CONFIDENTIAL

UNIVERSITI TUN HUSSEIN ONN MALAYSIA

FINAL EXAMINATION
SEMESTER I
SESSION 2011/2012

COURSE NAME	:	STATISTICS
COURSE CODE	:	BWM 11603 / BSM 1413
PROGRAMME	:	1 BIT 2 BIT 3 BIT 4 BIT
EXAMINATION DATE	:	JANUARY 2012
DURATION	:	3 HOURS
INSTRUCTION	:	ANSWER **ALL** QUESTIONS.

THIS EXAMINATION PAPER CONSISTS OF SEVEN (7) PAGES

CONFIDENTIAL

Q1 (a) Find the sample mean and sample variance for the following data:

6.1 5.7 5.8 6.0 5.8 6.3

Use the alternative formula, $s^2 = \dfrac{n\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)}$.

(6 marks)

(b) The time taken to perform a particular task, t hours, has the probability density function

$$f(t) = \begin{cases} 10ct^2, & 0 \le t < 0.6, \\ 9c(1-t), & 0.6 \le t < 1.0, \\ 0, & otherwise. \end{cases}$$

where c is a constant.

(i) Find the value of c.

(3 marks)

(ii) Determine the probability that the time will be between 0.4 and 0.8 hours.

(3 marks)

(iii)Find the expected value of time taken.

(3 marks)

(iv)Find the variance of time taken.

(5 marks)

Q2 (a) The height of boys at a particular age follow a normal distribution with mean 150.3 cm and variance 25 cm. Find the probability that a boy picked at random from this age group has height

(i) less than 153 cm.

(ii) between 150 cm and 158 cm.

(iii) at least 155 cm.

(iv) less than 154 cm or more than 157 cm.

(11 marks)

(b) A sample of n independent observations is taken from a normal population with mean 74 and standard deviation 6. The sample mean is denoted by $\overline{X}$. Find sample size if $P(\overline{X} > 75)$ is equal to 0.282.

(3 marks)

(c) Two independent experiments are being run in which two different types of paints are compared. A number of specimens n_A are painted using type A paint and the drying time in hours is recorded on each. The same is done with type B paint with n_B specimens. The population standard deviations are both known to be 1.0. Assuming that the mean drying time is equal for the two types of paint and both populations are normally distributed, find $P(\bar{X}_A - \bar{X}_B \leq 0.5)$, where $\bar{X}_A$ and $\bar{X}_B$ are average drying times for samples of size $n_A = 40$, and $n_B = 45$.

(6 marks)

Q3 (a) The average size of a farm in Genting Highlands is 191 acres, while the average size of a farm in Cameron Highlands is 199 acres. Assume that the data from Genting Highlands sample with size of 50 and Cameron Highlands sample with size of 55 give the standard deviation of 38 acres and 12 acres, respectively. Find the 90% confidence interval for the difference between two means.

(4 marks)

(b) The data below shows the price of an adult single-day ski lift ticket (in RM) from a selected sample of nationwide ski resorts.

59	54	53	52	51	39	49	46	49	48

Find the 90% confidence interval for the variance of the ticket price. Assume that the population is normally distributed.

(6 marks)

(c) A grading test for Mathematics was given to 25 male students and 13 female students. The result was obtained as below:

Male	$\bar{x}_1 = 82$	$s_1 = 8$	$n_1 = 25$
Female	$\bar{x}_2 = 78$	$s_2 = 7$	$n_2 = 13$

i) Find the 95% confidence interval for σ_1^2/σ_2^2, where σ_1^2 and σ_2^2 are grade population variances for all male and female students.

ii) Find the 98% confidence interval for σ_1/σ_2, where σ_1 and σ_2 are grade population standard deviation for all male and female students.

(10 marks)

Q4 (a) A promoter claimed that the mean amount of tar in filtered cigarettes is less than the mean amount of tar in non-filtered cigarettes. Refer to the data listed in **Table Q4 (a)**, use a 0.05 level of significance to test the promoter's claim. Assume σ unknown but equal and the two populations are normally distributed.

Table Q4 (a): Tar Contents (in mg) of Cigarettes

Filtered	16 15 16 14 16 1 16 18 10 14 12 11 14 13 13 13 16 16 8 16 11
Non-filtered	23 23 24 26 25 26 21 24

(12 marks)

(b) A medical researcher wishes to see whether the variances of the heart rates (in beats per minute) of smokers are different from the variances of heart rates of people who do not smoke. Two samples are selected, and the data are shown as below:

Smokers	**Nonsmokers**
$n_1 = 25$	$n_2 = 13$
$s_1^2 = 36$	$s_2^2 = 10$

Using $\alpha = 0.05$, is there enough evidence to support the claim?

(8 marks)

Q5 A mail-order catalog business that sells personal computer supplies, software and hardware maintains a centralized warehouse for the distribution of products ordered. Management is currently examining the process of distribution from the warehouse and is interested in studying the factors that affect warehouse distribution costs. Data have been collected over the past 10 months indicating the warehouse distribution costs (in thousands) and the number of orders received. The results are as follows:

Table Q5: The Warehouse Distribution Costs and The Number of Orders Received Over Past 10 Months.

Dist. Cost (Y)	52.95	71.66	85.58	63.69	72.81	68.44	52.46	70.77	82.03	74.39
No. of Order (X)	4015	3806	5309	4262	4296	4097	3213	4809	5237	4732

(a) Assuming a linear relationship, use the least squares method to find the simple linear regression model and interpret the meaning of β_1.

(9 marks)

(b) What is the approximate number of orders if the monthly warehouse distribution cost is 30,000 dollar?

(2 marks)

(c) Test the hypothesis against $\beta_1 \neq 0$ using $\alpha = 0.05$.

(6 marks)

(d) Compute the coefficient of determination, r^2 and interpret the meaning.

(3 marks)

FINAL EXAMINATION

SEMESTER / SESSION: SEM I / 2011/2012

COURSE: 1/2/3/4 BIT

SUBJECT : STATISTICS

CODE: BWM 11603 / BSM 1413

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 :

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

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

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

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

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 Testing :

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

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

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

$$F=\frac{s_1^2}{s_2^2},\quad \text{with } \frac{1}{f_{\alpha/2}(v_2,v_1)} \quad \text{and} \quad f_{\alpha/2}(v_1,v_2)$$

Simple Linear Regressions :

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

$$\hat{\beta}_1=\frac{S_{xy}}{S_{xx}},\ \hat{\beta}_0=\bar{y}-\hat{\beta}_1\bar{x},\ \hat{y}=\hat{\beta}_0+\hat{\beta}_1x,\ r=\frac{S_{xy}}{\sqrt{S_{xx}\cdot S_{yy}}},\ SSE=S_{yy}-\hat{\beta}_1S_{xy},\ MSE=\frac{SSE}{n-2},$$

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