UNIVERSITI TUN HUSSEIN ONN MALAYSIA

FINAL EXAMINATION
SEMESTER I
SESSION 2009/2010

SUBJECT	:	STATISTICS
CODE	:	DSM 2932
COURSE	:	3 DEE / DET / DFA / DFT / DDT / DDM
DATE	:	NOVEMBER 2009
DURATION	:	2 HOURS 30 MINUTES
INSTRUCTION	:	ANSWER **ALL** QUESTIONS IN **PART A** AND CHOOSE **THREE (3)** QUESTIONS IN **PART B**

THIS EXAMINATION PAPER CONSISTS OF 6 PAGES

PART A

Q1 The following **Table Q1** shows the earnings per share and dividends per share for eight electric utility companies in a recent year.

Table Q1 : The dividends per earnings

Earnings per share, x	2.78	1.41	2.74	0.92	2.44	3.50	3.68	1.97
Dividends per share, y	2.16	0.88	1.04	1.10	0.96	2.18	1.54	1.39

(a) Sketch a scatter plot for the data above.

(4 marks)

(b) Use the method of least squares to estimate the regression line and interpret the result.

(11 marks)

(c) Estimate the dividends per share if given the earnings per share are 4.02.

(2 marks)

(d) Calculate the sample coefficient of correlation and interpret the result.

(3 marks)

(e) Calculate the coefficient of determination and interpret the result.

(5 marks)

PART B

Q2 (a) Given that X is the random variable showing the number of boys in families with three children.

(i) Draw the tree diagram base on situation above.
(ii) State the possible values of X.
(iii) Construct a table showing the probability distribution of X.
(iv) State cumulative distribution function of X in a suitable table.
(v) Calculate an expected value of X.

(11 marks)

(b) Suppose X denotes the number of telephone received in a single family residential house. From an examination of the phone subscription records of 1000 residence in a city, the following probability distribution function of X is obtained

$$P(x) = \begin{cases} 0.17 & , \quad x = 0, 1 \\ 0.23 & , \quad x = 2, 3 \\ 0.2 & , \quad x = 4 \\ 0 & , \quad \text{otherwise.} \end{cases}$$

(i) Show that the function given above is probability distribution function of X.
(ii) Calculate the probability that the number of telephone received is at most one.
(iii) Calculate $Var(3X+1)$.

(14 marks)

PART A

Q1 The following **Table Q1** shows the earnings per share and dividends per share for eight electric utility companies in a recent year.

Table Q1 : The dividends per earnings

Earnings per share, x	2.78	1.41	2.74	0.92	2.44	3.50	3.68	1.97
Dividends per share, y	2.16	0.88	1.04	1.10	0.96	2.18	1.54	1.39

(a) Sketch a scatter plot for the data above.

(4 marks)

(b) Use the method of least squares to estimate the regression line and interpret the result.

(11 marks)

(c) Estimate the dividends per share if given the earnings per share are 4.02.

(2 marks)

(d) Calculate the sample coefficient of correlation and interpret the result.

(3 marks)

(e) Calculate the coefficient of determination and interpret the result.

(5 marks)

PART B

Q2 (a) Given that X is the random variable showing the number of boys in families with three children.

(i) Draw the tree diagram base on situation above.
(ii) State the possible values of X.
(iii) Construct a table showing the probability distribution of X.
(iv) State cumulative distribution function of X in a suitable table.
(v) Calculate an expected value of X.

(11 marks)

(b) Suppose X denotes the number of telephone received in a single family residential house. From an examination of the phone subscription records of 1000 residence in a city, the following probability distribution function of X is obtained

$$P(x) = \begin{cases} 0.17 & , \quad x = 0, 1 \\ 0.23 & , \quad x = 2, 3 \\ 0.2 & , \quad x = 4 \\ 0 & , \quad \text{otherwise.} \end{cases}$$

(i) Show that the function given above is probability distribution function of X.
(ii) Calculate the probability that the number of telephone received is at most one.
(iii) Calculate $Var(3X+1)$.

(14 marks)

Q3 (a) The probability that a student is accepted to a prestigious college is 0.3. If fifteen students from the same school apply,

(i) what is the probability that at most two are accepted ?

(ii) calculate mean and variance of the distribution.

(8 marks)

(b) In an English test, the score of the candidates were approximately normal distributed with a mean 210 points and standard deviation 45 points. Find the probability of the candidates who received the following scores

(i) greater than 160 points.

(ii) between 155 and 165 points.

(9 marks)

(c) The factory company in Perak had found out that the probability of cartridges failed to be sold is 40%. By consider the distribution as binomial, 500 cartridges were selected at random by the customers at the show room. Find the probability that less than 210 cartridges failed to be sold.

(8 marks)

Q4 (a) A shipment of steel bars will be accepted if the mean breaking strength of a random sample of ten steel bars is greater than 250 pounds per square inch. In the past, the breaking strength of such bars had a mean of 235 and a variance of 400. Assume that the breaking strength is normally distributed, what is the probability that the shipment will be accepted ?

(6 marks)

(b) A manufacturer of light bulbs claims that its light bulbs have a mean life of 700 hours and a standard deviation of 120 hours. Ayob purchased 144 of these bulbs and decided that he would purchase more if the mean life of his current sample exceeds 680 hours. What is the probability that he will not purchase again from this manufacturer ?

(7 marks)

(c) A study was designed to estimate the difference in diastolic blood pressure readings between men and women. The mean and standard deviation for sixteen men are 77.37 and 8.35, while for thirteen women are 71.08 and 9.22 respectively. Assume that the readings are normally distributed, find

(i) the sampling distribution of the different between diastolic blood pressure readings for men and women.

(ii) the probability that the different between diastolic blood pressure readings for men is greater than women.

(iii) the probability that the different between diastolic blood pressure readings from men is five less than women.

(12 marks)

Q5 (a) A nationwide survey of practicing physicians will be taken to estimate the true mean number of prescriptions written per day. The desired margin of sampling error is 0.75. A pilot study revealed that a reasonable planning value for the population standard deviation is five.

(i) If the desired level of confident is 99%, how many physicians should be contacted in the survey to estimate the true mean ?

(ii) If the desired confidence level is lowered to 95%, show whether the number of sample size is decreased or increased.

(8 marks)

(b) The data in **Table Q5 (b)** below show the typing rates (words per minute) achieved by each of ten secretaries working in two branches of an insurance company which use two different types of word processing programs.

Table Q5 (b) : Two branches insurance company

Branch 1	64	55	58	55	52	68	72	60	72	49
Branch 2	57	64	55	64	60	66	75	82	48	64

Construct a 90% confidence interval for the difference between two means of branch 2 and branch 1 by assuming that the **population variances are equal but unknown**.

(17 marks)

Q6 (a) A well known electrical company has published figures on the annual number of kilowatt-hours expended by various home appliances. It is claimed that a vacuum cleaner expends an average of 51 kilowatt-hours per year. If a random sample of 17 homes included in a planned study indicates that vacuum cleaners expend an average of 47 kilowatt-hours per year with a standard deviation of 12.3 kilowatt hours, does this suggest at the 0.05 level of significance that vacuum cleaners expend, on the average less than 51 kilowatt-hours annually ?

(8 marks)

(b) A local pizza restaurant located close to a college campus advertises that their delivery time to a college dormitory is less than the local branch of a national pizza chain. In order to determine whether this advertisement is valid, few students have decided to order six pizzas from the local pizza restaurant and six pizzas from the national chain, at all different times. The delivery times in minutes are shown in **Table Q6(b)** below.

Table Q6(b) : Pizza delivery times

Local	16.8	11.7	15.6	17.5	18.1	14.1
National	22.0	15.2	18.7	20.8	19.5	17.0

At the 0.01 level of significance, is there any evidence that the mean delivery times for the local pizza restaurant is lower than the mean delivery times for the national pizza chain ? Assume that the **population variances are not equal and unknown**.

(17 marks)

FINAL EXAMINATION

SEMESTER / SESSION: SEM I / 2009/2010 | COURSE : 3 DEE /DET /DFA /DFT /DDT/ DDM

SUBJECT : STATISTICS | CODE : DSM 2932

Formulae

Random Variable :

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

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

Special Probability Distributions :

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

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

Sampling Distributions :

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

Estimations :

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

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

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

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 $v = n_1 + n_2 - 2$,

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

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

FINAL EXAMINATION

SEMESTER / SESSION: SEM I / 2009/2010

COURSE : 3 DEE /DET /DFA /DFT /DDT/ DDM

SUBJECT : STATISTICS

CODE : DSM 2932

Hypothesis Testings :

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

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

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

Simple Linear Regressions :

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

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

$$R^2 = 1 - \frac{\text{SSE}}{S_{yy}}.$$