

Types of Data

Data is information is information collected usually in order to prove a hypothesis or scientific guess during an experiment.

A variable is a piece of data.

Numerical (Quantitative) Data:

This is data that is measureable and is always collected in numerical form such as weight, height etc. You should be able to carry out any maths operation on it such as adding it all together, finding the average etc.

There are two types of Numerical Data.

Discrete data: This is data that can be counted and can't be made more precise. You are counting whole indivisible entities. E.g. The number of pets different families have.

Continuous data: This is data that can be reduced and divided into finer and finer levels. There is meaning at all points between the numbers given. E.g. The gradual rise of temperature throughout the day.

Types of Variables

Nominal variables are variables that have no numerical value and there is no value to ordering them e.g. reason for going to the gym

Ordinal variables require a clear ordering of the variables. E.g. How often you use the weights machines.

Interval variables are variables in which the difference between the variables is evenly spaced.

Types of Data

Categorical (Qualitative) Data:

This is data that can represents characteristics and can be split into various groups. Examples of groups are age, race, gender.

Categorical data is generally analysed by using a data table divided into rows and columns. It can take on numerical values but those values don't have any mathematical meaning e.g. "1" indicates females and "2" indicates males.

Random Sampling

This means that each member of a 'population' has an equal chance of being selected.

It is a sampling technique where a group of subjects is selected for study from a larger group (population).



The most important advantage of random sampling is obtaining an unbiased sample. Obtaining a truly random sample can be difficult and expensive to achieve.

Always consider whether your sample size is robust and large enough. Do you think it truly represents the population?

Random Sampling

<https://www.youtube.com/watch?v=qYdwN4Ndz70>



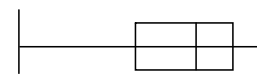
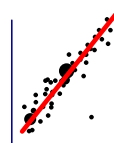
Bias: Imagine that you were testing a product on a group of people.

- Choose a product you would test
- Make a note of the population you would look at taking a sample from
- How could there be bias in your sample?

Outliers

An outlier is a piece of data that lies an abnormal distance from other values in a random sample. It is up to the analyst to decide whether that piece of information is abnormal. There are extreme outliers and mild outliers.

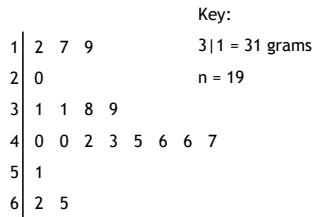
Outliers can be identified easily when drawing scatter plots and box plots.



Interpreting Tables and Graphs

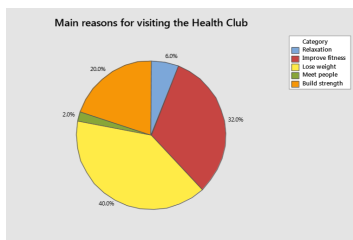
1. Stem and Leaf

This type of diagram enables the analyst to order the data. It provides information on location spread, outliers and the shape of distributions .



Interpreting Tables and Graphs

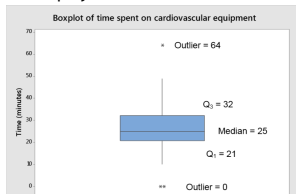
3. Pie Charts: These are easy to interpret when working with nominal and ordinal variables. They emphasise size comparisons for categories.



Interpreting Tables and Graphs

5. Box Plots

This is a chart that displays information about the median, range and quartiles.



Question: Create a box plot for the following test results

12 38 53 76 62 61 55 44 77 79 80 100

Interpreting Tables and Graphs

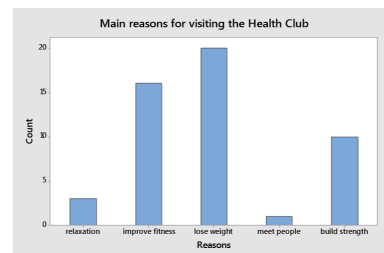
2. Frequency tables: These can be applied to both quantitative and qualitative variables.

This table shows the reasons members gave for visiting their health club.

Aim	Count	Percent
Relaxation	3	6.00
Improve Fitness	16	32.00
Lose Weight	20	40.00
Meet people	1	2.00
Build Strength	10	20.00
N=	50	

Interpreting Tables and Graphs

4. Bar Charts - Another way of emphasising size difference between the data.



Interpreting Tables and Graphs

6. Contingency Tables: This is a display format used to analyse and record the relationship between 2 or more categorical variables.

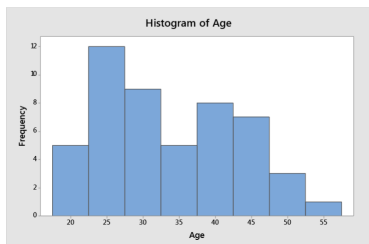
It is like a frequency table but allows the relationship between the variables to be summarised.

The table below examines the variables, gender and aim.

Aim		Male	Female	All
Relaxation	Number	0	3	6%
	Percentage		6%	
Improve fitness	Number	8	8	
	Percentage	16%	16%	32%
Lose weight	Number	4	16	
	Percentage	8%	32%	40%
Meet people	Number	1	0	1
	Percentage	2%	0%	2%
Increase Strength	Number	10	0	10
	Percentage	20%	0%	20%
N=50				

Interpreting Tables and Graphs

7. Histograms: These are useful for measuring the shape of the distribution.
In a histogram, the area of the bars represent the frequency.



This histogram can be considered bimodal. It has peaked in two places.