Proficience – 2018

Data Analytics - Fundamentals

Mid Term Test 1 – 3rd March 2018

Duration : 1:00 hour
Maximum Marks : 25

Section A **Each Question carries 2 marks**

1. What are the basic steps to understand the given set of data for prediction?
2. 18 values were observed of a variable. Their mean was found to be 24.11. Another observation was subsequently made and the value observed was 35. What is the new value of the mean? If the sum of the squares of the deviation of these 19 values from their mean is 514.11, what is the standard deviation?
3. If $\bar{x}$ is defined as $\frac{\sum_{i=1}^{n} x_i}{n}$, then $\bar{x}$ is an unbiased estimate of the population mean. True or False?
4. The standard deviation of two variables, based on 19 observations, is 5.34 and 5.4. The covariance between these two variables, based on the same observations, is -21.00. Estimate the correlation coefficient between these two variables. Are these two variables linearly related?
5. Write the expression for the total probability of an event.
6. A certain firm has plants A, B, and C producing, respectively, 35%, 15% and 50%, of total output. The probabilities of a non-defective product are, respectively, 0.75, 0.95, and 0.85. A customer receives a defective product. What is the probability that it came from plant C?
7. Give one example when logarithmic transformation is not a good option to apply on a non-linear relation
8. In a regression model, the errors follow which distribution?
9. What is Occam's razor in the context of predictive models?
10. In a bivariate data, write the equations for determining the value of the regression coefficients which will minimize the sum of the squared errors.

(Section B is in the next page)

Section B **This section carries 5 marks**

1. Skin cancer rates have been steadily rising over recent years. It is thought that this may be due to ozone depletion.

   The following data are ozone depletion rates in various localities and the rates of skin cancer.

Ozone depletion (%)	Skin cancer rate (%)
5	1
7	1
13	3
14	4
17	6
20	5
26	6
30	8
34	7
39	10
44	9

   a. Fit a straight-line regression model to the data

   b. What is the rate of skin cancer if ozone depletion is 40%?