Daffodil International University
Faculty of Science & Information Technology
Department of Computer Science & Engineering
Final Examination, Fall 2025
Course Code: STA101, Course Title: Statistics and Probability

Time: 02:00 Hours **Level:3** **Term:2** **Batch: 65** **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) **State** the conditions for applying the addition rule and multiplication rule of probability. [2]

b) A company uses an antivirus program to detect malware in users' computers. [3]
Case I: The software does not flag a file, but it is actually infected.
Case II: The software flags a file as malware, but it is actually safe.
i) **Clarify** which case is a Type I error and which is a Type II error.
ii) **Estimate,** what the power of the test means in the context of antivirus detection.

CO1

2. A data scientist records dataset size (thousands of samples) and model training time (minutes) for 11 experiments:

Dataset size (X1000 samples)	5	8	10	13	15	18	20	23	25	28	30
Model training time (minutes)	12	18	25	31	37	44	49	56	61	69	73

CO2

a) **Calculate** the prediction error (residual) for the dataset of size 15 thousand samples. Also interpret the regression parameter. [5] [1]

b) **Compute** the coefficient of determination and comment on its meaning in terms of the model's explanatory power. [3] [1]

3 a) At a DIU job fair, 200 students are evaluated for interviews at Company A and Company B. The data collected is as follows: **Attain** the probability of

CO3

	B Shortlisted	B Not Shortlisted
A Shortlisted	24	36
A Not Shortlisted	40	140

i. a student is shortlisted by at least one company. [1]
ii. a student is shortlisted by B given shortlisted by A. [2]
iii. Also **attain** whether being shortlisted by Company A and Company B are independent events; justify your answer. [2]

b) A cybersecurity operations center monitors how often malware intrusion attempts occur. Historical records show that, on average, 3 intrusion attempts occur per hour, and the attempts happen independently of one another at a steady rate. **Construct** the probabilities

i. that exactly 5 intrusion attempts occur in the next hour? [1]
ii. that at least 2 intrusion attempts occur in the next 30 minutes? [2]
iii. On average, how many hours would you expect to pass before seeing 10 total intrusion attempts? [2]

4 a) The time (in minutes) between consecutive cash withdrawals at an ATM follows an exponential distribution with a mean of 5 minutes.

i) Detect the probability of occurring the next withdrawal within 8 minutes. [2]
ii) Identify the probability that the next withdrawal takes more than 12 minutes. [2]

CO4

b)	A laptop manufacturer claims that the average boot-up time of its new laptops is 30 seconds. A quality control engineer tests a sample of 20 laptops and finds a sample mean boot-up time of 31.5 seconds with a variance of 4 second^2.		
	i) **Detect** the probability that a randomly selected laptop boots up between 29 and 33 seconds.	[1]	
	ii) Using the sample data, perform a hypothesis test at the 5% significance level to examine the manufacturer's claim that the average boot-up time is more than 30 seconds. **Diagnose** whether the test supports the claim.	[5]	
c)	Investigate whether there is a significant relationship between Dataset size and Model training time (using data from Question 2) at the 5% significance level. **Conclude** your findings and justify your conclusion using the provided tabulated values.	[5]	

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:

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

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

or,

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

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

Probability Distribution

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

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

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

Exponential distribution

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0, \\ 0 & x < 0. \end{cases}$$

Hypothesis Testing

$$Z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} \qquad t = \frac{\bar{X}-\mu}{s/\sqrt{n}}$$

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \qquad |t| = \left|\frac{\bar{X}_1 - \bar{X}_2}{s\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}\right| \quad \text{where, } s^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$$

$$|t| = \left|\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)}}\right| \quad \text{where, with} \quad \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2 - 1}} \; d.f$$

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

Tabulated value for Standard Normal Distribution:

$P(Z > 1.5) = \mathbf{0.0068}$, $P(Z < -0.5) = \mathbf{0.3085}$, $P(Z < -1.5) = \mathbf{0.0668}$, $P(Z < .5) = \mathbf{0.6915}$

$Z_{0.025} = \pm\mathbf{1.96}$; $Z_{0.05} = \pm\mathbf{1.645}$; $Z_{0.01} = \pm\mathbf{2.33}$; $Z_{0.005} = \pm\mathbf{2.575}$

Tabulated value for t Distribution:

1.725 for 20 df; 1. 729 for 19 df; 1.833 for 9 df; 1.812 for 10 df: 2.306 for 8 df