NATIONAL UNIVERSITY OF SINGAPORE

BT1101 – INTRODUCTION TO BUSINESS ANALYTICS

(Semester 2: 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 **TWENTY** Multiple Choice Questions and **SIX** Structured Questions, and comprises **Twenty-one** 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. Students are allowed to bring only a single sheet of A4 help-sheet for reference.
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
Question 1		
Question 2		
Question 3		
Question 4		
Question 5		
Question 6		
Total		

---- page left blank-----

Section B: Structured Questions (Total 70 marks, marks as indicated for each question)

Question 1 (6 marks)

(a) A financial advisor believes that the proportion of investors who are risk-averse (i.e. try to avoid risk in their investment decisions) is at least 0.7. A survey of 32 investors found that 20 of them were risk-averse. Formulate and test the hypotheses to determine whether his belief is valid. (3 marks)

(b) A management institute checked the past records of applicants and the mean score calculated was 350. The administration is interested to know whether the quality of new applicants has changed or not. From the recent scores of 100 applicants, the mean is 365 with a standard deviation of 38. Does this data provide statistical evidence that the quality of recent applicants has improved? (3 marks)

Question 2 (4 marks)

"Data used in business analytics need to be reliable and valid."

a) Explain what "reliable" and "valid" data mean (providing appropriate examples in your explanation). (2 marks)

b) Explain whether you agree with the above statement. (2 marks)

Question 3 (25 marks)

The AGym database contains 23 records. Fig 3.1 below shows the first 10 records of the data.

Fig3:1

	Gender	Body Type	BMI Classification	BMI Calculation	Time Spent in Gym	Pant size (inches)	Weight	Height (inches)	Weight Lift (Days)	Lifiting Session (Mins)	Running Times (Hours)	Distance of Run (Miles)
1	F	Thin	Obese	18.53613	0.00	25	108	64	0	0	0.6666667	3.0
2	M	Muscular	Obese	23.40106	198.00	30	145	66	3	60	0.5000000	3.0
3	M	Athletic	Obese	21.78719	198.00	30	135	66	3	60	0.5000000	4.0
4	F	Athletic	Obese	27.46094	128.00	32	160	64	2	30	0.6666667	4.0
5	M	Muscular	Muscular	28.29234	363.75	33	213	73	5	150	0.1166667	1.0
6	F	Athletic	Obese	21.15990	305.00	23	112	61	5	90	0.3333333	2.0
7	M	Average	Obese	30.41756	207.00	35	206	69	3	60	0.0000000	0.0
8	M	Average	Obese	24.21592	345.00	31	164	69	5	75	0.0000000	0.0
9	F	Average	Obese	21.91561	0.00	27	116	61	0	0	0.0000000	0.0
10	M	Muscular	Muscular	33.46955	71.00	36	240	71	1	60	0.5000000	1.0

Tom, the gym manager, conducted a series of descriptive analytics. Below are the R scripts and respective results obtained.

Fig3:2

```
> df2<- AGym
> shapiro.test(df2$`BMI Calculation`)

        Shapiro-Wilk normality test

data:  df2$`BMI Calculation`
W = 0.95144, p-value = 0.3133

> shapiro.test(df2$`Time Spent in Gym`)

        Shapiro-Wilk normality test

data:  df2$`Time Spent in Gym`
W = 0.93451, p-value = 0.1369

> shapiro.test(df2$`Weight Lift (Days)`)

        Shapiro-Wilk normality test

data:  df2$`Weight Lift (Days)`
W = 0.91251, p-value = 0.04615

> shapiro.test(df2$`Lifiting Session (Mins)`)

        Shapiro-Wilk normality test

data:  df2$`Lifiting Session (Mins)`
W = 0.90694, p-value = 0.03522

> shapiro.test(df2$`Running Times (Hours)`)

        Shapiro-Wilk normality test

data:  df2$`Running Times (Hours)`
W = 0.87275, p-value = 0.007237

> shapiro.test(df2$`Distance of Run (Miles)`)

        Shapiro-Wilk normality test

data:  df2$`Distance of Run (Miles)`
W = 0.92449, p-value = 0.08319
```

Fig3:3

```
> hist(df2$`BMI Calculation`)
> hist(df2$`Time Spent in Gym`)
> hist(df2$`Weight Lift (Days)`)
> hist(df2$`Running Times (Hours)`)
> hist(df2$`Distance of Run (Miles)`)
> hist(df2$`Lifiting Session (Mins)`)
```

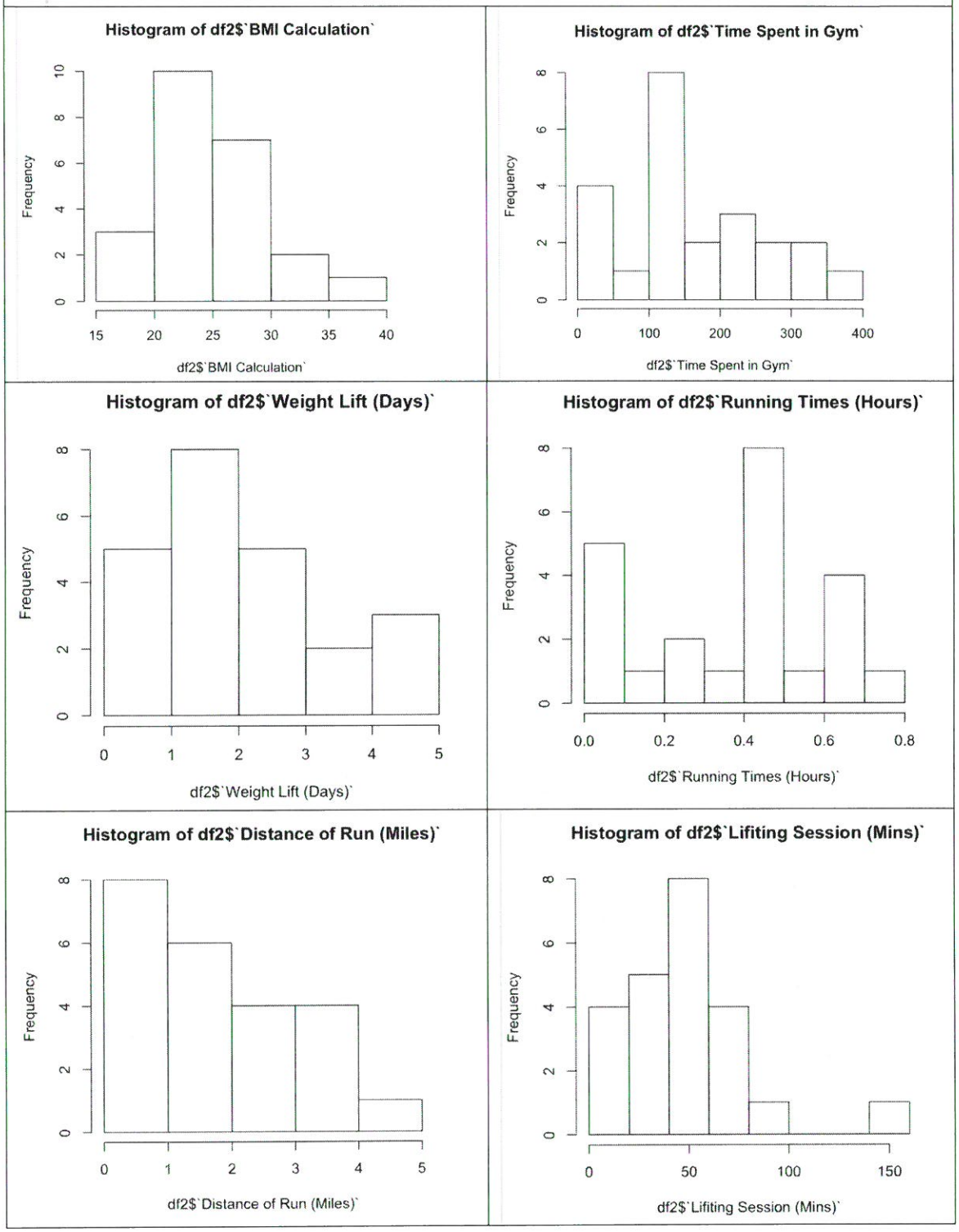

Fig3:4

```
> describe(df2)
                          vars  n   mean     sd median trimmed    mad   min    max  range  skew kurtosis    se
Gender*                      1 23    NaN     NA     NA     NaN     NA   Inf   -Inf   -Inf    NA       NA    NA
Body Type*                   2 23    NaN     NA     NA     NaN     NA   Inf   -Inf   -Inf    NA       NA    NA
BMI Classification*          3 23    NaN     NA     NA     NaN     NA   Inf   -Inf   -Inf    NA       NA    NA
BMI Calculation              4 23  24.89   4.48  24.13   24.55   4.08  18.3  35.73  17.42  0.67    -0.25  0.93
Time Spent in Gym            5 23 162.29 107.80 136.00  159.16 100.82   0.0 363.75 363.75  0.12    -0.88 22.48
Pant size (inches)           6 23  30.78   3.84  31.00   30.79   2.97  23.0  39.00  16.00 -0.14    -0.48  0.80
Weight                       7 23 162.87  40.20 160.00  160.95  37.06 105.0 240.00 135.00  0.28    -1.02  8.38
Height (inches)              8 23  67.43   3.70  68.00   67.42   2.97  61.0  75.00  14.00 -0.15    -0.65  0.77
Weight Lift (Days)           9 23   2.39   1.56   2.00    2.37   1.48   0.0   5.00   5.00  0.06    -0.88  0.33
Lifiting Session (Mins)     10 23  49.57  34.87  60.00   47.37  22.24   0.0 150.00 150.00  0.64     0.86  7.27
Running Times (Hours)       11 23   0.39   0.26   0.50    0.39   0.25   0.0   0.75   0.75 -0.41    -1.35  0.05
Distance of Run (Miles)     12 23   2.07   1.54   2.00    2.03   1.48   0.0   5.00   5.00  0.12    -1.23  0.32
```

Fig3:5

```
> describeBy(df2$`BMI Calculation`,group=df2$Gender)

 Descriptive statistics by group
group: F
   vars  n  mean   sd median trimmed  mad  min   max range skew kurtosis   se
X1    1 10 22.21 3.11  21.65   22.04 3.18 18.3 27.46  9.16 0.37     -1.3 0.98
------------------------------------------------------------------------------
group: M
   vars  n  mean   sd median trimmed  mad   min   max range skew kurtosis   se
X1    1 13 26.95 4.36   25.1   26.62 4.26 21.79 35.73 13.94  0.6    -0.98 1.21
```

Fig3:6

```
> describeBy(df2$`BMI Calculation`,group=list(df2$Gender,df2$`Body Type`))

 Descriptive statistics by group
: F
: Athletic
   vars n  mean   sd median trimmed  mad   min   max range skew kurtosis   se
X1    1 3 22.82 4.07  21.16   22.82 1.96 19.84 27.46  7.62 0.34    -2.33 2.35
------------------------------------------------------------------------
: M
: Athletic
   vars n  mean   sd median trimmed  mad   min   max range  skew kurtosis   se
X1    1 3 22.47 0.67   22.5   22.47 0.92 21.79 23.12  1.33 -0.05    -2.33 0.39
------------------------------------------------------------------------
: F
: Average
   vars n  mean  sd median trimmed  mad   min   max range skew kurtosis  se
X1    1 4 22.56 1.2  22.36   22.56 1.06 21.38 24.13  2.75 0.29    -2.02 0.6
------------------------------------------------------------------------
: M
: Average
   vars n  mean  sd median trimmed  mad   min   max range  skew kurtosis   se
X1    1 4 27.97 2.7  28.63   27.97 1.81 24.22 30.42   6.2 -0.46    -1.87 1.35
------------------------------------------------------------------------
: F
: Muscular
NULL
------------------------------------------------------------------------
: M
: Muscular
   vars n  mean   sd median trimmed  mad  min   max range skew kurtosis   se
X1    1 5 27.07 3.99   25.1   27.07 2.52 23.4 33.47 10.07 0.63    -1.51 1.78
------------------------------------------------------------------------
: F
: Round
   vars n  mean sd median trimmed mad   min   max range skew kurtosis se
X1    1 1 26.58 NA  26.58   26.58   0 26.58 26.58     0   NA       NA NA
------------------------------------------------------------------------
: M
: Round
   vars n  mean sd median trimmed mad   min   max range skew kurtosis se
X1    1 1 35.73 NA  35.73   35.73   0 35.73 35.73     0   NA       NA NA
------------------------------------------------------------------------
: F
: Thin
   vars n  mean   sd median trimmed  mad  min   max range skew kurtosis   se
X1    1 2 18.42 0.16  18.42   18.42 0.17 18.3 18.54  0.23    0    -2.75 0.12
------------------------------------------------------------------------
: M
: Thin
NULL
```

Fig3:7

```
> corr.test(df2[4:12])
Call:corr.test(x = df2[4:12])
Correlation matrix
                      BMI Calculation Time Spent in Gym Pant size (inches) Weight
BMI Calculation                  1.00              0.06                0.89   0.93
Time Spent in Gym                0.06              1.00                0.11   0.16
Pant size (inches)               0.89              0.11                1.00   0.93
Weight                           0.93              0.16                0.93   1.00
Height (inches)                  0.49              0.28                0.71   0.77
Weight Lift (Days)               0.03              0.99                0.05   0.10
Lifiting Session (Mins)          0.23              0.87                0.21   0.33
Running Times (Hours)           -0.23             -0.21               -0.22  -0.26
Distance of Run (Miles)         -0.42              0.00               -0.29  -0.32
                      Height (inches) Weight Lift (Days) Lifiting Session (Mins)
BMI Calculation                  0.49               0.03                    0.23
Time Spent in Gym                0.28               0.99                    0.87
Pant size (inches)               0.71               0.05                    0.21
Weight                           0.77               0.10                    0.33
Height (inches)                  1.00               0.20                    0.38
Weight Lift (Days)               0.20               1.00                    0.86
Lifiting Session (Mins)          0.38               0.86                    1.00
Running Times (Hours)           -0.20              -0.19                   -0.21
Distance of Run (Miles)         -0.01               0.00                   -0.06
                      Running Times (Hours) Distance of Run (Miles)
BMI Calculation                       -0.23                  -0.42
Time Spent in Gym                     -0.21                   0.00
Pant size (inches)                    -0.22                  -0.29
Weight                                -0.26                  -0.32
Height (inches)                       -0.20                  -0.01
Weight Lift (Days)                    -0.19                   0.00
Lifiting Session (Mins)               -0.21                  -0.06
Running Times (Hours)                  1.00                   0.77
Distance of Run (Miles)                0.77                   1.00
Sample Size
[1] 23
Probability values (Entries above the diagonal are adjusted for multiple tests.)
                      BMI Calculation Time Spent in Gym Pant size (inches) Weight
BMI Calculation                  0.00              1.00                0.00   0.00
Time Spent in Gym                0.77              0.00                1.00   1.00
Pant size (inches)               0.00              0.61                0.00   0.00
Weight                           0.00              0.47                0.00   0.00
Height (inches)                  0.02              0.19                0.00   0.00
Weight Lift (Days)               0.88              0.00                0.81   0.65
Lifiting Session (Mins)          0.30              0.00                0.33   0.13
Running Times (Hours)            0.28              0.33                0.31   0.23
Distance of Run (Miles)          0.05              1.00                0.19   0.14
                      Height (inches) Weight Lift (Days) Lifiting Session (Mins)
BMI Calculation                  0.49               1.00                    1.00
Time Spent in Gym                1.00               0.00                    0.00
Pant size (inches)               0.00               1.00                    1.00
Weight                           0.00               1.00                    1.00
Height (inches)                  0.00               1.00                    1.00
Weight Lift (Days)               0.37               0.00                    0.00
Lifiting Session (Mins)          0.07               0.00                    0.00
Running Times (Hours)            0.35               0.37                    0.33
Distance of Run (Miles)          0.98               0.99                    0.78
                      Running Times (Hours) Distance of Run (Miles)
BMI Calculation                           1                       1
Time Spent in Gym                         1                       1
Pant size (inches)                        1                       1
Weight                                    1                       1
Height (inches)                           1                       1
Weight Lift (Days)                        1                       1
Lifiting Session (Mins)                   1                       1
Running Times (Hours)                     0                       0
Distance of Run (Miles)                   0                       0
```

Fig3:8

```
> t.test(df2$`BMI Calculation`~df2$Gender, alt="two.sided")

        Welch Two Sample t-test

data:  df2$`BMI Calculation` by df2$Gender
t = -3.0439, df = 20.92, p-value = 0.006188
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -7.983411 -1.501651
sample estimates:
mean in group F mean in group M
       22.20999        26.95252
```

Fig3:9

```
> t.test(df2$`Time Spent in Gym`~df2$Gender, alt="two.sided")

        Welch Two Sample t-test

data:  df2$`Time Spent in Gym` by df2$Gender
t = -1.9416, df = 20.18, p-value = 0.06628
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -170.580484    6.065099
sample estimates:
mean in group F mean in group M
       115.8000        198.0577
```

(a) Describe what you think each record in the database describes. (1 mark)

(b) Explain with reference to the results generated,

(i) what type of distribution best describes the variable "BMI Calculation". (1 mark)

(ii) how many female records are there in the database? (1 mark)

(iii) what is the mean BMI calculation for members with 'Average' body type? [round off final answer to 2 decimal places] (1 mark)

(iv) what is the 95% confidence interval for 'time spent in gym'? [round off final answer to 2 decimal places] (2 mark)

(c) Draw a contingency table with rows = Gender, columns = Body Type, and cells = Frequency. (1 mark)

(d) Describe the linear relationship between "BMI Calculation" and other variables in the database? (2 marks)

(e)(i) State the hypotheses that Tom is testing in Fig 3.8 & Fig 3.9. (2 marks)

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

(f) Tom wants to know if the time spent in gym is significantly different for people of different body type. Suggest what descriptive analytics he could conduct and describe the different steps he needs to take to conduct this analytics. (5 marks)

(g) Tom needs help in computing the 95% prediction interval for the distance of run. Compute the prediction interval and explain the results to Tom. Round off the final answer to 2 decimal places. (5 marks)

Question 4 (7 marks)

ABC Nutrition Ltd manufactures baby formula milk power for infants. In developing a new product, in-house nutritionists have discovered that the mixture of baby formula should contain at least 11% protein and 28% fat, and no more than 4% fiber. Information on the ingredients of the new product is given below:

Ingredients	Protein %	Fat %	Fiber %	Cost/Kg
Ingredient A	15.7	28.8	25.7	$0.48
Ingredient B	13.1	4.3	5.6	$0.39
Ingredient C	8.6	4.7	2.8	$0.21
Ingredient D	17.2	6.2	3.1	$0.23
Ingredient E	6.7	3.9	1.7	$0.15
Ingredient F	12.4	1.5	2.5	$0.11
Ingredient G	17.4	14.7	23.4	$0.66
Ingredient H	13.9	4.6	13.2	$0.25
Ingredient I	15.3	12.4	15.3	$0.52

Develop an optimization model for minimum cost that meets the nutritional requirements. (7 marks)

Question 5 (12 marks)

ABC Nutrition Ltd currently has a number of factories in operation manufacturing a new type of baby formula milk powder. These factories are to deliver the finished new product to various distribution centers. The information of the delivery cost, production capacity tins per month and demand for the current year is given below:

	Distribution Centers					
Factories	**Center A**	**Center B**	**Center C**	**Center D**	**Center E**	**Capacity**
Factory A	$12.40	$13.75	$12.30	$12.15	$18.35	1300
Factory B	$8.50	$18.55	$8.80	$9.85	$16.45	900
Factory C	$9.30	$14.25	$6.45	$11.15	$14.95	400
Demand	160	330	600	590	880	

A linear optimization model is needed to minimize the transportation cost without exceeding production capacity, and still meet demand from distribution centers.

(a) Develop an optimization model to minimize the transportation cost without exceeding production capacity, and still meet demand from distribution centers. (5 mark)

(b) The demand forecasts for all distribution centers in the following year exceed the current year by 10%, and the company plans to add a new factory in one of two potential sites. Although similar in capacities of 1200 tins per month, the new factories will have a difference in transportation costs due to the difference in distance from the distribution centers. The information of the delivery cost and production capacity of the new factories is given below:

	Distribution Centers				
New Factories	**Center A**	**Center B**	**Center C**	**Center D**	**Center E**
Factory D	\$11.40	\$10.60	\$9.75	\$13.55	\$11.95
Factory E	\$14.10	\$15.50	\$13.85	\$9.45	\$12.25

Develop an optimization model to minimize the transportation cost without exceeding production capacity, and still meet demand from distribution centers, if only one new factory is to be built. (7 mark)

Question 6 (16 marks)

A supermarket is monitoring the sales performance of a new baby formula milk powder. The amount of tins sold for the last 10 weeks is given below:

Week	Tins Sold
1	15
2	18
3	37
4	40
5	39
6	38
7	35
8	41
9	55
10	30

(a) Develop a simple moving average table, determine the value of k observations by using Mean Square Error (MSE), and forecast the sale for Week 11. (8 mark)

(b) Develop an exponential smoothing table, determine the value of the smoothing constant (0.1, 0.3, 0.5, 0.7, 0.9) by using Root Mean Square Error (RMSE), and forecast the sale for Week 11. (8 mark)

END OF PAPER