Sri Lanka Institute of Information Technology

Year 02 – Semester I – 2025

Probability and Statistics – IT2120

Review Tutorial

1) 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

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

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.

2) 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.

   i.   Calculate the value of α.

   ii. Find the cumulative distribution function of W.

   iii. Calculate the probability that the experiment takes between 0.5 and 0.9 seconds to complete.
        iv. Calculate the mean and variance of W.

3) Waiting time for a bus X, measured in seconds has following probability density function.

$$f_X(x) = \begin{cases} k(x^2 - 5) & ; \ -2 \le x \le 2 \\ 0 & ; \ \text{otherwise} \end{cases}$$

   i.   Calculate the value of k.

   ii. Find the $F_X(1.25)$.

   iii. Find $E(X)$ and $V(X)$.

Department of Mathematics and Statistics – Faculty of Humanities and Sciences – SLIIT

4) The number of calls coming per-minute into a hotels reservation center is Poisson random variable with mean 3.

    i. Find the probability that no calls come in each 1minute period.

    ii. Assume that the number of calls arriving in two different minutes are independent. Find the probability that at least two calls will arrive in a given two-minute period.

5) Suppose that in late summer, the Fremantle Surf Life Saving club makes an average of seven surf rescues per day. Without using any approximation, find the probability that,

    i. More than two rescues are made on a particular day.

    ii. Three to five surf rescues are made on a particular day.

    iii. If club makes an average of 23 surf rescues per day, using suitable approximation, find the probability that fewer than seven rescues are made on a particular day.

6)

    i. State the conditions that should satisfy to approximate a Binomial distribution to Normal distribution.

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

7) A sample of 25 plastic hinges was subjected to repealed stress cycles until failure. The number of cycles which each survived is given below.

72	35	63	67	87	71	64	47	60	81	39	52	57
74	43	55	37	83	48	91	53	44	94	65	75	

    i. Calculate, to 2 decimal places, the mean and standard deviation for this data.

    ii. Construct 95% confidence interval for the true mean of number of stress cycles plastic hinges could survive.

8) In one of popular American journals, researchers reported that individuals who practice Transcendental Meditation (TM) lower their blood pressure significantly. After the mediation, if a random sample of 300 female TM practitioners shows a mean blood pressure of 115 *mm hg* with a standard deviation of 3 *mm Hg*, does it suggest that, on average, females who practice TM meditate lower their blood pressure than 120 *mm Hg*? Test at 5% level of significance. (Round up the test statistic up to 2 decimal places)

9) A manufacturer claims that the thickness of the spearmint gum it produces is 7.5 one-hundredths of an inch. A quality control specialist regularly checks this claim. On one production run, he took a random sample of n = 10 pieces of gum and measured their thickness. He obtained:

7.65	7.60	7.65	7.70	7.55
7.55	7.40	7.40	7.50	7.50

   Test at 5% level that the manufacturer's claim can be accepted.

10) A sample of 42 batteries of a newly produced brand was subjected to test their lifetimes before it is advertised for marketing. The lifetimes in hours which each survived is given below.

112	105	123	137	157	134	143	155	137	98	141	104	108	99
147	97	131	153	144	94	139	115	152	115	137	115	110	100
112	95	98	102	105	130	90	99	96	110	117	114	120	140

   For this data, $\bar{x}$ = 119.76 and S = 6.5123.

   i. Construct 95% confidence interval for the true mean lifetime of batteries.

11) The quality-control manager at a light bulb factory state that 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. At the 0.05 level of significance, is there evidence that the mean life is increased from 375 hours?

12) Suppose that we are told that the heights of adult males in a particular region of the world are normally distributed with a mean of 70 inches and standard deviation of 2 inches.

   i. Approximately what proportion of adult males are taller than 73 inches?
   ii. What proportion of adult males are between 72 and 73 inches?
   iii. What height corresponds to the point where 20% of all adult males are greater than this height?
   iv. If there are 2000 adult males in this region, how many males are there whose height is less than 73 inches?

13) A large group of students took a test in Physics and the final grades have a mean of 70 and a standard deviation of 10. If we can approximate the distribution of these grades by a normal distribution, what percent of the students,

    i.   Scored higher than 80?

    ii.   Should pass the test (grades≥60)?

    iii.   Should fail the test (grades<60)?

14) A large corporation is interested in determining whether a relationship exists between the commuting time of its employees and the level of stress-related problems observed on the job. A study of 116 workers reveals the following:

STRESS LEVEL

| COMMUTING TIME	High	Moderate	Low	Total
Under 15 min.	9	5	18	32
15–45 min.	17	8	28	53
Over 45 min.	18	6	7	31
Total	44	19	53	116

At the 0.05 level of significance, is there evidence of a significant relationship between commuting time and stress level?

15) Three candidates, Sunil, Nimal and Kamal are contesting for the presidency of the newly formed Mathematical Society of the City School. As the poling date approaches, a group of interested students conducted an opinion survey with 120 voters to predict the winning candidate at the elections. Following are the results.

Candidate	Sunil	Nimal	Kamal	Total
Votes	38	43	39	120

Do the above results suggest that all three candidates are equally popular to win at the elections?

16) The objective of the director of planning of Sunflowers Apparel Ltd., is to forecast annual sales for all new stores, based on store size. To examine the relationship between the store size in square feet and its annual sales, data were collected from a sample of 14 stores. Figure below displays the scatter plot for the data.

Scatter Diagram for Site Selection
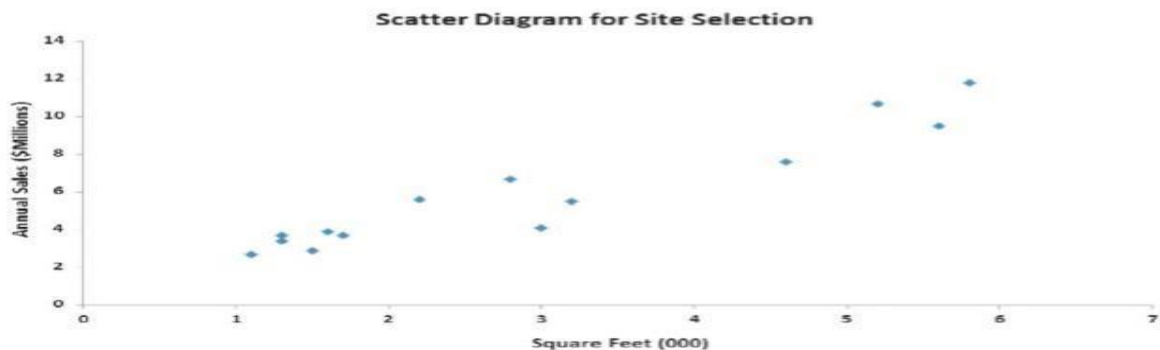
R outputs of the regression model are shown below

Regression Model

Coefficients

Intercept	Square feet
0.9645	1.6699

Analysis of Variance Table

Response: Annual Sales

	df	Sum Sq	Mean Sq	F value	Pr(>F)
Square feet	A	105.7476	E	G	0.000 ***
Residuals	B	11.2067	F		
Total	C	D			

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

i.   What can be concluded using the scatterplot?

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

iii.   State the estimated regression equation in the form of $\hat{Y} = \hat{\alpha} + \beta X$ and state how much more sales are expected if they increase the store size by 1,000 square feet.

iv.   Following information is given;

$$\sum X = 40.9$$
$$\sum Y = 81.8$$
$$\sum XY = 302.30$$

Standard deviations are $S_x = 1.71$ and $S_y = 2.98$
Where X is the square feet and Y is the annual sales

Calculate the Pearson's linear coefficient of correlation between the two variables.

v.   Use the regression equation to predict annual sales, if store size is 10,000 square feet.

17) 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
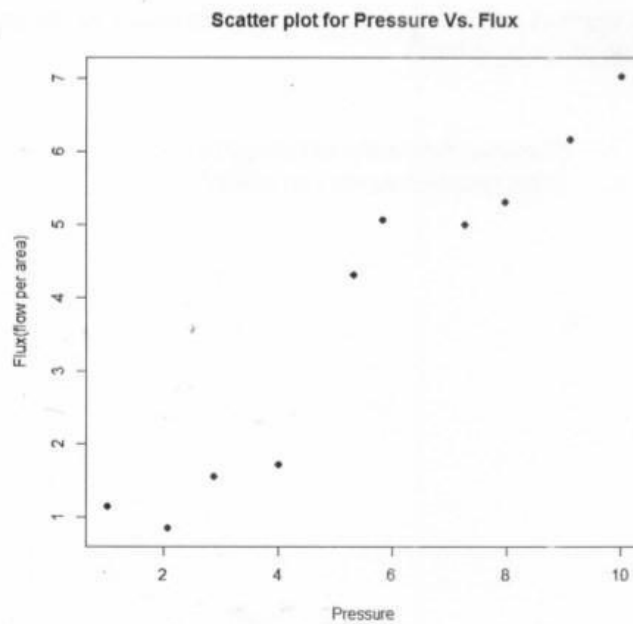
R Output

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

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

ii. State the estimated Regression equation. iii. Test whether the slope of the Regression line is significant and state the conclusions. iv. Find the estimated Flux when the pressure takes the value 5 units. v. Provide two meaningful conclusions from the above analysis.

18)

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 summarized 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.

ii. Test whether the correlation is significant, at 1% level of significance.

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

Scatterplot of Mileage vs. Age
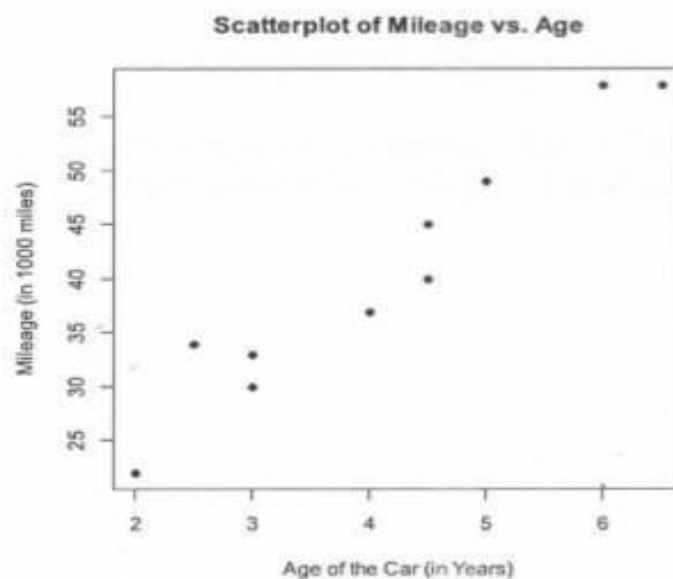
R Output

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?

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

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

19) A financial analyst wants to understand the drivers of company profitability (ROA: Return on Assets (%) - measure of profitability). The model uses following factors:

- **R_D_Spend:** Research & Development spending (in millions)
- **Employee_Productivity:** Revenue per employee (in thousands)
- **Debt_Ratio:** Total debt to total assets ratio
- **Asset_Turnover:** Sales revenue divided by total assets

Data was collected from 150 companies. Following are the R outputs for the regression model:

```
Call:
lm(formula = ROA ~ R_D_Spend + Employee_Productivity + Debt_Ratio +
    Asset_Turnover, data = company_data)

Residuals:
    Min      1Q  Median      3Q     Max
-4.1234 -0.9567 -0.0789  0.9234  4.0567

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)             7.89234    1.23456   6.394 1.23e-09 ***
R_D_Spend               0.24891    0.02345  10.614  < 2e-16 ***
Employee_Productivity   0.07923    0.00789  10.042  < 2e-16 ***
Debt_Ratio            -12.45678    1.56789  -7.945 9.87e-13 ***
Asset_Turnover          4.18901    0.45678   9.171  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.467 on 145 degrees of freedom
Multiple R-squared:  0.8123,    Adjusted R-squared:  0.8071
F-statistic: 156.7 on 4 and 145 DF,  p-value: < 2.2e-16
```

i.   Write the equation for the fitted regression model.
ii. Are all the regression coefficients significant in the fitted model? Justify your answer.
iii. What is the coefficient of determination? Interpret the value.
iv. Is the fitted regression model significant? Justify your answer.