Universiti Tun Hussein Onn Malaysia

UNIVERSITI TUN HUSSEIN ONN MALAYSIA

FINAL EXAMINATION
SEMESTER II
SESSION 2017/2018

COURSE NAME	:	STATISTICS
COURSE CODE	:	DAS 20502
PROGRAMME	:	DAA / DAE / DAM / DAU / DAT
EXAMINATION DATE	:	JUNE / JULY 2018
DURATION	:	2 HOURS 30 MINUTES
INSTRUCTIONS	:	ANSWER **ALL** QUESTIONS IN SECTION A AND ANSWER **THREE (3)** QUESTIONS IN SECTION B

THIS QUESTION PAPER CONSISTS OF **TEN (10)** PAGES

SECTION A

Q1 It is assumed that achievement test scores should be correlated with student's classroom performance. One would expect that students who consistently perform well in the classroom (tests, quizes, etc.) would also perform well on a standardized achievement test (0 - 100 with 100 indicating high achievement). A teacher decides to examine this hypothesis. At the end of the academic year, she computes a correlation between the student's achievement test scores (she purposefully did not look at this data until after she submitted student's grades) and the overall GPA for each student computed over the entire year. The data for her class are provided in **Table Q1**.

Table Q1

Achievement	GPA
98	3.6
88	4.0
86	3.8
59	3.0
84	1.7
75	2.6
72	2.9
62	1.6

(a) Find S_{xx}, S_{yy} and S_{xy}.

(9 marks)

(b) Compute the sample correlation coefficient, r.

(2 marks)

(c) Find $\widehat{\beta}_1$ and $\widehat{\beta}_0$.

(4 marks)

(d) Compute the slope and y-intercept for a regression line based on the data.

(3 marks)

(e) If Aishah scored 35 on the achievement test, estimate her predicted GPA.

(2 marks)

Q2 (a) According to the Digest of Educational Statistics, a certain group of preschool children under the age of one year each spends an average of 30.9 hours per week in non-parental care. A study of state university center-based programs indicated that a random sample of 32 infants spent an average of 32.1 hours per week in their care. The standard deviation of the population is 3.6 hours. At $\alpha = 0.01$, is there sufficient evidence to conclude that the sample mean differs from the national mean?

(10 marks)

(b) The average length of "short hospital stays" for men is slightly longer than that for women, 5.2 days versus 4.5 days. A random sample of recent hospital stays for both men and women revealed the following in **Table Q2 (b)**. At $\alpha = 0.01$, is there sufficient evidence to conclude that the average hospital stay for men is longer than the average hospital stay for women?

Table Q2 (b)

	Men	Women
Sample size	32	30
Sample mean	5.5 days	4.2 days
Population standard deviation	1.2 days	1.5 days

(10 marks)

SECTION B

Q3 **Table Q3** below show the length (in mm) of 50 carrots in a farm near Bukit Puteri.

Table Q3

Length (mm)	Frequency
150 - 154	5
155 - 159	2
160 164	6
165 - 169	8
170 - 174	9
175 - 179	11
180 - 184	6
185 - 189	3

(a) Copy and complete the table above with lower boundary, x, fx, fx^2 and F.

(8 marks)

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

(12 marks)

Q4 (a) A teacher asks each student to write his name on a piece of paper and put it inside the box for a lucky draw. There are 10 local girls, 20 foreign girls, 15 local boys and 30 foreign boys in the class.

(i) Please fill up the contingency table.

Student	**Boy (*B*)**	**Girl (*G*)**	**Total**
Local (*L*)			
Foreign (*F*)			
Total			

(2 marks)

(ii) Find $P(L \cap B)$ and $P(G \cap F)$.

(4 marks)

(ii) After making the draw, the teacher announces that the winner is a girl. What is the probability that winner is a local student?

(2 marks)

(iii) After making the draw, the teacher announces that the winner is a foreign student. What is the probability that winner is a boy?

(2 marks)

(b) Let X be a continuous random variables with probability density function.

$$f(x) = \begin{cases} \frac{5}{12}(x) + 3, & 0 \leq x \leq 2 \\ 0, & \text{otherwise} \end{cases}$$

(i) Find $P(0.25 < x \leq 1.75)$.

(2 marks)

(ii) Find $P(x > 1.25)$.

(2 marks)

(iii) Find the cumulative distribution function of X, $F(X)$.

(2 marks)

(iv) Find $P(0.5 \leq x \leq 2)$.

(2 marks)

(v) Find expected value, $E(X)$.

(2 marks)

Q5 (a) The length of time between charges of the battery for a certain type of machines is normally distributed with a mean of 50 hours and a standard deviation of 15 hours. Find

(i) $P(X < 45)$.

(3 marks)

(ii) $P(43 < X < 55)$.

(4 marks)

(iii) the value of X will fall five percent of the graph.

(3 marks)

(b) 25 percent of 150 hand phones manufactured by a company are defective.

(i) Show that approximating a binomial probability with a normal probability normal distribution can be used. Hence, find the mean and standard deviation.

(5 marks)

(ii) By using approximating with a normal distribution, find the probability that at least 35 of them are defective.

(5 marks)

Q6 (a) The random variable of X, representing the numbers of assignments for each college student for a particular semester has the following probability distribution.

x	1	2	3	4	5
$P(X=x)$	0.1	0.2	0.25	0.25	0.2

(i) Find the population mean and variance.

(5 marks)

(ii) Find the sample mean and variance if a random sample of 50 drawn from that population.

(3 marks)

(b) In a study of annual family expenditures for general health care, two group populations were surveyed as shown in **Table Q6 (b)**. 100 of families were selected randomly from each population. Assume that the population distribution is approximately normally distributed.

Table Q6 (b)

Population	*M*	*N*
Mean	RM 180	RM 168
Standard deviation	27	18

(i) Write the sampling distribution of both population.

(2 marks)

(ii) Find the probability that sample mean expenditure for general health care of population *M* will less about RM 20 than expenditure of population *N*.

(5 marks)

(iii) Find the probability that sample mean expenditure for general health care of population *M* will greater about RM 15 than expenditure of population *N*.

(5 marks)

Q7 (a) Your company wants to improve sales on men's shirt. Past 7 day's sales data are recorded as below. Find the 95% confidence interval for the mean of total sales on men's shirt recorded.

130	200	280	180	310	260	210

(8 marks)

(b) To compare customer satisfaction levels of two competing cable television companies, 174 customers of Company 1 and 355 customers of Company 2 were randomly selected and were asked to rate their cable companies on a five-point scale, with 1 being least satisfied and 5 most satisfied. The survey results are summarized in the **Table Q7 (b)**.

Table Q7 (b)

Company 1	**Company 2**
$n_1 = 174$	$n_2 = 355$
$\bar{x}_1 = 3.51$	$\bar{x}_2 = 3.24$
$s_1 = 0.51$	$s_2 = 0.52$

Construct a point estimate and a 99% confidence interval for $\mu_1 - \mu_2$, the difference in average satisfaction levels of customers of the two companies as measured on this five-point scale.

(12 marks)

- END OF QUESTIONS -

TERBUKA

FINAL EXAMINATION

SEMESTER / SESSION : SEM II / 2017/2018 PROGRAMME : DAT / DAA / DAM / DAE / DAU

COURSE : STATISTICS COURSE CODE : DAS 20502

Formula

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

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

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

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

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

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

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

$$Var(X)=E(X^2)-[E(X)]^2,$$

TERBUKA

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

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

FINAL EXAMINATION

SEMESTER / SESSION : SEM II / 2017/2018 PROGRAMME : DAT / DAA / DAM / DAE / DAU

COURSE : STATISTICS COURSE CODE : DAS 20502

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

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

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

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

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

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

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

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

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

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

$$v = \frac{\left(s_1^2/n_1 + s_2^2/n_2\right)^2}{\frac{\left(s_1^2/n_1\right)^2}{n_1 - 1} + \frac{\left(s_2^2/n_2\right)^2}{n_2 - 1}}$$