Daffodil International University
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Semester Final Examination, Spring 2024
Course Code: STA 227, Course Title: Statistics and Probability
Level: 2 Term: 2 Batch: 63

Time: 02:00 Hrs **Marks: 40**

Answer ALL Questions

[The figures in the right margin indicate the full marks and corresponding course outcomes. All portions of each question must be answered sequentially.]

1. *a)* A digital marketing agency manages multiple websites for various clients. They are interested in predicting website traffic to optimize their marketing strategies and resource allocation. They collect data on various factors that may influence website traffic, such as advertising expenditure, social media engagement, search engine ranking, and time of day. They want to build a predictive model that can accurately forecast website traffic based on these factors.
Define a multiple linear regression model to predict website traffic based on various IT-related factors. [2] CO1

b) Consider the experiment of rolling a pair of 6-sided dice simultaneously and show the Sample Space. Also **Find** the probability of same numbers. [2]

c) **Tell** the disadvantage of correlation coefficient 'r'. [1]

2. Suppose a student collected data regarding the height and weight of some students of his section. The data is shown below: CO2

Height (in inches)	65	68	70	63	72	66	69	71
Weight (in kg)	68	72	77	70	82	75	74	79

a) **Compare** which characteristic is showing more variability? [5]

b) **Show** the shape characteristics by the formula of skewness of more consistent variable from the above and comment on it. [5]

3. A large IT firm is interested to find a measure that can be used to fix the wages (yearly) of skilled engineers. On experimental basis, the data on the length of service and their yearly wages (in $00) from a group of 10 randomly selected skilled workers are given below: CO3

Length of service	**11**	**7**	**8**	**9**	**5**	**6**	**10**	**12**	**3**	**4**
Yearly wages	**14**	**11**	**13**	**10**	**9**	**10**	**14**	**16**	**6**	**7**

a) **Construct** the estimated regression model. [4]

b) **Identify** the Yearly wages of employees who have 15 years Length of service. [1]

c) **Develop** the "Coefficient of determination" of the model. [5]

4. *a)* A record of 200 engineers is collected from the HR department of a IT Firm and shown below: [5] CO4

Age/Degree	BSc only	MSc
Below 30	90	10
30-40	20	30
Above 40	40	10

Simplify the following probabilities of selected engineer **i)** has Msc) **ii)** MSc and age above 40) **iii)** BSc and age between 30 to 40) **iv**) MSc it is given that above 40.**v**) Age below 30 or MSc)

	b)	Probability that a patient recovers from a rare blood disease is 0.4. if 5 people are known to have this disease. **Simplify** the probabilities that, **i)** Exactly 3 people will survive **ii)** At least two survive **iii)** None will survive **iv)** find the mean and variance of the distribution.	[5]
	c)	**Justify** that there is a significant relationship between Length of service and yearly wages (use data question no-03) used at the 0.05 significance level, interpret your findings. [Tabulated Value: 2.306].	[5]

Formulas

Basic Concepts of Probability

General Rule of Addition

P (A or B) = P (A U B) = P(A) + P(B) – P(A∩B)

Special Rule of Addition

P (A or B) = P (A U B) = P(A) + P(B)

Conditional Probability

P(B|A) = P(A and B) / P(A)= $\frac{P(A\cap B)}{P(A)}$,

Hypothesis testing

Z test statistics $Z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$

t test statistics $t = \frac{\bar{X}-\mu}{s/\sqrt{n}}$

Testing the significance of the correlation coefficient

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$$

Measure of Dispersion

Range= X_{max} - X_{min}

Mean Deviation, M.D = $\frac{\sum_{i=1}^{n}|x_i-\bar{x}|}{n}$

Population variance

$\sigma^2 = \frac{\sum_{i=1}^{N}(X_i-\mu)^2}{N}$

Population standard deviation, $\sqrt{\sigma^2}$

Sample variance

$s^2 = \frac{\sum_{i=1}^{N}(X_i-\bar{X})^2}{n-1}$

Sample standard deviation, $\sqrt{s^2}$

Coefficient of variation for population, C.V= $\frac{\sigma}{\mu}\times 100$

Coefficient of variation for sample, C.V= $\frac{s}{\bar{x}}\times 100$

Correlation and Regression analysis

Regression Coefficient

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$$

Correlation Coefficient

$$r = \frac{N\Sigma xy-(\Sigma x)(\Sigma y)}{\sqrt{[N\Sigma x^2-(\Sigma x)^2][N\Sigma y^2-(\Sigma y)^2]}}$$

$$r = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}}$$

Binomial Distribution

$$f(x;n,p) = \begin{cases} \binom{n}{x} p^x q^{n-x} & \text{for } x = 0,1,2,\dots,n \\ 0; \text{ otherwise} \end{cases}$$

Poison Distribution

$$f(x;\lambda) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!} & \text{for } x = 0,1,2,\dots,\infty. \\ 0; \text{ otherwise} \end{cases}$$

Shape of the distribution

Coefficient of Skewness,

Sk= $\frac{3\times(Mean-Median)}{Standard\ deviation}$

Kurtosis $\beta_2 = \frac{\mu_4}{\mu_2^2}$