Daffodil International University

Faculty of Science & Information Technology
Department of Computer Science & Engineering
Final Examination, Summer 2025
Course Code: STA101, Course Title: Statistics and Probability
Level:3 Term:2 Batch: 64

Time: 02:00 Hours **Marks: 40**

Answer All the 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)	**Define** the conditions under which you can apply the Positive correlation and the Negative Correlation.	[2]	CO1
	b)	A group of computer science students is working on a project management tool. They want to help developers estimate how long a software module will take to build based on certain known factors. Where the independent variables are Lines of Code (LOC), Number of Functions, Developer Experience (in years). **Explain** the appropriate regression model from the above case in details.	[3]	
2.		The following data shows daily bandwidth usage (MB) and corresponding system latency (milliseconds) collected over 7 days:		CO2

Daily Bandwidth Usage (MB)	6.5	5	7	9.5	8	11	10.5
Processing Time (in milliseconds)	25	20	28	36	31	41	38

	a)	**Construct** the strength and direction of the relationship between daily bandwidth usage and processing time. **Interpret** the result as well.	[4]
	b)	**Build** a regression model using the given dataset. **Developed** the estimated coefficients represent in the given context with proper interpretation.	[3]
	c)	**Identify** the prediction error (residual) when the daily bandwidth usage is 11 GB.	[1]
	d)	**Construct** the coefficient of determination and **comment** on its meaning in terms of the model's explanatory power.	[2]
3.	a)	In a software company's code review system, 60 submissions are pending for review, with 40 correct and 20 containing errors. Two submissions are selected randomly one after another without replacement. **Examine** the probabilities of selecting: **i)** a second submission that contains errors, given the first submission also has errors. **ii)** two consecutive submissions that contain errors. **iii)** two consecutive submissions that are correct.	[5]	CO3
	b)	During the DIU Job Fair, several digital systems operate simultaneously. The career portal receives approximately 20 user requests per minute during peak hours. The CSE online judge handles about 6 code submissions every 10 minutes. The placement office email receives around 10 job applications per hour. **Simplify** the probabilities of a. receiving exactly 25 user requests per minute. b. no code submissions in the next 5 minutes c. receiving more than 7 job applications between 2 PM and 7 PM.	[5]	

$\frac{20}{60} \times \frac{20}{60}$

4.	*a)*	A company produces laptop batteries whose lifetimes are approximately normally distributed with a mean of 1000 hours and a standard deviation of 80 hours. **i) Inspect** the probability that a randomly selected battery lasts more than 1180 hours. **ii) Assume** the probability that a randomly selected laptop battery lasts between 890 and 1180 hours. **iii)** A quality control engineer tests a sample of 50 batteries with a sample mean lifetime of 970 hours. Assuming the population standard deviation is known to be 80 hours. **Test for** the 5% significance level to examine the validity of the company's claim that the average lifetime is at least 1000 hours. Conclude whether the test supports the claim.	[2] [3] [6]	CO4
	b)	Investigate whether there is a significant relationship between data usage and response time (using data from Question 2) at the 5% significance level. **Evaluate** your findings and justify your conclusion using the provided tabulated values.	[4]	

Useful Statistical Formulas

Correlation Analysis

Correlation Coefficient:

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

or,

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

Regression analysis

Regression Coefficients:

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

and $\hat{\beta}_1 = \frac{n\sum XY - (\sum x)(\sum y)}{n\sum x^2 - (\sum x)^2}$

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

Hypothesis Testing

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

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

Testing the significance of the correlation coefficient:

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

Probability Distribution

Binomial Distribution

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

Poison Distribution

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

Normal Distribution:

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

Standard Normal Variable: $Z = \frac{X - \mu}{\sigma}$

Tabulated value for Standard Normal Distribution:

$P(Z < 1.25) = 0.8944$ $\quad P(Z < -1.25) = 0.1056$

$P(Z > -2.25) = 0.9878$ $\quad P(Z > -1.375) = 0.9155$

$Z_{0.025} = \pm 1.96;$ $\quad Z_{0.05} = \pm 1.645;$

$Z_{0.01} = \pm 2.33;$ $\quad Z_{0.005} = \pm 2.575$

Tabulated value for t Distribution:

2.571 for 5 df, **2.447** for 6 df, **2.306** for 8 df

Basic Concepts of Probability

General Rule of Addition

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

Special Rule of Addition

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

Conditional Probability

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

$P(890 < x < 1180)$

$\frac{890 - \mu}{\sigma} < \frac{x - \mu}{\sigma} < \frac{1180 - \mu}{\sigma}$