Lesson 3

3.1 Random and Non-Random Samples

This course looks at three methods of taking a Random Sample from a population and two methods of taking a Non-Random Sample.

Each has advantages and disadvantages, plus the nature of the population itself or the situation may influence the method chosen of obtaining a suitable sample.

- **Simple** Random Sample
  Looked at in Exercise 2.2, Q2 where the calculator's RanInt button was used to randomly pick five pupils out of a class of twenty.
  Alternatively, draw tickets 'out of a hat'.
  Advantages : Free from bias
  Each item has a known, equal chance of being picked
  Cheap and easy for small populations and small samples
  Disadvantages : A sampling frame (a numbered list) is needed in advance
  Not practical if population or sample size is large

- **Systematic** Random sample
  The sample is formed by picking every tenth (for example) item from a list. In this example, the first element is picked by generating a random starting number between 1 and 10 using RanInt, then every tenth item after that.
  Advantages : Simple and quick to implement & therefore inexpensive
  Suitable for use with large populations or samples.
  Disadvantages : A sampling frame (a numbered list) is needed in advance
  Can introduce bias if the sampling frame has a structure

- **Stratified** Random Sample
  In a school, for example, there may be more boys than girls, say 600 boys and 400 girls. So that each gender is proportionally represented, a sample of ten would be arranged such that it comprised of 6 boys and 4 girls.
  Advantages : Sample mirrors the structure of the population
  Groups identified are proportionally represented in sample
  Disadvantages : A sampling frame is needed in advance
  Sorting out strata within the sampling frame is 'extra work'

- **Quota** Non-Random Sample

  A researcher decides in advance that, for example, he'd like to know the mean weight of the Lobsters and the mean weight of the Crabs caught in pots. He decides in advance to weigh five Lobsters and five Crabs so caught. On a fishing boat he watches the pots being raised one by one. Once he's got, say, five Lobsters and three Crabs, he ignores any further Lobsters caught and is only after the next couple of Crabs (to make up his quota).

  Advantages : No sampling frame is required; in fact it is not possible to have a list of all the Lobsters and Crabs in the sea.
  At a simple level, quick and easy and so inexpensive.

  Disadvantages : Bias can be introduced; For example, there may be many more Crabs than Lobsters in the sea so the 5:5 sample may not proportionally represent the population of all Lobsters and Crabs.
  Breaking down population into more groups increases complexity, time and cost. (There is a similarity with having more strata with stratified Random sampling)

- **Opportunity** Non-Random Sample (Also called Convenience Sampling)

  This is akin to simply asking the people who happen to be around you and who are willing to respond to your questionnaire.

  Advantages : No sampling frame is required.
  Quick and easy and so inexpensive.

  Disadvantages : Bias is very likely to be introduced, and the researcher themselves is likely to have an impact on the data collected.

3.2 Two Top Tips

- When saying a sampling method is cheap or quick or easy say in comparison to what.
  e.g. Opportunity Non-Random Sampling is less expensive than a census.
- With Stratified Sampling a good expression to use to describe the resulting sample is to say **it has the same structure** as the population from which it was taken.

3.3 Exercise

Question 1

The congestion of a roundabout is monitored for a month each evening at 6 pm.
If there is no traffic, the congestion is recorded as 0, up to grid-lock recorded as 5.
Here is the resulting data;

Sat 1,	Sun 2,	Mon 4,	Tue 5,	Wed 3,	Thu 3,	Fri 5,	Sat 2,	Sun 1,
Mon 5,	Tue 4,	Wed 5,	Thu 4,	Fri 5,	Sat 3,	Sun 2,	Mon 3,	Tue 3,
Wed 4,	Thu 4,	Fri 5,	Sat 3,	Sun 0,	Mon 5,	Tue 4,	Wed 4,	Thu 5,
Fri 5,	Sat 2,	Sun 1						

(i) The mean congestion on the roundabout is worked out by the council from all the data to be 3.4 which they claim is acceptable.

Explain how bias has been introduced into the calculation.
Hint : How many Sundays and how many Mondays are in the data ?

(ii) Work out the mean congestion when weekends are removed from the data.

(iii) More data is gathered over an entire year of 365 days.
As this is considered too much data a Systematic Random Sample is taken.
Peter does this by first numbering the list of data from 1 to 365.
With it in date order, he uses the RanInt button on his calculator to give a random number between 1 and 7.
It gives him the number 4 which happens to be a Sunday.
He now selects every 7th piece of data from the list thus getting a sample of size 51
He works out the mean congestion score for his sample and is surprised that it is only 1.2
What has gone wrong ?

Question 2

The Head Teacher of an infant school wishes to take a stratified sample of 20% of the pupils at the school which has the following numbers of pupils;

Year 1	Year 2	Year 3	Year 4
25	30	50	80

(i) Work out how many pupils in each Year will be in the sample.

(ii) Describe one benefit to the Head Teacher of using a stratified sample rather than a simple random sample.

Question 3

An interviewer wishes to know about the eating habits of all who live in London. He asks the first 50 people he sees outside a London fish and chip shop on a Friday evening about their eating habits.

(i) What type of sampling method is being used ?

(ii) Explain why the sampling method may not be representative of the eating habits of all who live in London.

(iii) Suggest two improvements that could be made to how the data is collected with a view to better capturing an impression of the eating habits of all who live in London.

Question 4

It is wished to take a Simple Random Sample of ten from a numbered list of 100 people. The RanInt function on a calculator is used to generate the following ten random numbers; 81, 94, 62, 6, 43, 69, 79, 81, 34, 58

Explain why the method has not generated a sample of the required size.

Question 5

A group of wildlife scientists wish to gather information on the masses of male and female wild deer in England. They set up a camp in the New Forest and set out to catch and weigh the first five male deer they come across and the first five female deer they come across.

(i) What type of sampling method are they using ?

(ii) Give an advantage of this method in comparison to taking a simple random sample.

The masses of the sampled deer are listed below;

Male, kg	75	80	90	85	82
Female, kg	67	72	75	68	65

(iii) Use the sample data to compare the masses of male and female wild deer.

(iv) Suggest two improvements the scientists could make to the sampling method.

Question 6

A factory manager wants to get information about the ways the employees travel to work. There are 480 employees in the factory, and each has a clocking-in number. The numbers go from 1 to 480.

Explain in detail how the manager could take a systematic sample of size 30.

Question 7

S3 Examination Question from May 2017, Q1

A company director decides to survey staff about changes to the company calendar. the company has staff in 4 different job roles

- 72 manages
- 108 drivers
- 180 administrators
- 360 warehouse staff

The director decides to take a stratified sample

(a) Write down one advantage of using a stratified sample rather than a simple random sample for this survey

[1 mark]

(b) Find the number of staff in each job role that will be included in a stratified sample of 40 staff.

[3 marks]

(c) Describe how to choose managers for the stratified sample

[2 marks]