UTHM
Universiti Tun Hussein Onn Malaysia

UNIVERSITI TUN HUSSEIN ONN MALAYSIA

FINAL EXAMINATION
SEMESTER I
SESSION 2020/2021

COURSE NAME	:	ENGINEERING STATISTICS
COURSE CODE	:	BDA 24103
PROGRAMME	:	BDD
EXAMINATION DATE	:	JANUARY/FEBRUARY 2021
DURATION	:	3 HOURS
INSTRUCTION	:	**SECTION A:** ANSWER **ALL** QUESTIONS. **SECTION B:** ANSWER **THREE (3)** FROM FOUR (4) QUESTIONS.

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

SECTION A

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

Q1 (a) The article "Testing the Influence of Climate, Human Impact and Fire on the Holocene Population Expansion of Fagus sylvatica in the Southern Prealps (Italy)" (V. Valsecchi, W. Finsinger, et al., The Holocene 2008:603 614) presents calculations of the ages (in calendar years before 1950) of several sediment samples taken at various depths (in cm) in Lago di Fimon, a lake in Italy. The results are presented in the Table Q1.

Table Q1: Results of ages at various depths

Depth (cm)	Age
284.5	1255
407.5	3390
512.0	5560
551.0	6670
578.5	7160
697.0	9820
746.5	11 030

(i) Construct a scatterplot of age *(y)* versus depth *(x)*. Verify that a linear model is appropriate.

(5 marks)

(ii) Compute the least-squares line for predicting age from depth of the lake.

(10 marks)

(iii) If two samples differ by 100 cm in depth, by how much would you predict their ages to differ?

(2 marks)

(iv) Predict the age for a specimen which depth is 600 cm.

(3 marks)

Q2 (a) A series of experiments were conducted to determine the effect of air voids on percentage retained strength of asphalt. For purposes of the experiment, air voids are controlled at three levels; low (1–3%), medium (4–6%), and high (7–9%). The data are shown in **Table Q2(a)**. Do the different levels of air voids significantly affect mean retained strength? Use $\alpha = 0.05$.

(6 marks)

Table Q2(a): Results of the air voids effect on percentage retained strength of asphalt

Air Voids	**Retained Strength (%)**							
Low	106	90	103	90	79	88	92	95
Medium	80	69	94	91	70	83	87	83
High	78	80	62	69	76	85	69	85

(b) An electronics engineer is interested in the effect on tube conductivity of five different types of coating for cathode ray tubes in a telecommunications system display device. The following conductivity data are illustrated in **Table Q2(b)**.

Table Q2(b): Data of conductivity

Coating Type	**Conductivity**			
	1	2	3	4
1	143	141	150	146
2	152	149	137	143
3	134	133	132	127
4	129	127	132	129
5	147	148	144	142

(i) Is there any difference in conductivity due to coating type? Use $\alpha = 0.01$.

(6 marks)

(ii) Analyze the residuals(estimated error) from this experiment. Conclude your findings.

(8 marks)

SECTION B

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

Q3 (a) Clearly state the differences between enumerative study and analitic study.

(4 marks)

(b) Explain clearly the basic methods for collecting data and provide an example for each method

(6 marks)

(c) Statistics is a science that helps us make decisions and draw conclusions in the presence of variability. Describe the application of statistical methodology in Engineering, Economic, Business, Quality Control and Health and Medicine

(10 marks)

Q4 (a) A homeowner has just installed 20 light bulbs in a new home. Suppose that each has a probability 0.2 of functioning more than three months.

(i) What is the probability that at least five of these function more than three months?

(3 marks)

(ii) What is the average number of bulbs the homeowner has to replace in three months?

(2 marks)

(b) During business hours, the number of calls passing through a particular cellular relay system averages five per minute. Find the probability that:

(i) No call will pass through the relay system during a 2-minute period.

(3 marks)

(ii) Three calls will pass through the relay system during a 2-minute period.

(3 marks)

(iii) At least four calls in the given minute.

(4 marks)

(c) The population of the usage per sheet of plastic for old and new certain products are distributed $N_1(2000, 60)$ and $N_2(2500, 40)$ respectively. Two random samples are taken from each population of size n_1 and n_2. Find the probability of mean usage of size $n_1 = 20$ and $n_2 = 25$ for new and old products with new products is at most 502 sheets more than old products.

(5 marks)

Q5 (a) 100 random samples of lubricant oil from a diesel engine lubricating were taken and the viscosity (centistokes (cSt)) measured. A 95% confidence interval on the mean of viscosity is $0.49 \leq \mu \leq 0.82$.

(i) Would a 99% confidence interval calculated from the same sample data been longer or shorter?

(1 marks)

(ii) Consider the following statement: There is a 95% chance that μ is between 0.49 and 0.82. Is this statement correct? Explain your answer

(2 marks)

(iii) Consider the following statement: If n = 100 random samples of lubricant oil from a diesel engine lubricating were taken and the 95% confidence interval on μ computed, and this process was repeated 1000 times, 950 of the confident interval s will contain the true value of μ. Is this statement correct? Explain your answer.

(2 marks)

(b) A study was conducted to estimate the average life of a large shipment of arc lamp. Based on the previous studies, it is suggested that the standard deviation is known to be 100 hours. A random sample of 50 light bulbs was selected and shown that the sample average life was 350 hours

(i) Construct a 95% confident interval estimate of the true average life for light bulbs in this shipment

(4 marks)

(ii) Determine how many samples of arc lamp must be selected if they want to be 99% confident that the error is less than 0.03.

(3 marks)

(c) A researcher hypothesizes that the average hours of men involves in sports and exercise that is greater than the average hours of women involves in sports and exercise. Number of hours spend on sports and exercise by men and women has been recorded in **Table Q5(a) and Table Q5(b)** . At a α= 0.10, is there enough evidence to support the claim? Assume standard deviation of hours spend by men is equal to standard deviation of hours spend by women.

(8 marks)

Table Q5(a): Hours spend by men

6	11	11	8	15
6	14	8	12	18
6	9	5	6	9
6	9	18	7	6
15	6	11	5	5
9	9	5	5	8
8	9	6	11	6
9	5	11	5	8
7	7	5	10	7
10	7	10	8	11

Table Q5(b): Hours spend by women

6	8	11	13	8
7	5	13	14	6
6	5	5	7	6
10	7	6	5	5
16	10	7	8	5
7	5	5	6	5
9	18	13	7	10
7	8	5	7	6
11	4	6	8	7
14	12	5	8	5

Q6 (a) The nine measurements that follow are furnace temperatures recorded on successive batches in a semiconductor manufacturing process as shown in **Table Q6(a)** (units are °F).

Table Q6(a)

Data
953, 950, 948, 955, 951, 949, 957, 954, 955

(i) Calculate the sample mean, sample variance, and standard deviation

(5 marks)

(ii) Find the median. How much could the highest temperature measurement increase without changing the median value?

(2 marks)

(iii) Construct a box plot of the data.

(3 marks)

(b) The data represent weight of the athlete that involve in different type of sports at School Sport at Bukit Jalil. Based on this data:

Table Q6(b): Weight of athletes(kg)

57	61	57	57	58	57	61	54	68
51	49	64	50	48	65	52	56	46
54	49	51	47	55	55	54	42	51
56	55	51	54	51	60	62	43	55
56	61	52	69	64	46	54	47	

(i) Were the data obtained from a population or a sample? Explain your answer.

(2 marks)

(ii) What was the lowest weight and highest weight of the athlete

(1 marks)

(iii) Based on the data, construct a histogram using 7 classes and brief summary of the nature of the data.

(6 marks)

(iv) Identify is there any possible outliers.

(1 marks)

END OF QUESTIONS

TERBUKA

FINAL EXAMINATION

SEMESTER / SESSION : SEMESTER I /2020/2021
COURSE : ENGINEERING STATISTICS

PROGRAMME : 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 :

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

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

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

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

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

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

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

(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 /2020/2021
COURSE : ENGINEERING STATISTICS

PROGRAMME : BDD
COURSE CODE: BDA 24103

EQUATIONS

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

Poisson Distribution	
Formula	$P(X=x)=\frac{e^{-\mu}\mu^x}{x!}$, $x=0,1,2,\dots,\infty$
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 $Z=\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$

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

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

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

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

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

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

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

FINAL EXAMINATION

SEMESTER / SESSION : SEMESTER I /2020/2021 PROGRAMME : BDD
COURSE : ENGINEERING STATISTICS COURSE CODE: BDA 24103

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

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

FINAL EXAMINATION

SEMESTER / SESSION : SEMESTER I /2020/2021
COURSE : ENGINEERING STATISTICS

PROGRAMME : 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

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

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

FINAL EXAMINATION

SEMESTER / SESSION : SEMESTER I /2020/2021
COURSE : ENGINEERING STATISTICS

PROGRAMME : BDD
COURSE CODE: BDA 24103

Simple Linear Regression Model

(i) Least Squares Method

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

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