**Activity Sheet**

1. Following are Kohli’s scores in 4 matches played between India and Australia, match 5 being the latest one played in a six match series.

|  |  |
| --- | --- |
| **Matches** | **Score** |
| 1 | 65 |
| 2 | 72 |
| 3 | 58 |
| 4 | 77 |

1. What are his median and mean scores for the first four matches played?
2. If he scores 70 in his fifth match, does his series mean score increase or decrease w.r.t. scores obtained for first four matches. Find his new mean score.
3. Which has increased more, his mean score or his median score after the fifth match?

Ans. R code:

kohli\_scores<- c(65, 72, 58, 77)

kohli\_scores

old\_median<- median(kohli\_scores)

old\_median

old\_mean<- mean(kohli\_scores)

old\_mean

new\_kohli\_scores<- c(kohli\_scores, 70)

new\_kohli\_scores

new\_median<- median(new\_kohli\_scores)

new\_median

new\_mean<- mean(new\_kohli\_scores)

new\_mean

2. Which measure of central tendency would be most useful in each of the following instances?

(a) The production manager for a manufacturer of glass jars, who is concerned about the proper jar size to manufacture, has sample data on jar sizes ordered by customers. Would the mean, median, or modal jar size be of most value to the manager?

(b) The sales manager for a quality furniture manufacturer is interested in selecting the regions most likely to purchase his firm’s products. Would he be most interested in the mean or median family income in prospective sales areas?

(c) A security analyst is interested in describing the daily market price change of the common stock of a manufacturing company. Only rarely does the market price of the stock change by more than one point, but occasionally the price will change by as many as four points in one day. Should the security analyst describe the daily price change of the stock in terms of the mean, median, or modal daily market price change?

Ans: a) mean b) median c) mode

3. The following numbers represent the time in minutes that twelve employees took to get to work on a particular day.

1, 7, 8, 4, 12, 14, 22, 15, 18, 14.

Calculate the quartiles and find the interquartile range.

Ans: Note that here we consider the two 14's to be distinct elements and not representing the same item; consider this like you obtained a score of 14 on two different quizzes.

First, we write the data in increasing order: 1, 4, 7, 8, 12, 14, 14, 15, 18, 22.

The median(Q2) is 13 (it is the mean of 12 and 14 — the pair of middle entries).

Therefore, the lower half of the data is: {1, 4, 7, 8, 12}.

Notice that 12 is included in the lower half since it is below the median value.

Then *Q*1 = 7 (there are five values in the lower half, so the middle value is the median). Similarly, the upper half of the data is: {14, 14, 15, 18, 22}, so *Q*3 = 15.

The interquartile range is Q3 minus Q1, so IQR = 15 - 7 = 8.

4. Analyze the performance of your class in the first ROTe taken at INSOFE

Scores: 11, 7.5, 8.5, 10, 10, 10.5, 5.5, 10, 9, 9.5, 5.25, 8, 6.5, 10.5, 8.75, 0, 6, 6, 6.75, 8.75,

0, 9.5, 7.5, 8.5, 7

1. How is the spread of the scores?
2. Draw a histogram to visualize the data distribution.
3. Find the 25th percentile, 50th percentile and 75th percentile for this data.
4. Find outliers, if any. Do a boxplot to visualize the same.

Ans: R Code:

# (a) How is the spread of the scores?

rote\_scores<- c(11, 7.5, 8.5, 10, 10, 10.5, 5.5, 10, 9, 9.5, 5.25, 8, 6.5, 10.5, 8.75, 0, 6, 6, 6.75, 8.75, 0, 9.5, 7.5, 8.5, 7)

rote\_mean<- mean(rote\_scores)

rote\_mean

diff\_mean<- rote\_scores - rote\_mean

diff\_mean

diff\_mean\_sq<- diff\_mean^2

diff\_mean\_sq

diff\_mean\_sq\_sum<- sum(diff\_mean\_sq)/length(rote\_scores)

diff\_mean\_sq\_sum

sd\_rote<- sqrt(diff\_mean\_sq\_sum)

sd\_rote

range <- max(rote\_scores) - min(rote\_scores) #Range

range

sd\_rote<- sqrt(sum((rote\_scores - mean(x = rote\_scores))^2)/length(x = rote\_scores))

quantile(x = rote\_scores)

quantile(x = rote\_scores, prob = seq(from = 0, to = 1, by = 0.1),type = 1)

### (b) Draw a histogram to visualize the data distribution.

hist(x = rote\_scores)

### (c) Find the 25th percentile, 50th percentile and 75 percentile for this data.

quantile(x = rote\_scores, prob = c(0,0.25, 0.5, 0.75))

### (d) Find outliers, if any. Do a boxplot to visualize the same.

boxplot(x = rote\_scores)

5. The mean salary of male employees and female employees in a company are Rs. 56,000 and Rs. 46,000 respectively. If the mean salary of all the employees of that company is Rs. 52,000, find the percentage of male and female employees.

Ans: Let is the mean salary of m male employees, be the mean salary of n female employees and be the mean salary of all the employees in the company. Therefore,

,, and

= = 52000

or,

or, = .

Therefore, male employees are 60% and female employees are 40%.

6. Suppose you have measured the heights of your dogs are:

600mm, 470mm, 170mm, 430mm and 300mm.

Find out the Mean, the Variance, and the Standard Deviation.

Ans.

|  |  |  |
| --- | --- | --- |
| Mean | = | *600 + 470 + 170 + 430 + 300***5** |
|  | = | *1970***5** |
|  | = | 39 |

|  |  |  |
| --- | --- | --- |
| **Variance** | | |
| σ2 | = | *2062 + 762 + (−224)2 + 362 + (−94)2***5** |
|  | = | *42436 + 5776 + 50176 + 1296 + 8836***5** |
|  | = | *108520***5** |
|  | = | 21704 |

**Standard Deviation**σ=√21704 =147.32... =**147** *(to the nearest mm)*