Universiti Tun Hussein Onn Malaysia

UNIVERSITI TUN HUSSEIN ONN MALAYSIA

FINAL EXAMINATION
SEMESTER I
SESSION 2022/2023

COURSE NAME	:	ENGINEERING STATISTICS
COURSE CODE	:	BDA 24103
PROGRAMME	:	BDD
EXAMINATION DATE	:	FEBRUARY 2023
DURATION	:	3 HOURS
INSTRUCTION	:	1. ANSWER **ALL** QUESTIONS FROM SECTION A AND **THREE (3)** QUESTIONS FROM SECTION B. 2. THIS FINAL EXAMINATION IS CONDUCTED VIA **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 **THIRTEEN (13)** PAGES

SECTION A

Instruction: Please answer **ALL questions** in this section.

Q1 Thermal conductivity of a material is the quantity of heat, transmitted through a thickness in a direction normal to a surface of the area. The thermal conductivity is due to a temperature gradient under a steady state condition. As regards, the thickness of 7 materials were measured and the thermal conductivity of each material was recorded as in **Table 1**.

Table 1: Thermal conductivity of a material

Thickness	21	26	28	31	25	19	35
Thermal Conductivity	12	16	19	21	14	11	24

(a) Plot the data on a scatter diagram.

(4 marks)

(b) Estimate the regression line by using the method of least square. Interpret your result.

(8 marks)

(c) Calculate the average of thermal conductivity if the thickness of a material is 29 mm.

(2 marks)

(d) Calculate the coefficient of correlation, r and coefficient of determination, r^2. Interpret their values.

(6 marks)

Q2 (a) The service lives of insulating fluids are being studied. Test data have been obtained from an experiment for three types of fluids as shown in **Table 2**. The test is intended to know whether different types of fluids will affect the service life.

Table 2: Service life

Fluid type	Service Life (in hour)		
A	12	10	11
B	8	9	10
C	16	15	18

(i) State the hypothesis for the above experiment.

(2 marks)

(ii) Is there any indication to show that different types of fluids will affect the service life (use $\alpha = 0.05$)? Display your results in an ANOVA table and conclude.

(6 marks)

(iii) Check the variance, σ^2 of residuals at all levels and conclude whether it is constant or not.

(4 marks)

(b) An experiment to compare the spreading rates of five different brands of yellow interior latex paint available in a particular area used 4 gallons of each paint. The sample average spreading rates (ft^2/gal) for the five brands were tabulated in **Table 3**. The computed value of F was found to be significant at level $\alpha = 0.05$. Given MSE = 272.8 and n = 6.

Table 3: Average spreading rates (ft^2/gal) of paint for different brands

Brand	Average spreading rates (ft^2/gal)
$\bar{x}_1$	5462.0
$\bar{x}_2$	5512.8
$\bar{x}_3$	5437.5
$\bar{x}_4$	5469.3
$\bar{x}_5$	5532.1

(i) Is there any significant difference in the true average spreading rates between brand $\bar{x}_3$ and $\bar{x}_5$? Use LSD procedure to support your answer.

(6 marks)

(ii) Please identify which average spreading rate is greater between $\bar{x}_3$ and $\bar{x}_5$.

(2 marks)

SECTION B

Instruction: Please answer **THREE (3) questions** from FOUR (4) questions provided in this section.

Q3 (a) (i) The field of statistics deals with collecting, presenting, analyzing, and using data to make decisions, solve problems, and design products and processes. State two types of methods to present the data.

(2 marks)

(ii) A group of researcher study how depression affects performance of an activity. The data is taken in a week among many groups of people. State what type of method used for this data collection.

(2 marks)

(b) Based on data from semester 2 2021/2022, the average heat transfer test score was 20.9, with a standard deviation of 4.6. Assuming that the scores are normally distributed.

(i) Define the random variable X, then write it distribution.

(2 marks)

(ii) Find the probability that a randomly selected student has a heat transfer score between 20 and 26.

(4 marks)

(c) A large chain retailer purchases a certain kind of printed circuit board (PCB) from a manufacturer. The manufacturer indicates that the defective rate of the PCB is 3%.

(i) The inspector randomly picks 20 PCBs from a shipment. What is the probability that there will be at least one defective item among these 20?

(3 marks)

(ii) Suppose that the retailer receives 10 shipments in a month and the inspector randomly tests 20 PCBs per shipment. What is the probability that there will be exactly 3 shipments each containing at least one defective PCB among the 20 that are selected and tested from the shipment?

(7 marks)

Q4 (a) A random sample of size $n_1 = 16$ is selected from a normal population with a mean of 75 and a standard deviation of 8. A second random sample of size $n_2 = 9$ is taken from another normal population with mean 70 and standard deviation 12. Let K_1 and K_2 be the two sample means. Find:

(i) The probability that $K_1 - K_2$ exceeds 4.

(4 marks)

(ii) The probability that $K_1 - K_2$ is greater by at least 3.5.

(4 marks)

(b) Twelve aluminium alloys of Company A contained an average magnesium content of 2.8 mg with a standard deviation of 0.2 mg. Ten aluminium alloys of Company B contained an average magnesium content of 2.4 mg with a standard deviation of 0.4 mg. Assumes the two sets of data are independent random samples from normal populations with equal variances.

(i) Construct a 95% confidence interval for the mean magnesium content from Company A.

(4 marks)

(ii) Construct a 98% confident interval for difference means of magnesium content from Company A and Company B.

(8 marks)

Q5 (a) The Test score for English paper has a normal distribution with mean 75 with sample standard deviation of seven. The lecturer claims that if the candidate learned more than 1 hours per day, the mean score will be different than 75. The test was given to a random sample of 65 candidates with the mean score was 82.

(i) Test the claim at 5% of significant level.

(6 marks)

(ii) Identify the type I error and type II error that correspond to the hypothesis above.

(4 marks)

(b) The data is about the average speed for two types of cars may achieved in 10s. The sample size of car type I is 18 with sample mean 120 km/h. While the sample size of car type II is 14 with sample mean 131 km/h. If sample standard deviation for both of cars are 1.7 and 1.9 respectively, test the hypothesis at

0.025 level of significance. The average of car speed type I is lower than the average of car speed type II. Assume that the variances population unknown but equal.

(10 marks)

Q6 (a) Sekolah Kebangsaan ABC scheduled the mid-term examination in June 2022. Therefore, **Table 4** shows the score marks for 30 students in class 5A for the Science subject.

Table 4: Score marks for Science subject

88	58	60	62	55
45	51	64	54	66
65	69	75	82	88
75	87	42	54	64
72	91	67	69	78
40	55	53	71	70

(i) Find the value of mode and mean.

(4 marks)

(ii) Calculate Quartile 1, Median, Quartile 3 and Interquartile range (IQR).

(4 marks)

(iii) Construct the boxplot.

(4 marks)

(b) Mr. Ahmad conducted the tensile testing on the metal rods at the Mechanical Laboratory. The tensile strength results in MPa of the metals rods is tabulated in **Table 5**.

Table 5: Tensile strength (MPa) of metal rods

33	75	67	109
67	117	77	67
90	50	36	45
120	78	112	74
56	49	53	99
30	102	100	52
108	34	48	81
45	89	96	52
77	77	82	110
88	57	118	35

(i) Calculate the number of class interval, k and width of the class interval, w.

(2 marks)

(ii) Calculate the lower, l and upper limit, u of the histogram.

(2 marks)

(iii) Construct the histogram (frequency versus classes).

(4 marks)

- END OF QUESTIONS -

FINAL EXAMINATION

SEMESTER / SESSION : SEMESTER I /2022/2023
COURSE NAME : ENGINEERING STATISTICS

PROGRAMME CODE : BDD
COURSE CODE: BDA 24103

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 :

- $E(aX + b) = a\,E(x) + b.$
- $Var(aX + b) = a^2\,Var(x)$

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

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

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

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

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

(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)	$P(k < X \le l) = P(X \ge k+1) - P(X \ge l+1)$

FINAL EXAMINATION

SEMESTER / SESSION : SEMESTER I /2022/2023
COURSE NAME : ENGINEERING STATISTICS

PROGRAMME CODE : BDD
COURSE CODE: BDA 24103

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 $Z=\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$.

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

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

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

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

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

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

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

FINAL EXAMINATION

SEMESTER / SESSION : SEMESTER I /2022/2023
COURSE NAME : ENGINEERING STATISTICS

PROGRAMME CODE : BDD
COURSE CODE: BDA 24103

Confidence Interval for Single Mean

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

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

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

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

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

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

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

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

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

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

FINAL EXAMINATION

SEMESTER / SESSION : SEMESTER I /2022/2023
COURSE NAME : ENGINEERING STATISTICS

PROGRAMME CODE : BDD
COURSE CODE: BDA 24103

Confidence Interval for Single Population Variance

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2,v}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,v}} \quad ; v = n-1$$

Confidence Interval for Ratio of Two Population Variances

$$\frac{s_1^2}{s_2^2}\frac{1}{f_{\alpha/2,v1,v2}} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{s_1^2}{s_2^2} f_{\alpha/2,v2,v1} \quad ; v_1 = n_1 - 1 \text{ and } v_2 = n_2 - 1$$

Case	Variances	Samples size	Statistical Test
A	Known	$n_1, n_2 \ge 30$	$Z_{Test} = \dfrac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$
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	$n_1, n_2 \ge 30$	$Z_{Test} = \dfrac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$
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)	$n_1, n_2 < 30$	$T_{Test} = \dfrac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$ $v = \dfrac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^2}{\dfrac{\left(\dfrac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \dfrac{\left(\dfrac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$

FINAL EXAMINATION

SEMESTER / SESSION : SEMESTER I /2022/2023
COURSE NAME : ENGINEERING STATISTICS

PROGRAMME CODE : BDD
COURSE CODE: BDA 24103

Simple Linear Regression Model

(i) Least Squares Method

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

$\hat{\beta}_1 = \frac{Sxy}{Sxx}$ (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

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

(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

FINAL EXAMINATION

SEMESTER / SESSION : SEMESTER I /2022/2023
COURSE NAME : ENGINEERING STATISTICS

PROGRAMME CODE : BDD
COURSE CODE: BDA 24103

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

$$k = \sqrt{n}$$

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

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

Limit of upper outliers = $q_3 + 1.5(IQR)$
Limit of lower outliers = $q_1 - 1.5(IQR)$

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

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

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

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

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

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

$$SS_E = SS_T - SS_F$$

$SS_T = N - 1 = an - 1$

$SS_F = a - 1$

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

$$MS_E = \frac{SS_E}{a(n-1)}$$

$$MS_F = \frac{SS_F}{a-1}$$

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

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		

$$t = \frac{\bar{y}_{i.} - \bar{y}_{j.}}{\sqrt{2MS_E / n}}$$

Control limit: F_{α, v_1, v_2}