

THE OPEN UNIVERSITY OF SRI LANKA
BSC IT DEGREE PROGRAMME - LEVEL 05
FINAL EXAMINATION - 2026/2027
ADU 5330 – INFERENCE TECHNIQUES AND STATISTICAL MODELING



DURATION: Two hours

Date: 28-06-2026

Time: 1.30 to 3.30 p.m.

Answer four questions only. Statistical tables are provided.
Non programmable calculators are permitted.

1.

(a) Briefly explain the following terms :

- (i) Point Estimators
- (ii) Interval Estimators

(b) Sample of lifetimes of 12 batteries produced by the *ABC Company* given below. From the past experience it is reasonable to assume that the lifetime of the batteries follows normal distribution with known standard deviation 4 hours.

102 105 98 110 108 101 99 104 107 103 106 100

- (i) Construct 95% confidence interval and comment on the claim that “mean lifetime of batteries equal to 100 hours”
- (ii) A company claims that the mean lifetime of its batteries exceeds 100 hours. Test the company’s claim at the 5% significance level.

2.

(a) Briefly explain the following terms:

- (i) Type I and Type II error
- (ii) Significance level

(b) A researcher claims that the average time taken to complete a specific task of students is 170 minutes. A sample of 10 students observed and taken the following times to complete the task. From the past experience it is reasonable to assume that the time taken to complete the task follows normal distribution.

168 172 169 171 173 167 170 174 166 169

- (i) Construct 95% confidence interval and comment on the claim that “the average time taken to complete a specific task of students is 170 minutes”
- (ii) Using a suitable statistical test check the validity of the claim “the average time taken to complete a specific task of students is less than 170 minutes”. Use 5% level of significance.

3.

(a) Briefly explain the following

- (i) Null and alternative hypotheses
- (ii) The steps involved in hypothesis testing

(b) Students' test scores before and after using an online learning platform given below. It is reasonable to assume that scores are follows normal distribution.

Student	Before	After
A	68	74
B	72	78
C	70	75
D	65	71
E	74	80
F	69	76

- (i) Construct 95% confidence interval and comment on the claim that "test scores before and after using an online learning platform are same"
- (ii) Using a suitable statistical test check whether the online learning platform improved the student scores. Clearly state your findings and conclusions.

4.

(a) A researcher aimed to examine whether a newly introduced teaching method enhances students' academic performance. After implementing the new method, the scores of 10 students were collected. Before the introduction of this teaching method, the median score of the students was 78. To evaluate the effectiveness of the new teaching method, the researcher performed a one-sample signed-rank test for the median. The objective was to determine whether the new teaching method led to an improvement in student performance.

- (i) State whether the test conducted by the researcher is a one-tailed test or a two-tailed test.
- (ii) The following Table-1 presents the Minitab output for the test. Clearly state the hypotheses tested and provide an appropriate conclusion with justification based on the output.

Table-1: Minitab output for the test

Sign Test for Median: Method

Sign test of median = 78.00 versus > 78.00

	N	Below	Equal	Above	P	Median
Method	10	5	1	4	0.7461	75.00

- (b) A study was conducted to determine whether the type of defects found in electronic devices manufactured at a production plant is associated with the production shift. A total of 309 defective electronic devices were observed, and each defect was categorized into one of four types: A, B, C, or D. In addition, every defective device was identified according to the production shift during which it was produced. A Chi-Square test was performed to analyze the relationship between defect type and production shift, and the corresponding Minitab statistical output is presented in Table-2.

Table-2: Minitab output for Chi-Square test

Tabulated statistics: Shift, Type of defect					
Using frequencies in Number of defects					
Rows: Shift Columns: Type of defect					
	A	B	C	D	All
1	15	21	45	13	94
	22.51	20.99	<u>a</u>	11.56	
2	26	31	34	5	96
	22.99	<u>b</u>	<u>c</u>	11.81	
3	33	17	49	20	119
	28.50	26.57	49.29	14.63	
All	74	69	128	38	309
Cell Contents: Count					
Expected count					
Pearson Chi-Square = <u>d</u> , DF = <u>e</u> , P-Value = 0.004					

- State the null hypothesis and the alternative hypothesis relevant to the given problem.
- Determine the expected cell frequencies denoted by a, b, and c.
- Compute the Pearson Chi-Square statistic represented by d.
- Determine the corresponding degrees of freedom represented by e.
- Using a 5% level of significance, provide a clear conclusion for the test.

5.

The time required to complete two specific tasks, X and Y, measured in hours, was recorded for 15 employees at ABC Company. A regression analysis was performed to examine the relationship between the time taken to complete task X and task Y. The corresponding Minitab outputs are presented in Table 3 and Table 4.

Table-3: Descriptive Statistics for X, Y & XY

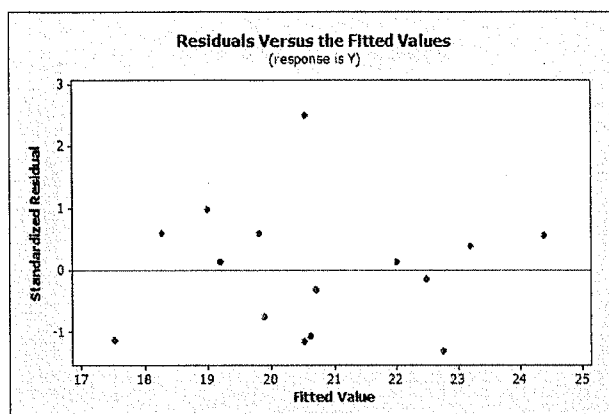
Descriptive Statistics					
Variable	N*	Mean	StDev	Sum	Sum of Squares
X	0	25.692	1.717	385.377	9942.304
Y	0	20.725	2.174	310.879	6509.232
X*Y	0	535.5	89.5	8033.2	4414348.5

Table 4: Minitab output for Regression analysis of Y vs X

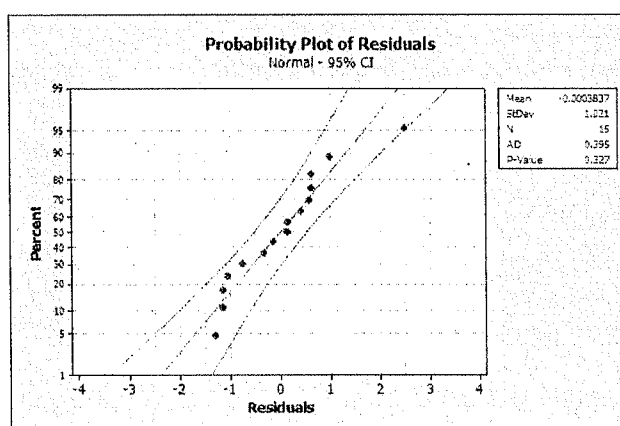
Regression Analysis: Y versus X					
The regression equation is					
$Y = \alpha + \beta X$					
Predictor	Coef	SE Coef	T	P	
Constant	α	4.257	-1.87	0.084	
X	β	0.1653	6.76	0.000	
S = 1.06212 R-Sq = 77.8% R-Sq(adj) = 76.1%					
Analysis of Variance					
Source	DF	SS	MS	F	P
Regression	1	51.499	51.499	45.65	0.000
Residual Error	13	14.665	1.128		
Total	14	66.164			

- (i) Estimate the intercept and slope coefficients of the regression model, and interpret their meanings in the context of the problem.
- (ii) Examine whether the fitted regression model is statistically significant.
- (iii) Interpret the proportion of total variation in the response variable that is explained by the fitted regression model.
- (iv) Test whether the slope coefficient of the regression model is statistically significant.
- (v) Residual analysis was performed and the corresponding graphs are provided below. Interpret the results shown in Graph-1 and Graph-2.

Graph-1



Graph – 2



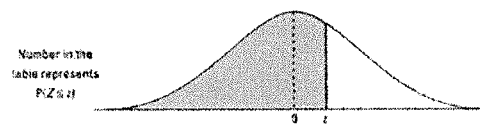
6.

- (a) Briefly describe the assumptions in regression analysis.
 (b) Following table gives the data on study hours (X) vs average exam marks (Y) of a certain subject.

Hours Studied (X)	Average Exam Marks (Y)
1	40
2	48
3	50
4	58
5	60
6	62
7	63
8	62
9	75
10	76
11	73
12	85

- (i) Construct an appropriate plot for the data and interpret the observed relationship between the variables.
- (ii) Fit a simple linear regression model to examine the relationship between study hours (X) and average examination marks (Y).
- (iii) Interpret the estimated slope and intercept of the fitted regression model.
- (iv) Estimate the expected examination marks when the study hours are equal to 13.

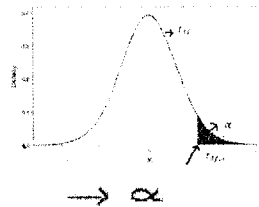
Standard normal distribution table



STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z score.

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449
1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327
1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856	.96926	.96995	.97062
1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558	.97615	.97670

T-distribution table - t – Distribution table - right tail probabilities



Degrees of freedom (df)	.2	.15	.1	.05	.025	.01	.005	.001
8	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501
9	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297
10	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144
11	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025
12	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930
13	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852
14	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787
15	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733
16	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686
17	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646
18	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610
19	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579
20	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552

Some useful formulas

$(1 - \alpha)\%$ confidence interval for μ when σ known $\left[\bar{X} - Z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}, \bar{X} + Z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}} \right]$

$(1 - \alpha)\%$ confidence interval for μ when σ unknown $\left[\bar{X} - t_{(n-1), \alpha/2} * \frac{\sigma}{\sqrt{n}}, \bar{X} + t_{(n-1), \alpha/2} * \frac{\sigma}{\sqrt{n}} \right]$

$(1 - \alpha)\%$ confidence interval for θ for large samples $\left[P - Z_{\frac{\alpha}{2}} * \sqrt{\frac{P(1-P)}{n}}, P + Z_{\frac{\alpha}{2}} * \sqrt{\frac{P(1-P)}{n}} \right]$

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

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

$$\text{Pearson Chi-square } (\chi^2) = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

One Sample Z-Test: $z = (\bar{X} - \mu_0) / (\sigma/\sqrt{n})$

One-Sample t-Test: $t = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$

Two sample t-test

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

Where:

- $\bar{x}_1$ and $\bar{x}_2$ are the sample means of the two groups being compared
- n_1 and n_2 are the sample sizes of the two groups being compared
- s_p^2 is the pooled sample variance, calculated as:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Where s_1^2 and s_2^2 are the sample variances for each group.

Expected frequency $E_{ij} = \frac{(\text{total of row } i) \times (\text{total of column } j)}{\text{grand total}}$

$$\text{Pearson Chi-square } (\chi^2) = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

$DF = (r - 1)(c - 1)$

Simple linear regression model $Y = \beta_0 + \beta_1 X$

$$\beta_0 = \bar{Y} - \beta_1 \bar{X}$$

$$\beta_1 = \frac{\sum X_i Y_i - \frac{(\sum X_i)(\sum Y_i)}{n}}{\sum X_i^2 - \frac{(\sum X_i)^2}{n}}$$