CONFIDENTIAL

UNIVERSITI TUN HUSSEIN ONN MALAYSIA

FINAL EXAMINATION
SEMESTER 2
SESSION 2022/2023

COURSE NAME	:	STATISTICS
COURSE CODE	:	BIT11603
PROGRAMME CODE	:	BIT
EXAMINATION DATE	:	JULY/AUGUST 2023
DURATION	:	3 HOURS
INSTRUCTION	:	1. ANSWER ALL QUESTIONS. 2.THIS FINAL EXAMINATION IS CONDUCTED VIA **CLOSED BOOK.** 3. STUDENTS ARE **PROHIBITED** TO CONSULT THEIR OWN MATERIAL OR ANY EXTERNAL RESOURCES DURING THE EXAMINATION CONDUCTED VIA CLOSED BOOK.

THIS QUESTION PAPER CONSISTS OF **SEVEN (7)** PAGES

CONFIDENTIAL

Q1 (a) Two balls are drawn from a bag containing 5 white and 7 black balls at random. What is the probability that they would be of different colors?

(4 marks)

(b) Let X be a random variable with the following probability density function (pdf)

$$f(x) = \begin{cases} \left(\frac{1+x}{2}\right)^2, & -1 < x \leq 0 \\ c + 2x(1-x), & 0 < x \leq 1 \\ \frac{2-x}{3}, & 1 < x \leq 2 \\ 0 & otherwise \end{cases}$$

Compute

(i) the value of c.

(4 marks)

(ii) $E(X^2)$.

(6 marks)

(ii) cumulative distribution function, $F(x)$ of X.

(6 marks)

Q2 (a) A factory produces components of which 1% are defective. The components are packed in boxes of 10. A box is selected at random.

(i) Find the probability that the box contains exactly one defective component.

(2 marks)

(ii) Find the probability that there are at least 2 defective components in the box.

(3 marks)

(iii) Using a suitable approximation, find the probability that a batch of 250 components contains between 1 and 4 (inclusive) defective components.

(5 marks)

(b) A gas supplier maintains a team of engineers who are available to deal with leaks reported by customers. Most reported leaks can be dealt with quickly but some require a long time. The time (excluding travelling time) taken to deal with reported leaks is found to have a mean of 65 minutes and a standard deviation of 60 minutes. Assuming that the times may be modelled by a normal distribution, calculate the probability that:

(i) it will take more than 185 minutes to deal with a reported leak.

(4 marks)

(ii) it will take between 50 minutes and 125 minutes to deal with a reported leak. (6 marks)

Q3 (a) A large local restaurant management conducts a study to find the amount spent on lunch at one of their branches. Random samples of 50 men and 100 women were gathered. For men, the average expenditure was RM20 with a standard deviation of RM3. For women, it was RM15 with a standard deviation of RM2. Assume that the populations are independent and normally distributed.

(i) Let μ_1 and μ_2 be the mean lunch expenditures for men and women. What are the point estimate of $\mu_1 - \mu_2$? (2 marks)

(ii) Find a 99% confidence interval for the lunch spending difference between men and women. Interpret the interval. (8 marks)

(b) **Table Q3** shows the body mass indexes (BMI) for recent Miss World winners. Use a 0.01 significance level to test the claim that recent Miss World winners are from a population with a standard deviation of 1.34, which was the standard deviation of BMI for winners from the 2010s. Do recent winners appear to have variation that is different from that of the 2010s?

Table Q3

19.5	20.3	19.6	20.2	17.8	17.9	19.1	18.8	17.6	16.8

(10 marks)

Q4 (a) In a manufacturing plant, plastic sheathing is specified to be at least two mils thick by one of the many quality measures. Set up the null and alternative hypothesis for a quality monitoring system that ensures the desired level of quality. (2 marks)

(b) An ice cream company claimed that its product contain on average 500 calories per pint.

(i) Test the claim if 24 pint containers were analyzed, giving the mean is 507 calories and a standard deviation of 21 calories at 1% level of significance. (6 marks)

(ii) Test the claim if 42 pint containers were analyzed, giving the mean is 509 calories and a variance of 18 calories at 1% level of significance. (7 marks)

Q5 **Table Q5** gives information on ages and cholesterol levels for a random sample of 10 men.

Table Q5

Age	58	69	43	39	63	52	47	31	74	36
Cholesterol level	189	235	193	177	154	191	213	165	198	181

(a) Draw the scatter plot of the data and explain if the graph indicates whether simple linear regression is appropriate for these data.

(4 marks)

(b) Taking age as an independent variable and cholesterol level as a dependent variable, compute S_{xx}, S_{yy}, and S_{xy}.

(8 marks)

(c) Find the regression equation of cholesterol level on age.

(6 marks)

(d) Calculate the R square, r value and explain what does it indicate about the simple linear regression model?

(5 marks)

(e) Predict the cholesterol level of a 60-year-old man.

(2 marks)

-END OF QUESTIONS –

FINAL EXAMINATION

SEMESTER/ SESSION: SEM II / 2022-2023 **PROGRAMME CODE: BIT**

COURSE NAME: STATISTICS **COURSE CODE: BIT 11603**

Formulae

Special Probability Distributions :

$P(x=r)={}^{n}C_r \cdot p^r \cdot q^{n-r}$, $r=0,1,\ldots,n$, $X \sim B(n,p)$, $P(X=r)=\dfrac{e^{-\mu}\cdot\mu^r}{r!}$, $r=0,1,\ldots,\infty$,

$X \sim P_0(\mu)$, $Z=\dfrac{X-\mu}{\sigma}$, $Z \sim N(0,1)$, $X \sim N(\mu,\sigma^2)$.

Sampling Distributions :

$\overline{X} \sim N(\mu, \sigma^2/n)$, $Z=\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$, $T=\dfrac{\overline{x}-\mu}{s/\sqrt{n}}$, $\overline{X}_1-\overline{X}_2 \sim N\left(\mu_1-\mu_2, \dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}\right)$.

Estimations :

$n=\left(\dfrac{Z_{\alpha/2}\cdot\sigma}{E}\right)^2$, $\overline{x} \pm z_{\alpha/2}(\sigma/\sqrt{n})$; $\overline{x} \pm z_{\alpha/2}(s/\sqrt{n})$, $\overline{x} \pm t_{\alpha/2,v}\left(\dfrac{s}{\sqrt{n}}\right)$

$$\left(\bar{x}_1-\bar{x}_2\right)-Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} < \mu_1-\mu_2 < \left(\bar{x}_1-\bar{x}_2\right)+Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}},$$

$$\left(\bar{x}_1-\bar{x}_2\right)-Z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}} < \mu_1-\mu_2 < \left(\bar{x}_1-\bar{x}_2\right)+Z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}},$$

$$\left(\bar{x}_1-\bar{x}_2\right)-t_{\alpha/2,v}\cdot S_p\sqrt{\frac{2}{n}} < \mu_1-\mu_2 < \left(\bar{x}_1-\bar{x}_2\right)+t_{\alpha/2,v}\cdot S_p\sqrt{\frac{2}{n}}; v=2n-2$$

$$\left(\bar{x}_1-\bar{x}_2\right)-t_{\alpha/2,v}\cdot S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} < \mu_1-\mu_2 < \left(\bar{x}_1-\bar{x}_2\right)+t_{\alpha/2,v}\cdot S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

where Pooled estimate of variance, $S_p^2=\dfrac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$ with $v=n_1+n_2-2$,

$$\left(\bar{x}_1-\bar{x}_2\right)-t_{\alpha/2,v}\sqrt{\frac{1}{n}\left(s_1^2+s_2^2\right)} < \mu_1-\mu_2 < \left(\bar{x}_1-\bar{x}_2\right)+t_{\alpha/2,v}\sqrt{\frac{1}{n}\left(s_1^2+s_2^2\right)} \text{ with } v=2(n-1),$$

FINAL EXAMINATION

SEMESTER/ SESSION: SEM II / 2022-2023 **PROGRAMME CODE: BIT**

COURSE NAME: STATISTICS **COURSE CODE: BIT 11603**

Formulae

$$\left(\bar{x}_1-\bar{x}_2\right)-t_{\alpha/2,v}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}<\mu_1-\mu_2<\left(\bar{x}_1-\bar{x}_2\right)+t_{\alpha/2,v}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}} \text{ with } v=\frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1-1}+\frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2-1}}$$

$$\frac{(n-1)\cdot s^2}{\chi^2_{\alpha/2,v}}<\sigma^2<\frac{(n-1)\cdot s^2}{\chi^2_{1-\alpha/2,v}} \text{ with } v=n-1,$$

Hypothesis Testing :

$$Z_{Test}=\frac{\bar{x}-\mu}{\sigma/\sqrt{n}},\quad Z_{Test}=\frac{\bar{x}-\mu}{s/\sqrt{n}},\quad T_{Test}=\frac{\bar{x}-\mu}{s/\sqrt{n}}$$

$$Z=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}},\quad T=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{S_p\cdot\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \text{ with } v=n_1+n_2-2,$$

$$Z=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}},\quad T=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{1}{n}\left(s_1^2+s_2^2\right)}},\quad T=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}} \text{ with}$$

$$v=\frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1-1}+\frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2-1}}\cdot\ ;\ S_p^2=\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}\ ;\ \chi^2=\frac{(n-1)s^2}{\sigma^2}$$

$$F=\frac{s_1^2}{s_2^2},\quad \text{with } \frac{1}{f_{\alpha/2}(v_2,v_1)} \text{ and } f_{\alpha/2}(v_1,v_2)$$

FINAL EXAMINATION

SEMESTER/ SESSION: SEM II / 2022-2023 **PROGRAMME CODE: BIT**

COURSE NAME: STATISTICS **COURSE CODE: BIT 11603**

Simple Linear Regressions :

$$S_{xy} = \sum x_i y_i - \frac{\sum x_i \cdot \sum y_i}{n},\ S_{xx} = \sum x_i^{\ 2} - \frac{\left(\sum x_i\right)^2}{n},\ S_{yy} = \sum y_i^{\ 2} - \frac{\left(\sum y_i\right)^2}{n},\ \bar{x} = \frac{\sum x}{n},\ \bar{y} = \frac{\sum y}{n},$$

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}},\ \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x},\ \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x,\ r = \frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}},\ SSE = S_{yy} - \hat{\beta}_1 S_{xy},\ MSE = \frac{SSE}{n-2}.$$