UNIVERSITY OF RUHUNA

Faculty of Engineering

End-Semester 4 Examination in Engineering: December 2015

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

[Three Hours]

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

Q1.

a) Explain clearly the following ~~three~~ Two sampling methods through at least one example:

i Simple Random Sampling,

ii Stratified Sampling.

[4.0 Marks]

b) A researcher is investigating a new method for applying the nickel layer onto the bond pads in the substrate and the thickness of the nickel layer is of particular interest. An assembly with 16 bond pads is examined and the nickel layer thickness is measured for each pad, resulting in the data set shown below.

2.72, 2.79, 2.81, 2.75, 2.77, 2.76, 2.75, 2.75, 2.81, 2.75, 2.74, 2.77, 2.79, 2.78, 2.80, 2.76

i Find the five-number summary for the nickel layer thickness.

ii Display the five-number summary in a box-and-whisker-plot.

iii What does the boxplot tell you about the nickel layer thickness?

[5.0 Marks]

c) A factory has two assembly lines, each of which is shut down (S), operating at partial capacity (P), or at full capacity (F). The sample space and its probabilities are given below.

$$\{(SS),(SP),(SF),(PS),(PP),(PF),(FS),(FP),(FF)\}$$

$$P(SS)=0.02,\quad P(SP)=0.06,\quad P(SF)=0.05,$$
$$P(PS)=0.07,\quad P(PP)=0.14,\quad P(PF)=0.20,$$
$$P(FS)=0.06,\quad P(FP)=0.21,\quad P(FF)=0.19,$$

where, (SP) denotes that the first assembly line is shut down and the second one is operating at partial capacity. What is the probability that

i both assembly lines are shut down?

ii neither assembly line is shut down?

[2.0 Marks]

d) A company sells five type of wheelchairs, with type A being 12% of the sales, type B being 34% of the sales, type C being 7% of the sales, type D being 25% of the sales, and type E being 22% of the sales. In addition, 19% of the type A wheelchair sales are motorized, 50% of the type B wheelchair sales are motorized, 4% of the type C wheelchair sales are motorized, 32% of the type D wheelchair sales are motorized, and 76% of the type E wheelchair sales are motorized. If a motorized wheelchair is sold, what is the probability that it is of type C?

[3.0 Marks]

Q2.

a) A manager supervises the operation of three power plants, X, Y, and Z. At any given time, each of the three plants can be classified as either generating electricity (1) or being idle (0). With the notation $(0,1,0)$ used to represent the situation where plant Y is generating electricity but plants X and Z are both idle.

i Define the sample space for the status of the three plants at a particular point in time.

ii If the manager's interest is directed only at the number of plants that are generating electricity, then define a random variable and list it all the values.

iii The probability values for the three power plants are given here.

$$P(0,0,0) = 0.07 \quad P(1,0,0) = 0.16 \quad P(0,0,1) = 0.04 \quad P(1,0,1) = 0.18$$
$$P(0,1,0) = 0.03 \quad P(1,1,0) = 0.21 \quad P(0,1,1) = 0.18 \quad P(1,1,1) = 0.13$$

Find the probability mass function of the random variable.

iv Find the expected number of power plants generating electricity.

[8.0 Marks]

b) The discrete random variable X has a probability distribution given by

$P(X = x) = x/10$ for $x = 1,2,3,4$.

Find:

i $E(X)$

ii $E(X^2)$

iii $E(X^2 + 2X - 3)$

[2.0 Marks]

c) The resistance X of an electrical component has a probability density function $f(x) = Ax(130 - x^2)$ for resistance values in the range $10 \leq x \leq 11$.

i Calculate the value of the constant A.

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

[4.0 Marks]

Q3.

a) A quality inspector at a glass manufacturing company inspects sheets of glass to check for any slight imperfections. Suppose that the number of these flaws in a glass sheet has a Poisson distribution with parameter λ. The expected number of flaws per sheet is only 0.5.

i Find the probability that there are no flows in a sheet.

ii Find the probability that there are exactly two flaws in a sheet.

iii Find the probability that there are two or more flaws which are scrapped by the company.

[7.0 Mark]

b) A company manufactures concrete blocks that are used for construction purposes. Suppose that the weights of the individual concrete blocks are normally distributed with a mean value of $\mu = 11.0Kg$ and a standard deviation of $\sigma = 0.3Kg$.

i Find the probability that a concrete block weight less than $10.5Kg$ and explain your result.

ii Suppose that a wall is constructed from 24 concrete blocks. What is the distribution of the total weight of the wall?

iii If the average weight of a randomly selected sample of concrete blocks is $10.6Kg$, then construct a 95% confidence interval for the population mean weight of the concrete blocks.

[7.0 Mark]

Q4.

a) Explain the each of the following with an example:

i A Statistic,

ii A Parameter.

[2.0 Marks]

b) Let $X_1, X_2, \ldots, X_n$ be a random sample from the Poisson distribution with the probability mass function

$$f(x \mid \mu) = \frac{e^{-\mu}\mu^{x}}{x!} \quad ; x = 0, 1, \ldots$$

i Use the maximum likelihood estimation method to estimate the parameter μ.

ii If the following sample represents the number of flaws per glass sheet recorded by a student in his experiment. By assuming that the distribution of the number of flaws per sheet has a Poisson distribution, estimate the mean of the distribution.

0, 1, 1, 1, 0, 0, 0, 2, 0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 2, 0, 0, 3, 1, 2, 0, 0, 1, 0, 0

[6.0 Marks]

c) Summary statistics for a sample of 60 metal cylinder diameters are given below.

Sample size: 60	Mean: 49.999	Standard deviation: 0.134
Maximum: 50.360	Minimum: 49.740	Median: 50.010

Construct

i 95%, and

ii 99%,

two-sided confidence intervals for the mean cylinder diameter. Interpret your answers.

[3.0 Marks]

d) A sample of 300 cars having cellular phones and one of 400 cars without phones were tracked for 1 year. The Table Q4 gives summary of the number of these cars involved and not involved in accidents over that year. Test the null hypothesis that having a cellular phone in a car and being involved in an accident are independent. Use the 5% level of significance.

[3.0 Marks]

Q5.

a) Describe each of the following relationships through scatter diagrams.

i Strong, positive linear relationship

ii Strong, negative linear relationship

iii No relationship

[2.0 Marks]

b) A manager of the car plant wishes to investigate how the plant's electricity usage depends upon the plant's production. The data set given in Table Q5 is compiled and provides the plant's production and electrical usage for each month of the previous year. The electrical usage is in units of a million kilowatt-hours, and the production is measured as the value in million-rupees units of the cars produced in that month.

i Plot the Electricity usage against plant's production.

ii What is the sample correlation coefficient between the plant's production and Electricity usage? Interpret your answer.

iii Fit a linear regression model with electricity usage as the response variable and the production as the predictor (explanatory) variable.

iv If a production level of 5.5 million rupees worth of cars is planned for next month, then predict that the electricity usage.

[12.0 Marks]

Table Q4: Number of Cars Involved in Accidents over the Year

	Accident	No Accident
With Cellular Phone	22	278
Without Cellular Phone	26	374

Table Q5: Plant's Production and Electrical Usage for each Month

Month	Production (million rupees)	Electricity Usage (million Kwh)
January	4.51	2.48
February	3.58	2.26
March	4.31	2.47
April	5.06	2.77
May	5.64	2.99
June	4.99	3.05
July	5.29	3.18
August	5.83	3.46
September	4.70	3.03
October	5.61	3.26
November	4.90	2.67
December	4.20	2.53

UNIVERSITY OF RUHUNA

Faculty of Engineering

End-Semester 4 Examination in Engineering: November/December 2016

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

[Three Hours]

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

Q1. a) A certain city divides naturally into ten subdivisions. How might a real estate appraiser select a sample of single family homes that could be used as a basis for developing an equation to predict appraised value from characteristics such as age, size, number of bathrooms, and distance to the nearest school?

[4.0 Marks]

b) If the appraiser has selected two samples for the characteristic "age" using different sampling methods as follows:

Sample A: 56, 32, 62, 66, 49, 53, 50, 65, 45, 56, 68, 38, 75, 57, 70

Sample B : 39, 42, 47, 40, 34, 35, 42, 36, 47, 49, 50, 37, 41, 48, 32

i Construct separate stem-and-leaf displays of the data samples.

ii Find the sample mean, minimum value, maximum value and sample range of each sample.

iii Describe similarities and differences for the two samples.

[6.0 Marks]

c) The system in Figure represents a configuration of solar photovoltaic arrays consisting of crystalline silicon solar cells. There are two sub systems connected in parallel, each one containing three cells. If components work independently of one another and the probability of component works is 0.9, calculate the probability of system works.

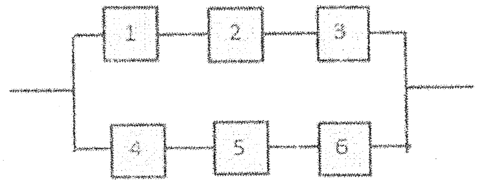


Figure: A system of solar cells connected in series - parallel

[4.0 Marks]

Q2. a) Consider an experiment in which each of three vehicles taking a particular freeway exit turns left (L) or right (R) at the end of the exit ramp.

i What are the possible outcomes that comprise the sample space?

ii Give the possible sample points of the compounds events given bellow.

A – the event that exactly one of the three vehicles turns right

B – the event that at most one of the vehicles turns right

C – the event that all three vehicles turn in the same direction

iii Assuming that the probability of randomly selected car taking a particular freeway exit turns left (L) is 0.25. If the random variable X denotes the number of vehicles taking exit turns left (L), then find the probability distribution of X.

[6.0 Marks]

b) An appliance dealer sells three different models of upright freezers having 13.5, 15.9, and 19.1 cubic feet of storage space, respectively. Let X = the amount of storage space purchased by the next customer to buy a freezer. Suppose that X has probability mass function as given in Table.

Table: Probability mass function of X

X	13.5	15.9	19.1
$P(X=x)$	0.2	0.5	0.3

i Compute $E(X)$, $E(X^2)$, and $Var(X)$.

ii If the price of a freezer having capacity X cubic feet is $25X - 8.5$, what is the expected price paid by the next customer to buy a freezer?

iii What is the variance of the price $25X - 8.5$ paid by the next customer?

iv Suppose that although the rated capacity of a freezer is X, the actual capacity is $h(X) = X - 0.01X^2$. What is the expected actual capacity of the freezer purchased by the next customer?

[4.0 Marks]

c) Suppose that the random variable X has a Normal distribution with mean μ and standard deviation σ. Then the random variable Y has a linear function of X such that $Y = aX + b$.

i Find the distribution of the random variable Y.

ii In an industrial process the diameter of a ball bearing is an important component part. The buyer sets specifications on the diameter to be 3.0 ± 0.01 *cm*. The implication is that no part falling outside these specifications will be accepted. It is known that in the process the diameter of a ball bearing has a normal distribution with mean 3.0 and standard deviation 0.005. On the average, how many manufactured ball bearing will be scrapped?

[4.0 Marks]

Q3. a) Let $X_1, X_2, ..., X_n$ be a random sample from a distribution with mean μ and standard deviation σ.

i Find the distribution of sample mean.

ii If the random sample represents the length of life of light bulbs those collected from an electrical firm. Based on the past data, the light bulbs manufactured by the firm have a length of life that is approximately normally distributed, with mean equal to 800 hours and a standard deviation of 40 hours. Find the probability that a random sample of 16 bulbs will have an average life of less than 775 hours.

[5.0 Marks]

b) Each of newly manufactured items is examined and the number of scratches per item is recorded, yielding the data given in Table.

Table: Number of scratches per item and frequencies

Number of scratches per item	0	1	2	3	4	5	6	7
Observed frequency	18	37	42	30	13	7	2	1

Let X = the number of scratches on a randomly chosen item, and assume that X has a Poisson distribution with parameter λ.

i Use Maximum Likelihood Method to find an estimator of λ.

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

iii Compute the estimate of λ for the data.

iv What is the standard deviation (standard error) of the estimator? Compute the estimated standard error.

[9.0 Marks]

Q4. a) Specimens of soil were obtained from a site both before and after compaction. Tests on 10 pre-compaction specimens gave a mean porosity of 0.413 and a standard deviation of 0.0324. Tests on 20 post-compaction specimens gave a mean porosity of 0.340 and a standard deviation of 0.0469. These standard deviations are not significantly different. Porosity follows a Normal distribution.

i At the 5% level of significance, did the compaction correspond to a significant reduction in mean porosity?

ii At the 5% level of significance, is the reduction in mean porosity significantly less than the desired reduction of 0.1?

[6.0 Marks]

b) A sample of 300 cars having cellular phones and one of 400 cars without phones were tracked for 1 year. The Table gives the number of these cars involved in accidents over that year.

Table: Number of cars involved in accidents

	Accident	No Accident
Cellular phone	22	278
No phone	26	374

Use the above data to test the hypothesis that having a cellular phone in a car and being involved in an accident are independent. Use the 5% level of significance.

[8.0 Marks]

Q5. Corrosion of steel reinforcing bars is the most important durability problem for reinforced concrete structures. Representative data on X = carbonation depth (mm) and Y = strength (MPa) for a sample of core specimens taken from a particular building are given in Table.

Table: Strength (Y) versus carbonation (X) depth for a sample of core specimens

x	8	15	16.5	20	20	27.5	30	30	35	38
y	22.8	27.2	23.7	17.1	21.5	18.6	16.1	23.4	13.4	19.5

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

[3.0 Marks]

b) Calculate the value of sample correlation coefficient and compare the result with part a).

[2.0 Marks]

c) If the answer for part a) is "yes", then

i determine the equation of the estimated regression line.

ii predict strength for carbonation depth value of 45.

iii calculate a 95% prediction interval for a strength value that would result from selecting a single core specimen whose carbonation depth is 45 mm (Hint: assume that the standard deviation of strength value is 2.68).

[9.0 Marks]

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]

UNIVERSITY OF RUHUNA

Faculty of Engineering

End-Semester 4 Examination in Engineering: December 2018

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 Descriptive Statistics

ii Inferential Statistics

[2.0 Marks]

b) Classify the following variables as either categorical or numerical.

i Length (in hours) of baseball games

ii Colors of paint in a paint company's inventory

iii Ranks of personnel in the military

iv Age of students entering a college

[2.0 Marks]

c) A tire manufacturer wants to determine the inner diameter of a certain grade of tire. Ideally, the diameter would be 570 mm. The data are as follows.

$$572, 572, 573, 568, 569, 575, 565, 570$$

i Find the sample mean and the median.

ii Find the sample variance, the standard deviation, and the range.

iii Using the calculated statistics in parts i) and ii), comment on the quality of the tires.

[6.0 Marks]

d) An experiment is proposed to test the three types of antimissile systems. From the design point of view, each of these systems has an equally likely chance of detecting and destroying an incoming missile within a range of 250 miles with a speed ranging up to nine times the speed of sound. However, in actual practice it has been observed that the precisions of these antimissile systems are not the same; that is, the first system will usually detect and destroy the target 10 of 12 times, the second will detect and destroy it 9 of 12 times, and the third will detect and destroy it 8 of 12 times. It is observed that a target has been detected and destroyed. What is the probability that the antimissile defense system was of the third type?

[4.0 Marks]

Q2. a) Briefly explain the followings.

i Random variable

ii Probability distribution

[2.0 Marks]

b) The length of time that an individual talks on a long-distance telephone call has been found to be of a random nature. Let X be the length of the talk; assume it to be a continuous random variable with probability density function given by

$$f(x) = \begin{cases} \alpha e^{-(1/5)x} & ; \quad x > 0 \\ 0 & ; \quad elsewhere \end{cases}$$

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

ii Find the probability that this individual will talk

a) between 8 to 12 minutes.

b) less than 8 minutes.

c) more than 12 minutes.

[5.0 Marks]

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

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

Then

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

Let X be a standard normal random variable has a probability density function

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

i Find the moment-generating function of X.

ii Find the mean and the variance of the random variable X.

iii Given that $M_Y(t) = e^{bt} M_X(at)$ where $Y = aX + b$. Hence, by using the answer in part i), find the moment-generating function of a normal random variable Y.

[7.0 Marks]

Q3. a) Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two unbiased estimators of θ. Let $\hat{\theta}_3$ is a convex combination of $\hat{\theta}_1$ and $\hat{\theta}_2$ such that

$$\hat{\theta}_3 = a\hat{\theta}_1 + (1-a)\hat{\theta}_2 \ ; \ 0 \le a \le 1$$

i Show that $\hat{\theta}_3$ is an unbiased estimator of θ.

ii If $\hat{\theta}_1$ and $\hat{\theta}_2$ are independent, and variance of $\hat{\theta}_1$ and $\hat{\theta}_2$ are σ_1^2 and σ_2^2 respectively, then find the variance of $\hat{\theta}_3$.

iii How should the constant a be chosen in order to minimize the variance of $\hat{\theta}_3$?

[5.0 Marks]

b) A chemist has two different methods for measuring the concentration level C of a chemical solution. Methods A and B produce measurements X_A and X_B respectively and those are distributed as follows.

$$X_A \sim N(C, 2.97) \quad , \quad X_B \sim N(C, 1.62)$$

i Find 99.7% confidence intervals for the measurements X_A and X_B separately. Hence explain that how does the Chemist select more accurate method for measuring the concentration level C?

ii With the knowledge of estimation theory, explain how the Chemist arrives at an optimum point estimate of the concentration level C?

[4.0 Marks]

c) Consider a set of independent data observations $x_1, x_2, \ldots, x_n$ that have a gamma distribution with $k = 5$ and unknown parameter $\lambda > 0$ has a probability density function

$$f(x) = \begin{cases} \dfrac{\lambda^k x^{k-1} e^{-\lambda x}}{\Gamma(k)} & ; \ x \geq 0 \\ 0 & ; \ x < 0 \end{cases}$$

where $\Gamma(k) = (k-1)\Gamma(k-1)$, $\Gamma(1) = 1$ and $E(X) = \frac{k}{\lambda}$.

i Use Maximum Likelihood Estimation Method to find estimators for λ.

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

[5.0 Marks]

Q4. a) An experimenter is interested in the hypothesis testing problem

$$H_0 : \mu = 3.0\, mm \qquad \text{versus} \qquad H_1 : \mu \neq 3.0\, mm$$

where μ is the average thickness of a set of glass sheets. Suppose that a sample of $n = 21$ glass sheets is obtained and their thicknesses are measured. Suppose that the sample mean is $3.04\, mm$ and the sample standard deviation is $0.124\, mm$.
(Assume that the thickness of a glass sheet has a normal distribution)

i What is the critical region that the experimenter accept the null hypothesis with a size $\alpha = 0.10$?

ii What is the critical region that the experimenter reject the null hypothesis with a size $\alpha = 0.01$?

iii Determine whether the null hypothesis accepted with $\alpha = 0.10$ and $\alpha = 0.01$.

iv Write down an expression for the p value to make the decision at $\alpha = 0.10$ level of significance.

[7.0 Marks]

b) In a test of the ability of a certain polymer to remove toxic wastes from water, experiments were conducted at three different temperatures. The data in the following table give the percentages of the impurities that were removed by the polymer in 21 independent attempts. Test the hypothesis that the polymer performs equally well at all three temperatures at the

i 5 percent level of significance

ii 1 percent level of significance.

Low Temperature	Medium Temperature	High Temperature
42	36	33
41	35	44
37	32	40
29	38	36
35	39	44
40	42	37
32	34	45

[7.0 Marks]

Q5. a) Infrared spectroscopy is often used to determine the natural rubber content of mixtures of natural and synthetic rubber. For mixtures of known percentages, the infrared spectroscopy gave the following readings.

Percentage	0	20	40	60	80	100
Reading	0.734	0.885	1.050	1.191	1.314	1.432

i Plot a scatter diagram to see if a linear relationship is indicated.

ii Find the least squares estimates of the regression coefficients.

iii If a new mixture gives an infrared spectroscopy reading of 1.15, estimate its percentage of natural rubber.

[8.0 Marks]

b) The percent survival of a certain type of animal semen, after storage, was measured at various combinations of concentrations of three materials used to increase chance of survival. The summarized data in the usual matrix notation are given by the least squares estimating equations, $(X'X)\boldsymbol{b} = X'\boldsymbol{y}$ and the inverse matrix, $(X'X)^{-1}$ as follows.

$$\begin{bmatrix} 13 & 59.43 & 81.82 & 115.40 \\ 59.43 & 394.7255 & 360.6621 & 522.0780 \\ 81.82 & 360.6621 & 576.7264 & 728.3100 \\ 115.40 & 522.0780 & 728.3100 & 1035.9600 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 377.5 \\ 1877.567 \\ 2246.661 \\ 3337.780 \end{bmatrix}$$

$$(X'X)^{-1} = \begin{bmatrix} 8.0648 & -0.0826 & -0.0942 & -0.7905 \\ -0.0826 & 0.0085 & 0.0017 & 0.0037 \\ -0.0942 & 0.0017 & 0.0166 & -0.0021 \\ -0.7905 & 0.0037 & -0.0021 & 0.0886 \end{bmatrix}$$

By using the above relations, estimate the multiple linear regression equation.

[6.0 Marks]

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

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

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

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

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

[7.0 Marks]

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

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

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.

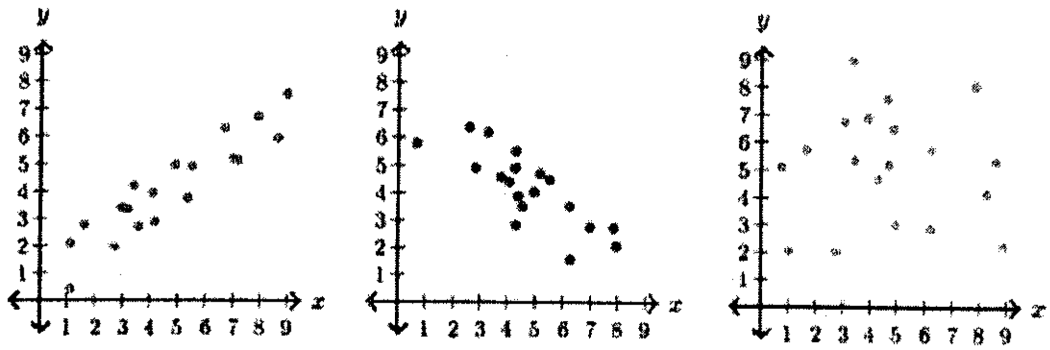
Total sum of squares: $SST = \sum_{i=1}^{I}\sum_{j=1}^{J} y_{ij}^2 - \frac{y_{..}^2}{IJ}$

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.



i) ii) iii)

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:

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

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

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]

UNIVERSITY OF RUHUNA

Faculty of Engineering

End-Semester 4 Examination in Engineering: February 2020

Module Number: IS4301 **Module Name: Probability and Statistics (New Curriculum)**

[Three Hours]

[Answer all questions, each question carries fourteen 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 descriptive statistics for the times (in seconds) of the 28 swimmers in the final round of the male students' 100-meter backstroke event in the school aquatic competition.

Table 1.1

Mean	Median	St Dev	Minimum	Maximum	Q1	Q3
33.36	34.5	8.73	11	46	28.25	39.75

i Calculate $1.5(IQR)$ and subtract it from Q_1 and add it to Q_3 to determine outliers.

ii From the descriptive statistics, construct a box-and-whisker plot and comment on the shape of the distribution.

iii Based on the shape, give measure of the center of the distribution.

[7.0 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 mean and variance of the defective glass sheets.

[8 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 random variable X having a Poisson distribution.

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

[6.0 Marks]

Q3. a) Let $x_1, x_2, \ldots, x_n$ are a random sample from a normal distribution with parameters μ and σ^2. The probability density function is given by

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

i Find the maximum likelihood estimators of μ and σ^2.

iii Use the following data to find the estimate for μ and σ^2.
2, 5, 7, 43, 18, 19, 16, 11, 22, 4, 34, 19, 21, 23, 6, 21, 7, 12

[7 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?

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

iv If the population variance is 6, then Construct a 95% confidence interval for the true mean of the thickness of sheets produced on this shift.

[7.0 Marks]

Q4. a) A student union of a college A claims that the mean student's food expenses is 450 rupees a day because there is no canteen in the college premises. To verify this claim, a study of 20 randomly selected students was investigated and collected the daily expenses for foods. The data as follows:

440, 456, 432, 462, 418, 444, 486, 402, 422, 460,
462, 454, 440, 472, 438, 475, 464, 455, 424, 470

Do the data contradict the students' union claim?

[5.0 Marks]

b) A study was carried to test the relationship between facility conditions at gasoline stations and aggressiveness in the pricing of gasoline. The corresponding observed and expected counts (given in brackets) 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 (21.56)	35 (30.94)	15 (22.5)
	Standard	40 (50.31)	60 (72.19)	75 (52.5)
	Modern	50 (43.13)	70 (61.88)	30 (45.0)

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 of significance.

[4.0 Marks]

c) 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 measured and the relevant data are summarized in Table 4.2.

Table 4.2

Type of Pot (i)	n_i (i^{th} sample size)	$y_{i.}$	y_{ij}^2
Aluminum	4	8.24	17.16
Clay	4	8.72	20.16
Iron	4	18.72	88.8

Use this data and a 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: $SSTr = \frac{1}{J}\sum_{i=1}^{I} y_{i.}^2 - \frac{y_{..}^2}{IJ}$

$$SST = SSTr + SSE$$

[5.0 Marks]

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

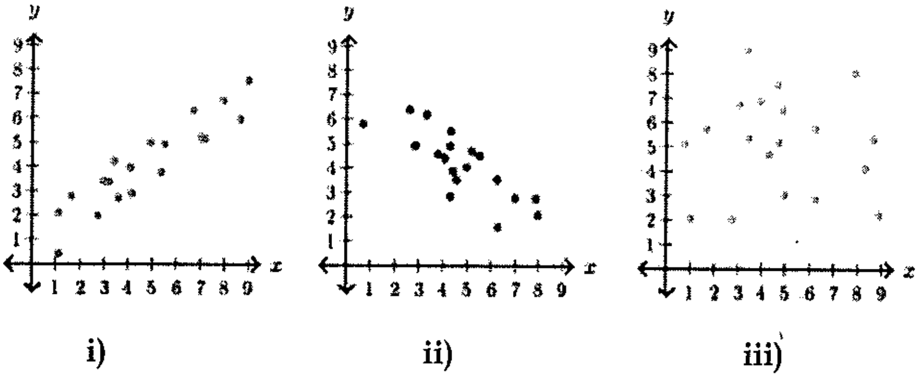


Figure 5.1: Scatter Diagrams

[1.0 Marks]

b) Table 5.1 shows the data on x – normal stress on a certain type of metal test specimen and y = shear resistance.

Table 5.1

x	26.8	25.4	28.9	23.6	27.7	23.9	24.7	28.1	26.9	27.4
y	26.5	27.3	24.2	27.1	23.6	25.9	26.3	22.5	21.7	21.4

i 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.

ii Estimate the shear resistance for a normal stress of 24.5 kilograms per square centimeter.

[8.0 Marks]

c) Table 5.2 represents the data on wear of a bearing y and its relationship to oil viscosity x_1 and load x_2.

Table 5.2

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 5.2 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]

UNIVERSITY OF RUHUNA

Faculty of Engineering

End-Semester 4 Examination in Engineering: January 2022

Module Number: IS4305 **Module Name: Probability and Statistics**

[Three Hours]

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

Q1. a) Specimens of two different types of clay were selected and the percent shrinkage on drying was measured for each specimen, resulting in the following data.

Type I 16 29 16 15 24 20 24 21 17 11 18

Type II 12 23 19 24 17 20 18 19 11 18 8

i Construct a comparative box plot (a box plot for each sample with a common scale).

ii Comment on similarities and differences.

[6.0 Marks]

b) A subway system in city A has five inbounds and five outbound gates. The number of gates open in each direction is observed at a particular time of day. Assume that each outcome of the sample space is equally likely.

i What is the sample space?

ii Find the probability that at least two gates are open in each direction.

iii What is the probability that the number of gates is open in both directions is the same?

[6.0 Marks]

Q2. a) To find out the walking condition of a particular type of machine produced by a particular company, an investigation was carried out and observed a number of times a randomly selected 200 machines aged 3 and over from the date of production had shown errors. The investigation shows 17 machines with no errors, 30 machines with one error, 58 with two errors, 5~~1~~0 with three errors, 38 with four errors, and 7 with five errors. Assuming these proportions continue to hold exhaustively for the population of that production of the company,

i find the probability distribution of the number of times a machine aged 3 and over of this company has shown errors.

ii what is the expected number of times those machines had shown errors?

iii what is the standard deviation of the number of times those machines had shown errors?

[6.0 Marks]

b) Consider the point estimators for μ.

$$\widehat{\mu_1} = \frac{X_1}{3} + \frac{X_2}{3} + \frac{X_3}{3}$$
$$\widehat{\mu_2} = \frac{X_1}{2} + \frac{X_2}{3} + \frac{X_3}{6}$$
$$\widehat{\mu_3} = \frac{X_1}{2} + \frac{X_2}{3} + \frac{X_3}{3} + 2$$

Suppose that $E(X_1) = E(X_2) = E(X_3) = \mu$, $Var(X_1) = 7$, $Var(X_2) = 13$, and $Var(X_3) = 20$. Assume that the random variables X_1, X_2 and X_3 are independent.

i Calculate the bias of each point estimator. Is any one of them unbiased?

ii Calculate the variance of unbiased point estimators. Which point estimator has the smallest variance?

iii Obtain an expression to find the mean square error of a point estimator. Then use it to find the mean square error of $\widehat{\mu_2}$.

[6.0 Marks]

Q3. a) Component parts for an engine produced by company A are shipped to customers in lots of 100. If the specifications of the parts suggest that 95% of items meet the specifications, use a suitable approximation to find the probability that

i more than 3 items will be defective in a given lot.

ii less than 10 items will be defective in a lot.

[6.0 Marks]

b) Table 3.1 shows the number of errors found in 85 software products of a particular company.

Table 3.1

No. of Errors found in a soft-ware product	1	2	3	4	5	6
Frequency	17	20	25	14	6	3

Is there any evidence that the number of errors is modeled with a Poisson distribution with mean $\lambda = 3$? Use 5% level of significance.

[6.0 Marks]

Q4. a) The viscosity of two different brands of car oil is measured and their summary statistics are given in Table 4.1.

Table 4.1

Sample Statistics	Brand A	Brand B
Size	6	7
Mean	10.57	10.54
Variance	0.027	0.002

i Is the mean viscosity of the two brands is equal? Assume that the populations have normal distributions with equal variances. Use $\alpha = 0.05$.

ii Find the 95% confidence interval for the difference in mean viscosity.

[6.0 Marks]

b) An experiment was conducted to measure water pollution based on the quantity of dissolved oxygen. Four different locations were identified adjacent to a farm, a factory, a city, and a village and randomly selected water specimens. Table 4.2 displays the dissolved oxygen readings of the specimens from the four locations.

Table 4.2

Location	Readings
Location 1	6.5 5.4 5.0 4.8 6.0
Location 2	6.1 6.3 5.9 6.0 4.4
Location 3	5.1 6.4 6.3 5.0
Location 4	6.9 6.8 6.9 6.8 7.0

i Construct the one-way Analysis of Variance Table (ANOVA).

ii Do the data provide sufficient evidence to indicate the difference in mean dissolved oxygen content for the four locations? Use 5% level of significance.

[6.0 Marks]

Q5. a) An experiment was conducted to examine the effect of varying the water/cement ratio on the strength of concrete that had been aged 30 days and the obtained data presented in Table 5.1.

Table 5.1

Water/Cement Ratio	1.1	1.3	1.2	1.4	1.8	1.9	1.6	1.5	1.7	1.5
Strength	1.2	1.3	1.2	1.3	1.6	1.7	1.4	1.4	1.4	1.3

i Calculate the value of the correlation coefficient and interpret the result.

ii Consider the model; $\ln w = \ln \alpha + \gamma \ln k + \varepsilon$; where w and k are variables, $\ln \alpha$ and γ are regression coefficients.
By assuming the model fits data given in Table 5.1, find the regression coefficient to predict the strength of concrete of specified water/cement ratio.

iii Predict the strength of concrete when the water/cement ratio is at 2.0.

[6.0 Marks]

b) An experiment was carried out to assess the impact of the several variables force (gm), power (mW), temperature (°C), and time (msec) on ball bond shear strength (gm). By assuming that the ball bond shear strength is linearly related with all the other variables;

i State assumptions that use to construct multiple linear regression model to predict ball bond shear strength.

ii use matrix approach to explain the finding of regression coefficients and the regression equation for the given set of variables.

[6.0 Marks]

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)**.

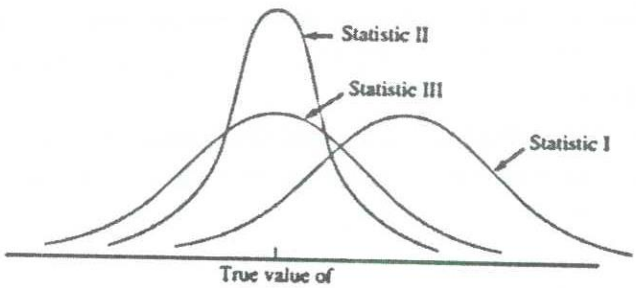


Figure Q3. a)

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.

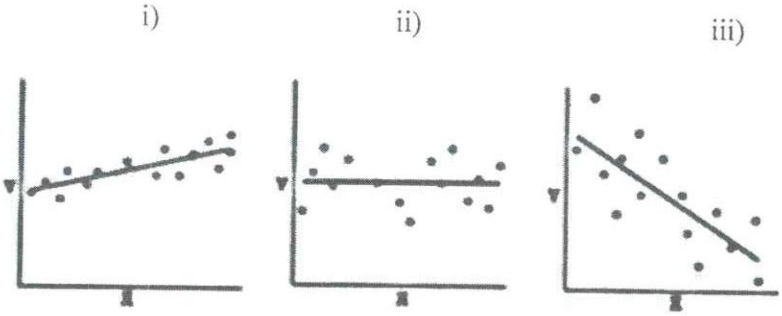


Figure Q5. a)

[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]

UNIVERSITY OF RUHUNA

Faculty of Engineering

End-Semester 4 Examination in Engineering: September 2023

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

[Three Hours]

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

Q1. a) A particular company wanted to compare and contrast the two different types of glass bottles that it manufactures, for future changes. Therefore, a sample of 20 glass bottles of each type was selected and the internal pressure strength of each bottle was measured, resulting in the following data.

Type 1: 84 88 90 92 94 94 94 96 98 100 100
100 100 102 104 106 107 107 107 115

Type 2: 97 98 99 98 98 99 99 99 100 100 100
100 101 101 101 101 102 102 102 108

i Obtain five number summaries of each sample.

ii Construct a comparative boxplot (plotting both boxplots against the same scale), and use it to compare and contrast the internal pressure strength of the two types of glass bottles.

[8 Marks]

b) Consider the following data on the thickness of the floor plate of an aboveground tank that is used to store crude oil.

43, 46, 52, 55, 55, 56, 58, 60, 62, 63, 64, 66, 66, 72, 74, 74, 75, 77, 77, 78, 83, 85, 85, 87, 88, 90, 91, 94

i Find the sample mean and the sample standard deviation of these data.

ii Display the data in a Stem-and-Leaf diagram. Do the data appear to be approximately normal?

iii What percentage of data values are within 2 standard deviations of the mean?

[4 Marks]

Q2. a) A particular company manufactures transistors and on a given day, the production consists of 10 defective transistors (that immediately fail when put in use), 15 partially defective (that fail after a couple of hours of use), and 75 acceptable transistors. A transistor is chosen at random and put into use. If it does not immediately fail, what is the probability that it is acceptable?

[3 Marks]

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

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

ii 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 in finding the probability that a metal cylinder has a diameter between 49.8 and 50.1 mm.

[2 Marks]

c) A certain gas station has six pumps. If a random variable X denotes the number of pumps that are in use at a particular time of day. If Table Q2.c) gives the probability distribution of X, find the following probabilities.

Table Q2.c)

X	0	1	2	3	4	5	6
P(x)	0.05	0.10	0.15	0.25	0.20	0.15	0.10

i $\Pr(X \leq 2)$

ii $\Pr(X \geq 3)$

[2 Marks]

d) Consider a random variable X has a discrete uniform distribution with the probability mass function $P(x)$, where

$$P(x) = \begin{cases} \frac{1}{n} ; & x = 1,2,3, \dots, n \\ 0 ; & otherwise \end{cases}$$

Compute $E(X)$ and $V(X)$.

(Hint: The sum of the first n positive integers is $\frac{n(n+1)}{2}$, whereas the sum of their squares is $\frac{n(n+1)(2n+1)}{6}$).

[5 Marks]

Q3. a) A random sample of a particular type of concrete cylinders were selected and measured their strength, resulting in the following data.

7.4, 5.8, 6.5, 8.4, 9.3, 10.0, 5.9, 7.3, 6.3, 8.1, 7.0, 7.6, 6.5, 9.0, 8.2, 8.7, 7.8, 9.7, 11.6, 11.3

i Calculate a point estimate of the mean value of strength for the conceptual population of all concrete cylinders manufactured in this particular type.

ii Calculate a point estimate of the strength value that separates the weakest 50% of all such concrete cylinders from the strongest 50%, and state which estimator you used.

iii Calculate a point estimate of the population standard deviation σ. Which estimator did you use?

[4 Marks]

b) A random sample of size 17 was obtained on the breakdown voltage of electrically stressed circuits and found that its variance was 137324.3. Find the 95% confidence interval for σ.

[3 Marks]

c) Random samples were taken on the tensile strength of liner specimens both when a certain fusion process was used and when the process was not used and the relevant data are summarized in Table Q3.b).

Table Q3.b)

	n	$\bar{x}$	s
No fusion	40	290.2	277.3
Fused	35	310.8	205.9

Does the data suggest that the fusion process increased the true average tensile strength by more than 10 units?

[5 Marks]

Q4. a) A computer manufacturer used four different designs for manufacturing a particular electrical circuit board. And he identified that there are three possible failure modes for the circuit board. He obtained the data on the number of failures in each mode in each design and the data is summarized in Table Q4.a).

Table Q4.a)

		Failure Mode 1	Failure Mode 2	Failure Mode 3
Design	1	16	40	11
	2	8	17	7
	3	10	31	13
	4	9	12	6

Does the design type appear to have an effect on failure mode?

[6 Marks]

b) An experiment was carried out to investigate the water quality in tap water in a particular city. Four tap water samples were obtained from four different locations of the city and their heavy metal content (ppm) was measured. The relevant data are displayed in Table Q4.b).

Table Q4.b)

Heavy Metal Content (ppm)			
Location 1	**Location 2**	**Location 3**	**Location 4**
5	8	6	7
7	7	8	8
8	6	7	8
9	7	9	6

Test for the equality of mean metal content at $\alpha = 0.05$.

(Hint: In the usual notations, $N = \sum_{i=1}^{I} n_i$; $SSTr = \frac{\sum_{i=1}^{s} y_{i\bullet}^2}{n_i} - \frac{y_{\bullet\bullet}^2}{N}$;
$SST = \sum_{i=1}^{s}\sum_{j=1}^{n_i} y_{ij}^2 - \frac{y_{\bullet\bullet}^2}{N}$; $SST = SSTr + SSE$)

[6 Marks]

Q5. Table Q5.a) gives the monthly data of a chemical plant in the year 2022 on electric power consumption (y), average ambient temperature (x_1), the number of days in the week (x_2), the average product purity (x_3), and the tons of product produce (x_4).

a) Calculate Karl Pearson Correlation coefficient (r) between the electric power consumption and the average ambient temperature and explain the relationship.

[2 Marks]

b) Based on the part a), determine the equation of the estimated regression line.

[2 Marks]

c) Based on part b), calculate R^2 and comment.

[2 Marks]

d) If Table Q5.b) gives the partial correlations between y and x_i's, explain each of the relationships.

[2 Marks]

e) Use the given information in Table Q5.c) to state the multiple linear regression model using all independent variables.

[2 Marks]

f) Predict power consumption for a month in which $x_1 = 75^0F$, $x_2 = 24$ days, $x_3 = 90\%$, and $x_4 = 98$ tons.

[Hint: $\hat{\beta}_1 = \frac{n\sum_{i=1}^{n} x_i y_i - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}$, $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$,
$r = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{(n-1)\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2}$]

[2 Marks]

Table Q5.a)

Y	$X_1(^0F)$	X_2 (days)	X_3 (%)	X_4 (tons)
240	25	24	91	100
236	31	21	90	95
290	45	24	88	110
274	60	25	87	88
301	65	25	91	94
316	72	26	94	99
300	80	25	87	97
296	84	25	86	96
267	75	24	88	110
276	60	25	91	105
288	50	25	90	100
261	38	23	89	98

Table Q5.b)

	Y
X_2	0.803
X_3	0.049
X_4	-0.009

Table Q5.c)

Predictor	Coefficient
x1	0.6054
x2	8.924
x3	1.437
x4	0.0136
Constant	-102.7

UNIVERSITY OF RUHUNA

Faculty of Engineering

End-Semester 4 Examination in Engineering: September 2024

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

[Three Hours]

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

Q1. a) Explain the following.

i How do you select a sample of 10 students out of 75 for a particular research study? Assume that all 75 students represent one specific field of study.

[2 Marks]

ii If the number of students in a population which consists of three fields 1, 2, and 3 is 135, 45, and 90 respectively, how do you select 30 students from that population of three different fields of study?

[2 Marks]

b) A study of the effect of brand of batteries on the lifetime of an electric component was conducted. The recorded measures of lifetime in minutes of two brands of batteries are as follows:

Brand A	6.9	5.6	2.2	4.7	5.3	4.8	5.2	3.4	6.0	4.3
Brand B	2.8	2.5	2.6	3.4	2.9	2.8	3.8	3.0	3.0	3.1

i Draw a Stem-and-leaf diagram to represent "Brand A" data, and give comments.

[2 Marks]

ii Compute the sample mean and sample variance of each brand.

[4 Marks]

iii Comment on the impact of the brand of batteries on the lifetime of an electric component.

[2 Marks]

Q2. a) A certain company manufactures washing machines and dryers. It is known that under the warranty period, 30% of these washing machines require services, whereas only 10% of its dryers need service. Assume that the two machines function independently of one another. If a person purchases both a washing machine and a dryer from this company, find

i the probability that both machines require warranty service.

[2 Marks]

ii the probability that neither machine requires warranty service.

[2 Marks]

b) Consider an experiment in which each of three vehicles taking a particular freeway exit turns left (L) or right (R) at the end of the exit ramp.

i What are the possible outcomes that comprise the sample space?

[1 Mark]

ii List all possible sample points for compound events A and B, as specified below.

$A-$ the event that exactly one of the three vehicles turns right
$B-$ the event that at most one of the vehicles turns right

[2 Marks]

c) The gas mileage (Km/L) of a particular type of car is normally distributed, with a mean of 25 and a standard deviation of 2. If a person randomly selects a car, find the following.

i The probability that the car will get a mileage of more than 28 kilometers per liter.

[2 Marks]

ii The probability that the car will get a mileage of between 26 and 28 kilometers per liter.

[2 Marks]

iii If the tank of the randomly selected car contains only one liter of gas and the driver has to drive 30 kilometers, should the driver first stop at the gas station? Justify your answer by calculations.

[1 Mark]

Q3. a) The following data represents the thickness measurements (in millimeters) of 10 specimens of plastic sheets produced by a machine on a particular shift.

7.5	9.4	8.3	8.6	7.7	9.2	8.0	10.1	6.8	8.4

i Calculate a point estimate of the population standard deviation σ.

[2 Marks]

ii Calculate a 90% confidence interval for σ.

[2 Marks]

iii If a sample of $n = 10$ thickness measurements yields the 95% confidence interval $(1.63, 11.50)$ for σ^2, find the standard deviation S for the sample.

[2 Marks]

b) The cube compressive strength (N/mm^2) of two different types of concrete specimens was measured, resulting in the following data.

Type I	10.6	10.5	10.3	10.7	10.4	10.7
Type II	10.5	10.5	10.5	10.6	10.5	10.5

Assuming that the populations have normal distributions with equal variances, test the hypothesis that the mean compressive strengths of the two types of concrete specimens are equal.

[6 Marks]

Q4. a) An experiment was conducted to investigate the improvement of the tensile strengths of a particular type of bag at different concentration levels of a specific element "A". The tensile strengths (N/mm^2) were measured at 5%, 10%, 15%, and 20% levels of the specific element "A" and summarized in Table Q4.a).

Table Q4.a)

Hardwood Concentration (%)			
5%	10%	15%	20%
12	18	15	24
14	12	18	25
15	17	10	24
11	13	19	15
10	18	16	13
	19	18	
	15		

Stating your assumptions, construct a one-way analysis of variance (ANOVA) table and, obtain the relevant conclusions at a 0.05 significance level.
(Hint: In the usual notations,

$N = \sum_{i=1}^{l} n_i$; $SSTr = \frac{\sum_{i=1}^{s} y_{i.}^2}{n_i} - \frac{y_{..}^2}{N}$; $SST = \sum_{i=1}^{s} \sum_{j=1}^{n_i} y_{ij}^2 - \frac{y_{..}^2}{N}$;
$SST = SSTr + SSE$)

[6 Marks]

b) An experiment was conducted to investigate two types of resin mixtures of different expansion and contraction properties that affect the chances of the tiles becoming cracked. For this, a combination of two surveys of tiles on a group of buildings A and B were done. The two groups of buildings were constructed about the same time and have exterior walls composed of the same type of tiles. However, the tiles in Building B were cemented into place with a different resin mixture than that used in buildings A.
The group of construction engineers found a total of 406 cracked tiles out of 6000 in building A. In building B, they found 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]

Q5. a) An experiment was conducted to examine the effect of varying the water/cement ratio on the strength of concrete that had been aged 30 days and the obtained data is presented in Table Q5.a).

Table Q5.a)

Water/Cement Ratio	1.1	1.3	1.2	1.4	1.8	1.9	1.6	1.5	1.7	1.5
Strength (N/mm^2)	1.2	1.3	1.2	1.3	1.6	1.7	1.4	1.4	1.4	1.3

i Calculate the value of the correlation coefficient (r) and interpret the result.

[2 Marks]

ii By assuming the model fits the data given in Table Q5.a), find the regression coefficients to predict the strength of concrete of specified water/cement ratio.

[2 Marks]

iii Predict the strength of concrete when the water/cement ratio is at 2.0.

[2 Marks]

$$\left[\text{Hint: } r = \frac{Cov\,(x,y)}{S_x S_y}\;;\; \hat{\beta}_1 = \frac{n\sum_{i=1}^{n} x_i y_i - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2},\quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}\right]$$

b) A study was conducted to identify the influence of blade speed (x_1) and voltage measuring sensor extension (x_2) on the voltage output of engines(y). Assume that all the measures are given in standard units. Table Q5.b) i) and Table Q5.b) ii) give the correlation matrix and the regression output, respectively.

Table Q5.b) i)

	y	x_1
x_1	0.607	
x_2	-0.530	0.325

Table Q5.b) ii)

Regression Analysis: Voltage Output versus Speed, x_1, Extension, x_2

```
Predictor                    Coef      SE Coef          T        P
Constant                    -1.64         0.24      -6.64     0.00
Speed, x1 (in./sec)       0.00055     0.000034      16.11     0.00
Extension, x2 (in.)        -67.39         4.48     -15.04     0.00
R-Sq = 96.1%
```

i Comment on the relationship between y and x_1 and y and x_2.

[2 Marks]

ii State the regression model using all independent variables.

[1 Mark]

iii Comment on R-squared value; Coefficient of Determination of the model.

[1 Mark]

iv State and test appropriate hypotheses to decide whether the model fits the data specifies a useful linear relationship between y and x_1.

[2 Marks]

UNIVERSITY OF RUHUNA

Faculty of Engineering

End-Semester 4 Examination in Engineering: September 2024

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

[Three Hours]

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

Q1. a) Explain the following.

i How do you select a sample of 10 students out of 75 for a particular research study? Assume that all 75 students represent one specific field of study.

[2 Marks]

ii If the number of students in a population which consists of three fields 1, 2, and 3 is 135, 45, and 90 respectively, how do you select 30 students from that population of three different fields of study?

[2 Marks]

b) A study of the effect of brand of batteries on the lifetime of an electric component was conducted. The recorded measures of lifetime in minutes of two brands of batteries are as follows:

Brand A	6.9	5.6	2.2	4.7	5.3	4.8	5.2	3.4	6.0	4.3
Brand B	2.8	2.5	2.6	3.4	2.9	2.8	3.8	3.0	3.0	3.1

i Draw a Stem-and-leaf diagram to represent "Brand A" data, and give comments.

[2 Marks]

ii Compute the sample mean and sample variance of each brand.

[4 Marks]

iii Comment on the impact of the brand of batteries on the lifetime of an electric component.

[2 Marks]

Q2. a) A certain company manufactures washing machines and dryers. It is known that under the warranty period, 30% of these washing machines require services, whereas only 10% of its dryers need service. Assume that the two machines function independently of one another. If a person purchases both a washing machine and a dryer from this company, find

i the probability that both machines require warranty service.

[2 Marks]

ii the probability that neither machine requires warranty service.

[2 Marks]

b) Consider an experiment in which each of three vehicles taking a particular freeway exit turns left (L) or right (R) at the end of the exit ramp.

i What are the possible outcomes that comprise the sample space?

[1 Mark]

ii List all possible sample points for compound events A and B, as specified below.

$A-$ the event that exactly one of the three vehicles turns right
$B-$ the event that at most one of the vehicles turns right

[2 Marks]

c) The gas mileage (Km/L) of a particular type of car is normally distributed, with a mean of 25 and a standard deviation of 2. If a person randomly selects a car, find the following.

i The probability that the car will get a mileage of more than 28 kilometers per liter.

[2 Marks]

ii The probability that the car will get a mileage of between 26 and 28 kilometers per liter.

[2 Marks]

iii If the tank of the randomly selected car contains only one liter of gas and the driver has to drive 30 kilometers, should the driver first stop at the gas station? Justify your answer by calculations.

[1 Mark]

Q3. a) The following data represents the thickness measurements (in millimeters) of 10 specimens of plastic sheets produced by a machine on a particular shift.

7.5	9.4	8.3	8.6	7.7	9.2	8.0	10.1	6.8	8.4

i Calculate a point estimate of the population standard deviation σ.

[2 Marks]

ii Calculate a 90% confidence interval for σ.

[2 Marks]

iii If a sample of $n = 10$ thickness measurements yields the 95% confidence interval $(1.63, 11.50)$ for σ^2, find the standard deviation S for the sample.

[2 Marks]

b) The cube compressive strength (N/mm^2) of two different types of concrete specimens was measured, resulting in the following data.

Type I	10.6	10.5	10.3	10.7	10.4	10.7
Type II	10.5	10.5	10.5	10.6	10.5	10.5

Assuming that the populations have normal distributions with equal variances, test the hypothesis that the mean compressive strengths of the two types of concrete specimens are equal.

[6 Marks]

Q4. a) An experiment was conducted to investigate the improvement of the tensile strengths of a particular type of bag at different concentration levels of a specific element "A". The tensile strengths (N/mm^2) were measured at 5%, 10%, 15%, and 20% levels of the specific element "A" and summarized in Table Q4.a).

Table Q4.a)

Hardwood Concentration (%)			
5%	10%	15%	20%
12	18	15	24
14	12	18	25
15	17	10	24
11	13	19	15
10	18	16	13
	19	18	
	15		

Stating your assumptions, construct a one-way analysis of variance (ANOVA) table and, obtain the relevant conclusions at a 0.05 significance level.
(Hint: In the usual notations,

$N = \sum_{i=1}^{l} n_i$; $SSTr = \frac{\sum_{i=1}^{s} y_{i.}^2}{n_i} - \frac{y_{..}^2}{N}$; $SST = \sum_{i=1}^{s}\sum_{j=1}^{n_i} y_{ij}^2 - \frac{y_{..}^2}{N}$;
$SST = SSTr + SSE$)

[6 Marks]

b) An experiment was conducted to investigate two types of resin mixtures of different expansion and contraction properties that affect the chances of the tiles becoming cracked. For this, a combination of two surveys of tiles on a group of buildings A and B were done. The two groups of buildings were constructed about the same time and have exterior walls composed of the same type of tiles. However, the tiles in Building B were cemented into place with a different resin mixture than that used in buildings A.
The group of construction engineers found a total of 406 cracked tiles out of 6000 in building A. In building B, they found 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]

Q5. a) An experiment was conducted to examine the effect of varying the water/cement ratio on the strength of concrete that had been aged 30 days and the obtained data is presented in Table Q5.a).

Table Q5.a)

Water/Cement Ratio	1.1	1.3	1.2	1.4	1.8	1.9	1.6	1.5	1.7	1.5
Strength (N/mm^2)	1.2	1.3	1.2	1.3	1.6	1.7	1.4	1.4	1.4	1.3

i Calculate the value of the correlation coefficient (r) and interpret the result.

[2 Marks]

ii By assuming the model fits the data given in Table Q5.a), find the regression coefficients to predict the strength of concrete of specified water/cement ratio.

[2 Marks]

iii Predict the strength of concrete when the water/cement ratio is at 2.0.

[2 Marks]

$$\left[\text{Hint: } r = \frac{Cov\,(x,y)}{S_x S_y} \;;\; \hat{\beta}_1 = \frac{n\sum_{i=1}^{n} x_i y_i - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2},\; \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}\right]$$

b) A study was conducted to identify the influence of blade speed (x_1) and voltage measuring sensor extension (x_2) on the voltage output of engines(y). Assume that all the measures are given in standard units. Table Q5.b) i) and Table Q5.b) ii) give the correlation matrix and the regression output, respectively.

Table Q5.b) i)

	y	x_1
x_1	0.607	
x_2	-0.530	0.325

Table Q5.b) ii)

Regression Analysis: Voltage Output versus Speed, x_1, Extension, x_2

```
Predictor                  Coef      SE Coef         T        P
Constant                  -1.64         0.24     -6.64     0.00
Speed, x1 (in./sec)     0.00055     0.000034     16.11     0.00
Extension, x2 (in.)      -67.39         4.48    -15.04     0.00
R-Sq = 96.1%
```

i Comment on the relationship between y and x_1 and y and x_2.

[2 Marks]

ii State the regression model using all independent variables.

[1 Mark]

iii Comment on R-squared value; Coefficient of Determination of the model.

[1 Mark]

iv State and test appropriate hypotheses to decide whether the model fits the data specifies a useful linear relationship between y and x_1.

[2 Marks]