Bringing British Education to You

www.nccedu.com

Foundation Mathematics

Topic 6 – Lecture 1: Presentation of Data

Organising Data

Gathering Data

Scope and Coverage

This topic will cover:

- Different methods for organising data
- Techniques used to gather data

Learning Outcomes

By the end of this topic students will be able to:

- Understand the principles underpinning the organisation of data
- Understand the principles surrounding data gathering

Data

- Data is a term that you will come across time and again within the business world. What does the term actually mean?
- Definition
  - ‘Data’ is simply a scientific name for facts, figures, information and measurement.
  - Data may be of several types. The first distinction to make is between attributes and variables.

Data - Attributes and Variables

- Definition
  - An attribute is something an object either has or has not.
  - An attribute cannot be measured. For example an individual is either male or female. There is no measure of how male or how female somebody is. In this case the gender of a person is an attribute.
- Definition
  - A variable is something that can be measured.
  - For example, the height of a person is a variable which can be measured according to some scale such as centimetres or inches.

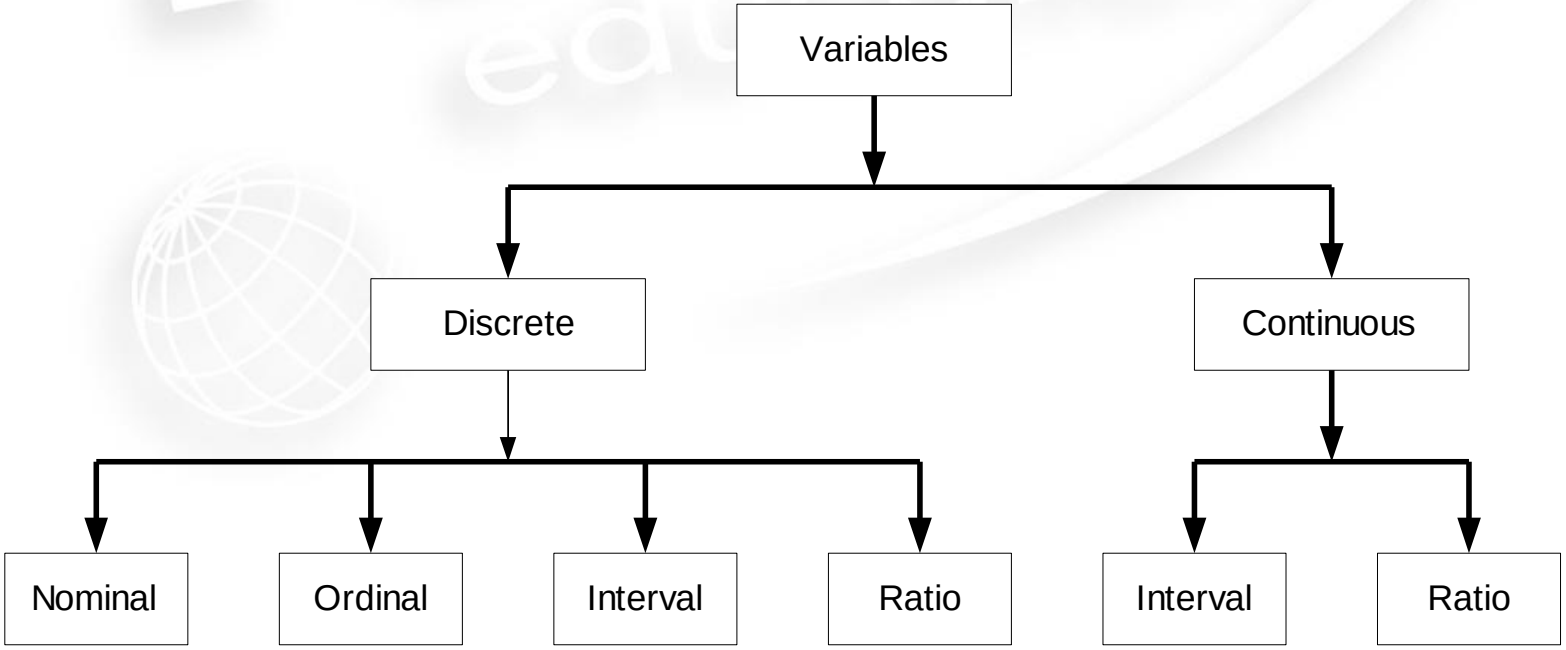
Data – Discrete Variables

- When we are dealing with data we need to be able to make sense of what the data is telling us. To do this we need to classify our data.
- Variables can be classified as either discrete or continuous.
- Discrete Variables
  - Discrete variables can only take a finite or countable number of values within a given range.
    - Example - the number of goals scored by a football team can only be equal to whole numbers. Therefore, it is possible to score 2 goals but not 2.134 or 2.456

Data – Continuous Variables

- ***Continuous variables*** can take on any value.
- Continuous variables are measured rather than counted.
- For example the height of a man from a given population may be 1.876metres or the weight of an individual woman may be 53.234kg.
- So far we have classified our data types into two main categories, however as we are trying to find the meaning within any data set we need to be able to describe data in more meaningful ways.

Classification of Data



- As we can see in our diagram it is possible to categorise data into different groups

Classification of Data - Grouping

- ***Nominal*** data refers to data which is non-numeric, for example gender
- ***Ordinal*** data is data which is arranged in a meaningful order, but refers to rank only (1st, 2nd, 3rd).
- ***Interval*** numbers where the differences between them has value (1-3 on a scale of agreement, 3 is stronger than 2 which is stronger than 1).
- ***Ratio*** data is data in which we are making a comparison between different values for example a salad dressing is made from a recipe with oil and vinegar in a ratio of 4:1 (therefore 4 parts oil to every 1 part vinegar)

Primary and Secondary Data

- Primary Data (Raw Data) and Secondary Data

- Definition
  - Primary data is data collected especially for the purpose of whatever survey is being conducted.
  - Secondary data is data which has already been collected elsewhere, but which can be used or adapted for another purpose.

Gathering Data

- Secondary data can come from existing sources such as Government documents and technical manuals.
- Primary data must be gathered.
- There are two basic methods of collecting primary data from individuals:
  - You ask them questions
  - You observe their behaviour
- For the purposes of this course we will concentrate on the first of these techniques which involves the collection of data using surveys.

Gathering Data - Surveys

- There are two main types of survey:
  - Interviews
  - Questionnaires
- Although surveys offer a relatively quick, efficient and cost-effective way of obtaining required data, they must be effectively planned, constructed, delivered and received back.
- The most popular way of gaining survey information is by using a questionnaire, this has the advantage of being cheaper and can involve a greater number of people.

Questionnaires

- When all of the population is examined, such an investigation is called a **census**.
- This is often an expensive and difficult process to manage, often taking place over a long time period.
- To overcome these difficulties many organisations use techniques that extract data from samples of a population.
- Such samples may be representative of the entire population and give sufficient accuracy to support the decision making processes.
- It is possible to ask more questions with a sample from a population.

The Choice of Sample

- One of the most important requirements of ***sample data*** is that they should be ***complete***. That is, the data should cover ***all areas*** of the ***population*** to be examined. If this requirement is not met, then the sample will be biased.
- Sampling methods fall into three main categories
  - Random sampling
  - Quasi-random sampling
  - Non-random sampling

Sampling Techniques

- Random Sampling
  - To ensure that the sample selected is free from **bias**, random sampling must be used.
- Definition
  - A simple random sample is a sample selected in such a way that every item in the population has an equal chance of being included.
  - To achieve this outcome in a formal and more scientific way it is often necessary to use a **sampling frame**

Sampling Frame

- A sampling frame is a numbered list of all the items in the population under investigation. For it to work the sampling frame must have the following characteristics
  - **Completeness** - All members of the target population must be included on the list.
  - **Accuracy** - Information on the population must be correct.
  - **Adequacy** - It must cover the entire population.
  - **Timeliness** - The list must be up to date.
  - **Duplication** - Each member of the population must appear only once.

Random Numbers

- Assuming that a sampling frame can be drawn up, then a random sample from the population can be achieved using a random number table.
- An example of a random number table:

93716	16864	98953	73231
32886	59780	09958	18065
95052	06831	19640	99413
39510	35905	85244	35159
27699	06494	03152	19121
92962	61773	22109	78508
10274	12202	94205	50380
75867	20717	82037	10268
85783	47619	87481	37220

Using Random Numbers - 1

- There are some simple rules that surround the use of random number tables
- The sample is found by selecting groups of random numbers with the number of digits depending on the total population size i.e.

Total Population Size	Number of Random Digits
1-10	1
1-100	2
1-1000	3

- The starting point on the table should be selected at random. After that, however, numbers must be selected in a consistent manner. In other words, you should use the table row-by-row or column-by-column.
- It is also possible to select random numbers generated by a pocket calculator or a computer.

Using Random Numbers - 2

- An investigator wishes to select a random sample from a population of 800 people who have been numbered 000, 001….799.
- As there are three digits in 799 the random numbers will be selected in groups of three.
- Working along the first line of the table given earlier, the first few groups are as follows:
  - 937 161 686 498 953 732
- Numbers that are over 799 are discarded and so the first four people in the sample will therefore, be those numbered 161, 686, 498 and 732.

Quasi Random Sampling

- **Stratified Sampling:**
- The population is divided into strata or categories.
- This approach helps to remove bias from our sampling activities.
- Bias can occur if, for example, our random approach generates samples that are too closely linked.
- For example too many people from a particular town included in our sample when we want to get representative data across the county.
- Stratification requires us to have some information about our target population so as to take into consideration any potential bias.

Stratified Sampling – Example - 1

- If we took a sample of all accountants in the UK it is possible that the entire sample might consist of those working only in public companies, therefore biasing our research.
- Stratified sampling removes this possibility as random samples could be taken from each type of possible employment.
- The number in each sample is proportional to the total number of accountants in each employment type:
  - Those in partnerships
  - Those in public companies
  - Those in private companies

Stratified Sampling – Example - 2

- The number of accountants in each type of work in a particular county are as follows:
  - Partnerships 500
  - Public companies 500
  - Private companies 700
  - Public practice 800
  - Total 2500
- To sample data accurately we need to take into account the different number of accountants in each employment type

Stratified Sampling – Example - 3

- If a sample of 20 was required the sample would be made up as follows:

		Sample
Partnerships	500 X 20 / 2500	4
Public companies	500 X 20 / 2500	4
Private compa nies	700 X 20 / 2500	6
Public practice	800 X 20 / 2500	6
Total Samples		20

Stratified Sampling – Advantages

- It ensures a representative sample
- The structure of the sample will reflect that of the population, if the same proportion of individuals is chosen from each stratum.
- Inferences can therefore be made about each stratum.
- Precision is increased as sampling takes place within strata and the range of variation is therefore reduced.

Sampling

- There are other sampling techniques that can be used including those which are non random such as
  - Quota Sampling
  - Cluster Sampling
- These are often implemented when cost of the survey is a very important factor and as such often sacrifice accuracy in order to keep costs down.

Topic 6 - Presentation of Data 1

Any Questions?