Sri Lanka Institute of Information Technology

B. Sc. Special Honours Degree / Diploma
in
Information Technology

Final Examination
Year 1, Semester 2 (2017)

IT 1070 – Probability and Statistics

Duration: 2 Hours

Instruction to Candidates:

- This paper contains 5 questions in 7 pages
- Answer **ALL** questions.
- Please show your work for full credit.
- Calculators are allowed.
- Use a 95% confidence level if not specified otherwise.
- Electronic devices retrieving text including electronic dictionaries and mobile phones are not allowed.

Question 1 **(20 Marks)**

a. A discrete numerical data set with N = 100 numbers has the following relative frequency table.

Relative frequency	0.25	0.1	0.2	0.2	m
Data Points	-1	0	1	3	4

i. What is the relative frequency 'm' of 4?
ii. What is the frequency of 0?
iii. What is the median of the data set?
iv. What is the first Quartile?

(12 marks)

b. The mean and standard deviation of a series of 8 data points (x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8) are 9.5 and $\sqrt{23.75}$.

Find;

i. the sum $\sum(x)$ of 8 data points,

ii. the sum of squares $\sum(x^2)$ of the 8 data points

(8 marks)

Question 2 **(20 Marks)**

a. Let A and B be two events associated with an experiment. Explain what is meant by "The Conditional probability of A given B" (4 marks)

b. Out of 50 people surveyed in a study, 35 smokes in which there are 20 males. What is the probability that if the person surveyed is a smoker then he is a male? (5 marks)

c. An experiment takes a random amount of time W, measured in seconds, to complete.
The probability density function of W is $f(w) = \alpha w + 1/2 \quad (0 < w < 1)$ where α is a constant.

   a. Calculate the value of α. (2 marks)
   b. Find the cumulative distribution function of W. (3 marks)
   c. Calculate the probability that the experiment takes between 0.5 and 0.9 seconds to complete. (3marks)
   d. Calculate the mean and variance of W. (3 marks)

Question 3 **(20 marks)**

a. State the conditions that should satisfy to approximate a Binomial distribution to Normal distribution. (2 marks)

In a box of 100 light bulbs, 10 are found to be defective. What is the probability that the number of defectives exceeds 13? (4 marks)

b. Suppose that we have a random sample of size $n = 40$ from a Normal distribution whose mean, μ has an unknown value but the variance, $\sigma^2 = 25.0$. From these data it has been calculated that the sample mean, $\bar{x} = 29.63$.

Show that there is no significant evidence at the 5% level to reject $H_0 : \mu = 30.0$ in favour of the alternative $H_1 : \mu < 30.0$. (6 marks)

c. You are off to soccer, and want to be the Goalkeeper, but that depends who is the Coach today:
- with Coach Sam the probability of being Goalkeeper is 0.5
- with Coach Alex the probability of being Goalkeeper is 0.3

Sam is Coach more often; about 6 out of every 10 games (a probability of 0.6).

What is the probability you will be the Goalkeeper today? (8 marks)

Question 4 **(20 Marks)**

a. State the general factors involved in determining sample size. (2 marks)

A study is to be performed to determine a certain parameter in a community. From a previous study a standard deviation of 46 was obtained. A sample error of up to 4 is to be accepted. How many subjects should be included in the sample of study at 95% level of confidence? (3 marks)

b. A company which sells mineral water has published the following data which depicts the sale of one-liter water bottles.

SALE OF ONE LT WATER BOTTLES '000S

Year	**Quarter 1**	**Quarter 2**	**Quarter 3**	**Quarter 4**
2010	20	32	62	29
2011	21	42	75	31
2012	23	39	77	48

a. Assuming the Additive Model, establish the trend values for the data using a suitable moving average. *5 marks*

b. Estimate seasonal variations for each quarter. *5 marks*

c. Forecast sales for each quarter of the year 2013, if the regression model is,
Estimated trend = 32.4 + 1.414*time index. *5 marks*

Question 5 **(20 Marks)**

a) A police authority conducts an eight-week experiment. In each week, it records the number of foot patrols (X) made in a small town and the number of reported crimes (Y) in that town. The data are summarised as follows.

$$\sum_i x_i = 52 \qquad \sum_i x_i^2 = 380 \qquad \sum_i x_i y_i = 1335$$

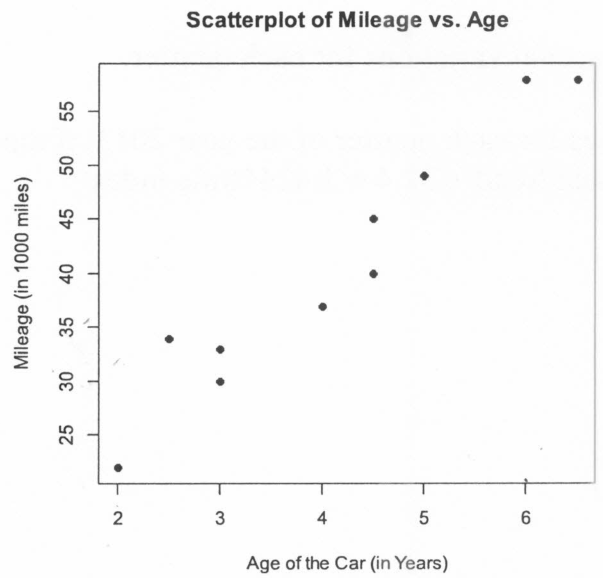
$$\sum_i y_i = 225 \qquad \sum_i y_i^2 = 7007 \qquad n = 8$$

i) Calculate the Pearson's product moment correlation coefficient. (4 Marks)

ii) Test whether the correlation is significant, at 1% level of significance. (5 Marks)

b) A second-hand car dealer has 10 cars for sale. She decides to investigate the link between the age of cars (in years) and the mileage (in thousand miles). The data collected, the scatterplot and the R outputs are given and the cars are shown in the table below.

Age	2	2.5	3	4	4.5	4.5	5	3	6	6.5
Mileage	22	34	33	37	40	45	49	30	58	58



R Output

Coefficients:

(Intercept)	Age
8.892	7.734

Analysis of Variance Table

Response: Age

	Df	Sum Sq	Mean Sq	F Value	Pr(>F)	
Mileage	1	1190.21	**D**	**G**	4.05e-06	***
Residuals	**A**	78.19	**E**			
Total	**B**	**C**	**F**			

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

i) What can be concluded using the scatterplot? (1 Mark)

ii) Find values marked A, B, C, D, E, F and G in the ANOVA table (Show workings). (8 Marks)

iii) State the estimated regression equation in the form of $\hat{Y} = \hat{\alpha} + \hat{\beta}X$ and state how much more will a car be driven in a 12-month period. (2 Marks)

End of the Question Paper

FORMULA SHEET

Business Statistics – IM209

Mean **Sample Variance** **Population variance**

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N}$$

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

$$\sigma^2 = \frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N}$$

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 Discrete 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 Varience of a continuius 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: $Z = \frac{X - \mu}{\sigma}$

PEARSON PRODUCT MOMENT CORRELATION

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

Linear regression Equation is shown below

Regression Formula:

Regression Equation(y) = a + bx

Where;

$Slope\ (b) = \frac{(N\Sigma XY - (\Sigma X)(\Sigma Y))}{(N\Sigma X^2 - (\Sigma X)^2)}$ Intercept(a) = (ΣY - b(ΣX)) / N

X – Independent variable and Y – Dependent variable