



Daffodil International University

Department of Computer Science and Engineering

Faculty of Science & Information Technology

Semester Final Examination, Spring 2023

Course Code: STA 221, Course Title: Statistics and Probability

Level: 2 Term: 2 Batch: All

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)	If you want to calculate the test scores on a final exam to better understand whether most students score close to the average or if there is a wide spread in test scores. Find the suitable measure of dispersion for the analysis and explain .	[1]	CO1							
	b)	Show the difference mutually exclusive and independent events? Illustrate with a hypothetical example.	[2]								
	c)	Meditation can produce a deep state of relaxation and a tranquil mind. During meditation, you focus your attention and eliminate the stream of jumbled thoughts that may be crowding your mind and causing stress. Choose the appropriate correlation is existing in this example. Explain	[2]								
2.	Software firm collected data for a sample of 9 programmer's firm's management wanted to predict salary using years of experience. The data are given below:			CO2							
	Salary(K)	24	43		47	51	35	38	42	40	49
	No. of experience (years)	4	7		6	7	5	5	6	5	6
	i) Construct a scatter diagram.		[2]								
	ii) Develop coefficient of determination using the above table.		[4]								
iii) Test the hypothesis that there is a significant relationship between salary and No. of experience (years) used at the 0.05 significance level, interpret your findings. [Tabulated Value: 1.86]		[5]									
3.	a)	By using the data 2		CO3							
	i) Build a regression model and interpret the results		[5]								
	ii) Build salary if number of experiences is 10 years		[1]								
	b)	If two dice are thrown solve the probabilities that the both would be odd, same number?	[3]								
4.	a)	The number of hits on a certain website follows poison distribution with a mean rate of 4 per minutes. Discover		CO4							
	i) the probability that at least 2 hits are received in 1 minutes		[2.5]								
	ii) the probability that between 2 to 3 hits are received in 2 minutes		[2.5]								
	b)	A package filling process at a cement fills a bag of cement to an average weight of 12 pounds, the standard deviation is 3 pounds. A sample of 15 bags have been taken and the mean was found to be 100 pounds. Assume that the weights of the bag follow normal distribution, and test the above process at 5% level of significance. [Z cal.=1.96]	[5]								
	c)	It is observed that 50% of mails are spam. There is a software that filters spam mail before reaching the inbox. Its accuracy for detecting a spam mail is 99% and chances of tagging a non-spam mail as spam mail is 5%. If a certain mail is tagged as spam discover the probability that it is not a spam mail.	[5]								

Formulas

Measure of Dispersion

$$\text{Range} = X_{\max} - X_{\min}$$

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

Population variance

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

Sample variance

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

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

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

Basic Concepts of Probability

General Rule of Addition

$$P(A \text{ or } B) = P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Special Rule of Addition

$$P(A \text{ or } B) = P(A \cup B) = P(A) + P(B)$$

Conditional Probability

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

Bayes theorem:

$$P[B_i | A] = \frac{P[B_i]P[A|B_i]}{\sum_{j=1}^n P[B_j]P[A|B_j]}; i = 1, 2, \dots, n$$

Hypothesis testing

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

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

Testing the significance of the correlation coefficient

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

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

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

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

Binomial Distribution

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

Poisson Distribution

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

Normal Distribution

$$f(x, \mu, \sigma^2) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}; \quad -\infty < x < \infty$$

Shape of the distribution

Coefficient of Skewness,

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

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