Universiti Tun Hussein Onn Malaysia

UNIVERSITI TUN HUSSEIN ONN MALAYSIA

FINAL EXAMINATION
SEMESTER I
SESSION 2023/2024

COURSE NAME	:	STATISTICS
COURSE CODE	:	DAC 21302
PROGRAMME CODE	:	DAA
EXAMINATION DATE	:	JANUARY / FEBRUARY 2024
DURATION	:	2 HOURS AND 30 MINUTES
INSTRUCTION	:	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 **Table Q1** contains data on the waiting times at a customer service hotline for a new product launch over 50 days.

Table Q1

Class limit	f
46 – 57	7
58 – 69	3
70 – 81	5
82 – 93	14
94 – 105	12
106 – 117	7
118 – 129	2

(a) If x is the midpoint, construct a table that contains lower boundary, cumulative frequency, midpoint, x^2, $f_i x_i$, $f_i x_i^2$, $\sum f$, $\sum f_i x_i$, and $\sum f_i x_i^2$.

(8 marks)

(b) Find the mean, median, mode, variance and standard deviation.

(12 marks)

Q2 (a) Several groups of workers scrutinized the budget allocated to sellers in the company's cafeterias for environmental maintenance and pollution control. After analyzing the data, they concluded that the distribution of this budget is as follows:

$$f(b) = \begin{cases} (1-b)^4 5; & 0 \le b \le 1, \\ 0; & \text{elsewhere.} \end{cases}$$

(i) Verify that $f(b)$ is a density distribution function.

(3 marks)

(ii) If a cafeteria is chosen at random, calculate the probability that it spends more than 45% of its budget on environment and pollution control.

(3 marks)

(iii) Calculate $E(B+5)$ and $Var(2B-1)$ if given that $E(B) = 0.167$ and $E(B^{2)} = 0.048$.

(4 marks)

(b) The average weight of individuals in a fitness club is 75.25 kg with a standard deviation of 6.2 kg.

(i) Determine the probability of obtaining a weight more than 77.84 kg.

(3 marks)

(ii) Determine the probability of obtaining a weight between 65 kg and 69 kg.

(3 marks)

(iii) Find the weight below which 25% of individual fall.

(4 marks)

Q3 The distribution of heights of a Terrier dogs has a mean height of 72 centimeters and a standard deviation of 10 centimeters, whereas the distribution of heights of Poodle dogs has a mean height 28 centimeters with a standard deviation of 5 centimeters. Assuming that the sample means can be measured to any degree of accuracy, find

(a) the probability that the height of Terrier dogs more than 55 centimeters.

(4 marks)

(b) the probability that the mean height of a random sample of 100 Poodle dogs falls between 27.5 centimeters and 30 centimeters.

(4 marks)

(c) the probability that the sample mean for a random sample of heights of 64 Terrier dogs exceeds the sample mean for a random sample of heights of 100 Poodle dogs by at most 44.2 centimeters.

(6 marks)

(d) the probability that sample mean for a random sample of heights of 50 Terrier dogs less than 50 Poodle dogs by 43.5 centimeters.

(6 marks)

Q4 (a) A workshop is manufacturing flower pots that are in circular shape. A sample of flowerpots is taken, and the diameters are 6, 8.5, 8, 10.2, 10, 7.4 and 9 centimeters. Find a 95% confidence interval for the mean diameter of flowerpots, assuming an approximate normal distribution.

(8 marks)

(b) Two independent fishing spots were selected for a study on the variety of fish species in Malaysian rivers. **Table Q4(b)** represent the monthly fish samples collected at each fishing spot.

Table Q4(b)

First spot	**Second spot**
$n_1 = 36$	$n_2 = 31$
$\overline{x_1} = 73.44$	$\overline{x_2} = 96.41$
$s_1^2 = 0.201$	$s_2^2 = 0.594$

Find a 90% confidence interval for the difference between the population means for the two fishing spots. Assume that the population are approximately normal distributed.

(12 marks)

Q5 A researcher wishes to explore the relationship between cost (Y) and delivery distance (X) for a courier company over 22 different routes, with both variables measured in dollars per mile.

Table Q5

Route	Cost	Distance	Route	Cost	Distance
1	2.139	3.147	12	2.027	3.141
2	2.126	3.16	13	1.985	2.928
3	2.153	3.197	14	1.956	3.063
4	2.153	3.173	15	2.004	3.096
5	2.154	3.292	16	2.001	3.096
6	2.282	3.561	17	2.015	3.158
7	2.456	4.013	18	2.132	3.338
8	2.599	4.244	19	2.195	3.492
9	2.509	4.159	20	2.437	4.019
10	2.345	3.776	21	2.623	4.394
11	2.059	3.232	22	2.523	4.251

By referring **Table Q5**,

(a) Calculate S_{xx}, S_{yy}, S_{xy}.

(9 marks)

(b) Determine and interpret sample correlation coefficient, r.

(3 marks)

(c) Evaluate $\hat{\beta}_0$ *and* $\hat{\beta}_1$.

(4 marks)

(d) Find the estimated regression line, $\hat{y}$.

(2 marks)

(e) Solve for the value of y if $x = 5$.

(2 marks)

- END OF QUESTIONS -

APPENDIX A

Formula

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

$$r=\frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}},\quad \hat{\beta}_1=\frac{S_{xy}}{S_{xx}},\quad \hat{\beta}_0=\bar{y}-\hat{\beta}_1\,\bar{x},$$

$$Z=\frac{\bar{x}-\mu}{\sigma/\sqrt{n}},\ Z=\frac{\bar{x}-\mu}{s/\sqrt{n}},\quad T=\frac{\bar{x}-\mu}{s/\sqrt{n}}$$

$$Z=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}},\ Z=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$$

$$\bar{x}=\frac{\sum f_i x_i}{\sum f_i},\quad M=L_M+C\times\left(\frac{n/2-F}{f_m}\right),\quad M_0=L+C\times\left(\frac{d_b}{d_b+d_a}\right)$$

$$s^2=\frac{1}{\Sigma f-1}\left[\sum_{i=1}^{n} f_i x_i^{\,2}-\frac{\left(\Sigma f_i x_i\right)^2}{\Sigma f}\right]$$

$$\sum_{i=-\infty}^{\infty} p(x_i)=1,\ E(X)=\sum_{\forall x} xp(x),\ \int_{-\infty}^{\infty} f(x)\ dx=1,\ E(X)=\int_{-\infty}^{\infty} xp(x)\ dx,\ Var(X)=E(X^2)-[E(X)]^2,$$

$$P(x)=\binom{n}{x}\cdot p^x\cdot(1-p)^{n-x}\quad x=0,1,\ldots,n,\ P(X=r)=\frac{e^{-\mu}\cdot\mu^r}{r!}\quad r=0,1,\ldots,\infty,$$

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

APPENDIX A

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

$$\overline{x} - z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \overline{x} + z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

$$\overline{x} - z_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right) < \mu < \overline{x} + z_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$

$$\overline{x} - t_{\alpha/2,\,v}\left(\frac{s}{\sqrt{n}}\right) < \mu < \overline{x} + t_{\alpha/2,\,v}\left(\frac{s}{\sqrt{n}}\right), \quad v = n - 1.$$

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

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

$$(\overline{x}_1 - \overline{x}_2) - t_{\alpha/2,\,v}\, S_P\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 < (\overline{x}_1 - \overline{x}_2) + t_{\alpha/2,\,v}\, S_P\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \text{ where}$$

$$S_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \text{ and } v = n_1 + n_2 - 2,$$

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

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