UNIVERSITI TUN HUSSEIN ONN MALAYSIA

FINAL EXAMINATION
SEMESTER I
SESI 2019/2020

COURSE NAME	:	ENGINEERING STATISTICS
COURSE CODE	:	BDA 24103
PROGRAMME	:	BDD
EXAMINATION DATE	:	DECEMBER 2019 / JANUARY 2020
DURATION	:	3 HOURS
INSTRUCTION	:	**SECTION A:** ANSWER **ALL** QUESTIONS. **SECTION B:** ANSWER **THREE (3)** QUESTIONS **ONLY.**

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

SECTION A

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

Q1 (a) In a study conducted in the Science Department of Faculty of Science, Technology and Human Development in a University; the researcher examined the influence of the drug *succinylcholine* on the circulation levels of androgens in the blood. Blood samples from wild, free-ranging deer were obtained via the jugular vein immediately after an intramuscular injection of *succinylcholine* using darts and a capture gun. Deer were bled again approximately 30 minutes after the injection and then released. The level of androgens at time of capture and 30 minutes later, measured in nanograms per milliliter (ng/ml), for 15 deers as in Table Q1.

Assuming that the populations of androgen at time of injection and 30 minutes later are normally distributed:

Table Q1

Deer	Time of injection	Androgen (ng/ml) 30 minutes after injection	Differences of androgen level d_i
1	2.76	7.02	4.26
2	5.18	3.10	-2.08
3	2.68	5.44	2.76
4	3.05	3.99	0.94
5	4.10	5.21	1.11
6	7.05	10.26	3.21
7	6.60	13.91	7.31
8	4.79	18.53	13.74
9	7.39	7.91	0.52
10	7.30	4.85	-2.45
11	11.78	11.10	-0.68
12	3.90	3.74	-0.16
13	26.00	94.03	68.03
14	67.48	94.03	26.55
15	17.04	41.70	24.66

(i) State the null and alternative hypotheses

(2 marks)

(ii) Find the average and standard deviation of this study

(4 marks)

(iii) Determine the critical region of this problem.

(4 marks)

(iv) Test at the 0.05 level of significance whether the androgen concentrations are altered after 30 minutes of restraint.

Given $T = \frac{\bar{d}-d_0}{s_d/\sqrt{n}}$

(6 marks)

(b) Indicate whether the following statements are Type I or Type II error:

(i) In a population, there is no relationship between alcohol consumption and hours of sleep. In a sample of 5000 individuals, we accept the null hypothesis that there is no relationship between alcohol and hours of sleep.

(2 marks)

(ii) In a population, there is a relationship between the number of times a person vacuum a carpet and mental disorders. In a sample of 700 households, we accept the null hypothesis that there is no relationship between carpet cleaning and mental disorders.

(2 marks)

Q2 An experiment in fluidized bed drying system concludes that the grams of solids removed from a material A (y) is thought to be related to the drying time (x). Ten observations obtained from an experimental study are given in the following **Table Q2**.

Table Q2 : Weight of solids removed

Drying time (min)	Weight (g)
2.5	4.3
3.0	1.5
3.5	1.8
4.0	4.9
4.5	4.2
5.0	4.8
5.5	5.8
6.0	6.2
6.5	7.0
7.0	7.9

(a) Illustrate a scatter diagram for these data

(4 marks)

(b) Find the estimated regression line by using the least square method. Interpret the result.

(8 marks)

(c) Test the hypothesis concerning $H_0 : \beta_1 = 3$ against $H_1 : \beta_1 \neq 3$ at the 0.05 level of significance.

(8 marks)

SECTION B

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

Q3 (a) The discrete random variable X has probability distribution given by the following **Table Q3**.

Table Q3 : Probability distribution of variable X

x	-1	0	1	2	3
$P(X = x)$	1/5	a	1/10	a	1/5

Where a is constant

(i) Find the value of a

(3 marks)

(ii) Write down E (X)

(2 marks)

(iii) Solve Var (X)

(5 marks)

(b) Now from **Table Q3** above, if the random variable Y= 6-2X,

(i) Solve Var (Y)

(4 marks)

(ii) Calculate P ($X \geq Y$)

(6 marks)

Q4 (a) An industrial firm supplies 10 manufacturing plants with a certain chemical. The probability that any one firm will call in an order on a given day is 0.2, and this probability is the same for all 10 plants. Demonstrate the probability that, on the given day, the number of plants calling in orders is as follows:

(i) At most 3

(2 marks)

(ii) At least 3

(2 marks)

(iii) Exactly 3

(2 marks)

(b) The manager of an industrial plant is planning to buy a new machine of either type A or type B. For each day's operation, the number of repairs X that machine A requires is a Poisson random variable with mean $0.10t$, where t denotes the time (in hours) of daily operation. The number of daily repairs Y for machine B is a Poisson random variable with mean $0.12t$. The daily cost of operating A is $C_A(t) = 20t + 40X^2$; for B, the cost is $C_B(t) = 16t + 40Y^2$. Assume that the repairs take negligible time and that each night the machines are to be cleaned so that they operate like new machines at the start of each day. Distinguish which machine minimizes the expected daily cost for the following times of daily operation?

(i) 10 hours

(5 marks)

(ii) 20 hours

(5 marks)

(c) A firm that manufactures and bottles apple juice has a machine that automatically fills bottles with 16 ounces of juice. (The bottle can hold up to 17 ounces.) Over a long period, the average amount dispensed into the bottle has been 16 ounces. However, there is variability in how much juice is put in each bottle; the distribution of these amounts has a standard deviation of 1 ounce. If the ounces of fill per bottle can be assumed to be normally distributed, evaluate the probability that the machine will overflow any one bottle.

(4 marks)

Q5 (a) The standard deviation of measurements of a linear dimension of a mechanical part is 0.14 mm. Solve the sample size required if the standard error of the mean must be not more than

(i) 0.04 mm

(3 marks)

(ii) 0.02 mm

(3 marks)

(b) An assembly plant has a bin full of steel rods, for which the diameters follow a normal distribution with a mean of 7.00 mm and a variance of 0.100 mm^2, and a bin full of sleeve bearings, for which the diameters follow a normal distribution with a mean of 7.50 mm and a variance of 0.100 mm^2. Demonstrate percentage of randomly selected rods and bearings will not fit together.

(6 marks)

(c) The effective life of a component used in a jet-turbine aircraft engine is a random variable with mean 5000 hours and standard deviation 40 hours. The distribution of effective life is fairly close to a normal distribution. The engine manufacturer introduces an improvement into the manufacturing process for this component that increases the mean life to 5050 hours and decreases the standard deviation to 30 hours. Suppose that a random sample of $n_1 = 16$ components is selected from the "old" process and a random sample of $n_2 = 25$ components is selected from the "improved" process. Differentiate the probability that the difference in the two samples means $\overline{X}_2 - \overline{X}_1$ is at least 25 hours. Assume that the old and improved processes can be regarded as independent populations.

(8 marks)

Q6 (a) A sample of 125 pieces of yarn had mean breaking strength 6.1 N and standard deviation 0.7 N. A new batch of yarn was made, using new raw materials from a different vendor. In a sample of 75 pieces of yarn from the new batch, the mean breaking strength was 5.8 N and the standard deviation was 1.0 N. Demonstrate a 90% confidence interval for the difference in mean breaking strength between the two types of yarn.

(6 marks)

(b) The temperature of a certain solution is estimated by taking a large number of independent measurements and averaging them. The estimate is 370^0C, and the uncertainty (standard deviation) in this estimate is 0.10^0C.

(i) Demonstrate a 95% confidence interval for the temperature.

(4 marks)

(ii) Solve if the confidence level of the interval 37 ± 0.1^0C?

(4 marks)

(iii) If only a small number of independent measurements had been made, what additional assumption would be necessary in order to compute a confidence interval? Write the type of population the measurements come from.

(2 marks)

(iv) Making the additional assumption, examine a 95% confidence interval for the temperature if 10 measurements were made.

(4 marks)

END OF QUESTIONS

Appendix I

FINAL EXAMINATION

SEMESTER / SESSION : SEMESTER I /2019/2020
COURSE : ENGINEERING STATISTICS

PROGRAM : 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.
- $\text{Var}(aX + b) = a^2 \text{ Var}(x)$

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

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

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

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

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

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

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

Appendix I

FINAL EXAMINATION

SEMESTER / SESSION : SEMESTER I /2019/2020
COURSE : ENGINEERING STATISTICS

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

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

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

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

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

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

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

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

Appendix I

FINAL EXAMINATION

SEMESTER / SESSION : SEMESTER I /2019/2020
COURSE : ENGINEERING STATISTICS

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

Appendix I

FINAL EXAMINATION

SEMESTER / SESSION : SEMESTER I /2019/2020 PROGRAM : BDD
COURSE : ENGINEERING STATISTICS 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	$n_1, n_2 \ge 30$	$Z_{Test} = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$
B	Known	$n_1, n_2 < 30$	$Z_{Test} = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$
C	Unknown	$n_1, n_2 \ge 30$	$Z_{Test} = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$
D	Unknown (Equal)	$n_1, n_2 < 30$	$T_{Test} = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{S_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ $v = n_1 + n_2 - 2$
E	Unknown (Not equal)	$n_1 = n_2 < 30$	$T_{Test} = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{1}{n}(s_1^2 + s_2^2)}}$ $v = 2(n-1)$
F	Unknown (Not equal)	$n_1, n_2 < 30$	$T_{Test} = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ $v = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$

Appendix I

FINAL EXAMINATION

SEMESTER / SESSION : SEMESTER I /2019/2020 PROGRAM : 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 $v = n-2$