UNIVERSITY OF RUHUNA

Faculty of Engineering

End-Semester 4 Examination in Engineering: February 2020

Module Number: IS4305 **Module Name: Probability and Statistics (Curriculum 2018)**

[Three Hours]

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

Q1. a) The scores of a sample of 20 students on college entrance examination are

36 44 78 84 66 48 50 69 74 70 52 54 59 61 57 56 60 58 64 65

i Construct a relative frequency histogram of the data (Consider equal class width and use 35-45 as the beginning class interval).

ii If the college wants to accept the top 35% of the applicants, what should the minimum score be?

iii If the university sets the minimum score at 45, what percent of the applicants will be accepted?

[3.0 Marks]

b) Table 1.1 shows the times (in seconds) of the top 8 finishers in the final and semi-final rounds of the male students' 100-meter backstroke event in the school aquatic competition respectively.

Table 1.1

Final Round	46	46.6	46.6	47	47.1	47.2	47.7	47.9
Semifinal Round	46.7	47.2	47.3	47.3	47.4	47.5	47.7	47.8

i Find the mean and the variance of time taken by students for each final and semi-final rounds.

ii Use part b) i to compare the student's performances in final and semi-final rounds.

[5.0 Marks]

c) A new analytical method is used to detect three different contaminants: organic pollutants, volatile solvents, and chlorinated compounds. The makers of the test claim that it can detect high levels of organic pollutants with 99.7% accuracy, volatile solvents with 99.95% accuracy, and chlorinated compounds with 89.7% accuracy. If a pollutant is not present, the test does not signal. Samples are prepared for the calibration of the test and 60% of them are contaminated with organic pollutants, 27% with volatile solvents, and 13% with traces of chlorinated compounds. A test sample is selected randomly.

i What is the probability that the test will signal?

ii If the test signals, what is the probability that chlorinated compounds are present?

[4.0 Marks]

Q2. a) In a quality control program, a manufacturer randomly selects two glass sheets from each lot of seven for inspection.

i List the different possible outcomes.

ii If the first, second and fourth glass sheets are the only defectives in a lot of seven, find the probability distribution of the number of defective glass sheets observed among those inspected.

iii Find the cumulative distribution function $F(x) = P(X \leq x)$ for all x and use it to calculate $P(2 \leq X \leq 5)$.

iv Find the expected number of defective glass sheets.

[5 Marks]

b) The moment–generating function of the random variable X is given by,

$$M_X(t) = E(e^{tX}).$$

Then

$$\left.\frac{d^r M_X(t)}{dt^r}\right|_{t=0} = \mu_r', \text{ where } \mu_r' = E(X^r),\ r = 1,2,3,\ldots.$$

i Find the moment-generating function of the random variable X having a normal probability distribution $f(x)$

$$f(x) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, & -\infty < x < \infty \\ 0, & otherwise \end{cases}$$

ii Hence, find the mean and variances of the random variable X.

[7.0 Marks]

Q3. a) Let $x_1, x_2, \ldots, x_n$ are a random sample from a geometric distribution with parameter $p, 0 \leq p \leq 1$. The probability mass function is given by

$$f(x,p) = p(1-p)^{x-1}, \qquad 0 \leq p \leq 1, \qquad x = 1,2,3,\ldots$$

The expected value of the geometrically distributed random variable is $1/p$.

i Find the maximum likelihood estimator of p.

iii Use the following data to find the estimate for p.

2, 5, 7, 43, 18, 19, 16, 11, 22, 4, 34, 19, 21, 23, 6, 21, 7, 12

[6.0 Marks]

b) A study was carried to test the thickness of plastic sheets produced by a machine as the viscosity of the liquid mold makes some variation in thickness measurements. Thickness measurements (in millimeters) of ten plastic sheets produced on a particular shift are 226, 228, 226, 225, 232, 228, 227, 229, 225, 230.

It is stated that the true standard deviation of thickness exceeds 1.5 millimeters, there is cause to be concerned about the product quality.

i State the assumption can be made about the population distribution.

ii Do the data substantiate the suspicion that the process variability exceeded the stated level on this particular shift? Use $\alpha = 0.05$.

iii Construct a 95% confidence interval for the true standard deviation of the thickness of sheets produced on this shift.

[6.0 Marks]

Q4. a) A study was carried to test the relationship between facility conditions at gasoline stations and aggressiveness in the pricing of gasoline. The corresponding observed counts based on a sample of 400 stations are given in Table 4.1.

Table 4.1

		Observed Pricing Policy		
		Aggressive	Neutral	Nonaggressive
Condition	Substandard	25	35	15
	Standard	40	60	75
	Modern	50	70	30

Does the data suggest that the facility conditions and pricing policy are independent of one another? Use a chi-square test at 0.05 level.

[5.0 Marks]

b) A study was designed to investigate the iron content of some of the foods cooked in Aluminum, Clay and Iron pots. The iron content (mg/100g food) of the food cooked in each of the three types of pots is summarized by the Table 4.2.

Table 4.2

Type of Pot (i)	n_i (i^{th} sample size)	$\bar{y}_{i.}$	s_i
Aluminum	4	2.06	0.25
Clay	4	2.18	0.62
Iron	4	4.68	0.63

Use this data and a significance level of 0.01 to test the null hypothesis of no difference in mean iron content of foods for three types of pots.

Total sum of squares: $SST = \sum_{i=1}^{I}\sum_{j=1}^{J} y_{ij}^2 - \frac{y_{..}^2}{IJ}$

Treatment sum of squares: $SSTr = \frac{1}{J}\sum_{i=1}^{I} y_{i.}^2 - \frac{y_{..}^2}{IJ}$

$$SST = SSTr + SSE$$

[7.0 Marks]

Q5. a) Describe the relationships between the two variables x and y given by the scatter diagrams in Figure 5.1.

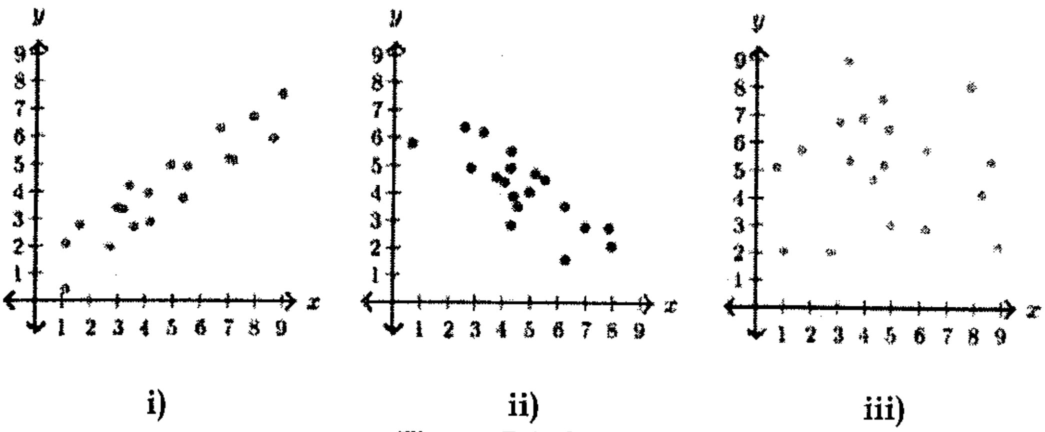

Figure 5.1: Scatter Diagrams

[1.0 Mark]

b) The data on x - shear force (kg) and y = percent fiber dry weight is summarized as:

$$n = 18, \quad \sum_{i=1}^{18} x_i = 1950, \quad \sum_{i=1}^{18} x_i = 1950, \quad \sum_{i=1}^{18} x_i^2 = 251{,}970,$$

$$\sum_{i=1}^{18} y_i = 47.92, \quad \sum_{i=1}^{18} y_i^2 = 130.6074, \quad \sum_{i=1}^{18} x_i y_i = 5530.92$$

i Calculate the value of the sample correlation coefficient and hence describe the nature of the relationship between the two variables.

ii The least square estimates of the slope β_1 and the intercept β_0 of the true regression line respectively are:

$$\hat{\beta}_1 = \frac{\sum_i^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_i^n (x_i - \bar{x})^2} \text{ and } \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

If the simple linear regression model is suitable for the data, find the regression equation.

iii Find the proportion of observed variation in percent fiber dry weight could be explained by the model relationship (use $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = 0.2033$).

[6.0 Marks]

c) Table 5.1 represents the data on wear of a bearing y and its relationship to oil viscosity x_1 and load x_2.

Table 5.1

y	x_1	x_2
193	1.6	851
172	22	1058
113	33	1357
230	15.5	816
91	43	1201
125	40	1115

i If the linear regression model is suitable for the data, state the regression equation in matrix notation.

ii Use the data in Table 5.1 to represent the matrices for $\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{\beta}$ and $\boldsymbol{\varepsilon}$.

iii If $\underline{\hat{\beta}} = (350.9943 \quad -1.272 \quad -0.1539)'$, write down the estimated regression equation.

iv Hence, predict wear when oil viscosity is 20 and load is 1200.

[5.0 Marks]