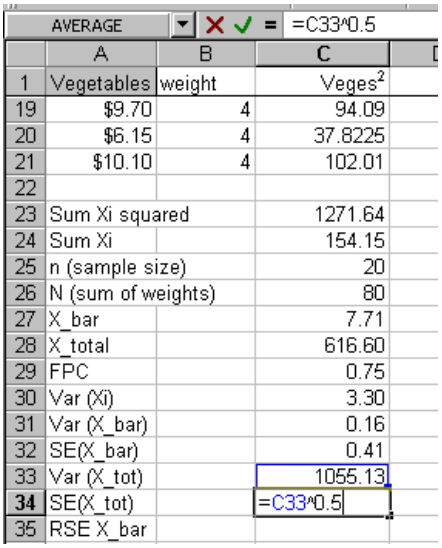
	A	B	C
1	Vegetables	weight	Veges^2
18	$8.70	4	75.69
19	$9.70	4	94.09
20	$6.15	4	37.8225
21	$10.10	4	102.01
22			
23	Sum Xi squared		1271.64
24	Sum Xi		154.15
25	n (sample size)		20
26	N (sum of weights)		80
27	X_bar		7.71
28	X_total		616.60
29	FPC		0.75
30	Var (Xi)		3.30
31	Var (X_bar)		0.16
32	SE(X_bar)		=C31^0.5
33	Var (X_tot)		

12. Calculate the variance of the estimated population total: the sample estimate multiplied by the estimated population squared (N^2). Enter the formula in Excel **=C31*(C26 ^2)**.



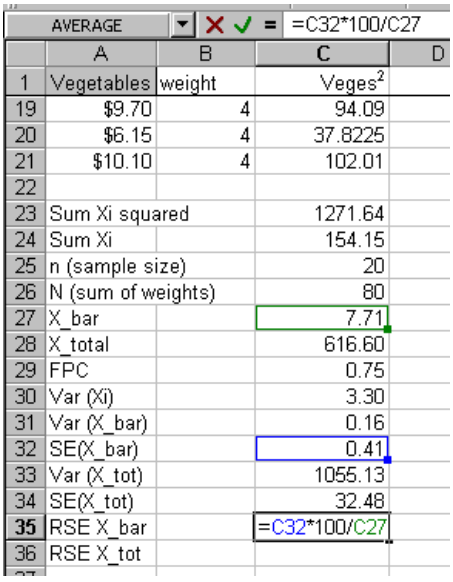
	A	B	C	D
1	Vegetables	weight	Veges^2	
18	$8.70	4	75.69	
19	$9.70	4	94.09	
20	$6.15	4	37.8225	
21	$10.10	4	102.01	
22				
23	Sum Xi squared		1271.64	
24	Sum Xi		154.15	
25	n (sample size)		20	
26	N (sum of weights)		80	
27	X_bar		7.71	
28	X_total		616.60	
29	FPC		0.75	
30	Var (Xi)		3.30	
31	Var (X_bar)		0.16	
32	SE(X_bar)		0.41	
33	Var (X_tot)		=C31*(C26^2)	
34	SE(X_tot)			

13. **Calculate the standard error (SE) of the estimated population total:** the standard error is the square root of the estimated population variance. Enter the formula in Excel **= C33 ^ 0.5**. This is the same as typing =sqrt(C33). Your worksheet should now look like this:



	A	B	C
1	Vegetables	weight	Veges^2
19	$9.70	4	94.09
20	$6.15	4	37.8225
21	$10.10	4	102.01
22			
23	Sum Xi squared		1271.64
24	Sum Xi		154.15
25	n (sample size)		20
26	N (sum of weights)		80
27	X_bar		7.71
28	X_total		616.60
29	FPC		0.75
30	Var (Xi)		3.30
31	Var (X_bar)		0.16
32	SE(X_bar)		0.41
33	Var (X_tot)		1055.13
34	SE(X_tot)		=C33^0.5
35	RSE X_bar		

14. **Calculate the relative standard error (RSE) of the sample mean:** the Relative Standard Error Percent is commonly used as a measure of reliability which can be compared across estimates and across surveys. It is the Standard Error of the mean divided by the sample mean * 100. Enter the formula in Excel **= C32 * 100 /C27**. Your worksheet should now look like this:



	A	B	C	D
1	Vegetables	weight	Veges^2	
19	$9.70	4	94.09	
20	$6.15	4	37.8225	
21	$10.10	4	102.01	
22				
23	Sum Xi squared		1271.64	
24	Sum Xi		154.15	
25	n (sample size)		20	
26	N (sum of weights)		80	
27	X_bar		7.71	
28	X_total		616.60	
29	FPC		0.75	
30	Var (Xi)		3.30	
31	Var (X_bar)		0.16	
32	SE(X_bar)		0.41	
33	Var (X_tot)		1055.13	
34	SE(X_tot)		32.48	
35	RSE X_bar		=C32*100/C27	
36	RSE X_tot			

15. **Calculate the relative standard error (RSE) of the estimated population total:** you calculate this as a check. It equals the SE of the population total divided by the estimated total *100. **The RSE X_bar should always equal the RSE X_Total**. If the two are NOT the same there is an error somewhere in your formula and you have to go back and find it. Enter the formula in Excel **=C34*100/C28**. Your worksheet should now look like this:

AVERAGE ✕ ✓ = =C34*100/C28

	A	B	C	D
1	Vegetables	weight	Veges^2	
18	$8.70	4	75.69	
19	$9.70	4	94.09	
20	$6.15	4	37.8225	
21	$10.10	4	102.01	
22				
23	Sum Xi squared		1271.64	
24	Sum Xi		154.15	
25	n (sample size)		20	
26	N (sum of weights)		80	
27	X_bar		7.71	
28	X_total		616.60	
29	FPC		0.75	
30	Var (Xi)		3.30	
31	Var (X_bar)		0.16	
32	SE(X_bar)		0.41	
33	Var (X_tot)		1055.13	
34	SE(X_tot)		32.48	
35	RSE X_bar		5.27	
36	RSE X_tot		=C34*100/C28	
37				

You have now calculated the Standard Errors and the Relative Standard Errors for your sample data.

WARNING!

If RSE X_bar DOES NOT EQUAL RSE X_Total there is an error in a formula and you have to go back and fix it.

16. **To calculate results when you have different weights:** often the data will have different weights. These weights may correspond to different regions or different parts of the population. These different components are often called strata (or stratum).

The overall total can be found by multiplying the sum of the x_i's (Sum Xi) by the weight for that stratum (the weight equals N/n) and then adding these components for all the different strata.

The overall mean is calculated by dividing the overall total by the population total (N).

To work out the overall variance for the total for this type of data it is necessary to work out the individual variances for each stratum. The individual variance of each total (Var X_tot) then have to be added together to get the overall variance. The square root of this number then gives the overall standard error (SE X_tot).

The variance of the overall X_bar can be found by dividing Var X_tot by the population total squared ($N*N$).