Research Methodology and Statistics (MSC 502)
Session: Monsoon Semester, 2022-2023

Correlation & Regression Analysis

1. Two judges mark the performances given by 7 contestants in a talent competition as given below. Does the data above exhibit any kind of consensus among the two judges.

Contestant	**A**	**B**	**C**	**D**	**E**	**F**	**G**
Judge 1 rank	3	1	4	7	3	6	5
Judge 2 rank	5	6	3	2	7	1	4

2. As part of a television phone-in, viewers were asked to rank 7 songs (A, B, …, G) in order of preference. Tanya's order of preference (favourite song first) was C, E, B, A, F, D, G. Jim's order of preference was E, A, C, F, D, B, G. Does Tanya and Jim has similar preferences of songs.

3. Let's say that we want to track the progress of a group of new employees of a large service organisation. We think we can judge the effectiveness of our induction and initial training scheme by analysing employee competence in weeks one, four and at the end of the six months. Let's say that Human Resource managers in their organisation have been urging the company to commit more resources to induction and basic training. The company now wishes to know which of the two assessments - the new employee's skills on entry or after week four - provides a better guide to the employee's performance after six months. Although there is a small sample here, let's assume that it is accurate. The raw data is given in the table below:

Name	**Skills on entry % score**	**Skills at week 4 % score**	**Skills at 6 months % score**
ab	75	75	75
bc	72	69	76
cd	82	76	83
de	78	77	65
ef	86	79	85
fg	76	65	79
gh	86	82	65
hi	89	78	75
ij	83	70	80
jk	65	71	70

4. A company wants to assess the impact of R&D expenditure (in Rs 000's) on its annual profit (in Rs 000's). The following table presents the information for the last eight years.

Year	R&D Expenditure	Annual Profit
2002	9	45
2003	7	42
2004	5	41
2005	10	60
2006	4	30
2007	5	34
2008	3	25
2009	2	20

Estimate the regression equation and predict the annual profit for the year 2012 for an allocated sum of Rs 4 50 000 as R&D Expenditure.

5. A random sample of eight drivers insured with a company and having similar auto insurance policies was selected. The following table lists their driving experiences (in years) and monthly auto insurance premiums.

Driving Experience (years)	Monthly Auto Insurance Premium
5	$64
2	87
12	50
9	71
15	44
6	56
25	42
16	60

i) Does the insurance premium depend on the driving experience or does the driving experience depend on the insurance premium? Do you expect a positive or a negative relationship between these two variables?
ii) Compute SS_{xx}, SS_{yy}, and SS_{xy}.
iii) Find the least squares regression line by choosing appropriate dependent and independent variables based on your answer in part (i).
iv) Interpret the meaning of the values of *a* and *b* calculated in part (iii).
v) Plot the scatter diagram and the regression line.
vi) Calculate R^2 and explain what it mean.
vii) Predict the monthly auto insurance premium for a driver with 10 years of driving experience.