FACULTY OF INDUSTRIAL SCIENCES & TECHNOLOGY
FINAL EXAMINATION

COURSE	**:**	**APPLIED STATISTICS**
COURSE CODE	**:**	**BUM2413**
LECTURER	**:**	**MOHD KHAIRUL BAZLI BIN MOHD AZIZ** **KU MUHAMMAD NA'IM BIN KU KHALIF** **NORATIKAH BINTI ABU** **NORYANTI BINTI MUHAMMAD** **NUR ZAHIRAH BINTI MD NOOR** **ROSLINAZAIRIMAH BINTI ZAKARIA** **SITI ROSLINDAR BINTI YAZIZ** **SITI ZANARIAH BINTI SATARI** **WAN NUR SYAHIDAH BINTI WAN YUSOFF**
DATE	**:**	**3 JANUARY 2018**
DURATION	**:**	**3 HOURS**
SESSION/SEMESTER	**:**	**SESSION 2018/2019 SEMESTER I**
PROGRAMME CODE	**:**	**BAA/BCG/BCN/BCS/BEE/BEP/BFF/BFM/BHA/ BHM/BKC/BMA/BMM/BPN/BPP/BPS/BPT/ BSB/BSK/BSP/BTC/BTE/BTF/BTK/BTM/BTP/ BTV**

INSTRUCTIONS TO CANDIDATES:

1. This question paper consists of **SEVEN (7)** questions. Answer **ALL** questions.
2. All answers to a new question should start on new page.
3. All the calculations and assumptions must be clearly stated.
4. All calculations must be in **FOUR (4) decimal places**.

EXAMINATION REQUIREMENTS:

1. Statistical Tables and Formulae

DO NOT TURN THIS PAGE UNTIL YOU ARE TOLD TO DO SO

This examination paper consists of **THIRTEEN (13)** printed pages including front page.

QUESTION 1

ACME Co. is studying the effects of artificial lighting on growth of chickens. In a laboratory test, 37 chicks were separated at random into two groups, namely Group 1 and Group 2, which both raised in windowless coops with the artificial lighting. Group 1 and Group 2 are placed in a 20-hour and a 24-hour per day artificial lighting cycles, respectively. The weekly growth of chicks is assessed by weight and the data is summarised in **Table 1**.

Table 1: Summarised Data for Group 1 (20-hour) and Group 2 (24-hour)

Artificial lighting cycle	Number of chicks	Mean (in grams)	Standard deviation (in grams)
20-hour	20	62	2.5
24-hour	17	51	1.8

Assume that the data for both groups are normally distributed with the same population variances.

(i) Identify the variable in the study.

(1 Mark)

(ii) Apply the confidence interval approach to validate the assumption of population variances at 90% confidence level.

(9 Marks)

(iii) At 10% level of significance, test whether the mean weight of chicks is higher with the 20-hour per day as compared to a 24-hour per day of artificial lighting cycles.

(9 Marks)

QUESTION 2

An investor wants to compare the risks associated with two different stocks, Stock X and Stock Y. One way to measure the risk of a given stock is to measure the variation in the stock's daily price changes. The investor obtains a random sample of 20 daily price changes from each stock, respectively. **Figure 1** presents the *Microsoft Excel* output for the related statistical test at 5% significance level.

F-Test Two-Sample for Variances

	Price Change for Stock X	*Price Change for Stock Y*
Mean	0.327	0.1105
Variance	0.720306316	0.279983947
Observations	20	20
df	19	19
F	2.57267005	
P(F<=f) one-tail	0.022911954	
F Critical one-tail	2.168251601	

Figure 1: *Microsoft Excel* Output

(i) What is the appropriate parameter that the investor should use in this study?

(1 Mark)

(ii) Conduct a necessary test in comparing either Stock X or Stock Y will provide the minimum risk using the *P*-value approach.

(5 Marks)

(iii) Based on your answer in **(ii)**, which stock should the investor choose for investment?

(1 Mark)

(iv) In the previous record, Stock X has lower risk as compared to Stock Y. Based on your answer in **(ii)**, identify the type of error involved and give a reason.

(3 Marks)

QUESTION 3

A research study was conducted to examine the impact of eating breakfast with high and low protein on teenager's performance during a physical fitness test. Half of the teenagers received a high protein breakfast and half were given a low protein breakfast. All teenagers, both male and female, were given a fitness test with high scores representing better performance. The physical fitness test scores are recorded in **Table 2**. Based on the assumption that the physical fitness test scores are normally distributed, the analysis of variance test is conducted on the data. The corresponding *Microsoft Excel* output is presented in **Figure 2**.

Table 2: Physical Fitness Test Score

Gender	Level of Protein	
	High	Low
Male	10	5
	7	4
	9	7
	6	4
	8	5
Female	5	3
	4	4
	6	5
	3	1
	2	2

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	45	1	45	20.0000	0.0004	4.4940
Columns	20	1	20	8.8889	0.0088	4.4940
Interaction	5	1	5	2.2222	0.1555	4.4940
Within	36	16	2.25			
Total	106	19				

Figure 2: *Microsoft Excel* Output of ANOVA Table

Justify that there is no interaction effect between teenager's gender and level of protein consumed. Hence, proceed with necessary test.

(15 Marks)

QUESTION 4

(a) Give **ONE (1)** objective of linear regression model?

(1 Mark)

(b) A production supervisor in a cooking oil company believes that the yield (in liter) of peanut oil extracted can be improved if the pressure (in bars) is properly applied during the extraction process. Various pressures are tested and the yield of peanut oil extracted is recorded in **Table 4**.

Table 4: Pressure and Yield of Peanut Oil

	1	2	3	4	5	6	7	8
Pressure	415	550	415	550	415	550	415	550
Yield	63	21	36	99	24	66	71	54

	9	10	11	12	13	14	15	16
Pressure	415	550	415	550	415	550	415	550
Yield	23	74	80	33	63	21	44	96

Given that

$$\sum x = 7720,\ \sum x^2 = 3797800,\ \sum y = 868,\ \sum y^2 = 57452,\ \sum xy = 422860$$

$$\bar{x} = 482.5,\ \bar{y} = 54.25,\ \hat{y} = 27.4230 + 0.0556x,\ se\left(\hat{\beta}_1\right) = 0.0997$$

Based on the given information, answer the followings.

(i) Identify the independent and dependent variable(s).

(1 Mark)

(ii) If 50% of the average pressure is applied during the extraction process, estimate the yield of peanut oil.

(3 Marks)

(iii) Test whether there is a significant relationship between the yield of peanut oil and pressure applied at 2% significance level by using *t*-test approach.

(7 Marks)

QUESTION 5

The Chief Executive Officer (CEO) of the same cooking oil company in **Question 4** wants the research department to conduct further investigation to improve the yield of peanut oil extracted. Based on the previous study, two more variables are considered which are temperature (in ^{0}C) and particle size (in mm) of peanut. The data is shown in **Table 5**.

Table 5: Pressure, Temperature, Particle Size and Yield of Peanut Oil

Pressure	Temperature	Particle size	Yield
415	25	1.28	63
550	25	4.05	21
415	95	4.05	36
550	95	1.28	99
415	25	4.05	24
550	25	1.28	66
415	95	1.28	71
550	95	4.05	54
415	25	4.05	23
550	25	1.28	74
415	95	1.28	80
550	95	4.05	33
415	25	1.28	63
550	25	4.05	21
415	95	4.05	44
550	95	1.28	96

The multiple regression analysis is conducted using *Microsoft Excel* and the outputs are given by **Figure 3 – Figure 9**. Let $x_1 =$ pressure, $x_2 =$ temperature and $x_3 =$ particle size.

Regression Statistics	
Multiple R	0.1473
R Square	0.0217
Adjusted R Square	-0.0482
Standard Error	26.9099
Observations	16

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	225	225	0.310712	0.586046084
Residual	14	10138	724.1429		
Total	15	10363			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	27.4230	48.5573	0.5652	0.5809	-76.7006	131.5895
Pressure (Bars)	0.0556	0.0997	0.5574	0.5860	-0.1582	0.2693

Figure 3: *Microsoft Excel* Outputs for Intercept and x_1

Regression Statistics	
Multiple R	0.3880
R Square	0.1506
Adjusted R Square	0.0899
Standard Error	25.0752
Observations	16

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	1560.25	1560.25	2.48144	0.137518436
Residual	14	8802.75	628.7679		
Total	15	10363			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	37.3214	12.4413	2.9998	0.0096	10.6375	64.0054
Temperature (Celcius)	0.2821	0.1791	1.5753	0.1375	-0.1020	0.6663

Figure 4: *Microsoft Excel* Outputs for Intercept and x_2

Regression Statistics	
Multiple R	0.8743
R Square	0.7644
Adjusted R Square	0.7475
Standard Error	13.2071
Observations	16

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	7921	7921	45.41114	9.48984E-06
Residual	14	2442	174.4286		
Total	15	10363			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	97.0632	7.1600	13.5563	0.0000	81.7065	112.4199
Particle size (mm)	-16.0650	2.3840	-6.7388	0.0000	-21.1781	-10.9519

Figure 5: *Microsoft Excel* Outputs for Intercept and x_3

Regression Statistics	
Multiple R	0.4151
R Square	0.1723
Adjusted R Square	0.0449
Standard Error	25.6871
Observations	16

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	1785.25	892.625	1.352817	0.29259794
Residual	13	8577.75	659.8269		
Total	15	10363			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	10.5159	47.6402	0.2207	0.8287	-92.4046	113.4363
Pressure (Bars)	0.0556	0.0951	0.5840	0.5692	-0.1500	0.2611
Temperature (Celcius)	0.2821	0.1835	1.5377	0.1481	-0.1142	0.6785

Figure 6: *Microsoft Excel* Outputs for Intercept, x_1 and x_2

Regression Statistics	
Multiple R	0.8866035
R Square	0.7860658
Adjusted R Square	0.7531529
Standard Error	13.059038
Observations	16

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	8146	4073	23.88318	4.4343E-05
Residual	13	2217	170.5385		
Total	15	10363			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	70.2576	24.3872	2.8809	0.0129	17.5722	122.9430
Pressure (Bars)	0.0556	0.0484	1.1486	0.2714	-0.0489	0.1600
Particle size (mm)	-16.0650	2.3572	-6.8152	0.0000	-21.1575	-10.9725

Figure 7: *Microsoft Excel* Outputs for Intercept, x_1 and x_3

Regression Statistics	
Multiple R	0.9565112
R Square	0.9149136
Adjusted R Square	0.9018234
Standard Error	8.2357102
Observations	16

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	9481.25	4740.625	69.89297	1.1069E-07
Residual	13	881.75	67.82692		
Total	15	10363			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	80.1346	5.6915	14.0798	0.0000	67.8389	92.4303
Temperature (Celcius)	0.2821	0.0588	4.7962	0.0003	0.1551	0.4092
Particle size (mm)	-16.0650	1.4866	-10.8066	0.0000	-19.2766	-12.8534

Figure 8: *Microsoft Excel* Outputs for Intercept, x_2 and x_3

Regression Statistics	
Multiple R	0.9678
R Square	0.9366
Adjusted R Square	0.9208
Standard Error	7.3979
Observations	16

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	9706.25	3235.416667	59.11686334	1.84761E-07
Residual	12	656.75	54.72916667		
Total	15	10363			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	53.3290	14.1745	3.7623	0.0027	22.4456	84.2125
Pressure (Bars)	0.0556	0.0274	2.0276	0.0654	-0.0041	0.1153
Temperature (Celcius)	0.2821	0.0528	5.3393	0.0002	0.1670	0.3973
Particle size (mm)	-16.0650	1.3354	-12.0304	0.0000	-18.9745	-13.1555

Figure 9: *Microsoft Excel* Outputs for Intercept, x_1, x_2 and x_3

(i) If all the three predictors are included in the analysis, state the coefficient of determination value. Interpret the value.

(2 Marks)

(ii) Based on the regression model in **Figure 9**, interpret the coefficient of x_3.

(2 Marks)

(iii) Based on **Figure 7**, test the hypothesis that at least one of the independent variables is related to the dependent variable.

(5 Marks)

(iv) Using the *Microsoft Excel* outputs provided, copy and complete the table.

Predictor variable	P-value	r^2	Adjusted r^2	Regression Model

Hence, suggest the best model which has the most significant variables and justify your answer.

(9 Marks)

(v) Based on your answer in **(iv)**, predict the yield of peanut oil if the pressure is 400 bars, temperature is 60^0C and particle size is 2.28 mm.

(2 Marks)

QUESTION 6

A principal of a boarding school in Kuantan is planning a field trip for his students. He samples a group of 100 students to see if they prefer sports event, activity at a local college, science museum visit or play at amusements parks. He records the following data in **Table 6**.

Table 6: Students Preference on Field Trip

Type of field trip	Number of students
Sports event	35
Activity at a local college	5
Science museum visit	10
Play at amusements parks	50

(i) What can you observe from the preference of the field trip in **Table 6**?

(1 Mark)

(ii) Test the hypothesis at 2.5% significance level if there is a preference on the type of field trip.

(10 Marks)

(iii) Give **ONE (1)** characteristic that demonstrate the data is a good fit to a model.

(1 Mark)

(iv) Based on your answer in **(ii)**, is your data is a good fit?

(1 Mark)

QUESTION 7

Aminah is a sales advisor from Toyota and she is interested to investigate type of Toyota's car distributions in Malaysia. She wants to know whether families and singles in Malaysia have the same car distribution. Then, she randomly selects 100 families and 200 singles and asked what type of Toyota's car that they drive: sport, sedan, hatchback, truck, van/SUV. The results are shown in **Table 7**.

Table 7: Type of Toyota's Car Distributions

	Sport	Sedan	Hatchback	Truck	Van/SUV
Family	5	15	35	17	28
Single	45	65	37	46	7

Do families and singles have the same distribution of Toyota's car? Test at 0.5% level of significance.

(11 Marks)

END OF QUESTION PAPER