UNIVERSITI TUN HUSSEIN ONN MALAYSIA

FINAL EXAMINATION
SEMESTER II
SESSION 2023/2024

COURSE NAME	:	ENGINEERING STATISTICS
COURSE CODE	:	BDA 24103
PROGRAMME CODE	:	BDD
EXAMINATION DATE	:	JULY 2024
DURATION	:	3 HOURS
INSTRUCTIONS	:	1. PART A : ANSWER ALL QUESTIONS PART B : ANSWER **THREE (3)** FROM FOUR QUESTIONS ONLY 2. THIS FINAL EXAMINATION IS CONDUCTED VIA ☐ Open book ☒ Closed book 3. STUDENTS ARE **PROHIBITED** TO CONSULT THEIR OWN MATERIAL OR ANY EXTERNAL RESOURCES DURING THE EXAMINATION CONDUCTED VIA CLOSED BOOK

THIS QUESTION PAPER CONSISTS OF **TWELVE (12)** PAGES

PART A

Q1 The resident of Taman Inspira is worried about a rise in housing costs in the area. The head of the people thinks home prices fluctuate with the land values. Data on 10 recently sold houses and the cost of the land on which they were built are seen here in thousands of ringgit as illustrated in **Table Q1.1**

Table Q1.1 Data on Sold Houses and Cost of the Land

Land values, x	7.0	6.9	5.5	3.7	5.9	3.8	8.9	9.6	9.9	10.0
Cost of the house, y	67	63	60	54	58	36	76	87	89	92

(a) Draw a scatter plot for the variables

(4 marks)

(b) Distinguish the regression line using the least squares method. Interpret the result by showing the least squares line in the scatter plot in **Q1** (a).

(8 marks)

(c) Estimate the cost of the house when the land value is RM 73000

(1 mark)

(d) Test the hypothesis concerning $H_o : B_1 = 1$ against the $H_1 : B_1 > 1$ at the 0.05 level of significance

(7 marks)

Q2 (a) State the relationship between the number of factors (k) and the total number of runs (N) in a Two-Level Full Factorial Design?

(2 marks)

(b) Suppose we are conducting a two-level fractional factorial experiment with four factors. Identify the total number of runs if we choose to run only a quarter of the complete factorial experiments?

(3 marks)

(c) Removing ammoniacal nitrogen is an essential aspect of leachate treatment at landfill sites. The rate of removal (in per cent per day) is recorded for several for each of the several treatment methods. The results are presented in **Table Q2.1.**

Table Q2.1 The rate of ammoniacal nitrogen removal (in per cent per day)

Treatment	Rate of Removal		
A	5.21	4.65	4.85
B	5.59	2.69	7.57
C	6.24	5.94	6.41
D	6.85	9.18	4.94
E	4.04	3.29	4.52

(i) Develop an ANOVA table.

(8 marks)

(ii) Conclude the treatment methods differ with their removal rates

(2 marks)

(d) An electronics engineer conducted a study to determine the effect of five different types of coating for cathode ray tubes on tube conductivity in a telecommunications system display device. The average of conductivity data is shown in **Table Q2.2**. The computed value of F was found to be significant at $\alpha = 0.01$, given MSE = 16.217 and n = 4. Are coating types 1 and 3 resulted in different conductivity? Use a Least Significant Difference (LSD) to test it.

(5 marks)

Table Q2.2 Conductivity of different coating types

Coating Type	Average Conductivity
1	145.00
2	145.25
3	131.5
4	129.25
5	145.25
Grand	696.25

PART B

Q3 (a) State the difference between enumerative study and analytic study

(2 marks)

(b) State three basic methods of collecting data in engineering environments

(3 marks)

(c) On year 2020, a survey was conducted to study the performance of chemotherapy treatment on cancer patients. A general hospital recorded that 75% of the patients died after the chemotherapy treatment. Assume that the distribution of the cancer patients who died after the chemotherapy treatment is binomially distributed. If five patients were selected randomly, identify the probability that:

(i) All of them were died

(3 marks)

(ii) Only two of the patients were recovered

(3 marks)

(d) Consider two populations of foreigner students at University Tun Hussein Onn Malaysia who participated in reading programmed prior to taking a Malay course. The populations are those who earn an A grade and who earn B grade. Eight students has been chosen randomly from the population of A Grade and six students have been selected from the population B Grade. The sample mean for A Grade is 37 and sample standard deviation is 8.7014. While, for the B Grade, the sample mean is 25 and the sample standard deviation is 8.5264. Let X be the number of books read by the students who participated in the programmed. Calculate the probability that the mean number of books read by the students who earn A grade is greater than the students who earn B grade

(9 marks)

Q4 (a) There are type X and type Y of Durian trees in Mr. Ahmad's farm. The mean height of type X is 15.2 m with a standard deviation of 1.8 m. Meanwhile, type B has a mean height of 11.8 m with a standard deviation of 1.3 m. Two samples of sizes 20 and 12 are randomly selected from Druian trees of type A and B, respectively.

(i) Determine the mean of type X is more than 15 m

(4 marks)

(ii) Calculate the mean of type X is 3 m more than the mean of type Y

(8 marks)

(b) Data in **Table Q4.1** shows the tensile strength of stainless steel rods for Batch A and Batch B. Determine 90% confidence interval of μ_A-μ_B if the population variances are not equal.

Table Q4.1 Tensile Strength of Stainless Steels

	Batch A	Batch B
Sample Mean	600	530
Standard Deviation	28	32
Sample Size	25	16

(8 marks)

Q5 (a) A garment company recorded the time (minutes) to repair ten pairs of trousers (**Table Q5.1**). The sample data suggest that the average time to repair the trousers is less than 13 minutes. Assume the measurements were taken from the population with a normal distribution.

Table Q5.1 Time taken (minutes) for repairing ten trousers

10.2	9.5	11.9	13.1	10.5
11.3	15.1	13.0	12.5	14.2

(i) State the null and alternative hypothesis

(2 marks)

(ii) Test the hypothesis by using a 0.025 significance level

(8 marks)

(a) In the Science quiz, the sample size from School A and School B is 15 and 25, respectively. For School A, the mean score was 80 with a standard deviation of 3; for School B, the mean score was 75 with a standard deviation of 2. The test performance between School A and School B was different. Assume the population variances are unknown and not equal.

(i) State the null and alternative hypothesis

(2 marks)

(ii) Test the difference performance between School A and School B using 0.005 significance level

(8 marks)

Q6 (a) The data in **Table Q6.1** represent the monthly sales revenue (in thousands of dollars) for two different products, A and B, over a period of six months. Identify whether there is a positive or negative correlation between the sales revenue of product A and the sales revenue of product B over these six months.

Table Q6.1 Monthly sales revenue (in thousands of dollars)

Month	**Product A**	**Product B**
Jan	120	150
Feb	130	160
Mar	140	170
Apr	150	180
May	160	190
Jun	170	200

(4 marks)

(b) The following **Table Q6.2** represents the heights (in inches) of 25 students in a classroom.

Table Q6.2 : Heights of students in a classroom

64	68	72	70	66	68	69	65	71	67
70	66	67	70	68	73	69	71	65	67
68	70	71	72	66					

(i) Calculate the sample mean height

(2 marks)

(ii) Calculate the sample variance height

(4 marks)

(iii) Illustrate the distribution of heights using a histogram

(8 marks)

(iv) Identify whether the height data set exhibits a normal distribution.

(2 marks)

- END OF QUESTIONS -

APPENDIX A

EQUATIONS

- $P(X \le r) = F(r)$
- $P(X > r) = 1 - F(r)$
- $P(X < r) = P(X \le r-1) = F(r-1)$
- $P(X = r) = F(r) - F(r-1)$
- $P(r < X \le s) = F(s) - F(r)$
- $P(r \le X \le s) = F(s) - F(r) + f(r)$
- $P(r \le X < s) = F(s) - F(r) + f(r) - f(s)$
- $P(r < X < s) = F(s) - F(r) - f(s)$
- $f(x) \ge 1$.
- $\int_{-\infty}^{\infty} f(x)\, dx = 1$.
- $P(a < x < b) = P(a \le x < b) = P(a < x \le b) = P(a \le x \le b) = \int_a^b f(x)\, dx$

$$\mu = E(X) = \sum_{\text{all } X_i} X_i P(X_i)$$

$$\sigma^2 = Var(X) = E(X^2) - [E(X)]^2$$

$$E(X^2) = \sum_{\text{all } X_i} X_i^2 . P(X_i)$$

Note :

- $E(aX + b) = a\,E(x) + b$.

  $ar(aX + b) = a^2\,Var(x)$

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(x)\, dx \text{ for } -\infty < x < \infty.$$

$$\mu = E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\, dx$$

$$\sigma^2 = \text{Var}(X) = E(X^2) - [E(X)]^2$$

$$E(X^2) - \int_{-}^{x} x^2 \cdot f(x)\, dx$$

$$\sigma = \text{Sd}(X) = \sqrt{\text{Var}(X)}$$

(a)	$P(X \ge k)$ = from table
(b)	$P(X < k) = 1 - P(X \ge k)$
(c)	$P(X \le k) = 1 - P(X \ge k-1)$
(d)	$P(X > k) = P(X \ge k-1)$
(e)	$P(X = k) = P(X \ge k) - P(X \ge k+1)$
(f)	$P(k \le X \le l) = P(X \ge k) - P(X \ge l+1)$
(g)	$P(k < X < l) = P(X \ge k+1) - P(X \ge l)$
(h)	$P(k \le X < l) = P(X \ge k) - P(X \ge l)$
(i)	$P(k < X \le l) = P(X \ge k+1) - P(X \ge l+1)$

TERBUKA

APPENDIX A

EQUATIONS

Binomial Distribution	
Formula	$P(X=x)=\frac{n!}{x!\cdot(n-x)!}\cdot p^x\cdot q^{n-x}={}^nC_x\cdot p^x\cdot q^{n-x}$
Mean	$\mu=np$
Variance	$\sigma^2=npq$

Poisson Distribution	
Formula	$P(X=x)=\frac{e^{-\mu}\cdot\mu^x}{x!}$, $x=0,1,2,\ldots,\infty$
Mean	$\mu=\mu$
Variance	$\sigma^2=\mu$

Normal Distribution	
Formula	$P\left(Z=\frac{x-\mu}{\sigma}\right)$

Poisson Approximation to the Binomial Distribution	
Condition	Use if $n\geq 30$ and $p\leq 0.1$
Mean	$\mu=np$

Normal Approximation to the Binomial Distribution	
Condition	Use if n is large and $np\geq 5$ and $nq\geq 5$
Mean	$\mu=np$
Variance	$\sigma^2=npq$

Sampling error of single mean : $e=\left|\bar{x}-\mu\right|$.

Population mean, $\mu=\frac{\sum x}{N}$.

Sample mean, is $\bar{x}=\frac{\sum x}{n}$.

Z-value for sampling distribution of $\bar{x}$ is $Z=\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$.

$\sigma_{\bar{x}}=\sigma/\sqrt{n}$

$\bar{x}\sim N\left(\mu_{\bar{x}},\sigma^2_{\bar{x}}\right)$

$P\left(\bar{x}>r\right)=P\left(Z>\frac{r-\mu_{\bar{x}}}{\sigma_{\bar{x}}}\right)$

$\mu_{\bar{x}_1-\bar{x}_2}=\mu_1-\mu_2$

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

$\bar{x}\sim N\left(\mu_{\bar{x}_1-\bar{x}_2},\sigma^2_{\bar{x}_1-\bar{x}_2}\right)$

$P\left(\bar{x}_1-\bar{x}_2>r\right)=P\left(Z>\frac{r-\mu_{\bar{x}_1-\bar{x}_2}}{\sigma_{\bar{x}_1-\bar{x}_2}}\right)$

TERBUKA

APPENDIX A

EQUATIONS

Confidence Interval for Single Mean

Maximum error : $E = Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$, Sample size : $n = \left(\frac{Z_{\alpha/2}(\sigma)}{E}\right)^2$

(a) $n \geq 30$ **or** σ **known**

(i) σ is known : $\left(\bar{x} - z_{\alpha/2}(\sigma/\sqrt{n}) < \mu < \bar{x} + z_{\alpha/2}(\sigma/\sqrt{n})\right)$

(ii) σ is unknown : $\left(\bar{x} - z_{\alpha/2}(s/\sqrt{n}) < \mu < \bar{x} + z_{\alpha/2}(s/\sqrt{n})\right)$

(b) $n < 30$ **and** σ **unknown**

$\left(\bar{x} - t_{\alpha/2,v}(s/\sqrt{n}) < \mu < \bar{x} + t_{\alpha/2,v}(s/\sqrt{n})\right)$: $v = n - 1$

Confidence Interval for a Difference Between Two Means

(a) ***Z*** **distribution case**

(i) σ is known : $(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\left(\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$

(ii) σ is unknown : $(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\left(\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\right)$

(b) ***t*** **distribution case**

(i) $n_1 = n_2,\ \sigma_1^2 \neq \sigma_2^2$: $(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2,v}\left(\sqrt{\frac{1}{n}(s_1^2 + s_2^2)}\right)$: $v = 2n - 2$

(ii) $n_1 = n_2,\ \sigma_1^2 = \sigma_2^2$: $(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2,v} S_p\left(\sqrt{\frac{2}{n}}\right)$: $v = 2n - 2$

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

(iii) $n_1 \neq n_2,\ \sigma_1^2 = \sigma_2^2$: $(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2,v} S_p\left(\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)$: $v = n_1 + n_2 - 2$

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

(iv) $n_1 \neq n_2,\ \sigma_1^2 \neq \sigma_2^2$: $(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2,v}\left(\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\right)$

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

APPENDIX A

EQUATIONS

Hypothesis for Single Mean

Population Standard Deviation σ (known)

Case A : n ≥ 30 with statistics test : $Z_{Test} = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}}$.

Case B : n < 30 also with statistics test : $Z_{Test} = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}}$.

Population Standard Deviation σ (Unknown)

Case C : n ≥ 30 with statistics test : $Z_{Test} = \dfrac{\bar{x}-\mu}{s/\sqrt{n}}$.

Case D : n < 30 with statistics test : $T_{Test} = \dfrac{\bar{x}-\mu}{s/\sqrt{n}}$.

Hypothesis for Different Two Means

Case	Variances	Samples size	Statistical Test
A	Known	$n_1, n_2 \geq 30$	$Z_{Test} = \dfrac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}$
B	Known	$n_1, n_2 < 30$	$Z_{Test} = \dfrac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}$
C	Unknown	$n_1, n_2 \geq 30$	$Z_{Test} = \dfrac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\dfrac{s_1^2}{n_1}+\dfrac{s_2^2}{n_2}}}$
D	Unknown (Equal)	$n_1, n_2 < 30$	$T_{Test} = \dfrac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{S_p \cdot \sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}$ $v = n_1 + n_2 - 2$
E	Unknown (Not equal)	$n_1 = n_2 < 30$	$T_{Test} = \dfrac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\dfrac{1}{n}(s_1^2+s_2^2)}}$ $v = 2(n-1)$
F	Unknown (Not equal)	$n_1, n_2 < 30$	$T_{Test} = \dfrac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\dfrac{s_1^2}{n_1}+\dfrac{s_2^2}{n_2}}}$ $v = \dfrac{\left(\dfrac{s_1^2}{n_1}+\dfrac{s_2^2}{n_2}\right)^2}{\dfrac{\left(\dfrac{s_1^2}{n_1}\right)^2}{n_1-1}+\dfrac{\left(\dfrac{s_2^2}{n_2}\right)^2}{n_2-1}}$

APPENDIX A

EQUATIONS

Simple Linear Regression Model

(i) Least Squares Method

The model : $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$

$\hat{\beta}_1 = \dfrac{Sxy}{Sxx}$ (slope) and $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$. (y-intercept) where

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

$$Sxx = \sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^{\,2} - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2.$$

$$Syy = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^{\,2} - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2$$

and n = sample size

Inference of Regression Coefficients

(i) Slope

$$SSE = Syy - \hat{\beta}_1 S_{xy} \quad . \quad MSE = \frac{SSE}{n-2} \, . \qquad T_{test} = \frac{\hat{\beta}_1 - \beta_C}{\sqrt{MSE/S_{xx}}}$$

(ii) Intercept

$$T_{test} = \frac{\hat{\beta}_0 - \beta_C}{\sqrt{MSE(1/n + \bar{x}^2/Sxx)}}$$

Coefficient of Determination. r^2.

$$r^2 = \frac{Syy - SSE}{Syy} = 1 - \frac{SSE}{Syy}$$

Coefficient of Pearson Correlation. r.

$$r = \frac{Sxy}{\sqrt{Sxx \cdot Syy}}$$

Confidence Intervals of the Regression Line

(i) Slope. β_1

$$\hat{\beta}_1 - t_{\alpha/2,v}\sqrt{MSE/Sxx} < \beta_1 < \hat{\beta}_1 + t_{\alpha/2,v}\sqrt{MSE/Sxx}\,.$$

where $v = n$-2

(ii) Intercept. β_0

$$\hat{\beta}_0 - t_{\alpha/2,v}\sqrt{MSE\left(\frac{1}{n} + \frac{\bar{x}^2}{Sxx}\right)} < \beta_0 < \hat{\beta}_0 + t_{\alpha/2,v}\sqrt{MSE\left(\frac{1}{n} + \frac{\bar{x}^2}{Sxx}\right)}$$

where $v = n$-2

APPENDIX A

EQUATIONS

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

$$k = \sqrt{n}$$

$$F_o = \frac{MS_F}{MS_E} \qquad P_k = \frac{k-0.5}{N}$$

$$w > \frac{r}{k} = \frac{\max - min}{k}$$

$$l \cong min - \frac{kw-r}{2}, \quad u = l + kw$$

Limit of upper outliers = $q_3 + 1.5(IQR)$

Limit of lower outliers = $q_1 - 1.5(IQR)$

Source of Variation	Sum of Squares	Degree of Freedom	Mean Square	F_0
Factor	SS_F	a-1	MS_F	$\frac{MS_F}{MS_E}$
Error	SS_E	a(n-1)	MS_E	
Total	SS_T	an-1		

$$r_{xy} = \frac{\sum_{i=1}^{n} y_i(x_i - \bar{x})}{\left[\sum_{i=1}^{n}(y_i - \bar{y})^2 \sum_{i=1}^{n}(x_i - \bar{x})^2\right]^{1/2}}$$

$$t = \frac{y_{i\cdot} - y_{j\cdot}}{\sqrt{2MS_E / n}}$$

$$\mu_i \approx \bar{y}_{i.}$$

Control limit: F_{α, v_1, v_2}

$$\tau_i = \mu_i - \mu \approx \bar{y}_{i.} - \bar{y}_{..}$$

$$\varepsilon_{ij} = y_{ij} - \mu_i \approx y_{ij} - \bar{y}_{i.}$$

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}^2 - \frac{y_{..}^2}{an}$$

$$SS_T = N - 1 = an - 1$$

$$SS_F = \sum_{i=1}^{a}\frac{y_{i.}^2}{n} - \frac{y_{..}^2}{an}$$

$$SS_F = a - 1$$

$$SS_E = SS_T - SS_F$$

$$SS_E = R\,(n - 1) = a\,(n - 1)$$