UNIVERSITI TUN HUSSEIN ONN MALAYSIA

FINAL EXAMINATION
SEMESTER I
SESSION 2010/2011

COURSE	:	STATISTICS
COURSE NAME	:	BSM 1413
PROGRAMME	:	1 BIT
EXAMINATION DATE	:	NOVEMBER / DECEMBER 2010
DURATION	:	3 HOURS
INSTRUCTION	:	ANSWER ALL QUESTIONS IN **PART A** AND **THREE (3)** QUESTIONS IN **PART B**

THIS EXAMINATION PAPER CONSISTS OF 7 PAGES

PART A

Q1 Executives of a video rental chain want to predict the success of a potential new store. The company's researcher begins by gathering information on number of rentals and average family income from several of the chain's present outlets. The **Table Q1** below shows the number of rentals and the average family income.

Table Q1 : The number of rentals and average family income (in thousands)

Rentals	Average family income (RM'000)
710	65
529	43
314	29
504	47
619	52
428	50
317	46
205	29
468	31
545	43
607	49
694	64

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

(3 marks)

(b) Calculate and interpret the coefficient of Pearson correlation.

(5 marks)

(c) Using the least square method, develop a regression model to predict the number of rentals by average family income.

(5 marks)

(d) Calculate the value of *SSE* and *MSE*.

(2 marks)

(e) Test the null hypothesis $\beta_1 = 1$ against the alternative hypothesis $\beta_1 > 1$ at the 0.05 level of significance.

(5 marks)

Q2 (a) The average price per square meter for warehouse has been RM322.80. A national real estate investor claims that the average price has changed now. The investors hires a researcher who randomly selected the samples of 36 warehouses that are for sale and finds that the mean price per square meter is RM316.70, with a standard deviation of RM12.90. Test the claim at 5% level of significance.

(6 marks)

(b) The power of supercomputers derives from the idea of parallel processing. Engineers at Wira Research are interested in determining whether one of two parallel processing designs produces faster average computing time. The **Table Q2** below shows the results (in seconds) of independent random computation times using the two designs.

Table Q2 : The following are the results (in seconds) of independent random computation times using the two designs.

Design 1	2.1	2.2	1.9	2.0	1.8	2.4	2.0	1.7	2.3	2.8	1.9
Design 2	2.6	2.5	2.0	2.1	2.6	3.0	2.3	2.0	2.4	2.8	3.1

Assume that the two populations of computing time are normally distributed and the population variances are equal. Is there any evidence that one parallel processing design (Design 1) allows for faster average computation than the other (Design 2) with 1% of significance level?

(8 marks)

(c) A large department store wants to test whether the variance waiting time in counter 1 is less than the variance waiting time in counter 2. Two independent random samples of 25 waiting times in each of the counters gives s_1=2.5 minutes and s_2=3.1 minutes. Test the hypothesis that the variance waiting time in counter 1 is longer than counter 2 with $\alpha = 0.02$.

(6 marks)

PART B

Q3 Given the probability density function of X.

$$f(x) = \begin{cases} \frac{1}{8}(x-3), & 3 \le x \le 7, \\ 0, & otherwise. \end{cases}$$

(a) Show that $f(x)$ is a density function.

(2 marks)

(b) What is the probability that X
  (i) is greater than 5.7?
  (ii) is less than 4.3?
  (iii) is between 2.9 and 3.2?

(6 marks)

(c) Find $E(X)$ and $E\,(3X\text{-}10)$.

(5 marks)

(d) Find $Var(X-25)$.

(7 marks)

Q4 (a) A high percentage of people, who fracture or dislocate a bone, will see a doctor for that injury. Suppose that the percentage for people who see a doctor for that injury is 99%. There are 300 people are randomly selected who have fractured or dislocated a bone. Using the appropriate approximation, find the

(i) probability that exactly five of them did not see a doctor.
(ii) probability that more than three of them did not see a doctor.
(iii) standard deviation of people who would not see a doctor.

(8 marks)

(b) The International Data Corporation reports that Compaq is number one PC market share in Hong Kong with 16% of the market. A researcher randomly selects 130 recent purchasers of PCs. Using the appropriate approximation, find the

(i) mean and variance of PC purchasers bought a Compaq.
(ii) probability that less than 25 PC purchasers bought a Compaq.
(iii) probability that between 15 and 23 inclusively PC purchasers bought a Compaq.

(12 marks)

Q5 (a) The average price for Shimano mountain bikes in Kuala Lumpur is RM2100. Suppose the standard deviation of bike price is RM450. If a random sample of 160 bikes are selected, what is the probability that the average price for a Shimano mountain bike in this sample will be

(i) exceed RM2200.
(ii) between RM2000 to RM2150.
(iii) not more than RM2000.

(8 marks)

(b) A survey shows that the average insurance cost to a company per employee per hour is RM3.40 for managers and RM3.80 for professional specialty workers. These figures were obtained from 35 managers and 41 professional specialty workers and their population standard deviation are RM0.38 and RM0.51.

(i) Write down the sampling distribution between managers and professional specialty workers.
(ii) Find the probability that the average insurance cost for managers is less than that the average insurance cost for professional specialty workers.
(iii) Given that the figures were changed from 35 managers to 15 managers. While the figures for the professional specialty was changed from 41 to 10 professional specialty. Find the probability that the average insurance cost for professional specialty workers is at least RM0.10 more than the average insurance cost for managers.

(12 marks)

Q6 Eleven people began a new diet to reduce their cholesterol. However, there were two people discontinue their new diet before the time ended. **Table Q6** below show the cholesterol reading of the people both before the new diet and a month after use the new diet.

Table Q6 : The cholesterol reading before and after the new diet

Dieter	1	2	3	4	5	6	7	8	9	10	11
Before	255	230	290	242	300	250	215	230	225	219	236
After	197	225	215	215	240	235	190	240	200		

(a) Construct 95% confidence interval for the difference between two means of the cholesterol reading before and after the new diet when the population variances for the cholesterol reading are not equal.

(13 marks)

(b) Construct 90% confidence interval for the ratio of the population variances, $\dfrac{\sigma^2 before}{\sigma^2 after}$.

(7 marks)

FINAL EXAMINATION

SEMESTER / SESSION: SEM I / 2010/2011 COURSE: 2 BIT

SUBJECT : STATISTICS CODE: BSM1413

Formulae

Random variables:

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

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

Special Probability Distributions :

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

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

Sampling Distributions :

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

Estimations :

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

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

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

where Pooled estimate of variance, $S_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}$ with $v = n_1 + n_2 - 2$,

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

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

$$\frac{(n-1)\cdot s^2}{\chi^2_{\alpha/2,v}} < \sigma^2 < \frac{(n-1)\cdot s^2}{\chi^2_{1-\alpha/2,v}} \text{ with } v = n-1,$$

$$\frac{s_1^2}{s_2^2}\cdot\frac{1}{f_{\alpha/2}(v_1,v_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2}\cdot f_{\alpha/2}(v_2,v_1) \text{ with } v_1 = n_1-1 \text{ and } v_2 = n_2-1.$$

Hypothesis Testings :

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

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

$$v = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1-1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2-1}}\ ;\ S_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}\ ;\ \chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

Simple Linear Regressions :

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

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

$$T = \frac{\hat{\beta}_1 - \beta_1^*}{\sqrt{\frac{MSE}{S_{xx}}}} \sim t_{n-2},\ T = \frac{\hat{\beta}_0 - \beta_0^*}{\sqrt{MSE\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}} \sim t_{n-2}.$$