UNIVERSITI TUN HUSSEIN ONN MALAYSIA

FINAL EXAMINATION
SEMESTER II
SESI 2015/2016

COURSE NAME	:	ENGINEERING STATISTICS
COURSE CODE	:	BDA 24103
PROGRAMME	:	2BDD
EXAMINATION DATE	:	JUNE / JULY 2016
DURATION	:	3 HOURS
INSTRUCTION	:	**SECTION A:** ANSWER **ALL QUESTIONS** IN THIS SECTION. **SECTION B:** ANSWER **FOUR (4) QUESTIONS** FROM **FIVE (5) QUESTIONS** PROVIDED IN THIS SECTION.

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

SECTION A

Instruction: Please answer all questions in this section.

Q1 (a) Choose the right answer

(i) Which of these is NOT continuous data:
A. A person's height each year
B. The volume of water in a swimming pool each day
C. Cars finished in a factory each day
D. A person's weight on each birthday

(ii) Which of these is NOT discrete data:
A. Height of a sunflower as measured each day
B. How many students are absent from school each day
C. How many widgets a business sell each day
D. The number of people who drive through a red light each hour during rush hour

(iii) Two dice are rolled and the sum of the face value is six. What is the probability that at least one of the dice came up a 3?
A. 1/2
B. 1/5
C. 2/3
D. 5/6

(iv) iv. If you toss a dice, what's the probability that you roll a 3 or less?
A. 1/6
B. 1/3
C. 1/2
D. 5/6

(v) The number of times that a cyclist rides over 45 km each day is what sort of data:
A. Discrete
B. Continuous

(5 marks)

(b) Use the graph of the uniform distribution in Figure Q1b (all values in the range have the same probability of occurring) to answer questions i until iii.

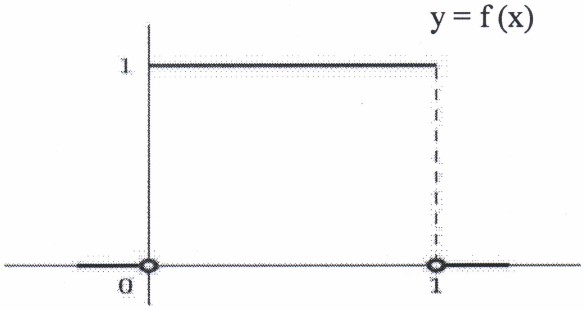

Figure Q1b

(i) Find P(0<X<1), the probability that X will be between 0 and 1.
(2 marks)

(ii) Find P(X>0.65), the probability that X will be greater than 0.65.
(2 marks)

(iii) Find (X<.40), the probability that X will be less than 0.40.
(2 marks)

(c) The mileage (in 1000's of miles) for which a certain type of tyre will last is a random variable

$$f(x) \begin{cases} \frac{1}{20} e^{-x/20} & \text{for all } x > 0 \\ 0 & \text{for all } x \leq 0 \end{cases}$$

Find the probability that the tyre will last:

(i) at most 10,000 miles
(3 marks)

(ii) between 16,000 and 24,000 miles
(3 marks)

(iii) at least 30,000 miles.
(3 marks)

SECTION B

Instruction: Please answer FOUR (4) questions from FIVE (5) questions provided in this section.

Q2 (a) Of all the registered automobiles in a certain state, 10% violate the state emissions standard. Twelve automobiles are selected at random to undergo an emissions test.

(i) Find the probability that exactly three of them violate the standard.
(4 marks)

(ii) Find the probability that fewer than three of them violate the standard.
(4 marks)

(iii) Find the probability that none of them violate the standard.
(4 marks)

(b) (i) Grandma bakes chocolate chip cookies in batches of 100. She puts 300 chips into the dough. When the cookies are done, she gives her grandchildren one. What is the probability that the cookie contains no chocolate chips?
(3 marks)

(ii) Grandma's grandchildren have been complaining that Grandma is too stingy with the chocolate chips. Grandma agrees to add enough chips to the dough so that only 1% of the cookies will contain no chips. How many chips must she include in a batch of 100 cookies to achieve this?
(5 marks)

Q3 Ali and Abu are both traveling for Batu Pahat to Alor Star on a weekly basis. The arrival time for both them is 98.4 minutes with a second derivation of 7.8 minutes and 110.7 minutes with standard derivation of 29.8 minutes respectively. Assume that the traveling population are approximately normal distributed and random samples of traveling time of Ali and Abu are 36 and 49 respectively.

(i) Find probability of Ali's average traveling time to be more than 90 minutes.
(5 marks)

(ii) Find the probability of Abu's mean average time to be in between 105 and 115 minutes.
(5 marks)

(iii) Calculate the mean of traveling of Abu will have a mean of case 13 mins more than traveling time of Ali.
(10 marks)

Q4 A study was conducted by the Department of Biology at the UTHM to estimate the difference in the amount of the chemical orthophosphorus measured at two different stations on the Sungai Batu Pahat. Orthophosphorus is measured in milligrams per liter. Fifteen samples were collected from station 1 and twelve samples were obtained from station 2. The samples from station 1 had an average orthophosphorus content of 3.84 milligrams per liter and a standard deviation of 3.07 milligrams per liter, while the samples from station 2 had an average content of 1.49 milligrams per liter and a standard deviation of 0.80 milligram per liter.

(i) Find a 95% confidence interval for the difference in the true average orthophosphorus contents at these two stations, assuming that the observations came from normal populations with different variances.

(10 marks)

(ii) A confidence interval for the difference in the mean orthophosphorus contents, measured in milligrams per liter, at two stations on the Sungai Batu Pahat was constructed by assuming the normal population variance to be unequal. Justify this assumption by constructing a 98% confidence interval for $\frac{\sigma_1^2}{\sigma_2^2}$ and for $\frac{\sigma_1}{\sigma_2}$, where σ_1^2 and σ_2^2 are the variances of the populations of orthophosphorus contents at station 1 and station 2, respectively.

(10 marks)

Q5 (a) A study was done to determine if the students from the public universities take longer to graduate than the students from the private universities. 100 students from both the public universities and private universities were surveyed. Suppose that from the years of research, it was known that the population standard deviations were 1.58 years and 1 year respectively.

The following data were collected. The public universities students took an average of 4.1 years with a standard deviation of 0.3. Use a 0.01 level of significance to test the claim.

(i) State the null and alternative hypotheses.

(3 marks)

(ii) Establish the Decision Rule.

(5 marks)

(iii) Is there any evidence to support the claim at α =0.01?

(15 marks)

Q6 One of the more challenging problems confronting the water pollution control field is presented by the furniture industry. Furniture wastes are chemically complex. They are characterized by high values of biochemical oxygen demand, volatile solids, and other pollution measures. Consider the experimental data of Table 6, which was obtained from 33 samples of chemically treated waste in the study conducted at the Universiti Tun Hussein Onn Malaysia. Readings on x, the percent reduction in total solids, and y, the percent reduction in chemical oxygen demand for the 33 samples, were recorded.

(i) Plot a scatter diagram for this population data.
(5 marks)

(ii) Estimate the regression line.
(5 marks)

(iii) Find 95% confidence interval for β in the regression line $\mu_{Y|x} = \alpha + \beta x$.
(5 marks)

(iv) Test the hypothesis that β = 1.0 against the alternative that β < 1.0.
(5 marks)

Table 6: Measures of solids and chemical oxygen demand

Solids reduction x (%)	Chemical oxygen demand, y (%)	Solids reduction x (%)	Chemical oxygen demand, y (%)
3	5	36	34
7	11	37	36
11	21	38	38
15	16	39	37
18	16	39	36
27	28	39	45
29	27	40	39
30	25	41	41
30	35	42	40
31	30	42	44
31	40	43	37
32	32	44	44
33	34	45	46
33	32	46	46
34	34	47	49
36	37	50	51
36	38		

END OF QUESTION

Appendix I

FINAL EXAMINATION

SEMESTER / SESSION : SEMESTER I /2015/2016 PROGRAM : 2BDD
COURSE : ENGINEERING STATISTICS COURSE CODE: BDA24103

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

Appendix I

FINAL EXAMINATION

SEMESTER / SESSION : SEMESTER I /2015/2016 PROGRAM : 2BDD
COURSE : ENGINEERING STATISTICS COURSE CODE: BDA24103

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

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

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

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

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

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

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

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

Appendix I

FINAL EXAMINATION

SEMESTER / SESSION : SEMESTER I /2015/2016 PROGRAM : 2BDD
COURSE : ENGINEERING STATISTICS COURSE CODE: BDA24103

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

Appendix I

FINAL EXAMINATION

SEMESTER / SESSION : SEMESTER I /2015/2016 PROGRAM : 2BDD
COURSE : ENGINEERING STATISTICS COURSE CODE: BDA24103

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)	$n_1, n_2 < 30$	$T_{Test} = \dfrac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$ $v = \dfrac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^2}{\dfrac{\left(\dfrac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \dfrac{\left(\dfrac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$

Appendix I

FINAL EXAMINATION

SEMESTER / SESSION : SEMESTER I /2015/2016 PROGRAM : 2BDD
COURSE : ENGINEERING STATISTICS COURSE CODE: BDA24103

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 $v = n-2$