Universiti Tun Hussein Onn Malaysia

UNIVERSITI TUN HUSSEIN ONN MALAYSIA

FINAL EXAMINATION
SEMESTER II
SESSION 2022/2023

COURSE NAME	:	ENGINEERING STATISTICS
COURSE CODE	:	BDA 24103
PROGRAMME	:	BDD
EXAMINATION DATE	:	JULY/AUGUST 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 **TWELVE (12)** PAGES

SECTION A

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

Q1 The modulus Young (GPa) of polymer composite panels were depend on the weight percentage (wt.%) of polycarbonate material. The results of these two parameters can be seen in **Table Q1**.

Table Q1: Modulus Young (GPa) of Polymer Composites Panels

Polycarbonate (wt.%)	100	97.5	95	90	80	50	30
Modulus Young (GPa)	2.12	2.26	2.57	3.26	3.46	4.54	8.5

(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 modulus Young if the polycarbonate:

(i) 25 wt.%

(ii) 10 wt.%

(4 marks)

(d) Calculate the coefficient of determination, r^2

(4 marks)

Q2 (a) Compare the one factor designs and factorial designs

(4 marks)

(b) A nutritionist randomly divided 15 cyclists into three groups of five each. The first group was given vitamin supplements to consume with each meal for the next three weeks. The second group was instructed to consume a specific variety of high-fiber, whole-grain cereal for the following three weeks. The third group was instructed to consume their meals as normally they do. After three weeks, the nutritionist instructed each cyclist to ride 6 miles. The following times were recorded in **Table Q2(b)**.

Table Q2(b): Time recorded for riding 6 miles

Vitamin group	15.6	16.4	17.2	15.5	16.3
Fiber cereal group	17.1	16.3	15.8	16.4	16.0
Control group	15.9	17.2	16.4	15.4	16.8

(i) Write the hypothesis that neither the vitamin nor the fiber cereal affects the speed of a bicyclist.

(2 marks)

(ii) By using 5% level of significance, test the hypothesis and conclude your findings.

(8 marks)

(c) Three standard chemical procedures are used to determine the copper content in a certain chemical compound. Each procedure was used 4 times on a given compound and the resulting data is shown in **Table Q2 (c)(i)**. Complete the ANOVA in **Table Q2(c)(ii)** and test the hypothesis that the mean readings are the same for all three methods. Use the 5 percent level of significance.

(6 marks)

Table Q2 (c)(i): Copper content

Method 1	76.43	78.61	80.4	78.22
Method 2	80.4	82.24	72.7	76.04
Method 3	82.16	84.14	80.2	81.33

Table Q2 (c)(ii): ANOVA of copper content measurements

Source of variation	df	Sum of squares (SS)	Mean of squares (MS)	F_0
Method			19.858	2.49
Error	9			
Total	11	111.53		

SECTION B

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

Q3 (a) A resistance welding process joins steel rods to form a grid. Resistance welding passes high levels of electrical current through an electrode and the pieces of metal to be joined. The heat resulting from the high current causes the two pieces to fuse together. However, the weld strength varies considerably during the process. Operators have control mainly on current, electrode force, part alignment and weld time. The other factors that are believed to contribute the weld strength variability are cleanliness of the parts and condition of the electrode. The rods are cleaned prior to welding. When welding problems occur, the cleaning process is often blamed. Engineers change electrodes every 50,000 cycles, but they want to know if this duration is excessive.

Table Q3: Influential factors in welding process

Factors	**Symbol**	**Description**
Current	A	Welding current (1.4, 1.7 A)
Force	F	Force applied to parts (550, 750 N)
Time	t	Weld time (0.10, 0.15 seconds)
Alignment	L	Part alignment (0.25, 0.50 mm)
Electrode	E	Condition of electrode (New, Used)
Clean	C	Cleanliness of parts (Basic, More)

(i) Identify the qualitative factors in this welding process.

(2 marks)

(ii) By excluding the qualitative factors, write the general empirical model of this process.

(4 marks)

(b) The grandchildren of Grandma have been complaining that she doesn't give them enough chocolate chips. Grandma agrees to add enough chips to the dough so that only 1% of the cookies won't have any. How many chips does she need to put in 100 cookies in order to get this result?

(6 marks)

(c) There are 30 restaurants in a particular city. Assume that four of them violate the health code. A health inspector visits ten restaurants chosen at random.

(i) What is the probability that two of the restaurants with health code violations will be visited?

(4 marks)

(ii) What is the probability that none of the restaurants that are visited will have health code violations?

(4 marks)

Q4 (a) What is the probability that a sample 75 books in library will have average 250 pages or less, if the average of the population is 265 with standard deviation of 55?

(4 marks)

(b) Two lorries bring the watermelon from the farm to the market. Lorry ABC and lorry XYZ can bring 315 and 289 watermelons respectively. If the mean and standard deviation of lorry ABC and lorry XYZ were 35 and 15 and 45 and 12 respectively, find the probability of mean by lorry XYZ is lower than lorry ABC.

(6 marks)

(c) Two different methods were used for producing the aluminium alloy bar which is powder metallurgy and extrusion. Twenty and twelve samples were taken from each method for inclusion detection. The mean and standard deviation for powder metallurgy and extrusion methods are 20 and 7, and 17 and 5 respectively with same population variance. Construct 95% confidence intervals between the means of powder metallurgy and extrusion.

(10 marks)

Q5 (a) Find the critical value for each situation using standard normal distribution Z and the appropriate figure to show the reaction region

(i) A right-tailed test with $\alpha = 0.1$

(2 marks)

(ii) A left-tailed test with $\alpha = 0.05$

(2 marks)

(iii) Two-tailed test with $\alpha = 0.01$

(2 marks)

(b) A group of UTHM researcher wants to study the performance of two different types of motorcycle engine which is engine Type A and engine Type B. They records the data of the average mileage for engine type A and engine type B. The sample mean for engine type A is 114, and the standard deviation is 1.6. While for engine type B, the sample mean is 123, and the sample standard deviation is 1.7. If the sample size for both engines is 18 and 14, respectively, test the hypothesis using $\alpha = 0.025$ and assume the two-population variances are unknown but equal.

(14 marks)

Q6 (a) A survey on spending money for household groceries per month by residents at Town XX was recorded in **Table Q6(a).**

Table Q6(a): Household Groceries (RM)

300	260	560	320	500	430	560
350	325	280	525	320	420	300
700	310	530	490	255	420	350
580	310	680	400	360	640	310
620	500	400	630	510	550	590
700	450	330	560	420	720	350
270	750	280	450	480	340	420

(i) Identify the lower and upper limit

(2 marks)

(ii) Construct the histogram relative frequency versus classes

(8 marks)

(b) **Table Q6(b)** shows the height of the seven years old boys in class Y1

Table Q6(b): Height of seven years old boys (mm)

122	105	112	108
111	98	100	120
110	106	125	109
105	112	104	125

(i) Identify first quartile, IQ1 and third quartile, IQ3

(2 marks)

(ii) Construct the box plot. Is that any outlier values?

(8 marks)

- END OF QUESTIONS -

FINAL EXAMINATION

SEMESTER / SESSION : SEMESTER II /2022/2023
COURSE NAME : ENGINEERING STATISTICS
PROGRAMME CODE : BDD
COURSE CODE: BDA 24103

EQUATIONS

- $P(X \le r) = F(r)$
- $P(X > r) = 1 - F(r)$
- $P(X < r) = P(X \le r-1) = F(r-1)$
- $P(X = r) = F(r) - F(r-1)$
- $P(r < X \le s) = F(s) - F(r)$
- $P(r \le X \le s) = F(s) - F(r) + f(r)$
- $P(r \le X < s) = F(s) - F(r) + f(r) - f(s)$
- $P(r < X < s) = F(s) - F(r) - f(s)$

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

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

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

Note :

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

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

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

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

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

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

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

(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 II /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 II /2022/2023 PROGRAMME CODE : BDD
COURSE NAME : ENGINEERING STATISTICS 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) $n_1 = n_2$, $\sigma_1^2 \neq \sigma_2^2$: $(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2,v}\left(\sqrt{\dfrac{1}{n}(s_1^2 + s_2^2)}\right)$; $v = 2n - 2$

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

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

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

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

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

FINAL EXAMINATION

SEMESTER / SESSION : SEMESTER II /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

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

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

FINAL EXAMINATION

SEMESTER / SESSION : SEMESTER II /2022/2023 PROGRAMME CODE : BDD
COURSE NAME : ENGINEERING STATISTICS COURSE CODE: BDA 24103

Simple Linear Regression Model

(i) Least Squares Method

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

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

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

(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 II /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 = $q_1 - 1.5(IQR)$

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

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

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

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

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

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

$$SS_E = SS_T - SS_F$$

$SS_T = N - 1 = an - 1$

$SS_F = a - 1$

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

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

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

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}

TERBUKA