Semester: 4 Subject: Statistics for AIDS Academic Year: 2023-2024

Exploring the Data Distribution:

The estimates that we have covered so far indicate that the data can be expressed in a single number to describe its location or variability.

We mention below the main terms used for exploring the distribution.

* Boxplot: Turkey has introduce Boxplot. It visualises distribution of data quickly.
* Frequency Table: It is the count of numeric data values in a set of intervals.
* Histogram: A plot of the frequency table with the given intervals on x-axis; and the corresponding count on y-axis.
* Density plot: It is the smoothed version of the histogram.

Boxplot:-

To draw a boxplot, we proceed as follows:

*Draw a rectangular box whose bottom is the lower quartile (i.e. 25th percentile) and whose top is the upper quartile (75th percentile).

*Draw a horizontal line segment inside the box to represent a median.

*Extend horizontal line segments from each end of the box out to the most extreme observations.

Semester : 4 Subject : Statistics for AIDS Academic Year: 2023-2024

Example 1:

Find Q_1, Q_2 and Q_3 for the following data set, and draw a box-and-whisker plot

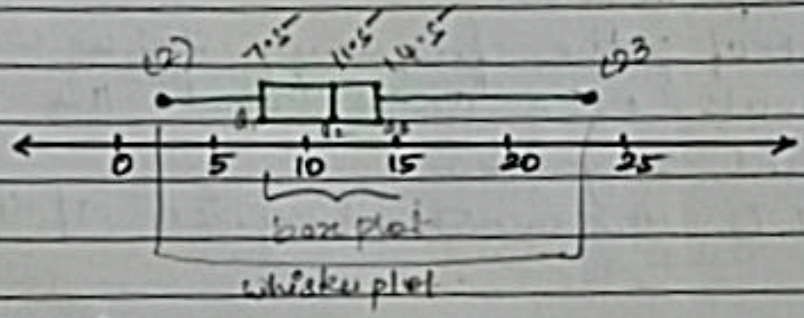
{2, 6, 7, 8, 8, 11, 12, 13, 14, 15, 22, 23}.

Solution:

2, 6, 7, 8, 8, 11, 12, 13, 14, 15, 22, 23

7·5 11·5 14·5

Q_1 Q_2 Q_3.



Example:

Find Q_1, Q_2 and Q_3 for the following data set. Identify any outliers and draw a box and whisker plot.

{5, 40, 42, 46, 48, 49, 50, 50, 52, 53, 55, 56, 58, 75, 102}.

There are total 15 values of arranged in increasing order.

Q_2 is the 8th data point.

$Q_2 = 50$

Semester: V Subject: Statistics for AIDS Academic Year: 2023-2024

Q_1 is 4th data point $\therefore Q_1 = 46$

Q_3 is 12th data point $Q_3 = 56$

Interquartile range $IQR = Q_3 - Q_1$

$= 56 - 46$

$= 10$

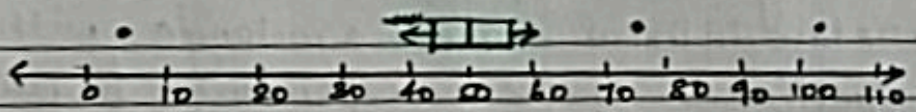
$Q_1 - (1.5 \times IQR)$ $\quad Q_3 + (1.5 \times IQR)$

$= 46 - (1.5 \times 10)$ $\quad = 56 + 15$

$= 46 - 15$ $\quad = 71$

$= 31$

Since 5 is less than 31 and 75, 102 are greater than 71, hence there are 3 outliers.



Frequency Table:

The frequency distribut

* Make the table self-explanatory provide a title, a brief description of a source of the data, state in what units the figures are expressed, label rows and columns where appropriate.
* Keep the table as simple as possible.
* Distinguish between zero values and missing observations.