UNIVERSITI TUN HUSSEIN ONN MALAYSIA

FINAL EXAMINATION
SEMESTER 1
SESI 2018/2019

COURSE NAME	:	ENGINEERING STATISTICS
COURSE CODE	:	BDA 24103
PROGRAMME	:	BDD
EXAMINATION DATE	:	DECEMBER 2018/JANUARY 2019
DURATION	:	3 HOURS
INSTRUCTION	:	**SECTION A:** ANSWER **ALL** QUESTIONS. **SECTION B:** ANSWER **THREE (3)** FROM FOUR (4) QUESTIONS.

THIS PAPER CONSISTS OF **ELEVEN (11)** PAGES

SECTION A

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

Q1 (a) Identify the type I error and the type II error that correspond to the given hypothesis below:

(i) The mean score of IQ test is less than 95

(3 marks)

(ii) The mean amount of soya bean drink in cans is equal to 325 ml.

(3 marks)

(b) UTHM has organized a physics quiz for secondary school in Batu Pahat. In the Physics quiz, the sample sizes of two parts are 10 students. For part 1, the mean score was 35 with standard deviation of 2.5. While, for part 2, the mean score was 24 with standard deviation of 2.1. Test the difference between the performances of the two parts using 0.01 of significance level. Assume that the variances of population are unknown but not equal.

(14 marks)

Q2 Quality engineers in an automotive company are trying to identify the key predictors of engine knock. They consider the air-fuel ratio (AFR) variable as one of the most significant factors. The engineers collect data from 13 randomly selected engines, all running on petrol with an octane rating of 95. The data is recorded in **Table Q2.**

Table Q2

AFR	**Knock**
13.9	83.7
13.8	84.0
13.7	84.1
13.8	84.2
13.9	84.4
15.2	88.4
14.9	89.4
15.3	90.1
15.2	90.2
14.8	90.5
16.1	93.0
15.9	93.1
15.5	93.3

(a) Using an appropriate plot, determine whether relationship exist between the response variable and the predictor.

(4 marks)

(b) Using the method of least squares, estimate the regression line in **Q2 (a).**

(8 marks)

(c) Test the slope, $\beta_1 = 2$ at 5% level of significance.

(6 marks)

(d) Estimate the knocking when the AFR is 19.

(2 marks)

SECTION B

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

Q3 (a) Let the X is a continuous random variable with the next probability density function

$$f(x) = \begin{cases} k(2x^3+5) & , 0 \leq x \leq 2 \\ 0 & , \text{otherwise} \end{cases}$$

(i) Evaluate value of k

(2 marks)

(ii) Find $P(0 < X \leq 1)$

(2 marks)

(b) Given that the distribution function for a random variables Y is as

$$F(y) = \begin{cases} 0 & , \quad y \leq 0 \\ \dfrac{y}{8} & , \quad 0 < y < 2 \\ \dfrac{y^2}{16} & , \quad 2 \leq y \leq 4 \\ 1 & , \quad y > 4 \end{cases}$$

(i) Find the probability distribution of y.

(4 marks)

(ii) Find the probability of y is larger or equal to 1 but equal or less than 3

(6 marks)

(iii) Find the probability of y is larger than or equal to 1.5

(6 marks)

Q4 Solid Mechanics quiz consists of 10 questions were given to a group of students. Each question has 4 choices. Grade A is given to the students who managed to get 80% marks and above. According to the compiled data, 45% of students managed to get grade A.

(a) Calculate the probability that exactly 3 of a random sample of 5 the students were grade A?

(6 marks)

(b) Calculate the probability that not more than 7 of a random sample of those 15 students were grade A?

(6 marks)

(c) Among the candidates, one student has not studied and decided to guess the answers to every questions. Find the probability that the candidate will get grade A in the quiz.

(8 marks)

Q5 (a) Explain this formula : $e = |\bar{x} - \mu|$

(2 marks)

(b) An electrical firm manufactures light bulbs that have a length of life that is approximately normally distributed, with mean equal to 800 hours and a standard deviation of 40 hours. Find the probability that a random sample of 16 bulbs will have an average life of less than 775 hours.

(8 marks)

(c) Two independent experiments are being run in which two different types of paints are compared. Eighteen specimens are painted using type A and the drying time in hours is recorded on each. The same is done with type B. The population standard deviations are both known to be 1.0. Assuming that the mean drying time is equal for two types of paint, find $P(\bar{X}_A - \bar{X}_B > 1.0)$, where $\bar{X}_A$ and $\bar{X}_B$ are average drying times for samples of size $n_A = n_B = 18$.

(10 marks)

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

(6 marks)

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

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

(4 marks)

(ii) What is the confidence level of the interval 37 ± 0.1 °C ?

(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?

(2 marks)

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

(4 marks)

END OF QUESTIONS

FINAL EXAMINATION

SEMESTER / SESSION : SEMESTER I /2018/2019
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 :

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

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

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

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

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

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

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

(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 /2018/2019
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	$P(X=x)=\frac{e^{-\mu}\cdot\mu^x}{x!}$, $x=0,1,2,\ldots,\infty$
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 $n\geq30$ and $p\leq0.1$
Mean	$\mu=np$

Normal Approximation to the Binomial Distribution	
Condition	Use if n is large and $np\geq5$ and $nq\geq5$
Mean	$\mu=np$
Variance	$\sigma^2=npq$

Sampling error of single mean : $e=\left|\bar{x}-\mu\right|$.

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

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

$$\bar{x}\sim N\left(\mu_{\bar{x}1-\bar{x}2},\sigma^2_{\bar{x}1-\bar{x}2}\right)$$

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

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

$$P(\bar{x}_1-\bar{x}_2>r)=P\left(Z>\frac{r-\mu_{\bar{x}1-\bar{x}1}}{\sigma_{\bar{x}1-\bar{x}2}}\right)$$

FINAL EXAMINATION

SEMESTER / SESSION : SEMESTER I /2018/2019
COURSE : ENGINEERING STATISTICS

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

FINAL EXAMINATION

SEMESTER / SESSION : SEMESTER I /2018/2019
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 I /2018/2019
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

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

(ii) Intercept

$$T_{test} = \frac{\hat{\beta}_0 - \beta_C}{\sqrt{MSE(1/n + \bar{x}^2 / Sxx)}}$$

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

Coefficient of Determination, r^2.

$$r^2 = \frac{Syy - SSE}{Syy} = 1 - \frac{SSE}{Syy}$$

Coefficient of Pearson Correlation, r.

$$r = \frac{Sxy}{\sqrt{Sxx \cdot Syy}}$$