# Types of Data

Many companies and organisations collect information to improve their products or services. They can do this through online surveys, via telephone, market research, census, etc. Have you been asked to provide feedback, e.g. rate the quality of a WhatsApp call? Then you have taken part in a survey!

The information stored from these surveys is called data and it is used to make analysis and judgements about what to do next.

There are two types of data **numerical** and **categorical**.

**Numerical (quantitative) –**

|  |
| --- |
| Describes data that can be counted and measured e.g. height, test score. |

**Categorical (qualitative) –**

|  |
| --- |
| Describes qualities, characteristics or categories e.g. hair colour, favourite football team. |

Each of these can be classified in two different ways.

**Categorical** can be **Nominal** or **Ordinal.**

**Nominal**

|  |
| --- |
| Data which cannot be ranked e.g. eye colour, voting preference. |

**Ordinal**

|  |
| --- |
| Data which can be put in order. |

**Numerical** can be **Discrete** or **Continuous**.

**Discrete**

|  |
| --- |
| Numerical data that can be counted, number of pupils in a class. |

**Continuous**

|  |
| --- |
| Numerical data that is measured e.g. height. |

**Example:** The following data is regarding a house. Decide what kind of data it is.

|  |  |  |
| --- | --- | --- |
|  | Categorical | Numerical |
| House | Nominal | Ordinal | Discrete | Continuous |
| Type of house | Y |  |  |  |
| Number of rooms |  |  | Y |  |
| Height of House |  |  |  | Y |
| Number of Windows |  |  | Y |  |
| Colour of front door | Y |  |  |  |
| What people think of the house (1 – very bad…. 5 very good) |  | Y |  |  |

**Population, Sampling and Bias**

**Population –** The whole set of individuals, items or data from which a statistical sample is drawn. It does not just have to involve people.

**Sample** – A small section of the population.

So a School may want to improve the quality of its food at lunch so they would take a survey from the pupils. The **population** in the example is all the pupils. But they wouldn’t ask all the pupils they would ask a select number who are the **sample**.

**Trials and Sampling**

These are the different ways of choosing the sample from the population.

**Random Sample –** Pick randomly from the sample.

**Systematic Sample –** Set it up so only every 4th person is task for example

**Stratified Sample –** This is random but in ratio of group size. So if a gym has 400 members, 300 are men and 100 are women. If they were doing a survey they would take a larger sample from men than women.

**Cluster Sample –** Whole group chosen at random.

**Example:** A company runs 3 restaurants in Glasgow

Each restaurant has 60 employees.

The owner of the company wants to conduct face to face interviews with its employees.

The owner wants to use a cluster sample

(a) Is this a suitable sampling method to use?

|  |
| --- |
| No because it means they will only ask employees from one restaurant. The other two restaurants might have very different opinions to the one being asked. \*However – would save time from going to different restaurants |

The owner then decides to try a different method that takes into account that of the 180 employees 72 are Male and 108 are female.

(b) What kind of sampling method should they use that takes this into account?

|  |
| --- |
| Random – Pick random employees from all 3 restaurants.Stratified – the same amount from each restaurant to ask |

(c) The owner requires a sample of 50 members. How many more Females than Males should be in the sample?

|  |
| --- |
| Men = 108/180 = 0.6, 50 x 0.6 = 30 MenWomen = 72/180 = 0.4, 0.4 x 50 = 20 Women |

**Example 2:** Look at the list of people below.

Mr Adams Miss Evans Mr Johnson Mr Rashid

Miss Booth Mrs Grant Miss Klein Miss Ryan

Mrs Carlisle Mr Graham Mr Khan Mr Thomson

Mr Carter Mr Henry Miss MacDonald Miss Violet

Miss Davidson Mr Hutton Mrs McCulloch Mr Zacharias

Give a way you could create a random sample using the names above.

|  |
| --- |
| Assign each person a number e.g. Mr Adams = 1, Miss Evans = 2. Then use a random number generator to pick someone. Repeat until enough people are chosen. |

**Exercise: Data and Sampling**

1. What type of data is each of the following?

a) Number of texts sent Numerical + Discrete

b) Colour of flowers Categorical + Nominal

c) Weight of a dog Numerical + Continuous

d) Exam Grade Categorical + Ordinal

e) Time asleep Numerical+ Continuous

f) Population of a country Numerical + Discrete

g) Area of a floor Numerical+ Continuous

h) Ticket prices Numerical+ Discrete

i) Favourite football team Categorical + Nominal

j) Nationality of football players Categorical + Nominal

k) Number of trees in a forest Numerical + Discrete

I) weight of a whale Numerical+ Continuous

m) Breeds of bird Categorical + Nominal

n) Number of eggs sold by a shop Numerical + Discrete

2. For each of the following what kind of sample would you do if you were trying to survey the following?

Random, Cluster or Stratified

a) Glasgow city is going to go ask people on the high street if they should add another subway stop. (Think about if they should ask anyone) **Cluster Sample of people who use the subway.** Might be no point asking people who are never going to use it. However. Random sample could be better because maybe more people would use it if there was another stop.

b) Newton Mearns local council is trying to find out what people think about Wi-Fi in the area. Cluster of people who live in Newton Mearns

c) Hamilton FC are taking a survey on the toilet facilities at their football stadium (What is the issue if they did a random sample and only asked 10 randomly selected fans? Stratified sample. Need to make sure men and women are asked.

d) A cinema is going to do a survey on what kind of films their customers prefer to watch. Random so everyone gets a chance

Stratified – Customers attending different types of film

3. A dentist wants to ask a random sample of patients how satisfied they are from the service they receive from her. Describe how she could select patients at random and what problems might be involved with doing this? (Potential bias) Bias you need to be careful of is that people will be less honest if you ask them directly. This could be fixed with an anonymous online survey. Send out a survey to random customers. Could ask every 4th patient who comes in to fill out a survey online.

4. A gardener wants to study the distribution of plants in a random sample of 20 square meters of land from a rectangular site that is 50m wide and 40m long. How could he select a random sample? Assign each square a number and letter A1, A2… B1, B2…

Ask someone to pick a letter A-D and a number 1-5.

Could number each square and inspect every 5th square.

5. A factory has the following employees

|  |  |  |
| --- | --- | --- |
|  | Full Time | Part Time |
| Male | 80 | 36 |
| Female | 24 | 44 |

If the boss wanted to do a random survey of 46 employees, how many of each subgroup should they select if they wanted to do a random sample?

|  |  |  |
| --- | --- | --- |
|  | Full Time | Part Time |
| Male | 20 | 9 |
| Female | 6 | 11 |

6. A shop looks at its customers and how much they spend per visit

|  |  |  |
| --- | --- | --- |
|  | Male | Female |
| Under £20 | 400 | 850 |
| £20+ | 700 | 750 |

If the shop wanted to do a survey of 216 random customers, then how many of each sub group should they select?

|  |  |  |
| --- | --- | --- |
|  | Male | Female |
| Under £20 | 32 | 68 |
| £20+ | 56 | 60 |

7. The table gives the number of entrants to higher education in one year.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Age | 18 | 19 | 20 | Other  | Total |
| Female | 94 775 | 42 640 | 11 950 | 33 980 | 183 345 |
| Male | 79 020 | 40 625 | 12 075 | 26 230 | 157 950 |
| Total | 173 795 | 83 265 | 24 025 | 60 210 | 341 295 |

a) A polling organisation wants to invite 10000 of the students to take part in a survey about their experiences of higher education. What is the minimum amount of students they use from each sub category?

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Age | 18 | 19 | 20 | Other  |
| Female | 2776 | 1244 | 350 | 995 |
| Male | 2315 | 1190 | 353 | 768 |

b) List other characteristics that might be relevant. Type of degree, nationality, what background they have.

c) Are there any issues with this poll? Categorey for other is too vague, should be split up more 21-25, 26-30, 30+ etc

8. A sample is a small survey. A census is a collection of the entire population.

An estate agency wants to know the following should they use a census or a sample.

a) The number of houses in England with an empty room. Census

b) How long an advert should appear on a website? Sample

c) The number of people who offer less than the asking price of a house. Census

9. In each of the following situations state which kind of sample should be used.

(a) A large airline wants to know what passengers think of the service on their flights Random

(b) A newsworthy event has occurred near a town centre. A TV news crew wants to include some representative local comments in its report. Cluster

(c) A town council wants to know how the various groups living there view the quality of its nursery provision. Stratified

(d) A scientist wants to investigate any damage that an invasive species may be causing to wildlife on our rivers. Random/Systematic

(e) A clothing manufacturer is considering developing a new range and wants to assess how popular it would be with both females and males across a range of ages. Stratified

(f) An election is coming up and the local branch of the political party want asses local opinion on a number of issues, they can only afford to interview 5% of the electorate. Cluster

10. We want to survey 3000 people in the USA. Stratified sampling is chosen. In the table below,

Calculate how many people of each age group would need to be asked.



125

525

297

798

792

390

11. ‘Do premiership footballers speak more than 1 language’ Think of 3 ways you could ask 20 different footballers how many languages they speak.

* Ask 20 footballers from a randomly selected club
* Ask 1 from each club
* Split into nationalities. Ask one footballer from 20 randomly chosen nationalities