UNIVERSITY OF RUHUNA

Faculty of Engineering

End-Semester 4 Examination in Engineering: November 2017

Module Number: IS4301 **Module Name: Probability and Statistics**

[Three Hours]

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

Q1. a) Briefly explain the followings.

i Sample Statistic

ii Simple Random Sampling

iii Central Limit Theorem

[3.5 Marks]

b) Assume that the followings are the Graduation Rates (GR) for Student Athletes and All Students who entered in 11 different Universities in a particular year.

GR for All Students: 75, 68, 62, 82, 64, 51, 92, 56, 80, 64, 74

GR for Student Athletes: 65, 66, 71, 68, 56, 65, 93, 50, 78, 72, 55

i Construct separate five-number summary diagrams and then construct the corresponding Box and Whisker plots in one diagram.

ii Based on the Boxplots in part (i), what can you say about the differences between the graduation rates of Student Athletes and All Students in 11 Universities?

iii How do you classify the shapes of the two distributions?

iv Would you use means or medians to compare the centers of the two distributions?

[5.5 Marks]

c) A certain auditorium has 30 rows of seats. Row 1 has 11 seats, while Row 2 has 12 seats, Row 3 has 13 seats, and so on to the back of the auditorium where Row 30 has 40 seats. A door prize is to be given away by randomly selecting a row (with equal probability of selecting any of the 30 rows) and then randomly selecting a seat within that row (with each seat in the row equally likely to be selected).

i Find the probability that Seat 15 was selected given that Row 20 was selected.

ii Show that

$$P(Seat\ 15)=\sum_{k=5}^{30}\left(\frac{1}{k+10}\right)\frac{1}{30}.$$

iii Find the probability that Row 20 was selected given that Seat 15 was selected.

[5.0 Marks]

Q2. a) Define appropriate random variables and list all the values for each of the following experiments.

i A Manager is interested to find the repair cost for a particular machine breakdowns due to an electrical failure within the machine, mechanical failure of some component of the machine or operator misuse. Assume that the electrical failures generally cost an average of 20,000 rupees, mechanical failures have an average repair cost of 35,000 rupees, and operator misuse failures have an average repair cost of only 5000 rupees.

ii For safety purposes, a factory manager is interested in how many factory floor accidents occur in a given year.

iii A company manufactures metal cylinders that are used in the construction of a particular type of engine. The metal cylinders it manufactures can have a diameter anywhere between 49.5 and 50.5 mm. Suppose that the company manager is interested to find the probability that a metal cylinder has a diameter between 49.8 and 50.1 mm.

[3.5 Marks]

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

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

Then

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

i Find the moment-generating function of the binomial random variable X, based on "n" number of trials and the success probability "p".

ii Then verify that the mean of the random variable X, $\mu = np$ and the variance of the random variable X, $\sigma^2 = np(1-p)$.

[5.0 Marks]

c) The resistance X of an electrical component has a probability density function given by

$$f(x) = \begin{cases} Kx(130 - x^2) & ; 10 \le x \le 11 \\ 0 & ; elsewhere \end{cases}$$

i Find the value of K that makes $f(x)$ a density function.

ii Find the cumulative distribution function.

iii What is the probability that the electrical component has a resistance between 10.25 and 10.5?

iv What is the expected value of the resistance?

v What is the standard deviation of the resistance?

[5.5 Marks]

Q3. a) Suppose that $E(X_1) = \mu$, $Var(X_1) = 10$, $E(X_2) = \mu$, and $Var(X_2) = 15$, and consider the following three point estimators

$$\hat{\mu}_1 = \frac{X_1}{2} + \frac{X_2}{2}$$

$$\hat{\mu}_2 = \frac{X_1}{4} + \frac{3X_2}{4}$$

$$\hat{\mu}_3 = \frac{X_1}{6} + \frac{X_2}{3} + 9$$

i Is any one of them unbiased?

ii Which one has the smallest variance?

[3.5 Marks]

b) The breaking strength of a long wire was tested, using a sample of 15 equal lengths. The results are shown below.

76, 75, 74, 72.5, 72, 69, 69, 65, 64, 63, 62, 61, 58, 52, 48

Assume that the breaking strengths are normally distributed with mean μ and the variance σ^2.

i Use Maximum Likelihood Estimation Method to find estimators for μ and σ^2.

ii Show that the estimator of μ is an unbiased estimator.

iii Compute the estimates of μ and σ^2 for the given data.

[5.0 Marks]

c) Two different brands of latex paint are being considered for use. Drying time in hours is being measured on specimen samples of the use of the two paints. Fifteen specimens for each were selected and the drying times are as follows:

Paint A					Paint B				
3.5	2.7	3.9	4.2	3.6	4.7	3.9	4.5	5.5	4.0
2.7	3.3	5.2	4.2	2.9	5.3	4.3	6.0	5.2	3.7
4.4	5.2	4.0	4.1	3.4	5.5	6.2	5.1	5.4	4.8

Assume that the drying time is normally distributed with $\sigma_A = \sigma_B$.
Find a 95% confidence interval for the difference of the true average drying time between two samples.

[5.5 Marks]

Q4. a) The Table 4.1 shows the radar detection distances in miles for 15 targets. The observations x_i are for the standard system and the observations y_i are for the new system. An initial look at the data indicates that the detectability of the targets varies from about 45 miles in some cases to over 55 miles in other cases, and this confirms the advisability of a paired experiment.

Table 4.1: Radar Detection Distances

Target	Standard Radar System (x_i)	New Radar System (y_i)
1	48.4	51.1
2	47.7	46.4
3	51.3	50.9
4	50.4	49.8
5	47.1	47.9
6	53.0	53.2
7	48.9	46.7
8	52.0	54.4
9	51.1	49.8
10	47.3	47.4
11	50.1	50.6
12	46.5	47.9
13	52.0	52.3
14	51.9	52.9
15	49.1	50.6

i Write down the null and alternative hypotheses for analyzing the data set if the experimenter is interested in ascertaining whether or not the new radar system can detect targets at a greater distance than the standard system.

ii Test the hypotheses in part (i).

iii Why are the two data samples paired? Why did the experimenter decide to do perform a paired experiment rather than an unpaired experiment?

[4.0 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 $n = 30$ sheets from each production line and measures their thicknesses. The glass sheets from the first production line have a sample average of $\bar{x}_{1\bullet} = 3.105$ mm with a sample standard deviation of $s_1 = 0.107$ mm. The glass sheets from the second production line have a sample average of $\bar{x}_{2\bullet} = 3.018$ mm with a sample standard deviation of $s_2 = 0.155$ mm, while the glass sheets from the third production line have a sample average of $\bar{x}_{3\bullet} = 2.996$ mm with a sample standard deviation of $s_3 = 0.132$ mm. What conclusions should the quality inspector draw?

[5.0 Marks]

c) Table 4.2 shows a data set of the number of errors found in a total of $n = 85$ software products. For example, 3 of the products had no errors, 14 had one error, and so on. Is it plausible that the number of errors has a Poisson distribution with $\lambda = 3$?

Table 4.2: Number of Errors

No. of Errors	0	1	2	3	4	5	6	7	8
Frequency	3	14	20	25	14	6	2	0	1
Exp. Frequency	4.23	12.70	19.04	14.28	8.57	4.28	1.84	0.69	0.33

[5.0 Marks]

Q5. a) Values of modulus of elasticity (MOE, the ratio of stress in, GPa) and flexural strength (a measure of the ability to resist failure in bending, in MPa) were determined for a sample of concrete beams of a certain type, resulting in the following data.

MOE : 33.0, 33.2, 33.7, 35.3, 35.5, 36.1, 36.2, 36.3, 37.5, 37.7, 38.7, 38.8, 39.6, 41.0, 42.8

Strength : 6.9, 7.2, 7.3, 7.4, 7.5, 7.6, 7.5, 7.5, 7.7, 7.8, 7.8, 7.7, 7.6, 7.7, 7.9

i Construct a Scatter plot. Does a scatter plot support the choice of the simple linear regression model? Explain.

ii If the answer for part (i) is "yes", then obtain the equation of the least squares line for predicting strength from modulus of elasticity and then predict strength for a beam whose modulus of elasticity is 40.

iii Find the coefficient of determination. Does this value suggest that the simple linear regression model effectively describes the relationship between the two variables? (use as SSE = 0.2587, SST = 0.9560)

iv Calculate a confidence interval with a confidence level of 95% for the slope β_1 of the population regression line, and interpret the resulting interval.

[9.0 Marks]

b) Consider the multiple linear regression model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ to the Car plant electricity usage data set given in Table 5.1.

Table 5.1: Car Plant Electricity Usage

Month	Electricity Usage (Million kWh)	Production (Million Rs.)	Cooling degree days
January	2	5	0
February	2	4	0
March	2	4	5
April	3	5	10

i Write down the vector of observed values of the response variable Y and the design matrix X.

ii Calculate X'X.

iii Find the estimates of β_0, β_1 and β_2.

[5.0 Marks]