PSG College of Technology, Coimbatore -641 004

Department of Applied Mathematics and Computational Sciences

8th Semester MSc TCS

18XT87 Data Mining Lab

Problem Sheet - 1

1. Understand and illustrate the following concepts (use your own datasets).
    a) Boxplots
    b) Quantile plot
    c) Quantile- Quantile plot
    d) Histograms
    e) Scatter Plots

2. In a study of how students use their mobile telephones, the phone usage of a random sample of 11 students was examined for a particular week.

    The total length of calls, y minutes, for the 11 students were

    17, 23, 35, 36, 51, 53, 54, 55, 60, 77, 110

    a. Find the median and quartiles for these data.

        A value that is greater than Q3 + 1.5 × (Q3 – Q1) or smaller than Q1 – 1.5 × (Q3 – Q1) is defined as an outlier.

    b. Show that 110 is the only outlier.

    c. Draw a box plot for these data indicating clearly the position of the outlier.

3. In a study of how much time students spend on social media, usage of a random sample of 15 students was examined for a particular day.

    The total time of usage, *x* minutes, for the 15 students were

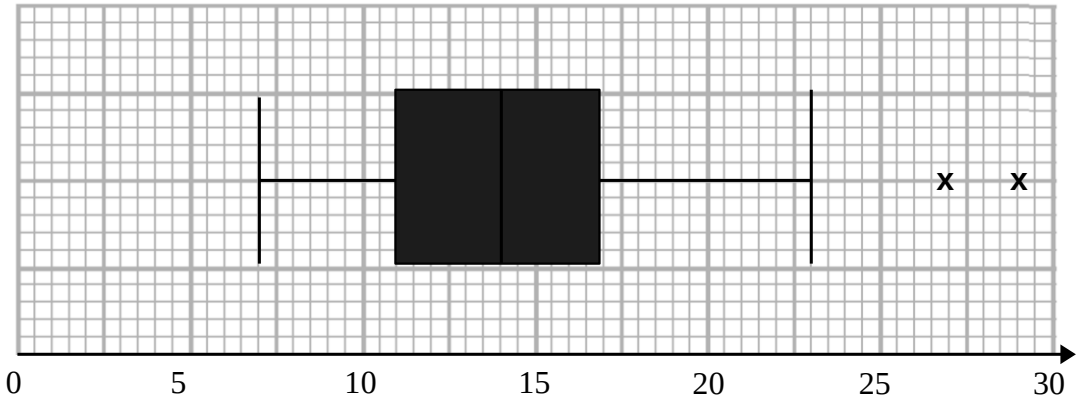
    6, 25, 39, 62, 65, 74, 80, 94, 125, 127, 154, 159, 184, 210, 251

    a. Find the median and quartiles for these data.

    b. Show that there are no outliers.

    c. Draw a box plot for these data.

4. Children from two schools A and B took part in a competition to complete a puzzle. The distribution of times taken, to the nearest minute, by children in school A are shown in the box plot below.



(a) Find the time by which 75% of the children in school A had completed the puzzle.

(b) State what the two crosses (**x**) represent on the box plot above. Interpret these in context.

For school B the shortest time taken was 6 minutes, the longest time taken was 22 minutes and the quartiles were 12, 15 and 17 minutes respectively.

(c) Determine if there are any outliers.

(d) Draw a box plot for this information.