UNIVERSITY OF RUHUNA

Faculty of Engineering

End-Semester 4 Examination in Engineering: November 2022

Module Number: IS4305 **Module Name: Probability and Statistics (C 18)**

[Three Hours]

[Answer **all questions**, each question carries twelve marks]

Q1. a) List two characteristics of a Bernoulli trial.

[1 Mark]

b) From a manufacturing process, a sample of 15 units is randomly chosen each day to check the percent defective of the process. Based on historical data it is known that the probability of a unit being defective is 0.05. If two or more defectives are found, the process will be stopped at any time. This procedure is used to alarm the process in case the number of defective units increases.

Let X be the number of defective units found from the sample.

- i Suggest a suitable probability distribution to model X.
- ii Write the probability density/ mass function of the above-identified distribution.
- iii What is/ are the value/s of the parameter/s of the distribution?
- iv What is the probability that on any given day the production process will be stopped?
- v Find the expected value and the variance of X.
- vi Suppose that the probability of a defective has increased to 0.07. What is the probability that on any given day the production process will not be stopped?

[9 Marks]

c) Consider 50 independent random variables each having a Poisson distribution with parameter $\lambda = 0.03$ and the sum of the variables, $S_{50} = X_1 + X_2 + \cdots . + X_{50}$.

- i Evaluate $P(S_n \geq 3)$ using central limit theorem.
- ii Compare your answer with the exact value of the probability. (Assume that S_n has a Poisson distribution, with parameter $n\lambda > 0$)

[2 Marks]

Q2. a) Maga Engineering Company is looking for a construction engineer from four candidates $W_1, W_2, W_3, and\ W_4$ for a specific project and their probabilities of being selected are 0.3, 0.2, 0.4, and 0.1, respectively. The probabilities of the project being approved are 0.35, 0.85, 0.45, and 0.15, depending on which of the four candidates is chosen. Using Bayes' theorem, calculate the probability of the project being approved.

[3 Marks]

b) A survey was conducted to analyze the factors that affect to improve the production unit of an industry. The age distribution of workers is recorded in **Table Q2.b)** as follows.

Table Q2. b)

Age(years)	Number of workers
15 – 23	15
24 – 32	46
33 – 41	49
42 – 50	32
51 – 59	28
60 – 68	10

i Draw a histogram for the above distribution.

ii Draw the frequency polygon on the histogram.

iii Comment on the shape of the age distribution of workers.

iv Calculate the mean, variance, and mode of the age distribution of workers.

v All the workers aged 60 years and above have retired as a result of the survey. Find the new mean.

[9 Marks]

Q3. a) Three different statistics are being considered for estimating a population characteristic. The density curves of sampling distributions of the three statistics are shown in **Figure Q3.a)**.

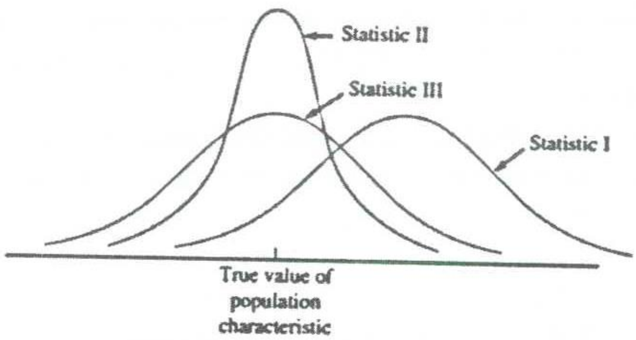

Figure Q3. a)

State the most suitable statistic and explain.

[4 Marks]

b) The breaking strengths of a random sample of 20 bundles of "Type A" fibers have a sample mean 436.5 and a sample standard deviation 11.9. Here, μ is the true average breaking strength of "Type A" fibers.

i Construct 95% and 99% two-sided confidence intervals for μ and compare the lengths of the confidence intervals.

ii How many additional data observations should be obtained to construct a 99% two-sided confidence interval for the true average breaking strength with a length no longer than 10.0?

iii If the experimenter is interested in testing the hypotheses,

$$H_0 : \mu = 430 \;\; Vs \;\; H_1 : \mu \neq 430$$

Test the experimenter's claim.

[8 Marks]

Q4. a) An experiment was conduct to investigate two types of resin mixtures of different expansion and contraction properties that affect the chances of the tiles becoming cracked. For thus, a combination of two surveys of tiles on a group of buildings A and B were done.

The two group of buildings were constructed about the same time and have exterior walls composed of the same type of tiles. However, the tiles on buildings B were cemented into place with a different resin mixture than that used on buildings A. The group of construction engineers were found that a total of 406 cracked tiles out of 6000 of buildings A. In buildings B, they found that a total of 83 cracked tiles out of 2000.

Does this experiment provide any evidence that the two types of resin mixtures of different expansion and contraction properties that affect the chances of the tiles becoming cracked?

[6 Marks]

b) A factory has three production lines producing glass sheets that are all supposed to be of the same thickness. A quality inspector takes a random sample of 30 sheets from each production line and measures their thicknesses. **Table Q3.b)** gives the respective summary statistics.

Table Q3.b)

	Production Line 1	Production Line 2	Production Line 3
$\bar{y}_{i.}$	3.015	3.018	2.996
s_i	0.107	0.155	0.132

What conclusions should the quality inspector draw?

(In the usual notations, $SSTr = \frac{\sum_{i=1}^{I} y_{i.}^2}{n_i} - \frac{y_{..}^2}{N}$; $N = \sum_{i=1}^{I} n_i$; $SSE = \sum_{i=1}^{I}(n_i - 1)s_i^2$

[6 Marks]

Q5. a) Scatter plots of three sets of samples were obtained and displayed in **Figure Q5.a)**. The correlation coefficients of three samples were computed and the values are given in ascending order below.

$$-0.76, 0.00, 0.82$$

Describe the nature of the relationship between the two variables x and y of each scatter plot with the appropriate correlation coefficient.

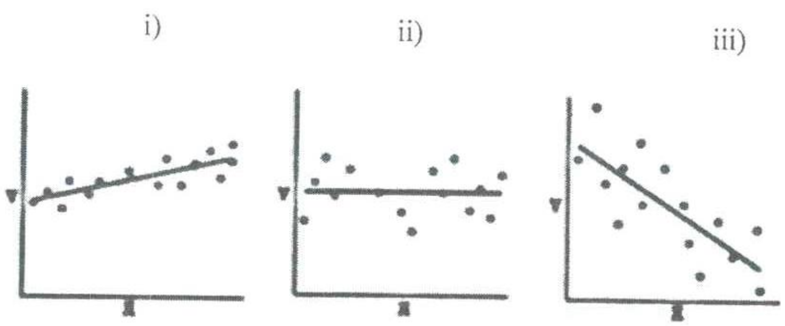

Figure Q5. a)

[2 Marks]

b) Consider the simple linear regression model; $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, that fits a straight line through a set of paired data observations $(x_1, y_1), \dots, (x_n, y_n)$. The error terms $\epsilon_1, \dots, \epsilon_n$ are taken to be independent observations from a $N(0, \sigma^2)$ distribution. The intercept parameter β_0 and the slope parameter β_1 are estimated from the data set as follows.

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

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

Values of modulus of elasticity (MOE in GPa) X_i and flexural strength (in MPa) Y_i were determined for a sample of concrete beams of a certain type and the results were summarized as follows.

$$n = 27, \sum_{i=1}^{n} x_i = 1217.9, \sum_{i=1}^{n} y_i = 219.8, \sum_{i=1}^{n} x_i^2 = 59512.81,$$

$$\sum_{i=1}^{n} x_i y_i = 10406.5, \sum_{i=1}^{n} y_i^2 = 1860.94$$

i Calculate $\hat{\beta}_1$ and $\hat{\beta}_0$.

ii Determine the equation of the estimated regression line (least square line) for predicting strength from modulus of elasticity.

iii Predict strength for a beam whose modulus of elasticity is 40.

iv Is it comfortable using the least square line to predict strength when modulus of elasticity is 100?

[5 Marks]

c) An experiment was conducted to determine the energy content of the waste be available for an efficient design of certain type of municipal waste incinerators. The accompanying data were analyzed and the output for obtaining the fitted line (regression line) using the software "MINITAB" is given below.

Regression Analysis: Energy Content versus Plastics, Paper, Garbage, Water

Predictor	Coef	SE Coef	T	P
Constant	2244.9	177.9	12.62	0.000
Plastics	28.925	2.824	10.24	0.000
Paper	7.644	2.314	3.30	0.003
Garbage	4.297	1.916	2.24	0.034
Water	-37.354	1.834	-20.36	0.000

S = 31.4828   R-Sq = 96.4%   R-Sq(adj) = 95.8%

i Find the values of the estimated regression coefficients.

ii Write down the estimated regression equation.

iii State and test the hypothesis to decide whether the model fit to the data specifies a useful linear relationship between energy content and one of the predictors "Plastics".

iv Does "Garbage" provide useful information about energy content?

[5 Marks]