SLIIT

Discover Your Future

Sri Lanka Institute of Information Technology

B. Sc. Special Honours Degree / Diploma

in

Information Technology

Final Examination

Year I, Semester II (2018)

June intake

IT 1070 – Probability and Statistics

Duration: 2 Hours

Instruction to Candidates:

- This paper has five questions in six pages
- Answer **ALL** questions.
- Please show your work for full credit.
- Calculators are allowed.
- Electronic devices retrieving text including electronic dictionaries and mobile phones are not allowed.

Question 1 (20 Marks)

A sample of 25 batteries of a newly produced brand was subjected for testing their lifetimes before it is advertised for marketing. The lifetimes in hours which each survived is given below.

112	105	123	137	157	141	104	147	97	131	139	152	137
134	143	155	137	98	108	99	153	144	94	115	115	

a) Produce a stem-and-leaf plot of these measurements. (4 marks)

b) Find the three quartiles for this data. (4 marks)

c) Calculate, to 2 decimal places, the mean and standard deviation for this data. (Marks will be given for calculations.) (6 marks)

d) Comment on the distribution of this data. (2 marks)

e) Construct 95% confidence interval for the true mean lifetime of batteries. (4 marks)

Question 2 **(20 Marks)**

a) Explain what is meant by;

i. Mutually exclusive events
ii. Sample space
iii. Collectively exhaustive events
iv. Independent events

(8 marks)

b) i. What is meant by conditional probability?
ii. If A and B are two independent events show that $P(A \cap B) = P(A) \times P(B)$ (4 marks)

c) A TV manufacturer buys TV tubes from three sources. Source A supplies 50% of all tubes and has a 1% defective rate. Source B supplies 30% of all tubes and has a 2% defective rate. Source C supplies the remaining 20% of tubes and has a 5% defective rate.

i. What is the probability that a randomly selected purchased tube is defective?
ii. Given that a purchased tube is defective, what is the probability it came from Source A?

(8 marks)

Question 3 **(20 marks)**

a) A random variable X has the following probability mass function defined in tabular form

X	-1	1	2
p(x)	2c	3c	4c

i. Find the value of c. (3 Marks)
ii. Compute p (−1), p (1), and p (2). (3 Marks)
iii. Find E(X) (3 Marks)
iv. Find Var (x). (3 Marks)

b) An officer is always late to the office and arrives within the grace period of ten minutes after the start. Let X be the time that elapses between the start and the time the officer signs in with a probability density function.

$$f(x) = \begin{cases} kx^2 & 0 \leq x \leq 10 \\ 0 & \text{otherwise} \end{cases}$$

where $k > 0$ is a constant.

i. Compute the value of k
ii. Find the probability that the he arrives less than 3 minutes after the start of the office.

(8 Marks)

Question 4 **(20 Marks)**

The quality-control manager at a light bulb factory needs to determine whether the mean life of a large shipment of light bulbs is equal to 375 hours. The population standard deviation is 100 hours. A random sample of 64 light bulbs indicates a sample mean life of 350 hours.

a) At the 0.05 level of significance, is there evidence that the mean life is different from 375 hours? (5 Marks)

b) Compute the p-value and interpret its meaning. (5 Marks)

c) Construct a 95% confidence interval estimate of the population mean life of the light bulbs. (5 Marks)

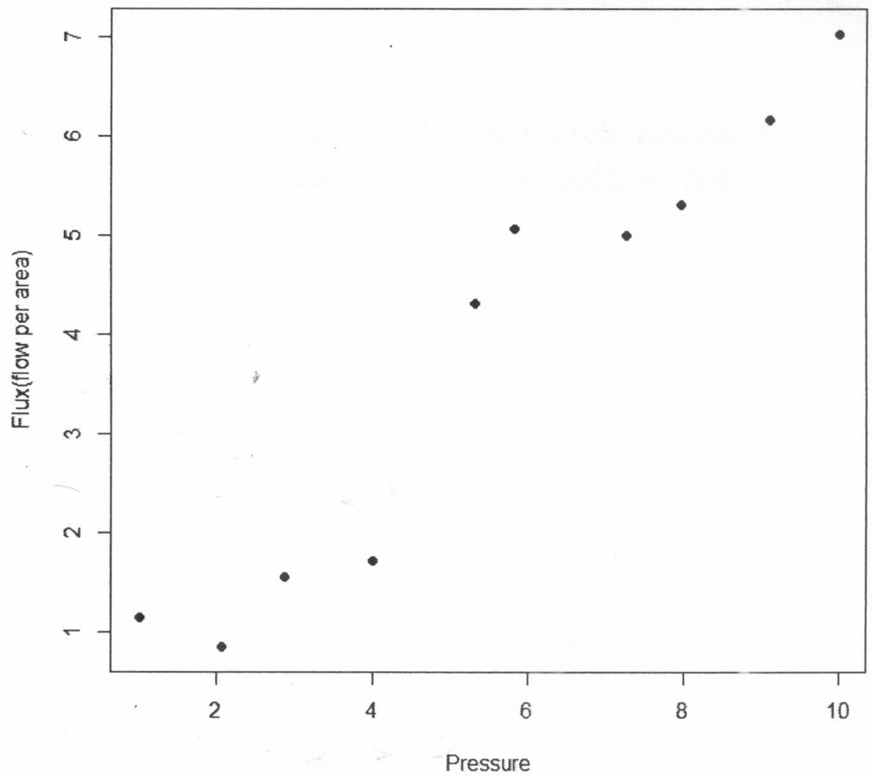
d)
   i. Compare the results of (a) and (c).
   ii. What conclusions do you reach? (5 Marks)

Question 5 (20 Marks)

When purifying drinking water you can use a so-called membrane filtration. In an experiment one wishes to examine the relationship between the pressure drop across a membrane and the flux (flow per area) through the membrane. We observe the following 10 related values of pressure (x) and flux (y)

Pressure (x)	1.02	2.08	2.89	4.01	5.32	5.83	7.26	7.96	9.11	9.99
Flux (y)	1.15	0.85	1.56	1.72	4.32	5.07	5.00	5.31	6.17	7.04

Scatter plot for Pressure Vs. Flux



R Output

Coefficients:

(Intercept)	Pressure
-0.1886	0.7225

Analysis of Variance Table

Response: Flux

	df	Sum Sq	Mean Sq	F Value	Pr(>F)	
Pressure	T	P	**S**	**104.59**	7.177e-06	***
Residuals	**U**	Q	**0.416**			
Total	**9**	**R**				

Significant Codes:	0 '***'	0.001 '**'	0.01 '*'	0.05 '.'	0.1 ' '	1

a) Find values marked P, Q, R, S, T and U in the ANOVA table (Show workings to get full marks) (6 Marks)

b) State the estimated Regression equation. (3 Marks)

c) Test whether the slope of the Regression line is significant and state the conclusions. (6 Marks)

d) Find the estimated Flux when the pressure takes the value 5 units (2 Marks)

e) Provide two meaningful conclusions from the above analysis (3 Marks)

End of the Question Paper

FORMULA SHEET
Probability & Statistics – IT 1070

	Mean	**Variance**
Sample	$\bar{X} = \dfrac{\sum_{i=1}^{n} X_i}{n}$	$S^2 = \dfrac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$
Population	$\mu = \dfrac{\sum_{i=1}^{N} X_i}{N}$	$\sigma^2 = \dfrac{\sum_{i=1}^{N}(X_i - \mu)^2}{N}$

Covariance

$$\operatorname{cov}(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{n-1}$$

Coefficient of Correlation

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}} = \frac{\operatorname{cov}(X,Y)}{S_X S_Y}$$

The conditional probability of A given that B has occurred

$$P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)}$$

Expected Value of a RV

$$\mu = E(X) = \sum_{i=1}^{N} X_i P(X_i)$$

Variance of a Discrete RV

$$\sigma^2 = \sum_{i=1}^{N}[X_i - E(X)]^2 P(X_i)$$

Mean and Variance of a continuous RV

$$\mu = E(X) = \int X f(X)dx \quad and \quad V(X) = \int (X - E(X))^2 f(X)dx$$

Translation to Z: $Z = \dfrac{X - \mu}{\sigma}$

The formula for r, the Pearson product moment correlation coefficient is given below

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n(\sum x^2) - (\sum x)^2]\,[n(\sum y^2) - (\sum y)^2]}}$$

Spearman rank correlation

$$r_s = 1 - \frac{6\sum d^2}{n(n^2 - 1)}$$

Simple Linear Regression Equation

Estimated (or predicted) Y value for Observation

Estimate of the regression intercept

Estimate of the regression slope

$$\hat{Y}_i = b_0 + b_1 X_i$$

X value for Observation

Regression Equation y = a + bx

$$Slope\ (b) = \frac{(N\Sigma XY - (\Sigma X)(\Sigma Y))}{(N\Sigma X^2 - (\Sigma X)^2)}$$

Intercept(a) = (ΣY - b(ΣX)) / N

X - Independent variable and Y - Dependent variable