Activity Sheet

1. A company has 10 employees and their average salary is 1 lakh per annum. Now they have recruited a CEO who salary is normally very high. Which of the following measure of central tendencies will be impacted
   a. Median
   b. Mode
   c. Mean

2. Here is the data of experience of individuals in a CPEE class. We have grouped individuals into 6 groups and here is the data. Compute the average and median values for each group and list your observations.

Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
0	3	14	9	16	13
0	12	9.5	4.5	12	10
0	3.5	4.5	9	6.5	10
0	14	7.5	8	1	1
0	3	5	4	11	3
0	2.8	2	6	5	4
1	5	4.8	5	3	6
1	9	3.6	3.5	8	3.8
1	5.5	6	2.8	3	4
16	9	8.5	12	4	8

   a. What is the average in each group?
   b. What is the median in each group?
   c. What is the average experience across all groups?
   d. Does the average of the each group average experience is same as the Grand average ?
   e. Add an extra person with experience of 17 to Group5 and calculate the mean.
   f. Add an extra person with experience of 20 to Group6 and calculate the median.

3. You and your friends regularly order food online and prefer door delivery services. Each one believes that their respective service providers are very prompt. To understand it better, you have started collecting the time to deliver food in 20 different occasions for all. Here is the data of delivery times.

Time taken to deliver the order in minutes

Eagle Boys	Food Panda	Swiggy	Pizza Hut	Dominos
30	39	33	30	35
35	37	31	35	23
23	35	25	23	35
12	33	37	12	33
15	31	28	15	30
16	25	36	16	31
19	37	20	19	25
31	28	30	31	37
35	10	35	35	28
21	46	23	0	36
39	30	12	60	20
37	35	15	37	12
35	23	16	35	15
33	12	19	33	16
31	15	31	31	19
25	16	35	25	31
37	19	0	37	35
28	31	60	28	21
36	35	37	36	44
20	21	35	20	32

Use "deliveryserivces". Csv to solve the problem in R language

a. Now that you know central measures help you understand data better you go ahead with computing the central measures. (mean, median, mode, quartiles, range, inter-quartile range, variance, standard deviation)
b. Plot the box plot for each of the delivery service.
c. What do you observe across delivery services
d. Do you still believe that all the service providers are prompt in their services?

4. Two people work in a factory making parts for cars. The table shows how many complete parts they make in one week.

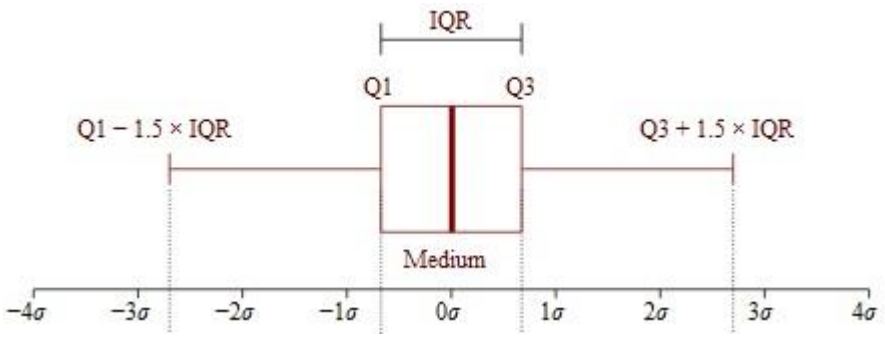
Worker	Mon	Tue	Wed	Thu	Fri
Philip	20	21	22	20	21
Mathews	30	15	12	36	28

(a) Find the mean, median and range for Philip and Mathews.
(b) Are the above measures enough to check the consistency?
(c) Who is more consistent?

5. Find the mode for 8,6,2,4,6,8,10,8 and also for apples,apples,apples,oranges,grapes

6. Analyze the performance of a class

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



a. How is the spread of the scores? Compute range, variance & standard deviation
b. Find the 25th percentile, 50th percentile and 75 percentiles for this data
c. Find the Q1, Q2 and Q3 and IQR for this data.
d. Find outliers, if any. Do a boxplot to visualize the same.

7. A large retailer store regularly orders cartons of Pineapples. The average weight of the cartons is supposed to be 22 kgs. Random samples of cartons from two suppliers were weighed. The weights in kgs of the cartons were

Supplier – I	17	22	22	22	27
Supplier – II	17	19	20	27	27

a. Compute the range of carton weights from each supplier
b. Compute the mean weight of cartons from each supplier.
c. Look at the two samples again. The samples have the same range and mean. How do they differ? The retailer store uses one carton of Pineapples in each Pineapples muffin recipe. It is important that the cartons be of consistent weight so that the muffins turn out right.

8. A quick recall – Solve the following in R

a. Create a vector with first 100 natural numbers. Compute its mean and median
b. Create a vector with 12,13,15,32,24,53,45,78,91. Compute its mean and median
c. Create a random vector of length 10 with values between 20 and 40. Compute its mean and median
d. Create a vector with ten 1s, twenty 2s, thirty 3s. Compute its mean and median

9. Do the following:

a. Import mtcars, an R in-built dataset
b. Observe rows and columns. See the summary statistics
c. Subset all numeric attributes in the data and assign it to a data frame num_data
d. Subset all categorical attributes in the data and assign it to a data frame cat_data

e. For each column of num_data, find the mean. Draw boxplots and what is your inference
f. How many 6-cylindered cars are there?
g. Find the mean mpg of 6-cylindered cars.

Assignment:

Create a neatly commented R script for the below question.
Read in the 'Batch_marks.csv' and find the following:

1. Explore the data and report all your findings with visualizations
2. Find mean, median and mode for the two batches.
3. What measures of variability can you think of?
4. Find range, variance, standard deviation and IQR for the two batches.
5. Do a boxplot to analyze the data for the two batches. Did you find any outliers?