Universiti Tun Hussein Onn Malaysia

UNIVERSITI TUN HUSSEIN ONN MALAYSIA

FINAL EXAMINATION
SEMESTER II
SESSION 2023/2024

COURSE NAME	:	STATISTICS
COURSE CODE	:	BWC 20603
PROGRAMME CODE	:	BWC
EXAMINATION DATE	:	JULY 2024
DURATION	:	3 HOURS
INSTRUCTIONS	:	1. ANSWER ALL QUESTIONS 2. THIS FINAL EXAMINATION IS CONDUCTED VIA ☐ Open book ☒ 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 **SIX (6)** PAGES

Q1 (a) In a semiconductor manufacturing processes, the occurrence of defects in microchips vary. An analysis reveals that out of all the microchips produced, 50% have no defects, 40% have one defect, 8% have two defects, and 2% have three defects. Let x represent the number of defects in a randomly selected microchip manufactured by the process.

(i) Compute $\mathrm{P}(x_i = 0), \mathrm{P}(x_i = 1), \mathrm{P}(x_i = 2)$ and $\mathrm{P}(x_i = 3)$

(4 marks)

(ii) Convert probability density function $\mathrm{f}(x_i)$ to cumulative distribution $\mathrm{F}(x_i)$ of the random variable x_i

(4 marks)

(iii) Identify F(2)

(2 marks)

(b) A continuous random variable X_i has the probability density function

$$f(x) = \begin{cases} ce^{(-|x|)}, & -\infty < x < \infty \\ 0, & \text{otherwise} \end{cases}$$

Determine

(i) the value of c

(5 marks)

(ii) $\mathrm{P}(-2 < X_i < 1)$

(5 marks)

Q2 (a) According to a recent survey by Safety Malaysia, 10% of Malaysians are afraid of heights. If a random sample of 50 Malaysians is selected, calculate the following probabilities using the binomial table:

(i) Probability that exactly **THREE (3)** people in the sample are afraid of heights.

(3 marks)

(ii) Probability that at most **FIVE (5)** people in the sample are afraid of heights.

(3 marks)

(iii) The mean, variance, and standard deviation of the number of Malaysians who are afraid of heights.

(6 marks)

(b) The weight of a certain brand of chocolate bars is assumed to be normally distributed with a mean of 100 grams and a standard deviation of 5 grams.

(i) Calculate the probability that a randomly selected chocolate bar weighs greater than 95.5 grams

(3 marks)

(ii) Calculate the probability that a randomly selected chocolate bar weighs between 90 and 105 grams

(4 marks)

Q3 (a) Selecting random samples of size n from an infinite population that has a standard deviation of two.

(i) Show that $\bar{x}$ would be a more precise estimator of μ if a sample size were increased from four to sixteen.

(2 marks)

(ii) Interpret the result.

(2 marks)

(b) The average monthly electricity consumption (in kilowatt-hours) between two types of households: households with traditional incandescent light bulbs and households with energy-efficient LED light bulbs are distributed $N_1(2000, 60)$ and $N_2(2500, 40)$, respectively. Two random samples are taken from each population, with sample sizes n_1 and n_2.

(i) Identify the sampling distribution of the difference in means of LED to incandescent households.

(2 marks)

(ii) Compute the probability of the mean monthly electricity consumption for sample size n_1=n_2= 30 for LED households compared to incandescent households, with the LED households consuming is at least 500.5 kWh more than traditional incandescent light bulb.

(4 marks)

(iii) Compute the probability of the mean monthly electricity consumption for sample size $n_1 = 20$ and $n_2 = 25$ for LED households with incandescent and LED households consuming is at most 502 kWh more than incandescent household.

(4 marks)

(c) Consider a study investigating the effectiveness of two different teaching methods, Method X and Method Y in improving students' test scores. The study involves collecting data from two independent random samples of students: one group taught using Method X and the other using Method Y. The sample sizes are $n_1 = 7$ for Method X and $n_2 = 13$ for Method Y. Let's denote the sample variances as s_1^2 and s_2^2 for Method X and Method Y, respectively. Determine the probability that the ratio of the sample variances, $\frac{s_1^2}{s_2^2}$ is less than 3.73.

(5 marks)

Q4 (a) **Table Q4(a)** below shows the difference between queuing system utilized at a grocery store checkout: A Single Queue System (SQS) and a Multiple Queue System (MQS). The waiting times for the SQS and MQS are as follows:

Table Q4(a)

SLS	6.5	6.6	6.7	6.8	7.1	7.3	7.4	7.7	7.7	7.7
MLS	4.2	5.4	5.8	6.2	6.7	7.7	7.7	8.5	9.3	10.0

(i) Construct a 95% confidence interval for σ under the SQS.

(4 marks)

(ii) Construct a 95% confidence interval for σ under the MQS.

(4 marks)

(iii) Explain which arrangement seems better: the SQS or MQS?

(2 marks)

(b) An officer wants to study the relationship between biomass productions of orange and cumulative intercepted solar radiation (Wh/m^2) over a six-week period following emergence. Biomass production is the mean dry weight in grams of independent samples of four plants which is collected at Sime Darby plantation. The data are shown in **Table Q4(b).**

Table Q4(b)

Solar radiation (x_i)	28.8	48.5	68.3	90.5	120.2	170.5
Plant Biomass (y_i)	15.8	48.2	71.1	95.7	150.4	210.5

(i) Sketch the scatter diagram, for the above data.

(4 marks)

(ii) Compute $\hat{\beta}_0$ and $\hat{\beta}_1$ for the linear regression of plant biomass on intercepted solar radiation. Write the regression equation and interpret the result.

(6 marks)

(iii) Calculate the plant biomass for 300 Wh/m^2 solar radiations.

(2 marks)

-END OF QUESTIONS-

APPENDIX A

FORMULAE

Random variables:

$\sum_{i=-\infty}^{\infty} P(x_i) = 1,\quad E(X) = \sum_i x_i \cdot P(x),\quad E(X^2) = \sum_i x^2 \cdot P(x),\quad \int_{-\infty}^{\infty} f(x)\, dx = 1,$

$E(X) = \int_{-\infty}^{\infty} x \cdot P(x)dx,\quad E(X^2) = \int_{-\infty}^{\infty} x^2 \cdot P(x)dx,\quad Var(X) = E(X^2) - [E(X)]^2.$

Special probability:

$P(x = r) = nC_r \cdot p^r \cdot q^{n-r}, r = 0,1,\ldots,n,\quad X \sim B(n,p),$

$P(X = r) = \frac{e^{-\mu} \cdot \mu^r}{r!},\quad r = 0,1,\ldots,\infty,\quad X \sim P_0(\mu),$

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

Sampling Distributions:

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

Estimations:

$\bar{x} - t_{\frac{\alpha}{2},v}\sqrt{\frac{s^2}{n}} < \mu < \bar{x} + t_{\frac{\alpha}{2},v}\sqrt{\frac{s^2}{n}},\quad \frac{(n-1)\cdot s^2}{\chi^2_{\frac{\alpha}{2},v}} < \sigma^2 < \frac{(n-1)\cdot s^2}{\chi^2_{1-\frac{\alpha}{2},v}}$ with $v = n - 1$.

Simple Linear Regression:

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

$s_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n},\quad s_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}.$