CONFIDENTIAL

UTHM
Universiti Tun Hussein Onn Malaysia

UNIVERSITI TUN HUSSEIN ONN MALAYSIA

FINAL EXAMINATION (ONLINE) SEMESTER II SESSION 2020/2021

COURSE NAME	:	ENGINEERING STATISTICS
COURSE CODE	:	BDA 24103
PROGRAMME	:	BDD
EXAMINATION DATE	:	JULY 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

CONFIDENTIAL

SECTION A

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

Q1 Foot ulcers are common problem for people with diabetes. Higher skin temperatures on the foot indicate an increased risk of ulcers. The article "An Intelligent Insole for Diabetic Patients with the Loss of Protective Sensation" (Kimberly Anderson, M.S. Thesis, Colorado School of Mines), reports measurements of temperatures, in $°F$, of both feet for 18 diabetic patients. The results are presented in the **Table Q1**.

Table Q1: Measurements of temperatures, in $°F$ of left foot Vs right foot for 18 diabetic patients

Left Foot	Right Foot	Left Foot	Right Foot
80	80	76	81
85	85	89	86
75	80	87	82
88	86	78	78
89	87	80	81
87	82	87	82
78	78	86	85
88	89	76	80
89	90	88	89

(a) Compute the least-squares line for predicting the right foot temperature from the left foot temperature.

(9 marks)

(b) If the left foot temperatures of two patients differ by 2 degrees, predict by how much would their right foot temperatures will differ.

(1 marks)

(c) Predict the right foot temperature for a patient whose left foot temperature is 81 degrees.

(1 marks)

(d) Test the slope, $\beta_1 = 1$ at 5% level of significance.

(6 marks)

(e) Calculate the coefficient of correlation r and r^2 and then interpret their values

(3 marks)

Q2 (a) The data as shown in **Table Q2(a)** are the modulus of elasticity (x 10^6 psi) for lumber of three different grades. Given the MS_E is 0.183.

Table Q2(a): Modulus of elasticity for different lumber grades

Grade	n	$\bar{y}_{i.}$	$s_{i.}$
1	10	1.63	0.27
2	10	1.56	0.24
3	10	1.42	0.26

(i) Using a significance level of 0.05, test the null hypothesis of no difference in mean modulus of elasticity for the three grades.

(6 marks)

(ii) Can Least Significant Difference be applied to test the differences between the individual level means? Explain your answers.

(4 marks)

(b) A study of the properties of metal plate-connected trusses used for roof support yielded the axial-stiffness index (kN/m) for plate lengths 0.10, 0.15, 0.20, 0.25, and 0.30 m as shown in **Table Q2(b).** Does variation in plate length have any effect on true average axial stiffness? State and test the relevant hypotheses using analysis of variance (ANOVA) with $\alpha = 0.01$. Display your results in an ANOVA table.

Table Q2(b): Data of axial-stiffness index (kN/m) for metal plate-connected trusses

Plate	Axial-stiffness index (kN/m)						
length (m)	1	2	3	4	5	6	7
0.10	34.93	46.27	35.14	36.89	35.79	39.52	34.99
0.15	45.43	39.23	40.79	45.70	37.40	39.42	43.13
0.20	44.34	41.38	39.66	40.35	46.31	41.50	43.16
0.25	39.17	51.17	52.13	48.93	46.39	43.41	40.97
0.30	46.03	49.92	47.44	46.40	53.49	49.85	52.63

(10 marks)

SECTION B

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

Q3 An automobile engineer claim that his team has successfully designed a new engine that saves more car fuel than the previous designs. He wishes to prove that by conducting on road experiment to compare both designs. A sample of 40 cars of previous and new engines version were involved in the experiment. The data of fuel consumption (liter/100km) were recorded in **Table Q3**.

Table Q3 Data of fuel consumption (liter/100km)

	Mean	Variance
Previous version	5.9	0.017
New version	5.4	0.022

(a) State the type of data collection involved in the given case study. Justify your answer.

(3 marks)

(b) Construct 90% confidence interval for the different mean between previous and new engine version.

(7 marks)

(c) Suppose the sampling for both engine versions were reduced to 15. Test the engineer claim at 0.05 significance level.

(10 marks)

Q4 (a) The number of accidents per day has the poisson distribution where the value of mean is 0.8. Determine:

(i) The number of accidents within eight days that are selected at random.

(1 marks)

(ii) The probability that at least three accidents happened within eight days that are selected at random.

(1 marks)

(iii) The probability that less thank four accidents happen within twelve days that are selected at random.

(2 marks)

(b) The weight, in grams, of red beans in a can is normally distributed with mean μ and standard deviation 7.6. Given that 5% of cans contains less than 200g, determine;

(i) The value of μ.

(4 marks)

(ii) The percentage of cans that contains more than 225 g of beans.

(3 marks)

(c) A factory produces two types of artificial lamp A and B. The mean durability of artificial lamp A is 850 hours and its standard deviation is 40 hours. While, mean durability of artificial lamp B is 820 hours and its standard deviation is 60 hours. Find the probbality of 35 artificial lamp A will have mean durabilirty at least 25 hours more than 45 artificial lamp B.

(9 marks)

Q5 (a) A manufacturing company claims that the mean lifetime of bulbs produced was at least 160 hours. In a random sample of 100 bulbs produced, the sample mean lifetime of bulbs was 150 hours and the sample standard deviation was 25 hours.

(i) Construct a 95% confidence interval for the mean lifetime of bulbs produced.

(4 marks)

(ii) Construct a 99% confidence interval for the mean lifetime of bulbs produced.

(4 marks)

(iii) Calculate the width of confidence interval in (a) and (b). Interpret the result.

(2 marks)

(b) A research has been conducted to study the amount of tar in filtered cigarrate and nonfiltered cigarated. The researcher claimed that the mean of tar in filtered cigarattes is less than the mean amount in nonfiltered cigarattes. By refer to the data listed in Table Q5, use a 0.05 level of significance to test the reseachers's claim. Assume σ unknown but equal, and the two populations are normally distributed.

Table Q5 Tar contents (in mg) of cigarettes

Filtered cigaretted	16 15 16 14 16 1 16 18 10 14 12 11 14 13 13 13 16 16 8 16 11
Nonfiltered cigaretted	23 23 24 26 25 26 21 24

(10 marks)

Q6 (a) **Table Q6(a)** presents male athlete running records for the 100 m events of athletes between the Year 1 athletes and Year 2 athletes.

Table Q6(a) Data of athlete running records

Year 1	43.5, 36.8, 46.2, 50.7, 52.7, 56.0, 49.1, 54.3, 54.4, 55.8, 58.5, 45.8
Year 2	40.8, 39.4, 33.2, 36.2, 37.6, 37.0, 40.6, 41.3, 42.1, 44.5, 50.3, 35.4

(i) Construct a double box plot of this data.

(5 marks)

(ii) Compare the performance of athlete between Year 1 and Year 2.

(2 marks)

(b) The following data in **Table Q6(b)** are the numbers of cycles to failure of aluminum test coupons subjected to repeated alternating stress at 10,000 psi, 10 cycles per second.

Table Q6(b): Number of cycles to failure of aluminum test

373	254	237	243	308	210	266	253	201	266
239	114	224	373	286	329	236	284	247	273
198	361	416	207	243	326	251	160	360	311
215	189	344	268	363	21	270	165	240	48
150	300	207	314	197	209	210	260	327	406

(i) Find the value of mean, standard deviation and variance.

(5 marks)

(ii) Based on the data in Table Q6(b), construct a histogram using 12 classes.

(7 marks)

(iii) Decsribe the shape of the histogram.

(1 marks)

- END OF QUESTIONS -

FINAL EXAMINATION

SEMESTER / SESSION : SEMESTER II /2020/2021 PROGRAMME : BDD
COURSE : ENGINEERING STATISTICS 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 :

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

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

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

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

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

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

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

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

FINAL EXAMINATION

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

PROGRAMME : 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	$P(X=x)=\frac{e^{-\mu}\cdot\mu^x}{x!}$, $x=0,1,2,....\infty$
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 : $e=\left|\bar{x}-\mu\right|$.

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

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

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

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

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

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

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

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

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

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

FINAL EXAMINATION

SEMESTER / SESSION : SEMESTER II /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 II /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

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

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 II /2020/2021 PROGRAMME : BDD
COURSE : 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

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

(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 $v = n-2$