Assignment I
Statistics Methods for Data Science (DS 212)
BTech (Data Science) Semester IV

- **All the questions are compulsory. Submit the scan copy of the assignment, solved on A4 size plain sheets. The date of Submission is 12/02/2022**

1. Portfolio Management: A portfolio manager keeps a close watch on the price- earnings ratio (defined as current market price divided by earnings for the most recent four quarters) of 200 common stocks. He reasons that, when the majority of stocks in a representative sample have low price earnings (P-E) ratios by historical standards, it is time to become an aggressive buyer. Low P-Es may mean that investors in general are unrealistically pessimistic. Moreover, stocks with low P-Es can benefit in a two- fold way when earnings increase: (a) higher earnings multiplied by a constant P-E ratio means a higher market price and (b) rising earnings are usually accompanied by rising P-E ratios.

Price-Earnings Ratios for 200 common Stocks

	12	10	9	8	18	29	20	13	7	7	21	5
23	22	22	11	10	14	8	14	10	29	8	15	
11	9	30	25	26	18	19	23	12	16	19	20	
10	9	12	17	19	22	18	17	12	15	19	29	
17	29	26	5	8	10	16	18	11	19	10	12	
17	18	9	7	16	18	18	17	21	22	27	18	
28	27	16	15	25	24	21	10	19	5	8	19	
15	17	18	10	21	22	27	29	10	11	17	19	
10	10	19	21	25	14	19	18	16	10	16	10	
21	16	28	29	5	8	19	10	11	22	16	26	
18	19	27	14	19	16	19	11	10	22	17	26	
10	11	19	11	10	10	9	9	11	19	18	10	
10	11	11	17	18	25	24	13	17	18	19	10	
18	19	12	17	19	17	16	18	19	10	13	12	
19	16	18	5	7	8	30	16	10	29	18	10	
19	17	19	15	17	18	19	15	14	18	19	10	
20	28	29	8	9	10	17	18					

Question for discussion

a. Organize the values of the variable.
b. Construct a frequency distribution table.
c. Mean, Median, Mode of the data

2. Following is the data represents the total number of wage-earners in the various age groups.

Age group:	14-15	16-17	18-20	21-24	25-29	30-34	35-39
No. of wage-earner:	120	140	150	110	110	100	90

Find P10, D7, Q1, Q3, P78, D3 and P45 for the given data.

3. A quality control inspector tested nine samples of each of the three designs A, B and C of certain bearing for a new electrical winch. The following data are the number of hours it took for each bearing to fail when the winch motor was run continuously at maximum output, with a load on the winch equivalent to 1.9 times the intended capacity.
   A: 16 16 53 15 31 17 14 30 20
   B: 18 27 23 21 22 26 39 17 28
   C: 31 16 42 20 18 17 16 15 19
   Calculate the mean and median for each group and suggest which design is best and why?

4. The following distribution gives the pattern of overtime work per week done by 100 employees of a company. Calculate median, first quartile, seventh decile and sixtieth percentile.

Overtime hours:	10-15	15-20	20-25	25-30	30-35	35-40
No. of employees:	11	20	35	20	8	6

5. The following table shows the age distribution of heads of families in a certain country during the year 1957. Find the coefficient of skewness for the following data.

Age of Head of the family (yrs)	Under 25	25-29	30-34	35-44	45-54	55-64	65-74	Above 74	Total
Number (million)	2	4	5	10	9	6	4	5	45

6. Calculate Karl Pearson's coefficient of correlation between age and playing habits from the data given below:

Age:	20	21	22	23	24	25
No. of students:	500	400	300	240	200	160
Regular players:	400	300	180	96	60	24

7. Given the following bivariate data:

x:	-1	5	3	2	1	1	7	3
y:	-6	1	0	0	1	2	1	5

(a) Fit a regression line of y on x and predict y if x=10.

(b) Fit a regression line of x on y and predict x if y =2.5.

8. The following data in 6 cities, calculate Pearson's coefficient of correlation between the density of population and death rate:

City	Area in kilometers	Population (in '000)	No. of Deaths
A	150	30	300
B	180	90	1440
C	100	40	560
D	60	42	840
E	120	72	1224
F	80	24	312

9. Calculate pairwise Spearman's coefficient of correlation between marks assigned to ten students by judges X, Y and Z in a certain competitive test as shown below:

S.No	1	2	3	4	5	6	7	8	9	10
Marks by judge X	52	53	42	60	45	41	37	38	25	27
Marks by Judge Y	65	68	43	38	77	48	35	30	25	50
Marks by Judge Z	34	45	50	46	67	34	36	25	20	56

Also, identify which of the two judges have similar opinion.

(Hint: Calculate the three rank correlation coefficients between judge X&Y, Y&Z and Z&X.)

10. Two automatic filling machines A and B are used to fill tea in 500 g cartons. A random sample of 100 cartons on each machine showed the following:

Tea contents (in g)	Machine A	Machine B
485-490	12	10
490-495	18	15
495-500	20	24
500-505	22	20
505-510	24	18
510-515	4	13

Comment on the performance of the two machines based on average filling and dispersion.

11. The share prices of a company in Mumbai and Kolkata markets during the last ten months are recorded below:

Month	Mumbai	Kolkata
January	105	108
February	120	117
March	115	120
April	118	130
May	130	100
June	127	125
July	109	125
August	110	120
September	104	110
October	112	135

Determine the mean and standard deviation of prices of share. In which market are the share prices more stable?

12. Calculate the coefficient of variation and Karl Pearson's coefficient of skewness from the data given below:

Marks (Less than):	20	40	60	80	100
No. of students: 15	30	48	52	60	

Comment on the results.