CONFIDENTIAL

UTHM
Universiti Tun Hussein Onn Malaysia

UNIVERSITI TUN HUSSEIN ONN MALAYSIA

FINAL EXAMINATION
SEMESTER II
SESSION 2021/2022

COURSE NAME	:	STATISTICS
COURSE CODE	:	BIC 10603
PROGRAMME CODE	:	BIS / BIP / BIW / BIM
EXAMINATION DATE	:	JULY 2022
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 **SIX (6)** PAGES

CONFIDENTIAL

Q1. (a) The number of successful projects (x) per day obtained by a small IT firm can be described by the following probability distribution function:

$$f(x) = \begin{cases} 0.5x - 1, & 2 < x < 4 \\ 0, & otherwise \end{cases}$$

(i) Find the cumulative distribution function.

(4 marks)

(ii) Find $P(X<2)$.

(1 mark)

(iii) Determine the mean and variance.

(5 marks)

(b) A Gardenia factory produces breads where 2% are broken. The bread is packed into a box of 12 and it selected randomly.

(i) Find the probability that the box contains exactly 2 broken breads.

(5 marks)

(ii) Find the probability that there are at least 3 broken breads in the box.

(5 marks)

Q2. (a) State **TRUE** or **FALSE** for each of the following statements.

(i) The Central Limit Theorem gives the exact probability of estimating the true mean.

(1 mark)

(ii) The Central Limit Theorem only applies when the population distribution is normal.

(1 mark)

(iii) The range of values for the sampling distribution of means is larger than the range of scores in the population.

(1 mark)

(iv) The Central Limit Theorem requires that all samples are randomly selected from a single population.

(1 mark)

(b) A random sample of 50 cars was taken from the population of 5000 cars. The average commuting time for the population is 45 minutes with standard deviation of 15 minutes.

(i) Determine the probability that the sample mean is within 5 minutes from the population mean.

(6 marks)

(ii) Explain whether the answer in **Q2(b)(i)** is exact or an estimate.

(2 marks)

(iii) Determine the probability in **Q2(b)(i)** by using the sample size of 200.

(8 marks)

Q3. The following data represent a random sample of 10 houses in a particular area, each of which is heated with natural gas, is selected and the amount of gas (therms) used during the month of March is determined for each house.

130 165 181 98 152 174 122 190 138 99

(a) Let μ denote the average gas usage during March by all houses in this area. Compute a point estimate of μ. What is the estimated standard error of the estimator?

(8 marks)

(b) Construct a 95% and 99% confidence interval about μ in **Q3(a)**.

(8 marks)

(c) Suppose there are 10,000 houses in this area that use natural gas for heating. Let M denote the total amount of gas used by all these houses during March. Estimate M using the data of **Q3(a)**. What estimator did you use in computing your estimate?

(4 marks)

Q4. A vote is to be taken among the residents of a town and the surrounding county to determine whether a proposed chemical plant should be constructed. The construction site is within the town limits, and for this reason many voters in the county believe that the proposal will pass because of the large proportion of town voters who favor the construction. To determine if there is a significant difference in the proportions of town voters and county voters favoring the proposal, a poll is taken. The poll results in 120 of 200 town voters favor the proposal and 240 of 500 county residents favor it. Use an $\alpha = 0.05$ level of significance.

(a) State the hypotheses.

(2 marks)

(b) Compute the proportion of town voters p_1, the proportion of country voters p_2, and the pooled estimate of proportion p.

(7 marks)

(c) Compute z-value and its corresponding probability P.

(5 marks)

(d) State the conclusion whether you would agree that the proportion of town voters favoring the proposal is higher than the proportion of country voters.

(6 marks)

Q5. A patient is given an 5 fluid drip, and the concentration of the fluid in his blood is measured in appropriate units at one-hour intervals. The doctors expected that a linear relationship exists between the variables. The following information was discovered:

Time, x (hours)	0	1	2	3	4	5	6	7	8
Concentration, y	2.4	4.3	5.0	6.9	9.1	11.4	13.5	15.1	16

(a) Calculate the linear regression coefficients b_0 and b_1 for this data.

(10 marks)

(b) Write the fitted simple linear regression model for the above data.

(1 mark)

(c) Find the concentration of the fluid at 3.5 hours.

(2 marks)

(d) Find how long it would be for the concentration reaches 8 units. State your answer in hours, minutes and seconds.

(2 marks)

(e) Compute the coefficient of correlation.

(3 marks)

(f) Interpret the sample correlation between time and concentration based on **Q5(e).**

(2 marks)

- END OF QUESTIONS –

Formula

$$\mu = \frac{\sum x}{N} \qquad \bar{x} = \frac{\sum x}{n} \qquad \bar{x} = \frac{\sum (x \cdot f)}{n}$$

$$\text{Median} = L_m + \left(\frac{\frac{n}{2} - F}{f_m} \right) i \qquad \sigma^2 = \frac{\sum (x-\mu)^2}{N} \qquad \sigma^2 = \frac{\sum m^2 f - \frac{(\sum mf)^2}{N}}{N}$$

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

$$\text{ıla, } s^2 = \frac{n\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)} \qquad E(X) = \mu = \Sigma x P(x) \qquad Var(X) = \sigma^2 = E(X^2) - [E(X)]^2$$

$$Var(X) = \sigma^2 = E(X^2) - [E(X)]^2 \qquad \mu = np \qquad \sigma^2 = npq \qquad E(X) = \int_{-\infty}^{\infty} x f(x)\, dx$$

$$P(x) = {}_nC_x p^x q^{n-x} = \frac{n!}{(n-x)!x!} p^x q^{n-x} \qquad P(X = x) = \frac{e^{-\mu} \cdot \mu^x}{x!}$$

$$\bar{x} - z_{\frac{\alpha}{2}} \left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \bar{x} + z_{\frac{\alpha}{2}} \left(\frac{\sigma}{\sqrt{n}}\right) \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \qquad \bar{X} \pm Z_{\frac{\alpha}{2}} \left(\frac{\sigma}{\sqrt{n}}\right) \qquad \bar{X} \pm Z_{\frac{\alpha}{2}} \left(\frac{s}{\sqrt{n}}\right) \qquad \bar{X} \pm t_{\frac{\alpha}{2}, n-1} \left(\frac{s}{\sqrt{n}}\right)$$

$$(\bar{X}_1 - \bar{X}_2) \pm Z_{\frac{\alpha}{2}} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Where $S_p{}^2 = \frac{(n_1 - 1)s_1{}^2 + (n_2 - 1)s_2{}^2}{n_1 + n_2 - 2}$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} X_i Y_i - \frac{(\sum_{i=1}^{n} X_i)(\sum_{i=1}^{n} Y_i)}{n}}{\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}} = \qquad r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$$