NATIONAL UNIVERSITY OF SINGAPORE

BT1101 – INTRODUCTION TO BUSINESS ANALYTICS

(Semester 1: AY2017/18)

Time Allowed: 2 Hours

INSTRUCTIONS TO STUDENTS

1. Please write your Student Number only. Do not write your name.
2. This assessment paper contains **THIRTY** Multiple Choice Questions and **FIVE** Structured Questions, and comprises **Twenty** printed pages including the cover page. The total mark is 90.
3. Students are required to answer **ALL** questions. Students should use the **OCR Form** for Multiple Choice Questions, and write the answers for each Structured Question in the space provided below each question.
4. This is a **CLOSED BOOK** assessment.
5. Students are provided with statistical tables for reference.
6. Students are permitted to use approved non-programmable electronic calculators only.

STUDENT NO: _____

This portion is for examiner's use only

Section B	Marks	Remarks
Q1		
Q2		
Q3		
Q4		
Q5		
Total		

Section B: Structured Questions (Total 60 marks, marks as indicated for each question)

1. An online apparel store captures clickstream data of all its online shoppers (including the pages they view, the items they click on, what they put in the shopping cart, what they eventually buy or delete from their shopping cart, and the time stamps of all these activities). Provide two examples of how the online apparel store selling apparel can provide value to its customers using analytics. For each example, state the type of analytics used (e.g. descriptive/predictive/prescriptive), describe the data/metrics used and the customer value that is generated. Please state any assumptions you make on any other data that the online apparel store may be capturing (e.g. customer data). (8 marks)

Answers:

Example 1: Type of Analytic:

Data/Metrics Used:

Customer Value:

Example 2: Type of Analytic:

Data/Metrics Used:

Customer Value:

2(a) A soft drink bottle filling machine is known to have a mean of 200ml and a standard variation of 10ml. The quality control manager took a random sample of the filled bottles and found the sample mean to be 215ml. She assumed the sample must not be representative. Do you agree with the conclusion made by the quality control manager? Justify your answer. (4 marks)

Answers:

(b) "Covariance and Correlation measures the same thing" Do you agree with this statement? Explain why. (3 marks)

Answers:

3. The Cell Phone Survey database contains 52 records. The first 10 records of the database are displayed below.

	Gender	Carrier	Type	Usage	Signal strength	Value for the Dollar	Customer Service
1	M	AT&T	Smart	High	5	4	4
2	M	AT&T	Smart	High	5	4	2
3	M	AT&T	Smart	Average	4	4	4
4	M	AT&T	Smart	Very high	2	3	3
5	M	AT&T	Smart	Very high	5	5	2
6	M	AT&T	Smart	Very high	4	3	5
7	M	AT&T	Smart	Very high	3	4	4
8	F	AT&T	Smart	Very high	3	2	3
9	F	AT&T	Smart	Very high	4	3	4
10	M	AT&T	Smart	Very high	3	3	1

Bernard would like to determine if the mean response for Value for the Dollar is the same for different types of cell phones ("Smart", "Camera", "Basic"). The following is the R codes and results he obtained.

```
> View(Cell_Phone_Survey)
>  df1<-Cell_Phone_Survey
> your.aov=aov(df1$`Value for the Dollar`~df1$Type)
> your.aov
Call:
   aov(formula = df1$`Value for the Dollar` ~ df1$Type)

Terms:
                df1$Type Residuals
Sum of Squares   5.26123  41.43108
Deg. of Freedom        2        49

Residual standard error: 0.9195282
Estimated effects may be unbalanced
```

```
> summary(your.aov)
             Df Sum Sq Mean Sq F value Pr(>F)
df1$Type      2   5.26  2.6306   3.111 0.0535 .
Residuals    49  41.43  0.8455
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(a) State the hypotheses that Bernard is testing. (2 mark)

Answers:

(b) What conclusions can Bernard draw from his analysis. Explain your answer with respect to the type of analyses he conducted and the results he obtained. (4 marks)

Answers:

(c) Based on the summary statistics of customer services, construct a 95% confidence interval for customer services. Round off your answers to 2 decimal places. (4 marks)

```
> summary(df1$`Customer Service`)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.000   3.000   3.000   3.231   4.000   5.000
> sd(df1$`Customer Service`)
[1] 0.9623375
> var(df1$`Customer Service`)
[1] 0.9260935
```

Answers:

(d) If Bernard would like to know if there is any significant difference in Customer Service ratings between Males and Females, what is the analysis that he should perform in R and how should he draw his conclusion based on the results? (5 marks)

Answers:

4. Singapore Sugar produces standard grain sugar from sugar cane. The factory produces two different sizes of sugar grain from the same raw material. Regular Grain can be produced at a rate of 5,000 kg per hour, and has a demand of 400 tons per week with a price of $900 per ton. Fine Grain can be produced at a rate of 3,000 kg per hour, and has a demand of 200 tons per week with a price of $2K per ton. A minimum of 800 tons of sugar cane has to be processed every week to clear the storage area for incoming raw materials. The factory operates 24 hours non-stop 7 days a week. (Note: 1 ton = 1000 kilograms)

(a) Develop and solve a linear optimisation model to determine the number of tons of each product to produce each week to maximise revenue. (5 marks)

Answers:

(b) What will be the impact if the minimum amount of sugar cane to be processed every week is increased? (3 marks)

Answers:

(c) If the price per ton for Regular Grain is increased to $1100, how will the model be affected? (3 marks)

Answers:

(d) Due to decreasing demand, the price per ton for Fine Grain is decreased to $1400. How will the model be affected? (4 marks)

Answers:

5. The following table shows a data sample from a recent survey at an automobile fair:

(c) (d)

Gender	Age	Driving Yrs	Type	Purchased	Mileage	Vehicle Age	MPG	Purchased C	Interaction
Male	32	12	Mid-size	Used	127233	15	28.7		
Female	34	2	Mid-size	New	23970	1	43.4		
Male	45	21	Small	New	77392	7	24		
Female	54	25	Large SUV	Used	185397	14	15.2		
Female	44	23	Small	New	26001	2	37		
Female	23	2	Minivan	New	180643	9	20		
Male	32	4	Small	Used	72083	6	45.7		
Male	19	1	Small	New	165353	11	42		
Male	54	24	Small	Used	205288	13	33		
Female	46	6	Small	New	142897	7	31		
Male	65	40	Minivan	Used	182584	14	12		
Male	38	13	Small SUV	Used	140479	13	20		
Female	26	5	Small	New	22114	2	28		
Female	59	30	Mid-size	New	3454	0.25	28.3		
Female	39	11	Large SUV	New	130905	7	21		
Female	23	3	Small	Used	105628	10	35		
Female	62	38	Small	New	48678	5	30.4		
Male	75	32	Mid-size	New	6849	0.5	40.2		
Female	53	18	Small	Used	137941	10	30		
Female	22	1	Small SUV	New	29823	4	24.9		
Male	53	28	Small SUV	Used	85763	14	21		
Female	47	25	Small	Used	134172	12	31		
Male	37	10	Mid-size	Used	86387	12	27		

(a) Vehicle Age has been thought to affect miles per gallon (MPG). Provide the model, conclusions and recommendations for further analysis from the output below. (3 marks)

```
> auto.a <- read.xlsx("mysurvey.xlsx", sheetIndex = 1, startRow = 1, header = TRUE)
> fit.auto.a <- lm(auto.a$MPG ~ auto.a$Vehicle.Age)
> summary(fit.auto.a)


Call:
lm(formula = auto.a$MPG ~ auto.a$Vehicle.Age)
```

```
Residuals:
     Min       1Q   Median       3Q      Max
-12.1393  -6.9653   0.8932   5.2844  15.3032

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)        36.0743     3.3129  10.889 4.27e-10 ***
auto.a$Vehicle.Age -0.8525     0.3488  -2.444   0.0235 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.996 on 21 degrees of freedom
Multiple R-squared:  0.2214, Adjusted R-squared:  0.1843
F-statistic: 5.972 on 1 and 21 DF,  p-value: 0.02345
```

Answers:

(b) To get a regression model to predict MPG as a function of vehicle age and mileage, the following results were generated. Provide conclusions and recommendations for further analysis from the model below. (3 marks)

```
> fit.auto.a <- lm(auto.a$MPG ~ auto.a$Vehicle.Age + auto.a$X.Mileage)
> summary(fit.auto.a)

Call:
lm(formula = auto.a$MPG ~ auto.a$Vehicle.Age + auto.a$X.Mileage)

Residuals:
    Min      1Q  Median      3Q     Max
-11.980  -7.066   0.712   5.069  15.582

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)         3.618e+01  3.453e+00  10.479 1.43e-09 ***
auto.a$Vehicle.Age -7.664e-01  6.306e-01  -1.215    0.238
auto.a$X.Mileage   -8.057e-06  4.863e-05  -0.166    0.870
---
```

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

Residual standard error: 8.188 on 20 degrees of freedom
Multiple R-squared:  0.2225, Adjusted R-squared:  0.1447
F-statistic: 2.861 on 2 and 20 DF,  p-value: 0.08075
```

Answers:

(c) To get a better regression model to predict miles per gallon (MPG), a categorical variable for Purchased was added to the model in part (b). Fill in the table under Purchased C, and provide conclusions and recommendations for further analysis from the model below. (4 marks)

```
> fit.auto.a <- lm(auto.a$MPG ~ auto.a$Vehicle.Age + auto.a$X.Mileage +
auto.a$Purchased.C)
> summary(fit.auto.a)

Call:
lm(formula = auto.a$MPG ~ auto.a$Vehicle.Age + auto.a$X.Mileage +
    auto.a$Purchased.C)

Residuals:
     Min       1Q   Median       3Q      Max
-12.9138  -5.9886   0.5966   4.2783  19.9841

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)         4.486e+01  7.503e+00   5.978 9.41e-06 ***
auto.a$Vehicle.Age -1.682e+00  9.399e-01  -1.790   0.0895 .
auto.a$X.Mileage    1.975e-05  5.241e-05   0.377   0.7105
auto.a$Purchased.C -7.604e+00  5.865e+00  -1.296   0.2104
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.052 on 19 degrees of freedom
```

```
Multiple R-squared:  0.2857, Adjusted R-squared:  0.1729
F-statistic: 2.533 on 3 and 19 DF,  p-value: 0.08765
```

Answers:

(d) To determine whether any significant interaction exists between Vehicle Age and Purchased, fill in the table under Interaction, and provide conclusions and recommendations for further analysis from the model below. (3 marks)

```
> fit.auto.a <- lm(auto.a$MPG ~ auto.a$Vehicle.Age + auto.a$Interaction)
> summary(fit.auto.a)

Call:
lm(formula = auto.a$MPG ~ auto.a$Vehicle.Age + auto.a$Interaction)

Residuals:
     Min       1Q   Median       3Q      Max
-12.4474  -6.9056   0.8093   4.9668  16.8393

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)         36.6505     3.7862   9.680 5.45e-09 ***
auto.a$Vehicle.Age  -0.8717     0.3609  -2.415   0.0254 *
auto.a$Interaction  -0.1729     0.5089  -0.340   0.7376
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.17 on 20 degrees of freedom
Multiple R-squared:  0.2259, Adjusted R-squared:  0.1485
F-statistic: 2.918 on 2 and 20 DF,  p-value: 0.07728
```

Answers:

(e) From the above iterations in parts (a) to (d), give conclusions on the best model for predicting MPG. (2 marks)

Answers:

END OF PAPER