**THE OPEN UNIVERSITY OF SRI LANKA**
**BSE DEGREE PROGRAMME - LEVEL 05**
**FINAL EXAMINATION - 2024/2025**
**ADU 5330 – INFERENTIAL TECNIQUES AND STATISTICAL MODELING**

**DURATION: TWO HOURS**

**Date: 12-06-2025** **Time: 1.30 to 3.30 p.m.**

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

1.

(a) Explain the terms "*sampling distribution*" and "*properties of an estimator*". (30 marks)

(b) Suppose the random variables $X$ denote the diameter (in cm) of a manufactured laptop screen. Assume that $X$ is normally distributed with unknown mean and variance. Randomly selected fifteen laptop screens were checked and diameters were recorded. Following table gives the data in cm.

| | | | | |
|---|---|---|---|---|
| 26.12 | 26.53 | 25.69 | 25.66 | 25.74 |
| 26.24 | 25.42 | 26.53 | 26.43 | 25.05 |
| 27.52 | 26.45 | 26.74 | 25.79 | 26.91 |

(i) Estimate the mean and standard deviation of diameter for a randomly selected laptop screen. Round your answer to two decimal places. (15 marks)

(ii) Find 95% confidence interval for mean length of diameter for laptop screen. (40 marks)

(iii) Manufacturer clam that the "mean diameter of a laptop screen is 26 cm". Using the answer in part (ii) justify the claim. (15 marks)

2.

(a) The booting times of two operating systems, A and B, are presented in the following contingency table. Using a suitable statistical test check the claim that "Booting time of machine brand *XY* is independent from the operating system". Use 0.05 level of significance and for significance level 0.05 table (Critical ) value for $\chi^2_{table} = 3.84$.

| | | Booting time of a machine brand *XY* | |
|---|---|---|---|
| | | 30 seconds or less than 30 seconds | Greater than 30 seconds |
| Operating System | A | 25 | 10 |
| | B | 40 | 20 |

(50 marks)

(b) A teacher wants to model the marks obtained by a student for a computer course using the time spent for study (in hours) for the course. Data, time spent for study for the course and the marks obtained were collected from 50 students. The Minitab output of the simple linear regression analysis of the collected data is given bellow.

**Regression Analysis: Marks versus Time spent for study**

```
The regression equation is
Marks = 08.2 + 0.8 Time spent for study

Predictor              Coef     SE Coef        T       P
Constant               08.2     2.002      10.09  0.0000
Time spent for study   0.08     0.02009    27.10  0.0000

S = 2.72271   R-Sq = 83.9%   R-Sq(adj) = 83.7%

Analysis of Variance
Source          DF      SS      MS        F        P
Regression       1  5442.4  5442.4   734.16    0.000
Residual Error  48   355.8     7.4
Total           49  5798.2
```

Write a brief report on findings based on the above output. Clearly state the fitted model, interpret the fitted model, and check whether the slope and the intercept are significant. Also explain the proportion of variation explained by the fitted model. State if any additional techniques which are needed to finalise the fitted model. (50 marks)

3.

(a) Explain the following terms and use of it in hypotheses testing:

(i) Significance level (15 marks)

(ii) Power of the test (15 marks)

(iii) Test statistic and critical region (20 marks)

(b) According to past experience a company that produces a certain electronic chip claims that the life time $X$ (in years) has a normal distribution with unknown mean and variance. A sample drawn from the above distribution is given below.

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 8.16 | 11.50 | 7.28 | 10.07 | 12.90 | 11.01 | 11.11 | 8.27 | 11.65 | 9.68 | 11.46 | 6.26 |
| 10.39 | 8.05 | 10.48 | 14.62 | 8.96 | 10.72 | 10.37 | 11.71 | | | | |

Using a suitable statistical test comment on the claim that "mean lifetime of a randomly selected electronic chip is less than 10.5 years". (50 marks)

4.

(a) Explain the following terms in hypotheses testing:

(i) Null Hypotheses and Alternative hypotheses (15 marks)

(ii) One tail tests and two tail tests (15 marks)

(b) A shoe company wants to compare two materials, $A$ and $B$, for use on the soles of boys' shoes. To test this, each of ten boys in a study wore a special pair of shoes with the sole of one shoe made from Material $A$ and the sole on the other shoe made from Material $B$. The sole types were randomly assigned to account for systematic differences in wear between the left and right foot. After three months, the shoes are measured for wear. Using a suitable statistical test check the validity of the claim that "wear of both materials $A$ and $B$ are same". Use 5% level of significance. It is reasonable to assume normal distribution for wear of both materials $A$ and $B$. Connect the below table with the above paragraph. (70 marks)

| Wear of Material-$A$ | Wear of Material-$B$ |
|---|---|
| 13.2 | 14 |
| 8.2 | 8.8 |
| 10.9 | 11.2 |
| 14.3 | 14.2 |
| 10.7 | 11.8 |
| 6.6 | 6.4 |
| 9.5 | 9.8 |
| 10.8 | 11.3 |
| 8.8 | 9.3 |
| 13.3 | 13.6 |

5.

(a) Briefly Explain the terms "point estimation and interval estimation". (20 marks)

(b) A software engineer collected a sample of processing times for a specific task using two algorithms, A and B. The table below shows the recorded times in seconds. It is reasonable to assume equal variances and normal distribution for time of processing for both algorithms.

| Algorithm A | Algorithm B |
|---|---|
| 17 | 22 |
| 14 | 19 |
| 14 | 18 |
| 15 | 19 |
| 21 | 24 |
| 15 | 18 |
| 15 | 22 |
| 13 | 24 |
| 13 | 25 |
| 13 | 22 |
| | 17 |
| | 21 |
| | 23 |
| | 25 |
| | 25 |

(i) Find 95% confidence interval for mean difference of processing time of two algorithms $A$ and $B$. Comment on the claim "processing time of two algorithms $A$ and $B$ are equal". (50 marks)

(ii) The following Minitab output is generated to test the claim that "lowest processing time of the highest 50% runs of the algorithm $A$ is less than 15 seconds". Write a report on the findings. Clearly mention the test which is used to test the claim, whether the test is parametric or nonparametric, the hypotheses tested and conclusion with justification. (30 marks)

| **Wilcoxon Signed Rank Test: Algorithm A** |
|---|

```
Test of median = 15.00 versus median < 15.00

                    N
                  for   Wilcoxon           Estimated
              N  Test  Statistic       P      Median
Algorithm  A  10    7       11.5  0.368       14.50
```

6.

(a) Briefly explain the "coefficient of determination". (20 marks)

(b) Assignment marks and final examination marks of a particular subject for 15 students are given below. The teacher wants to find a model to predict the Final Mark ($Y$) of a student using the Assignment mark.

| Student Name | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Assignment mark (X) | 60 | 47 | 60 | 56 | 47 | 27 | 45 | 61 | 68 | 62 | 35 | 53 | 57 | 25 | 31 |
| Final Mark (Y) | 67 | 54 | 53 | 49 | 47 | 35 | 30 | 77 | 57 | 54 | 42 | 60 | 63 | 42 | 28 |

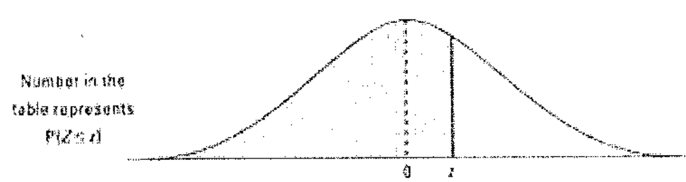
n = 15 $\sum X_i = 734$ $\sum Y_i = 758$ $\sum X_i Y_i = 39066$ $\sum X_i^2 = 38566$ $\sum Y_i^2 = 40924$

(i) Draw a scatter plot for the above data and interpret it. (25 marks)

(ii) Fit simple linear regression model $Y = \beta_0 + \beta_1 X$ to the above data. Clearly estimate the $\beta_0$ and $\beta_1$ and interpret the results. (30 marks)

(iii) Plot the residuals against the fitted values and interpret the results. (25 marks)

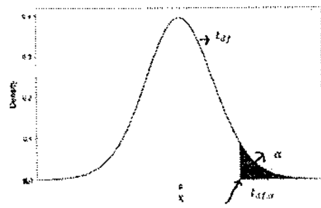
**Standard normal distribution table**




**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



$\longrightarrow \alpha$

| 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 $\mu$ when $\sigma$ 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 $\mu$ when $\sigma$ un known

$$\left[\bar{X} - t_{(n-1),\ \alpha/2} * \frac{\sigma}{\sqrt{n}} \, , \, \bar{X} + t_{(n-1),\ \alpha/2} * \frac{\sigma}{\sqrt{n}}\right]$$

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

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

Sample variance $S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n-1} = 1/(n-1)\,(\Sigma x_i^2 - n\bar{x}^2)$

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

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

**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.

Simple linier regression model

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X}$$

$$\hat{\beta}_1 = \frac{\sum X_iY_i - \frac{(\sum X_i \sum Y_i)}{n}}{\sum X_i^2 - \frac{(\sum X_i)^2}{n}}$$