

CONFIDENTIAL



UNIVERSITI TUN HUSSEIN ONN MALAYSIA

**FINAL EXAMINATION
SEMESTER I
SESSION 2014/2015**

COURSE NAME : STATISTICAL PACKAGES
COURSE CODE : BWB 20703
PROGRAMME : 2 BWQ
EXAMINATION DATE : DECEMBER 2014/JANUARY 2015
DURATION : 3 HOURS
INSTRUCTION : ANSWER ALL QUESTIONS

THIS QUESTION PAPER CONSISTS OF **SIX (6)** PAGES

CONFIDENTIAL

- Q1 (a)** Three hundred and ninety two Kedah motorists were studied during 3 month periods in 2012. One group of 202 motorists receiving a placebo each day and the other 190 receiving one gram of Vitamin C per day as treatment group. The study was double blind - neither the subjects nor the researchers knew who received what treatment. Let p_1 be the probability that a member of the Vitamin C group contracts a hot during the study period and p_2 be the corresponding probability for the placebo group. The researchers are interested in testing whether $p_1 = p_2$. The data are summarised as a two by-two table of counts such as in the Table **Q1(a)**.

Table Q1(a) : Treatment Group

Outcome	Vitamin C	Placebo
With hot	23	45
With no hot	167	157
Total	190	202

- (i) Find the sample proportion of motorists developing hot in the placebo and treatment groups. (4 marks)
- (ii) Calculate the pooled proportion of the developed hot in the study. (2 marks)
- (iii) State and prove the hypothesis that the probability of contracting a hot is the different if given a placebo or Vitamin C at 1% level of significance. (10 marks)
- (iv) Find the 98% confidence interval of $p_1 - p_2$ and state the appropriate conclusion. (5 marks)
- (b) The Minitab output (Table **Q1(b)**) of the cervical dysplasia data are as follows:

Table Q1(b) : The Cervical Dysplasia

Test and CI for Two Proportions			
Sample	X	N	Sample p
1	164	175	0.937143
2	130	308	0.422078
Difference = p (1) - p (2)			
Estimate for difference: 0.515065			
95% CI for difference: (0.449221, 0.580909)			
Test for difference = 0 (vs not = 0): Z = 11.15			
P-Value = 0.000			

- (i) Find the sample proportion for Sample 1 and Sample 2. (2 marks)
- (ii) State the hypothesis based on the output given above. (2 marks)

- Q2** (a) The Table **Q2(a)** is a linear regression output which has been covered up with a number of blanks, from **(i)** to **(x)**. Fill in these blanks.

Table Q2(a) : Linear Regression

Predictor	Coef	SE Coef	T	P
Constant	3571.8	158.567	(i)	0.0000
Acid	0.13099	(ii)	1.35	0.1801
Buffer	-0.56833	(iii)	-12.45	0.0000

Analysis of Variance				
Source	DF	SS	MS	F
Regression	(iv)	2452000	1226000	(v)
Residual Error	(vi)	(vii)	13028.0	
Total	107	3820000		

$S_e = \mathbf{(viii)}$ $R\text{-sq} = \mathbf{(ix)}$ $R\text{-sq(adj)} = \mathbf{(x)}$

(20 marks)

- (b) Please answer the following questions base on the **R** result given in the Table **Q2(b)**.

Table Q2(b) : Linear Model

```
lm(formula = divorce ~ year + unemployed + femlab
+ marriage + birth + military, data = divusa)

Residuals:
Min      1Q  Median      3Q      Max
-2.9087 -0.9212 -0.0935  0.7447  3.4689

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  380.14761    99.20371   3.832 0.000274 ***
year         -0.20312     0.05333  -3.809 0.000297 ***
unemployed   -0.04933     0.05378  -0.917 0.362171
femlab        0.80793     0.11487   7.033 1.09e-09 ***
marriage     0.14977     0.02382   6.287 2.42e-08 ***
birth       -0.11695     0.01470  -7.957 2.19e-11 ***
military    -0.04276     0.01372  -3.117 0.002652 **

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Analysis of Variance Table
```

```

Response: divorce
      Df Sum Sq Mean Sq F value    Pr(>F)
year   1 1888.22  1888.22  825.0759 < 2.2e-16 ***
unemployed 1    0.05    0.05   0.0223  0.881843
femlab   1  169.40  169.40  74.0231 1.413e-12 ***
marriage 1   57.12   57.12  24.9587 4.141e-06 ***
birth    1  145.31  145.31  63.4934 2.090e-11 ***
military 1   22.23   22.23   9.7142 0.002652 **
Residuals 70  160.20    2.29

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
    
```

- (i) What is the response of the fitted regression? (1 mark)
- (ii) Which is the explanatory variable should be remove from model? (1 mark)
- (iii) What is the sample size for this fitted regression? (1 mark)
- (iv) Give the value of R^2 . (2 marks)

Q3 In multiple regression, it is useful to examine the standard deviation ratio S_e/S_y , where S_y is the standard deviation of the independent variable and S_e is the standard error of regression. There is a corrected, or adjusted, form of the R^2 statistic, called R^2_{adj} which uses S_e/S_y . This adjusted R^2 is defined as $R^2_{adj} = 1 - S_e^2/S_y^2$. Now consider the following regression analysis of variance table in the Table Q3.

Table Q3 : Regression Analysis of Variance

Source	Degrees of freedom	Sum Squares	Mean Squares	F
Regression	5	150,000	30,000	30
Residual	70	70,000	1,000	
Total	75	220,000		

- (a) How many independent variables were used in this regression? List all. (2 marks)
- (b) What is the sample size? (2 marks)

- (c) Find the value of S_e . (3 marks)
- (d) Find the value of S_y . (4 marks)
- (e) Give the value of R^2 . Interpret. (5 marks)
- (f) Give the value of R^2_{adj} . (4 marks)
- (g) Would these regression results be considered significant at the 0.05 level? (5 marks)

- Q4** (a) Blueberry ice cream is made at Crema Ice Cream factory by squirting high quality blueberry syrup into an ice cream base. The resulting mixture is then stirred and frozen. Because the blueberry syrup is sticky, the dispenser cannot always squirt out the same quantity. The management is concerned about the concentration of blueberry syrup in the ice cream. A sample of 24 half gallon containers was taken and the blueberry syrup content of each was measured. The average quantity in this sample was 41.6 ounces, with a standard deviation of 3.5 ounces.
- (i) Find a 95% confidence interval for the mean amount of blueberry syrup in a half-gallon container and state the conclusion. (6 marks)
- (ii) Find a 98% confidence interval for the mean amount of blueberry syrup in a half-gallon container and state the conclusion. (6 marks)
- (b) The Test 2 result of Statistical Packages module for the second year student divide by group, known as Section 1 and Section 2 are normally distributed with $N(59, 3^2)$ and $N(62, 4^2)$ respectively. Two samples of size ten and fifteen are randomly selected from Section 1 and Section 2 respectively.
- (i) What is the different of sample mean between Section 2 and Section 1? (2 marks)

- (ii) Calculate the standard deviation of sample mean between Section 2 and Section 1.
(3 marks)
- (iii) State the distribution form for the different between two means in **Q4(b)(i)**.
(3 marks)
- (iv) Find the probability that the mean of Section 2 is less than the mean of Section 1.
(5 marks)

- END OF QUESTION -