CHAPTER 4: MEASURING SPREAD

1 Mean, median and mode

Means, medians and modes provide a one number summary of a set of data, but they do not tell us anything about the distribution of the data. Suppose that five students sit three exams in three different subjects, and the marks are as follows: Table 1

But look carefully at the number of marks out of ten

Table 1 Exam marks

Subject	Marks out of ten	Mean Average	Median average
French	2, 4, 5, 7, 7	5	5
Religious Studies	0, 5, 10, 7, 3	5	5
History	5, 5, 4, 6, 5	5	5

each student received. We can see that the distribution around the mean is different.

By putting the marks into dotplots we can see that the marks for Religious Studies are very spread out and the marks for History are very close together.

2 The standard deviation

The five figure summary and box plots provide a very useful way for summarising a large data set numerically and graphically. A more commonly used measure of spread is called the standard deviation. The standard deviation uses the mean rather than the median as its central point and provides a one number summary of how spread out the data is. In this book we use the abbreviation SD for standard deviation, but you may see it abbreviated to s or STDV.

Figure 1 Dotplot for Religious Studies 1

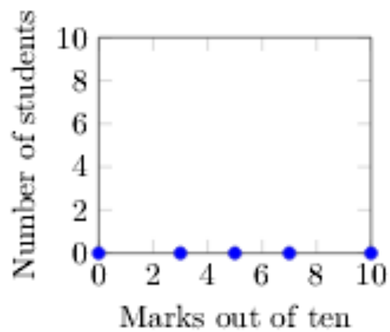


Figure 2 Dotplot for History

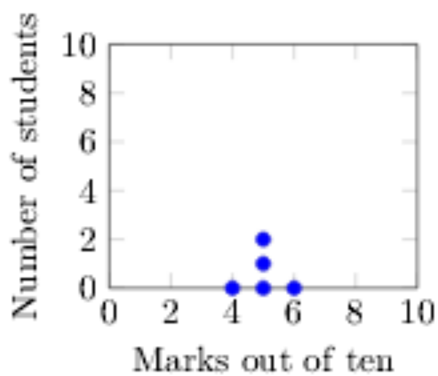


Figure 3 Dotplot for French

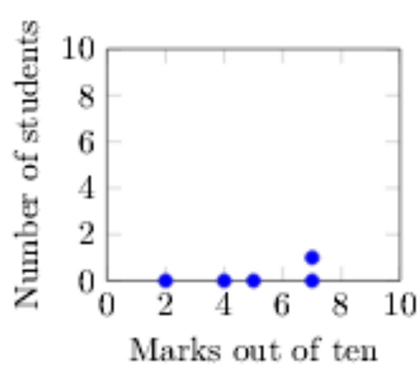


Table 2 Calculating the Standard Deviation

Observation. Each piece of data is called an observation. We write the age of each child in this column. We will call this observation	Deviation. The deviation is the difference between the observation and mean (which in this case is 8). We take the observation then subtract the mean. This means that some of our numbers will be negative (minus) numbers.	Squared deviation. We take the deviation then square it (by multiplying it by itself). These numbers are always positive as a minus number multiplied by a minus number is always a positive number.
4	4-8=-4	$-4\times-4=-4^2=16$
5	5-8=-3	$-3\times-3=-3^2=9$
8	8-8=0	$0\times0=0^2=0$
11	11-8=3	$3\times3=3^2=9$
12	12-8=4	$4\times4=4^2=16$

Example 1: Suppose a couple have five children aged 4, 5, 8, 11 and 15.

To find out the standard deviation we first need to find the mean age of the children.

$$\frac{4+5+8+11+15}{5} = 8\,years$$

The mean average is usually written as $\bar{x}$

Now we need to calculate the variance. The variance is a measure of how spread out the data are. This is best done by drawing a table:

2.1 Calculating the Standard Deviation

Add the squares of the deviation together to calculate

$$Variance = \frac{sum\ of\ the\ squared\ deviations}{number\ of\ observations}$$

the sum of the squares of the deviation.

16+9+0+9+16=50

Now we have the numbers we need to calculate the variance.

$$\frac{50}{5} = 10$$

The standard deviation is the square root of the variance

$$\sqrt{10} = 3.16$$

Standard deviation (SD) = 3.16 years

2.2 Interpreting the standard deviation

The SD is always given in the same units as your observations. In this case we have used the age of five children in years so the SD is also in years.

If we were examining heights in centimetres for our observations then the mean and SD would also be in centimetres. If we were examining weight in tonnes, the SD would be in tonnes. If we were examining bushels of wheat, the SD would be in bushels of wheat.

So what does a SD of 3.16 years actually tell us?

What we can know from this figure is that around 68% of the observations will be within one standard deviation of the mean. We will discuss where this 68% comes from in a later section. In other words we expect that 68% of the children will be been the age of 8 (the mean) and plus AND minus 3.16 years:

Therefore 8 (the mean) + 3.16 (the SD) = 11.16 years

AND 8 (the mean) -3.16 (the SD) = 4.84 years

Therefore 68% of the children will be between the ages of 4.84 and 11.16 years.

As we stated before the SD gives us an idea of how spread out our data is.

Another couple with five children who an average of eight years old, but their children are 8-year old quintuplets.

Like our first couple their children still have a mean age of 8, but their children are all exactly the same age.

If we calculate the SD we will see that the SD equals zero.

This is because there is no variation in their ages.

Figure 2 The Canadian Dionne quintuplets (born 1934), aged about 4. Mean =4, Median=4, Standard Deviation=0. A special Act of the Ontario Legislature, The Dionne Quintuplets' Guardianship Act, 1935 was passed to allow the Ontario Government to take them away from their parents and exhibited as a tourist attraction.

Example 3:

Suppose a third couple have children age 1, 2, 3, 14 and 20. Again this couple have five children with an average age of eight years but they have SD of 7.52 as their children are more spread out.

All three couples have five children with a mean age of 8, but the standard deviations are different. The SD gives us an idea of the spread of the children's ages.

2.3 Stem and leaf plots

A stem and leaf plot is not a graph as such but is a good way of checking out the shape of the distribution of a small sample. The stem and leaf plot can done by hand fairly quickly.
Table 4 records the age of death of 39 people buried in Accrington in 1839.

2.4 The Normal Distribution

If a dataset is normally distributed the mean, median and mode will coincide. Most of the observations will be near this central point with a smaller number far away from the central point.

Imagine a crowd of people and consider their heights. Most people seem to be similar height, give or take a few centimetres. A small number of the people are clearly shorter or taller than the average and an even smaller number of people seem to be very short or very tall.

Figure 3 Heights of Bavarian conscripts (1810-1840)

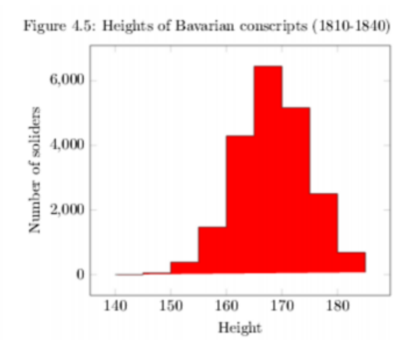


Figure 3 shows the heights of 5000 Bavarian men, conscripted between 1810 and 1840. [1] We can see from Figure 3 (called a histogram) that there were a small number of conscripts who were particularly short (less than 155 cm tall) and a small number who were very tall (taller than 180cm) but most of the conscripts were between around 160 and 172 cm.

So how do we describe the shape of this distribution?

This shape is often called the bell curve due to its resemblance to the shape of a bell. As the dataset is so large I have used a data analysis pack Minitab to calculate the mean and standard deviation of the sample.

The standard deviation is 6.439*cm* (we will say 6.4cm for convenience).

Creating a Stem and Leaf plot

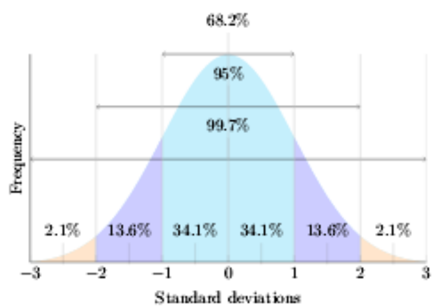
8	20	63	78	32
17	82	1	22	0
0	22	80	73	47
0	38	12	35	25
57	1	0	0	0
21	39	3	0	15
54	48	29	41	1
2	13	0	1	

0	8 0 0 2 1 1 0 3 0 0 0 1 0	Put numbers 0 to 9 in this row
1	1 2 3 7	Put the second number of 10-19 in this row
2	0 1 2 2 2 9	Put the second number of 20-29 in this row
3	5 8 9	Put the second number of 30-39 in this row
4	4 7 1 8	Put the second number of 40-49 in this row
5	7 5 5	Put the second number of 50-59 in this row
6	3	Put the second number of 60-69 in this row
7	8 3	Put the second number of 70-79 in this row
8	2 0 0	Put the second number of 80-59 in this row

STEM	LEAF
0	0 0 0 0 0 0 0 1 1 1 2 3 8
1	1 2 3 7
2	0 1 2 2 2 9
3	5 8 9
4	1 4 7 8
5	7 5 5
6	3
7	3 8
8	0 0 2

Figure 4 The normal distribution

Figure 4.6: The normal distribution



The mean is

$166.8cm$

So what are the heights for the conscripts 1 standard deviation (1 SD) from the mean?

As before we take our mean of

$168.8cm$

then add the SD.

$168.8cm+6.4cm=175.2$

Then we take our mean again and subtract the SD

$168.8cm-6.4cm=162.4$

Therefore we can state that the conscripts with heights of between 162.4 cm and 175.2 cm are within one standard deviation (1SD) of the mean. Approximately 68% of data is within 1 SD of the mean; this is a rule about the standard deviation.

Now we have calculated SD we can also identify the heights of conscripts within 2 standard deviations (2SD) of the mean. This will take into account the heights of conscripts who were taller or shorter than those within 1SD of the mean.

To find the extent of the second deviation we simply add or subtract the standard deviation twice instead of once.

$168.8cm+6.4cm+6.4cm=181.9$

$168.8cm-6.4cm-6.4cm=156.0$

Therefore we can say that conscripts between 156.0 cm and 181.9 cm are within 2 standard deviations (2SD) from the mean. Approximately 95% of data lies within 2SD of the mean.

We can keep on going to calculate the heights of conscripts within 3 standard deviations (3SD) of the mean.

$168.8cm+6.4+6.4+6.4=188.3cm$

$168.8cm-6.4-6.4-6.4=149.6cm$

Therefore we can say that conscripts between 149.6 cm and 188.3 cm are within 3 standard deviations (3SD) from the mean. Approximately 99.7% of data lies within 3SD of the mean.

We can see at this point that only 0.3% of conscripts remain outside 3 standard deviations from the mean (0.15% of whom are very very short and 0.15% are very very tall). We can keep on adding standard deviations with smaller and smaller percentages of conscripts, but by now we have a good idea of the distribution of our dataset.

3 Non-normal distributions

Not all samples are normally distributed. In the case of the burials in Accrington or the farms on Chile we can see that the mean, median and mode do not coincide; they are not even close to one another. Many of the statistical tests in Part 2 work on the assumption that the data is normally distributed– these are known as parametric tests meaning the tests are based on assumptions that the dataset lies within the parameters or expected pattern of the normal distribution. If the data is not normally distributed then these parametric tests will not be reliable. Non-parametric tests are those which do not assume a normal distribution. This will be noted in each section of the book.

3.1 Skewness

Skewness is a very important concept in dealing with social data. Normal distributions are common in natural phenomena such as the distribution of people's heights. However data from the human social world is often not normally distributed. As we have seen with the examples of the Chilean farms only four of the 24 are larger than the mean average.

In a normally distributed sample the mode, median and mean will coincide, meaning that the same number of observation will be below the mean as above the mean. In non-normal distributions, this is not the case. Non-normal distributions are particularly common in the case of income. A large number of people have below average incomes and a small number of people have very high incomes. This results in a mean which is misleading. Any report of income which states only an average should be treated with suspicion. In the above examples it is quite easy to see that the samples are skewed. The skewness test enables us to identify whether or not a distribution is skewed (not always as easy to spot as in the examples here), the size of the skew (is it near normal or far from normal) and the direction of the skew, (Is it positively skewed or negatively skewed?). A distribution which is perfectly normally distributed will have a skewness of zero. A positive skew (a number greater than zero) will occur where most of the observations are less than the mean and a negative skew (a number of less than zero) means that most of the observations are greater than the mean.

3.2 Calculating skewness

To calculate skewness we will use the same data as for calculating the standard deviation of the children's ages in Section 1. The SD for the children's ages in Example 1 was 3.16. We will need this to calculation skewness. The equation of measuring skewness is

$$Skewness = \frac{\Sigma(x-\bar{x})^3}{(n-1)SD^3}$$

n= the number of observations
SD = The standard deviation
$(x-\bar{x})^3$ = Sum of Squares to the power of 3 (see Table 7)

$$\frac{0}{(5-1)\times 3.16^3} = 0$$

$$Skewness = 0$$

Interpreting the skewness

Our skewness is 0, but what does this mean? Bulmer suggests the following guidelines (see Table 8). [2]

So with a skew of 0 we can say our data is not skewed at all.

Kurtosis

Kurtosis is often neglected in statistics books; Kurtosis is a measure of the `peakiness' of the distribution. Some distributions have a sharp peak and others a flatter peak. A distribution can be normally distributed (has a measure of skewness close to zero), yet have a high Kurtosis. The equation for measuring kurtosis is

$$Kurtosis = \frac{\Sigma(x-\bar{x})^4}{(n-1)SD^4}$$

n = the number of observations
SD=The standard deviation
$(x-\bar{x})^4$ = The Sum of the Squares to the power of 4 (see Table 7.

$$\frac{674}{(5-1)\times 1.364} =$$

$$\frac{674}{4\times 99.71} =$$

$$\frac{674}{398.8} = 1.69$$

Table 7 Table for calculating skewness and kurtosis

Age of child	$x-\bar{x}$	$(x-\bar{x})^2$	$(x-\bar{x})^3$	$(x-\bar{x})^4$
4	-4	16	-64	256
5	-3	9	-27	81
8	0	0	0	0
11	3	9	27	81
12	4	16	64	256
Totals		50	0	674

Figure 7: Measures of skew and kurtosis are two similar tests which help us identify how biased a distribution is below or above the mean and the shape of the distribution

Figure 4.7: Distribution of skew and kurtosis

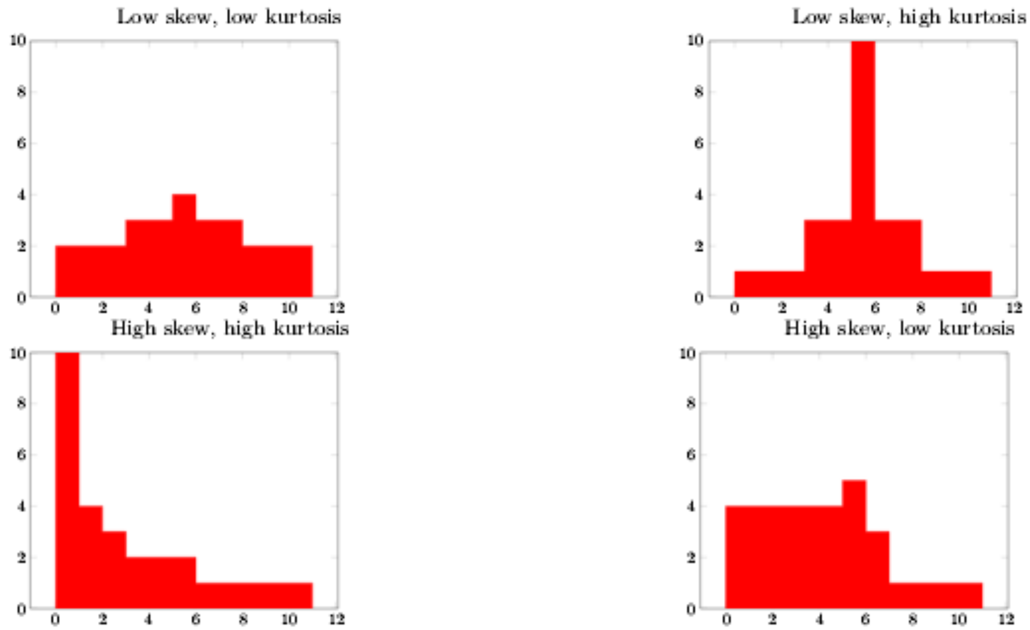


Table 8 Interpreting skewness

Description	Measure of skew
Highly skewed	More than 1 or less than -1
Moderately skewed	Between 0.5 and 1 or -0.5 and -1
Low skew	Between 0.5 and -0.5
No skew	Zero or approximately zero

Table 9: Interpreting kurtosis

Measure of kurtosis	Possible characteristics
Greater than 3	Leptokurtic distribution. `Peakier' and sharper than a normal distribution. Values concentrated around the mean and with high probability of extreme values
3	Mesokurtic. A normal distribution would have a kurtosis of 3.
Less than 3	Platykurtic distribution. Usually flatter than a normal distribution with a wider peak. The probability for extreme values is less than for a normal distribution. Values spread wider round the mean

Exercises

1. The following are results from a French speaking exam: 5, 6, 7, 10, 1, 2, 9, 8, 8
    1. Calculate the mean, median and mode.
    2. Calculate the standard deviation.
    3. Calculate the skew and kurtosis.
2. Draw a stem and leaf plot of the age at accession of English monarchs, Table 10

References

[1] Baton, J (1999) Heights of Bavarian male , 19th century, Data hub Heights and Biological Standard of Living Available from http://www.uni-tuebingen.de/uni/wwl/dhheight.html
[2] M G. Bulmer (1979) *Principles of statistics*. Dover}.
[3]These figures have been derived from those used by D. Ebdon (1985) *Statistics in Geography* (Oxford: Blackwell).
[4]Lawrence T. DeCarlo(1997) On the Meaning and Use of Kurtosis, *Psychological Methods* 2, pp.292-307 goes into some details about the complexities

Table 11: Top ten subsidised theatres (England) 2012-13: Source: Arts Council for England

Theatre name	Money each is receiving from ACE financial year 12-13, (£)
Royal National Theatre	17,462,920
Royal Shakespeare	15,675,270
Royal Exchange	2,318,609
English Stage Company (Royal Court Theatre)	2,297,916
Leicester Theatre Trust Ltd (Curve Theatre)	1,903,000
Birmingham Repertory Theatre	1,823,385
Young Vic	1,750,000
Liverpool Everyman and Playhouse	1,649,019
Chichester Festival Theatre	1,604,079
Northern Stage	1,551,976

Table 10 Age at Accession: English and (from 1603) British monarchs

Monarch	Age at Accession	Monarch	Age at Accession
King William IV	64	King Edmund II lronside	25
King Edward VII	59	Queen Elizabeth II	25
King George IV	57	King Henry V	25
King George I	54	Queen Elizabeth I	25
King James II	51	King Charles I	24
King Harold II	45	King Edward II	23
King George V	44	King Edred	22
King George II	43	King George III	22
King Edward VIII	41	King Harthacnut	21
King George VI	40	King Henry II	21
King Stephen	38	King Cnut (Canute)	21
King William I	38	King Harold I Harefoot	19
King William III and Queen Mary II	38	King Edward IV	18
King Edward The Confessor	38	King Edmund	18
Queen Mary I	37	Queen Victoria	18
Queen Anne	37	King Henry VIII	17
King James I	36	King Edgar	16
King Henry IV	33	King Edwy (Eadwig)	15
King Edward I	33	King Edward III	14
King John	32	King Edward V	12
King Henry I	31	King Edward The Martyr	12
King Richard I	31	King Richard II	10
King William II	31	King Aethelred II The Unready	9
King Richard III	30	King Edward VI	9
King Charles II	30	King Henry III	9
King Athelstan	29	King Henry VI	0
King Henry VII	28		

CHAPTER 5: SAMPLING

1 Sampling

It is rare that we are able to look at a set of data in its entirety. We do not have time to ask every Muslim in Britain to answer our questionnaire, read through the entire 1891 UK census or collect language use data from every single person who speaks English. When we are collecting data we need to take a **sample**. A sample is a part of something which gives an idea about the whole. We might buy a paint sample to get an idea of what the whole wall will look like when it has been painted. A carpet sample is a small piece of carpet which gives an idea of the colour, thickness and fluffiness of the carpet when it has been laid in a whole room.

2 Defining terms

2.1 Population

The entire group of objects about which information is wanted. See Table 1 for a list of examples.

Figure 1 A unit from a population of chickens

Table 1 Population for different research questions

What we want to find out about	Our population
Use of the Irish language amongst people living in Ireland in 1911	People living in Ireland in 1911
Use of the Irish language amongst people living in Castleffrench, Galway in 1911	People living in Castleffrench, Galway in 1911
Life expectancy of people in Accrington in the nineteenth century	People in Accrington in the nineteenth century
Heights of Bavarian soldiers in the nineteenth century	Bavarian soldiers in the nineteenth century
The occupations of people who sailed on the Titanic	People who sailed on the Titanic
Attitudes of school pupils learning French in Hampshire	School pupils learning French in Hampshire
Development of the use of the word 'noob' in the English language	A corpus of written English covering different years.
The cost of bags of wool in England in the 1500s	Bags of wool sold in England in the 1550s

2.2 Unit

Any individual member of the population. Depending on the population a unit could be:

- A person
- A bag of wool
- A cow or chicken
- A soldier
- An Irish speaking person

2.3 Sample

Part or sub-set of the population used to gain information about the whole. By definition a sample is not the whole population. Going back to the above examples a sample could be:

- Some, but not all, pupils studying French in Hampshire
- Some, but not all, people who sailed on the Titanic
- Some, but not all, people who lived in Accrington in the nineteenth century.
- Some, but not all, bags of wool sold in the 1500s.

2.4 Sampling frame

The source or list from which the sample is chosen. A sampling frame could be:

- The 1911 Census of Ireland
- A list of school pupils studying French in Hampshire
- The passenger list from the Titanic
- Records of wool sales in in England in the 1500s.

2.5 A variable

The characteristic of a unit, to be measured for all those units in the sample. We will be finding out specific information about the population of Ireland or the passengers on the Titanic.

Examples of variables might include:

- Sex
- Income
- Place of Birth
- Religion
- Language spoken
- The price of oats
- Number of windows
- Occupation
- Opinion on how enjoyable it is to study languages.

2.6 Census

Obtaining information about every unit in a population. The best known type of census are the population census when everybody who lives in particular country is asked questions. A census is when we look at every individual person, bag of wool, tax return, school pupil etc. By definition a census is not a sample.

3 Why we take a sample

When we take a sample we hope that our conclusions about the sample apply the whole population generally. When pollsters ask 1000 people how they plan to vote in an election their underlying purpose is to come to a conclusion about what the election result will be when everybody has voted. If we look at a sample of children learning French in Hampshire we are intending to draw conclusions about all children learning French in Hampshire, even if they were not included in the survey.

One of the most famous examples of sampling going wrong was the Chicago Tribune's headline"Dewey defeats Truman" in the 1948 US election. Sure of their sampling techniques journalists at the newspaper were confident enough to proclaim that Dewey had won. When all the votes at been counted they discovered that this was not the case. Their sample led them to believe that Dewey would win but it turned out that the electorate as a whole (the population) favoured Truman. Although this is well-known example of sampling going wrong the great thing about opinion polls of this kind is that we can see where the sampling went wrong (the sample suggested an outcome which turned out not be true). However when it comes to knowing whether the conclusions we draw from sampling bags of wool, Irish speakers or school pupils are representative of the population as a whole we are unlikely to get such quick feedback on our conclusions. So how do we go about taking a sample of our population?

4 A convenience sample

I used the example of Accrington in Chapter 3 because the Lancashire Parish Registers are available on the internet.[1] Of all the settlements in Lancashire I chose Accrington because it was near the top of the list which was in alphabetical order. 1838 was just a year that I guessed at. This is a pure convenience sample. I do not know whether the people of Accrington enjoyed a longer or shorter life than other residents of England or of Lancashire. I do not know whether or not 1838 was a year of a lot of deaths or few deaths. I do not know what diseases might have been going round Accrington in 1838 and what age of people would have been most effected. In no sense do I know if Accrington was representative of the country as a whole. To get an accurate picture of life expectancy in nineteenth century England we would need to get a lot more data. So why don't we just collect the ages of all the people who died in England between 1800 and 1899?

Put simply we do not have time to search through the records and to record the age of death of each person who died in England in the 100 years of the nineteenth century.

5 Random sampling

The advantage of selecting at random is that a random sample is free of researcher bias. If I was just picking and choosing from the 1911 Census of Ireland I might be inclined to choose individuals or households which I think might be interesting. I've just been searching the census for individuals over 100 years-old and came across a 104 year women called Ellen Hefferan working as a live —in domestic servant in County Wexford which is both unusual and interesting. I might conveniently choose to make her part of my sample.

You can take a random sample by giving each unit (person, household, bag, cow etc.) a number and use random numbers to select your sample. Traditionally you could have a used a printed page of random numbers or a random number generator on a scientific calculator. Today, however, the internet is your friend and www.random.org will generate random choices of all sorts including coin tosses, playing cards and numbers. This website even has a free list randomiser. For our Irish community example we can simply paste in the names then click the generate button and select the names at the top of the list for our sample.

6 Stratified sample

One of the problems with random samples is that some types of people (or communities or objects) might be underrepresented, overrepresented or not presented before in our sample. We might end up with a disproportionate number of men compared to women, urban residents to rural residents, English speakers to French speakers, Catholics to Protestants, or big towns to small towns. Stratified sampling ensures that our sample contains individual units (people, bags of wool, cows etc.) from different strata of the population. Suppose we want to take a sample of 100 people from the village of Castleffrench in Ireland with a population of 1000 people. We want to explore household size in the 1911 census. From the census we see that:

1. 5% of the population are Irish Speakers.
2. 2% are over 80
3. 48% are male and 52% female
4. 75% are Catholics and 25% are Protestants

We can use a stratified sample to ensure that people from each of these groups are included in the sample. A total random sample cannot guarantee that every group will be represented in the sample. We can see that just 2% are over 80. If we randomly select 100 people how many of those people will be over 80? 2% of 100 is 2 so we might say that we expect two people in the random sample to be over 80. We will explore randomness and probability later in this book, but can we be sure that we would get two people?The answer is that we can't be sure. We may get two, but we may get one or zero or six or ten. A stratified sample will ensure that we have some representatives from this group of people in our sample. If a strata of the population is small we might choose to oversample, that is to select more than 2% of the population over 80. We can then take a random sample with each section.

7 Important considerations when sampling

Samuel Johnston is reputed to have said "You don't have to eat the whole ox to know the meat is tough.". This is a good metaphor from sampling — we can draw a conclusion about the whole from a sample of the part. An ox is actually a good illustration for statistics. The toughness of the meat depends upon which part of the ox the cut came from. If our sample is a sirloin cut we will come to different conclusions about beef than if our sample comes from the shank. However carefully we select a sample there are certain issues which may arise: These are particularly common in questionnaire research, but they can appear in other situations too.

7.1 Oversampling

Units from outside the population are included These would need to be removed (if we know that they are there). For our survey of school pupils studying French in Hampshire we may find that some of our questionnaires have been filled in by pupils in Hampshire not studying French or pupils studying French in other counties.

7.2 Undersampling

Units from certain groups in the population are not included or are underrepresented. In the case of the Irish community we may find that no people over 80 are included if we took a random sample.

7.3 Non-response bias

A particular problem with questionnaires is that 100% response rates are very unusual. It is difficult to ascertain whether the people who did not answer your questionnaire have the same characteristics or opinions as the people who did respond. Even censuses have this problem to a certain extent — they are meant to survey the whole population, but some people are missed out or didn't respond for one reason or another.

7.4 Sampling frame error

A sample is only as good as the sampling frame from which the sample was derived. If the sampling frame is incomplete, inaccurate or excludes units which are present in the population you are researching then your conclusions are less likely to be true of the whole population. There were many shortcomings with polling in the 1948 "Dewey defeats Truman" election, but one was using the telephone for polling. Owners of telephones were more likely to support Dewey, whereas non-telephone owners were more likely to support Truman. However pollsters assumed that telephone owners were representative of the population as a whole, and called the election for Dewey. [2]

7.5 Voluntary response bias

People who volunteer to fill in your questionnaire or contact you to give their opinion are not necessarily representative of the whole population. If you ask people to give their views on the TV show Neighbours then the respondents are likely to be people who watch Neighbours. They might also be people who have time to respond to you or are particularly outspoken. They may not be representative of the population at large or of Neighbours fans in particular.

8 Exercises

1. A headteacher wishes to know whether or not parents of children at her school favour expanding the range of languages available at GCSE. She receives 33 letters of which 29 support an expansion of languages and four oppose. Can she assume that 87.9% of parents support expanding the range of languages? Why/ why not?
2. You are investigating the extent to which people who identify as 'Christians' believe in traditional doctrines such as the resurrection and the virgin birth. You have designed a questionnaire to find out. What sampling issues arise if you carry out your questionnaire:
    1. Face-to-face on the high street on a Saturday afternoon?
    2. Face-to-face outside a church on a Sunday morning?
    3. On your personal website?
3. You are researching life expectancy in Accrington for the whole of the nineteenth century.You don't have time to look at all burials in the town. How might you sample in order to get a picture of life expectancy over the course of the whole century?
4. You are doing a survey of television viewing habits in your neighbourhood. Who might be excluded from your survey if you carried it out:
    1. As a door-to-door survey on a Tuesday morning?
    2. As an online survey
    3. As a telephone survey on a Friday evening?

References

[1] Online Parish Clerks for the County of Lancashire http://www.lan-opc.org.uk/

[2] Polling errors in the Truman-Dewey election were not restricted to this one poll. See, for example Jeanne Curran and Susan R. Takata (2002)*Getting a Sample Isn't Always Easy* Online at http://www.csudh.edu/dearhabermas/sampling01.htm