18 Sampling

18.1 Random Samples

In this section we look at random samples and at the difference between a *census* and a *sample*.

In a census, information on every member of a population is considered. In the UK, a census is carried out every 10 years. The amount of work required to carry out a census means that it is an expensive process.

In a sample, a subset of the population is considered to try to obtain information about a particular problem or issue. Because a sample is normally much smaller than the whole population, it is quicker and easier to take and to analyse a sample than to carry out a census of the entire population. Sampling entails less effort and less expense. In some cases, it is essential to take a sample. For example, imagine a firm that uses quality control to test the light bulbs it manufactures, to see how long they last. (If every item was tested until it stopped working they would have no light bulbs left to sell!) They therefore take samples from the production and test these to see if the quality is up to standard.

In a *random sample*, every member of the population is *equally likely* to be included in the sample. One way of selecting a sample is to use random numbers, as demonstrated in the example below. You can find random numbers in books of statistical tables. You can also generate them using a calculator or a computer.

The diagram shows part of a table of random digits.

Random Digits

98859	09884	45275	09467	93026	32912
26604	95099	93751	00590	93060	64776
82984	65780	94428	30160	86023	52284
70888	14063	96700	83008	17579	71321
77803	61872	86245	68220	66267	01379
11304	01658	82404	46728	35228	49673
53552	51215	45611	83927	00772	99295

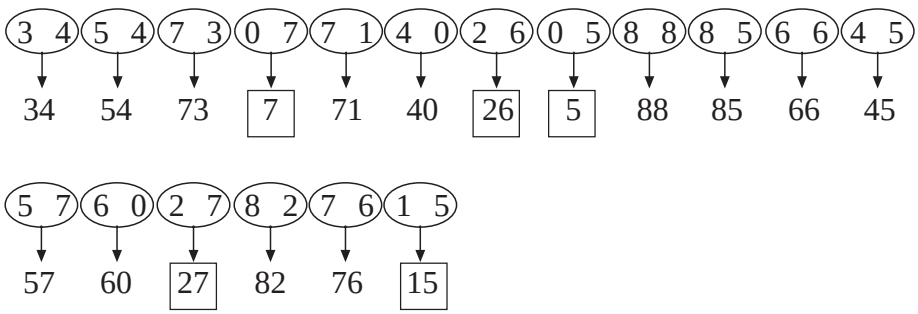
Example 1

In a class there are 30 pupils. The teacher decides to take a random sample of 5 pupils to estimate the mean height of the pupils in the class. Select a random sample of 5 pupils from the list.

1	Alan	10	Rachel	19	Sacha	28	Salif
2	Lucy	11	Ben	20	Halim	29	Annie
3	Tom	12	Emma	21	Daniella	30	Karen
4	Azar	13	Hannah	22	Joseph		
5	Jayne	14	Grace	23	Anna		
6	Nadima	15	Miles	24	Sophie		
7	Matthew	16	James	25	Kathryn		
8	Sushi	17	Joshua	26	Helen		
9	Mohammed	18	Lisa	27	Fatoumata		

Solution

To take a random sample you need to use a list of random digits, as follows:



The digits are taken in pairs to form 2-digit numbers, as shown above. All those numbers greater than 30 are discarded (as there are 30 pupils on the list). The process is continued until 5 different numbers between 1 and 30 have been obtained.

So the sample will be made up of the following pupils:

7	Matthew
26	Helen
5	Jayne
27	Fatoumata
15	Miles

Example 2

Discuss whether or not the following situations produce random samples.

(a) Mark is conducting a survey for a magazine. He stops people at random on a Saturday morning at his local shopping centre.

(b) Granny Taylor's National Lottery numbers.

Solution

(a) To produce a random sample, every member of the population must have an equal chance of being selected. In the case of Mark's sample, he is excluding people who are at work on that Saturday, as well as other people who haven't gone to that shopping centre. So although Mark is stopping people at random, he is not producing a random sample.

(b) Most people use personal reasons when they select their National Lottery numbers. Granny Taylor may, for example, have used the number of grandchildren that she has, their birthdays, the number of her house, etc. If that is the case then she has not selected a random sample. However, if she bought a 'Lucky Dip' from her local shop then the computerised National Lottery till should have produced a random sample.

Note: For small populations it is relatively easy to produce a random sample. Simply number every member of the population, write those numbers on pieces of paper and put them into a hat or tombola. Mix them well and ask someone to pick out as many numbers as you need for your sample, then take the corresponding items from the population. This process clearly becomes unmanageable when we investigate large populations, which is why we tend then to use random number generators.

Exercises

When using random number tables in the following questions, work from left to right along the top row of numbers, then similarly along subsequent rows.

1. Use the random digits below to select a second sample of 5 from the class in Example 1.

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

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

2. Use the table of random numbers shown at the start of this unit to select a sample of 10 pupils from the list in Example 1.

3. There are 10 competitors in an athletics event. Their names are:

Jimmy Jump	Harold Hammer	Tom Throw
Dick Discus	Harry Hop	Paul Putt
Sam Shot	Liam Long	Jake Javelin
Victor Vault		

   (a) Number the competitors from

   1 Jimmy Jump 2 Harold Hammer .....

   10 Victor Vault

(b) Use the following list of random digits to select a random sample of 3 of the competitors for drug testing.

26	60	49	50	99	93	75	10
05	90	93	06	06	47	76	82
98	46	57	80	94	42	83	01
60	86	02	35	22	84	70	88

4. A council wants to talk to the residents of a street to discuss a proposed traffic calming scheme. The houses in the street are numbered from 1 to 57.

   Use the list of random digits in question 1 to identify a random sample of 10 of the houses for the council to visit.

5. In another road the houses are numbered from 1 to 539.

   Use the list of random numbers in question 1 to identify a random sample of 6 houses for the council to visit.

6. A telephone directory has 250 pages. On each page there are 400 names.

   (a) Describe how you could use random numbers to select a random sample from the telephone book.

   (b) Explain why the sample is *not* a random sample from the whole population of the area.

7. The ages, in years, of the members of a computer club are listed below.

Dee	12	Max	16	Ollie	18
Denise	14	Nazir	15	James	11
Tom	16	Jane	17	Hannah	14
Holly	11	Ferdi	11	Gemma	13
Richard	15	Kim	14	Nadia	16
Jai	13	Grant	12	Hugh	14
Victor	13	Juliette	13	Ben	13
Peter	14	Nigel	14	Ali	15

   (a) Number the club members from 1 Dee, 2 Max, ... 24 Ali.

   (b) Use the list of random digits in question 1 to generate a random sample of 5 club members and calculate the mean age for your sample.

   (c) Use the list of random digits in question 1 *in reverse order* (i.e. 5 5 6 9, etc.) to generate a second random sample of 5 club members. Calculate the mean age for this new sample.

   (d) Compare the two samples and the two means.

8. (a) Describe the *advantages* of using a census rather than a sample.

   (b) Describe the *disadvantages* of using a census rather than a sample.

9. Mr May wants to know the mean IQ of the pupils in his class.
   Would you recommend that he uses a sample or a census?
   State which *you* would use, and explain why.

10. A large school has 1800 pupils. The headteacher wants to find out how far the pupils have to travel to school. Advise him whether to carry out a census or to use a sample. Explain why you give this advice.

18.2 Sampling Techniques

In this section we look at two further techniques for sampling: *systematic* sampling and *quota* sampling.

A systematic sample is taken by sampling at regular intervals.

A quota sample is when the different categories that make up the population are represented according to their proportion within the overall population.

A typical use of quota sampling is in opinion poll surveys where there is a need to reflect the way the population breaks down between the two genders, into different age groupings, into cultural and ethnic backgrounds, etc. The choice of people selected from each category is left to the person collecting the information.

We will use the class from Example 1 at the beginning of this unit to demonstrate how to use these sampling techniques. The names are listed again below:

1 Alan	10 Rachel	19 Sacha	28 Salif
2 Lucy	11 Ben	20 Halim	29 Annie
3 Tom	12 Emma	21 Daniella	30 Karen
4 Azar	13 Hannah	22 Joseph	
5 Jayne	14 Grace	23 Anna	
6 Nadima	15 Miles	24 Sophie	
7 Matthew	16 James	25 Kathryn	
8 Sushi	17 Joshua	26 Helen	
9 Mohammed	18 Lisa	27 Fatoumata	

Example 1

Select a systematic sample of size

(a) 6 (b) 5

from the class.

Solution

(a) As there are 30 pupils in the class, and we want a sample of 6, we calculate $30 \div 6 = 5$. We can then obtain our sample of size 6 by selecting every 5th member of the class.

This would give:

5	Jayne
10	Rachel
15	Miles
20	Halim
25	Kathryn
30	Karen

(b) For a sample of size 5 we could select every 6th member of the class:

6	Nadima
12	Emma
18	Lisa
24	Sophie
30	Karen

You can select a systematic sample of size 6 by choosing any of the first 5 items as a starting point and then taking every 5th item thereafter. For example, in part (a) we could have started with pupil number 2 (Lucy) and then selected pupils number 7, 12, 17, 22 and 27. This would have generated the sample

Lucy, Matthew, Emma, Joshua, Joseph, Fatoumata.

Similarly, starting with pupil number 4 (Azar) in part (b) would have generated the sample

Azar, Rachel, James, Joseph, Salif.

Example 2

Create a quota sample of size 10 from the class.

Solution

Note that the class contains 12 boys and 18 girls.

As the ratio of boys to girls is 2 to 3 in the class, they must be in the same ratio in the sample.
So for a sample of 10 we need 4 boys and 6 girls. These do not need to be chosen

at random, and so could be:

Alan	Lucy
Tom	Jayne
Azar	Nadima
Matthew	Sushi
	Rachel
	Emma

If the teacher is selecting the pupils then it is possible that they may choose their favourite 4 boys and 6 girls, which would introduce some bias to the sample. It is best to avoid this possibility by selecting 4 boys randomly from the group of 12 boys, likewise for the girls. This produces a *stratified random sample.* We will look at stratified random samples in more detail in section 18.3

Exercises

1. Select a systematic sample of size 10 from the class used in Example 1.

2. There are 400 trees in a plantation. All the trees have been planted in rows. Describe how to create a systematic sample of 25 trees.

3. The houses in a street are numbered from 1 to 340.
   Describe how to create a systematic sample of size 20.

4. A theatre group has 40 members of whom 15 are boys. A quota of size 8 is to be interviewed. How many girls and how many boys should be included in the sample?

5. Is a quota sample also a random sample?

6. Explain why a systematic sample is *not* a random sample.

7. A company employs the following numbers of staff in 3 categories:

Management	10
Technical	20
Administrative	20

   How many from each category should be included in a quota sample of size:

   (a) 10 (b) 5 (c) 25 ?

8. Refer to the members of the computer club in question 7 of Exercise 18.1. Create systematic samples of size:

   (a) 8 (b) 6 (c) 4

   from the list of members.

9. In a street the homes numbered 1 to 70 are houses and those numbered 71 to 90 are bungalows.

   (a) Which home numbers would you include in a systematic sample of size 18 ?

   (b) How many bungalows would you include in a quota sample of size 18 ?

   (c) Comment on the number of bungalows included in each sample.

10. For the homes described in question 9, use the random number generator on a calculator or computer to create a random sample of size 18.

    How does the number of bungalows included in *this* sample compare with the samples used in question 9 ?

18.3 Stratified Random Samples

A stratified random sample is similar to a quota sample. The key difference is that the representatives from each category are chosen at random. In the previous section we looked at a quota sample that was made up of 4 boys and 6 girls.

In the *quota sample* the first 4 boys from the list were chosen and then the first 6 girls.

In a *stratified random sample* the 4 boys should be selected randomly from the 12 boys in the class. In the same way, the 6 girls should be selected randomly from the 18 girls in the class.

Example 1

Take a stratified random sample of size 5 from the class considered in the earlier sections.

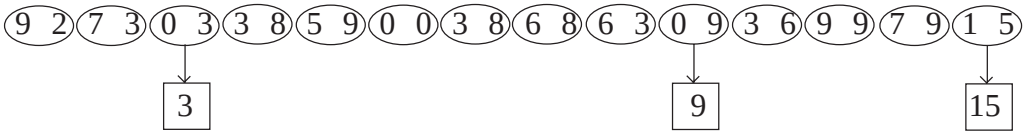
Solution

The class consists of 18 girls and 12 boys, 30 pupils altogether. We want a sample of size 5 so we need to include $\frac{5}{30} = \frac{1}{6}$ of the girls, and $\frac{1}{6}$ of the boys. This means that we need to choose 3 girls and 2 boys.

We choose a random sample of 3 girls. The girls in the class are:

1	Lucy	7	Hannah	13	Sophie
2	Jayne	8	Grace	14	Kathryn
3	Nadima	9	Lisa	15	Helen
4	Sushi	10	Sacha	16	Fatoumata
5	Rachel	11	Daniella	17	Annie
6	Emma	12	Anna	18	Karen

Using random digits gives:



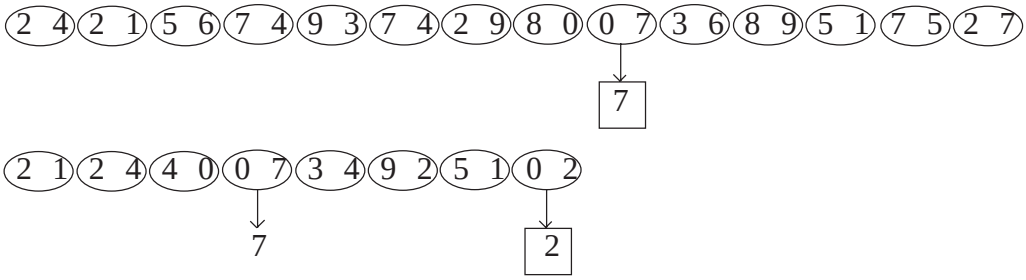
So the sample will contain:

3	Nadima
9	Lisa
15	Helen

The boys are:

1	Alan	5	Mohammed	9	Joshua
2	Tom	6	Ben	10	Halim
3	Azar	7	Miles	11	Joseph
4	Matthew	8	James	12	Salif

Using random digits gives:



Note that the second 7 is ignored.

So the sample will also contain:

7	Miles
2	Tom

The complete stratified random sample will therefore consist of

Nadima, Lisa, Helen, Miles and Tom.

Example 2

A company has a total of 360 employees in four different categories:

Managers	36
Drivers	54
Administrative Staff	90
Production Staff	180

How many from each category should be included in a stratified random sample of size 20 ?

Solution

To create a sample of size 20 we need $\frac{20}{360}$ or $\frac{1}{18}$ of the workforce. So we take this fraction of the number of employees in each category.

Managers	$\frac{1}{18} \times 36$	=	2
Drivers	$\frac{1}{18} \times 54$	=	3
Administrative Staff	$\frac{1}{18} \times 90$	=	5
Production Staff	$\frac{1}{18} \times 180$	=	10
	TOTAL	=	20

Exercises

1. Create a stratified random sample of size 10 for the class considered in the Examples.

2. A catering company employs the staff listed in the following table:

Delivery Drivers	12
Cooks	36
Cleaning Staff	4
Sales Staff	8

How many of each category of staff should be included in a stratified random sample of size 15 ?

3. A garage services VW and Audi cars. They want to carry out a customer satisfaction survey.

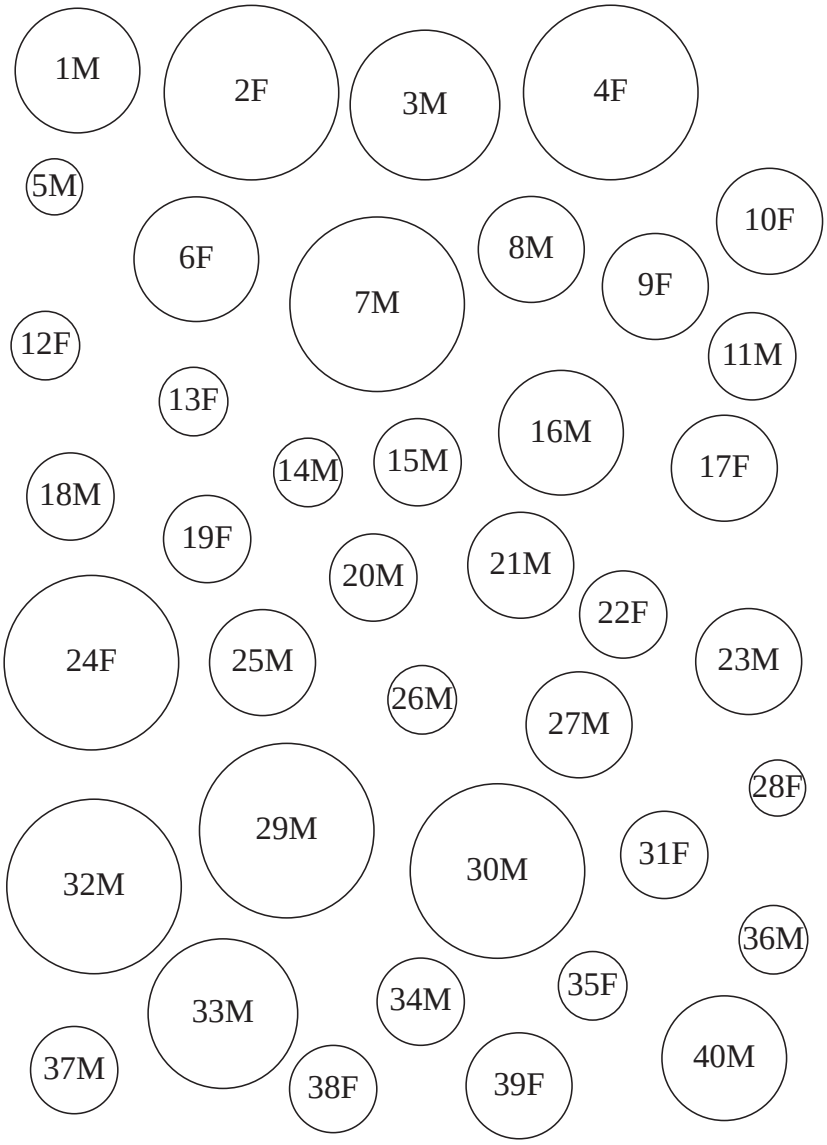
   (a) Explain why they might want to use a stratified random sample.

   (b) They have 2000 regular customers of whom 650 have Audi cars.
   Describe how to obtain a stratified random sample of size 40 of the customers to interview.

4. A farmer owns 120 Jersey cows and 180 Friesians. How many of each breed of cow should he include in a stratified random sample of 50 for a survey of milk quality?

5. A survey was carried out to determine the size people would prefer for a new coin. In the survey people were asked to select the size coin they preferred and indicate their sex by using the letter M or F.

   Forty people took part and the results are shown below.



(a) How many males and how many females would you include in a stratified random sample of size 10 ?

(b) Select a stratified random sample of size 10 and use it to estimate the mean diameter of the coins.

6. There are 1200 pupils in a school. Their methods of transport to school are given below.

Bus	720
Walk	310
Cycle	90
Car	80

How many pupils from each category should be included in a stratified random sample of size 50?

7. There are 1500 homes in one part of a city. These homes can be divided into 3 categories:

Privately owned homes	720
Homes rented from council	592
Homes rented from private landlords	188

The council is to carry out a survey into the quality of homes.

(a) What would be the advantage of using a stratified random sample?

(b) How many homes of each type should be included in a stratified random sample of size 200 ?

8. (a) Find out how each member of your class travels to school.

(b) Create a stratified random sample of size 10.

(c) Find the time that it takes each of the people in the sample to travel to school.

(d) Calculate the mean time for your sample.

(e) Compare your results with other people's samples.