CONFIDENTIAL

UNIVERSITI TUN HUSSEIN ONN MALAYSIA

FINAL EXAMINATION
SEMESTER I
SESSION 2019/2020

COURSE NAME	:	STATISTICS FOR TECHNOLOGY
COURSE CODE	:	BWD 21602
PROGRAMME CODE	:	BWD
EXAMINATION DATE	:	DECEMBER 2019/ JANUARY 2020
DURATION	:	2 HOURS
INSTRUCTION	:	ANSWER **ALL** QUESTIONS

THIS EXAMINATION PAPER CONSISTS OF **SEVEN (7)** PAGES

CONFIDENTIAL

Q1 (a) A binomial distribution is a type of discrete random variable. (True/False)

(1 mark)

(b) Give two properties of a normal distribution.

(2 marks)

(c) A factory produces light bulb that will be packed in boxes, which consists of two hundred light bulb each and there are 50 boxes during inspection. Suppose that 1 in 5000 light bulbs are defective. Let X denote the number of defective light bulbs.

(i) State the distribution type for the above statement.

(3 marks)

(ii) Obtain mean and variance for the number of defective light bulb.

(3 marks)

(iii) What is the chance that at most 3 out of 10000 light bulb is defective?

(5 marks)

(d) The mean time to complete a task is 727 milliseconds for 3rd graders and 532 milliseconds for 5th graders. The variances of the two grades are 12,000 milliseconds^2 for 3rd graders and 10,000 milliseconds^2 for 5th graders. The times for both graders are normally distributed. You randomly sample 12 3rd graders and 14 5th graders.

(i) What is the sampling distribution of difference for mean time to complete a task for 3rd graders and 5th graders?

(5 marks)

(ii) Give your comment about the mean difference to complete a task between 3rd graders and 5th graders in term of probability.

(6 marks)

Q2 (a) The following data in **Table Q2(a)** are the heat-producing capacities of copper from two factories (in millions of calories per ton):

Table Q2(a)

Factory *A*	1620	1555	1477	1524	1520	
Factory *B*	1417	1434	1473	1523	1462	1495

Assuming that the data constitute independent random samples from normal populations with unequal variances. Is the difference copper between the true average heat-producing capacities from factory *A* and *B* reliable with 0.05 alpha? The difference heat-producing of copper is reliable if *CV* for upper confidence level is less than 10% ($\% \, CV = \frac{s}{\bar{x}} \, x \, 100$).

(14 marks)

(b) In order to investigate the relationship between mean job tenure in years among workers who have a bachelor's degree or higher and those who do not, random samples of each type of worker were taken with the following results as in **Table Q2(b)**.

Table Q2(b)

Worker qualification	N	$\overline{X}$	s
Bachelor's degree or higher	153	5.3	1.4
Do not have bachelor's degree or higher	203	5.0	1.7

(i) State the 2 parameters of point estimate for job tenure in years among workers.
(2 marks)

(ii) At 1% level of significance, determine the confidence interval between the mean job tenure between worker with bachelor's degree or higher and worker without bachelor's degree or higher.
(9 marks)

Q3 (a) Rice pops (RP) are a type of breakfast cereal which are packed into boxes with a quoted net mass of 296 g by one of two different filling machines. The mass of RP delivered by filling machine A, an old machine, is known to be normally distributed with a standard deviation of 5 g. The mass of RP delivered by filling machine B, a new machine, is also known to be normally distributed but with a standard deviation of 3 g. **Table Q3(a)** shows the net masses in g of 7 boxes filled by machine A and B.

Table Q3(a)

A	300	297	299	302	306	307	308
B	297	302	299	296	299	303	301

(i) Abu claims that the average mass of RP delivered by machine A is more than 296 g. Test the hypothesis with 0.05 level of significance (T_{test} = 4.1221 and critical point, $t_{0.05,6}$ = 1.943).
(3 marks)

(ii) Abu also claims that there is no significant difference in the mean mass of RP delivered by the two filling machines. Using alpha 0.05, perform the hypothesis testing whether Abu's claim is true, if the variance population for 2 classes are equal.
(11 marks)

(b) A new drug is proposed to lower total cholesterol. A randomized controlled trial is designed to evaluate the efficacy of the medication in lowering cholesterol. Thirty participants are enrolled in the trial and are randomly assigned to receive either the new drug or a placebo. The participants do not know which treatment they are assigned. Each participant is asked to take the assigned treatment for 6 weeks. At the end of 6 weeks, each patient's total cholesterol level is measured and the sample statistics are as **Table Q3(b)**.

Table Q3(b)

Treatment	Sample Size	Mean	Standard Deviation
New drug	15	195.9	28.7
Placebo	15	227.4	30.3

Is there statistical evidence of a reduction in mean total cholesterol in patients taking the new drug for 6 weeks as compared to participants taking placebo, if the variance population for 2 treatment are not equal (using $\alpha = 1\%$).

(11 marks)

Q4 A student is investigating the relationship between the price, y (RM) of 100g chocolate and the percentage, x (%) of cocoa solids in the chocolate. The detail data are shown as in the **Table Q4**. The SPSS output is shown as in **Appendix A.**

Table Q4

Chocolate brand	A	B	C	D	E	F	G
X	10	20	30	35	40	50	60
Y	0.5	0.7	1.2	1.5	1.8	2.5	3.6

(a) State 2 assumptions that need to be fulfilled before we can use the linear regression model.

(4 marks)

(b) Analyze the linear regression model by calculating the regression equation for data in **Table Q4** without referring the SPSS output. Draw a conclusion about the regression parameter. Hence validate your calculation with the SPSS output.

(12 marks)

(c) If the percentage of chocolate is 80%, predict the price of chocolate in RM.

(2 marks)

(d) Referring to the SPSS output, draw a conclusion about the Pearson and determination coefficient.

(4 marks)

(e) Based on ANOVA Table, draw a conclusion about the regression model obtained.

(3 marks)

- END OF QUESTIONS –

FINAL EXAMINATION

SEMESTER / SESSION: SEM I / 2019/2020 PROGRAMME CODE: BWD
COURSE NAME : STATISTICS FOR TECHNOLOGY COURSE CODE: BWD 21602

FORMULA

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 :

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

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

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 :

$$n = \left(\frac{Z_{\alpha/2} \cdot \sigma}{E}\right)^2, \quad \bar{x} - Z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}} < \mu < \bar{x} + Z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}}, \quad \bar{x} - t_{\alpha/2,v}\sqrt{\frac{s^2}{n}} < \mu < \bar{x} + t_{\alpha/2,v}\sqrt{\frac{s^2}{n}}$$

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

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

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

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

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 $v = n_1 + n_2 - 2$,

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

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

FINAL EXAMINATION

SEMESTER / SESSION: SEM I / 2019/2020 PROGRAMME CODE: BWD
COURSE NAME : STATISTICS FOR TECHNOLOGY COURSE CODE: BWD 21602

FORMULA

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

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

Hypothesis Testing :

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

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

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

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

Simple Linear Regressions :

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

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

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

FINAL EXAMINATION

SEMESTER / SESSION: SEM I / 2019/2020 PROGRAMME CODE: BWD
COURSE NAME : STATISTICS FOR TECHNOLOGY COURSE CODE: BWD 21602

FORMULA

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

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

Hypothesis Testing :

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

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

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

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

Simple Linear Regressions :

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

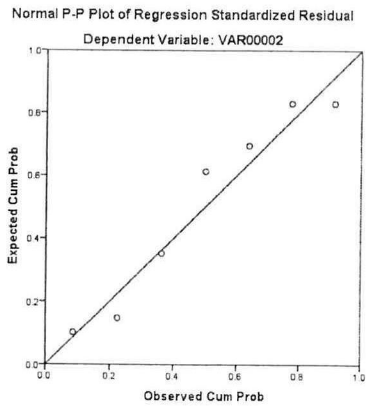
$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}},\ \hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x},\ \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x,\ r = \frac{S_{xy}}{\sqrt{S_{xx}\cdot S_{yy}}},\ SSE = S_{yy} - \hat{\beta}_1 S_{xy},\ MSE = \frac{SSE}{n-2},$$

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

FINAL EXAMINATION

SEMESTER / SESSION: SEM I / 2019/2020 PROGRAMME CODE: BWD
COURSE NAME : STATISTICS FOR TECHNOLOGY COURSE CODE: BWD 21602

Appendix A



Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.999[a]	.998	.997	.09008

a. Predictors: (Constant), VAR00001

b. Dependent Variable: VAR00002

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	18.928	1	18.928	2332.676	.000[b]
	Residual	.041	5	.008		
	Total	18.969	6			

a. Dependent Variable: VAR00002

b. Predictors: (Constant), VAR00001

Normal P-P Plot of Regression Standardized Residual

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	
		B	Std. Error	Beta			VIF
1	(Constant)	2.374	.083		28.710	.000	
	VAR00001	.104	.002	.999	48.298	.000	1.000

Expected Cum Prob