Sri Lanka Institute of Information Technology

B.Sc. Honours Degree in Information Technology

Final Examination
Year 2, Semester 2 (2019)

IT 2110 – Probability and Statistics

Duration: 2 Hours

October 2019

Instructions to Candidates:

- You have **10 minute** reading time.
- This paper has **4** questions.
- Answer **all** questions in the booklet given.
- This paper **8 pages**, including the cover page and the equation sheet.
- Please show your work for full credit.
- Calculators are allowed.
- Electronic devices capable of storing and retrieving text, including electronic dictionaries and mobile phones are not allowed.

Question 01 **25 Marks**

a) A continuous random variable X has probability density function given by,

$$f_X(x) = \begin{cases} k(3 - x^2) \; ; & -1 \leq x \leq 1 \\ 0 & ; \text{otherwise} \end{cases}$$

i. Find k value. **(3 marks)**

ii. Find $P(X>0.1)$. (Give your answer in 4 decimal places) **(3 marks)**

iii. Find $V(X)$. **(3 marks)**

iv. Find cumulative distribution function and $F_X(0.5)$. (Give your answer in 4 decimal places) **(3 marks)**

b) In an experiment, it is given that $P(A) = 0.36$, $P(B) = 0.44$ and $P(A \cup B) = 0.8$. Are A and B mutually exclusive? **(3 marks)**

c) A certain area of the eastern United States is, on average, hit by 6 hurricanes a year. Without any approximation, find the probability that in a given year that area will be hit by,

i. Fewer than 4 hurricanes. **(3 marks)**

ii. Anywhere from 6 to 8 hurricanes. **(3 marks)**

iii. Using a suitable approximation, find the probability that in a given year that area will hit by more than 8 hurricanes. **(4 marks)**

Question 02 — 25 Marks

a) In a survey of Catholic priests, each priest reported the total number of baptisms, marriages, and funerals conducted during the past calendar year. The responses are given in the following table.

32	33	26	31	53	29	37	38	40	30
31	26	31	41	51	50	59	38	32	28
29	34	32	40	40	54	48	41	35	45
38	39	43	42	36	41	60	43	46	45
40	35	28	50	40	47	57	34	50	53

For this data, $\bar{X} = 40.02$ and $S = 8.9157$.

i. Using above information, construct a 95% confidence interval on the mean (μ) number of baptisms, marriages and funerals conducted during the past calendar year. (Round up the answer to the nearest integer) **(10 marks)**

b)

i. Explain what is meant by Type I and Type II errors in hypothesis testing. **(5 marks)**

ii. A study at the University of Colorado at Boulder shows that running increases the resting metabolic rate (RMR) in older women where the current average RMR for older women is 0.3 with the standard deviation of 0.105. The average RMR of 30 elderly women runners was 0.34. Is there a significant increase in average RMR of the women runners at 5% level of significance? Assume the population to be approximately normally distributed. (Give your test statistic in 4 decimal places) **(10 marks)**

Question 03 — 25 Marks

a) A soft-drink machine is regulated so that it fills an average of 250 milliliters per cup. If the amount of drink is normally distributed with a standard deviation equal to 15 milliliters,

i. What is the probability that cups will contain more than 225 milliliters?

(3 marks)

ii. What is the probability that a cup contains between 195 and 209 milliliters?

(3 marks)

iii. How many cups will probably overflow if 260-milliliter cups are used for the next 1000 drinks? (Round up the answer to the nearest integer) **(3 marks)**

iv. Below what value do we get the smallest 25% of the drinks? (Give your answer in two decimal places) **(3 marks)**

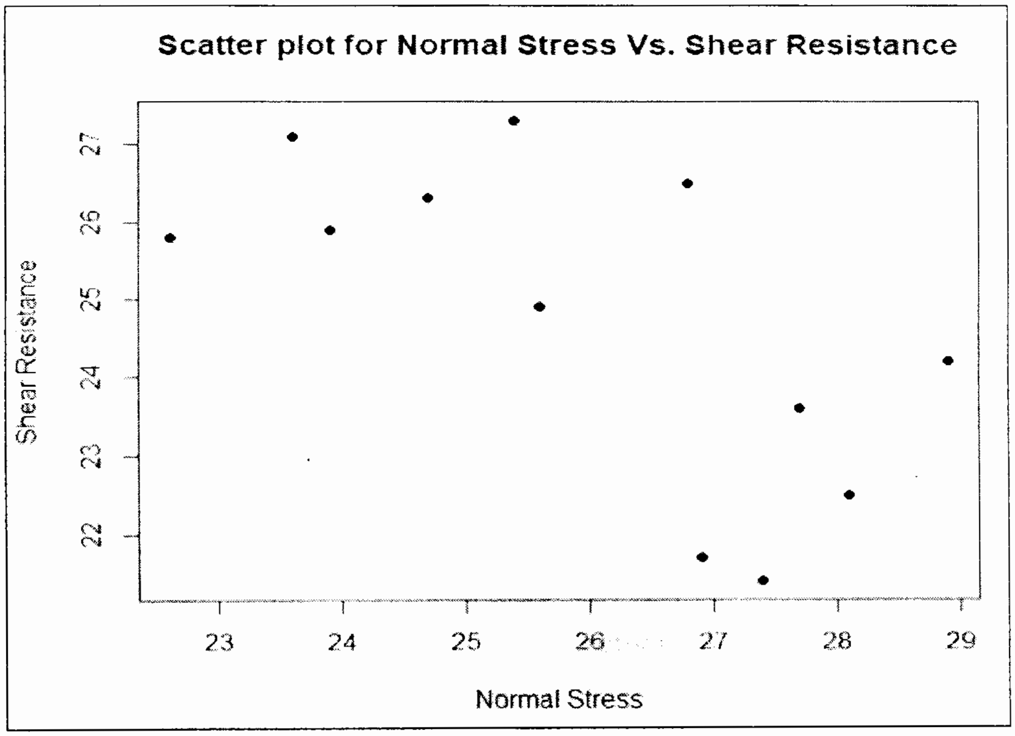
b) A cellular phone company conducts a survey to determine the ownership of cellular phones in different age groups. The results for 1044 households are shown in the table below. Test whether that the ownership of cellular phones are independent from the age group. Consider 5% level of significance. (Give your answer in 2 decimal places)

Cellular - Phone	Age Group				Total
	18 - 24	25 - 54	55 - 64	≥ 65	
Yes	50	80	78	56	264
No	200	180	210	190	780
Total	250	260	288	246	1044

(13 marks)

Question 04 — 25 Marks

In a certain type of metal test specimen, the normal stress on a specimen is known to be functionally related to the shear resistance. To examine the relationship between the normal stress and the shear resistance, data were collected from a sample of 12 specimen. Figure below displays the scatter plot for the data. (Give all of your answers in four decimal places)



R outputs of the regression model are shown below.

Regression Model

Coefficients

Intercept	Normal Stress
42.5818	-0.6861

Analysis of Variance Table

Response: Shear Resistance

	Df	Sum Sq	Mean Sq	F value	*Pr(>F)*
Normal Stress	A	D	20.2621	G	0.0206*
Residuals	B	26.884	F		
Total	C	E			

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

a) What can be concluded using the scatterplot? **(2 marks)**

b) Find values marked A, B, C, D, E, F and G in the ANOVA table (Show your work). **(10 marks)**

c) State the estimated regression equation in the form of $\hat{Y} = \hat{\alpha} + \hat{\beta}X$ and state in how much shear resistance will decrease if they increase the normal stress by 1unit. **(4 marks)**

d) Following information is given.

$\sum X = 311.6$ $\sum Y = 297.2$ $\sum XY = 7687.76$ $\sum X^2 = 8134.26$ $\sum Y^2 = 7407.8$

Where X is the normal stress and Y is the shear resistance.

i. Calculate Pearson's correlation coefficient between the two variables. **(5 marks)**

e) Use the regression equation to predict shear resistance if normal stress is 30. **(4 marks)**

End of the Question Paper

FORMULA SHEET

Probability & Statistics (IT2110)

Population Mean:

$$\mu = \frac{\sum X_i}{N}$$

Sample Mean:

$$\bar{X} = \frac{\sum X_i}{n}$$

Sample Variance:

$$S^2 = \frac{\sum(X_i - \bar{X})^2}{n-1}$$

Population variance:

$$\sigma^2 = \frac{\sum(X_i - \mu)^2}{N}$$

Pearson's Product moment correlation coefficient:

$$r = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sqrt{\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)\left(\sum_{i=1}^{n} y_i^2 - n\bar{y}^2\right)}}$$

The conditional probability of A given that B has occurred,

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Expected value of a discrete random variable:

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

Expected value of a continuous random variable:

$$E(X) = \int (x * f_X(x))dx$$

Variance of any random variable:

$$V(X) = E(X^2) - (E(X))^2$$

Translation to Z:

$$Z = \frac{X - \mu}{\sigma}$$

Chi square test statistic:

$$X^2 = \sum_{alli} \sum_{allj} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Simple linear regression equation:

$$\hat{Y} = \hat{\alpha} + \hat{\beta} X$$

Where,

$\hat{Y}$ = Estimated Y value

$\hat{\alpha}$ = Estimated intercept

$\hat{\beta}$ = Estimated regression coefficient

X = Predictor variable (Independent variable)