

Statistics

Created by
Graduate Bsc (Hons) MathsSci (Open) GIMA

There are few definitions that are fundamental to understanding basic Statistics.

The mean is the arithmetic average of a set of data.

e.g. For the data set 7, 2, 5

$$\text{mean} = \frac{7 + 2 + 5}{3} = 5$$

The median is the middle point of a data set when arranged in ascending order.

e.g. For the data set 1, 2, 3

$$\text{median} = 2$$

e.g. For the data set 1, 2, 3, 4

$$\text{median} = \frac{2 + 3}{2} = 2.5$$

The mode is the most frequently occurring point in a data set.

e.g. For the data set 1, 2, 3, 1

$$\text{mode} = 1$$

The range is the difference between the largest point and the smallest point when arranged in ascending order.

e.g. For the data set 1, 2, 3, 4

$$\text{range} = 4 - 1 = 3$$

Statistics

Created by
Graduate Bsc (Hons) MathsSci (Open) GIMA

The standard deviation is a measure of how spread out the data is around the mean.

NOTE: The standard deviation tells you nothing about the mean.

x = datapoints

$$\text{Standard}_{\text{dev}} = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}}$$

Give a set of data in ascending order we can divide it into Quartiles which divides the data into 4 equal pieces.

e.g. For the data set 1, 2, 3, 4

Quartiles are

$$Q_1 = \frac{1+2}{2} = 1.5 \quad Q_2 = \frac{2+3}{2} = 2.5 = \text{median} \quad Q_3 = \frac{3+4}{2} = 3.5$$

$$\text{Semi-interquartile} = \frac{1}{2} \cdot (Q_3 - Q_1)$$

The information above is the basic information you need to know for the exam. The following two examples illustrate the type of questions normally set in the exam. The first asks you draw a box plot for a set of data. Box plots are an effective way of getting you to work out the median upper and lower quartiles and the semi-interquartile.

Statistics

Created by
Graduate Bsc (Hons) MathsSci (Open) GIMA

Examples

1. Given below is rainfall data in mm for the year 2004. Draw a box plot for the data and find the range, mean, median and the semi-interquartile.

20, 10, 15, 8, 9, 4, 14, 8, 1, 22, 23, 24

Solution

Ascending order to get.

1, 4, 8, 8, 9, 10, 14, 15, 20, 22, 23, 24

$$\text{Range} = 24 - 1 = 23$$

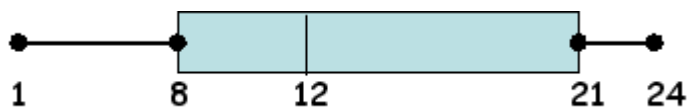
$$\text{Mean} = \frac{1 + 4 + 8 + 8 + 9 + 10 + 14 + 15 + 20 + 22 + 23 + 24}{12} = 13.2$$

Median = Q_2 See below

$$Q_1 = \frac{8 + 8}{2} = 8 \quad Q_2 = \frac{10 + 14}{2} = 12 \quad Q_3 = \frac{20 + 22}{2} = 21$$

$$\text{Semi-interquartile} = \frac{1}{2} \cdot (Q_3 - Q_1) = 21 - 8 = 13$$

Hence box plot is given below



Statistics

Created by
Graduate Bsc (Hons) MathsSci (Open) GIMA

2. A survey of milk prices is given below.

49 44 41 52 47 43

Calculate the mean and Standard deviation for the data.

Solution

$$(a) \quad \text{mean} = \frac{49 + 44 + 41 + 52 + 47 + 43}{6} = 46$$

- (b) The standard deviation is

$$\sum x^2 = 49^2 + 44^2 + 41^2 + 52^2 + 47^2 + 43^2 = 12780$$

$$(\sum x)^2 = 76176$$

$$\text{Standard}_{\text{dev}} = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}} = \sqrt{\frac{12780 - \frac{76176}{6}}{6-1}} = 4.1$$

- (c) The Standard deviation for sugar prices is 2.6. Make a valid comment between the milk and sugar prices.

Solution

There is less variation in sugar prices than milk.