Universiti Tun Hussein Onn Malaysia

UNIVERSITI TUN HUSSEIN ONN MALAYSIA

FINAL EXAMINATION
SEMESTER I
SESSION 2013/2014

COURSE NAME	:	STATISTICAL PACKAGES
COURSE CODE	:	BWB 20703
PROGRAMME	:	2 BWQ
EXAMINATION DATE	:	DECEMBER 2013/JANUARY 2014
DURATION	:	3 HOURS
INSTRUCTION	:	A) ANSWER **ALL (2)** QUESTIONS IN SECTION A
		B) ANSWER **ONLY TWO (2)** QUESTIONS IN SECTION B

THIS QUESTION PAPER CONSISTS OF **TWENTY (20)** PAGES

SECTION A

Q1 (a) Two hundred and seventy nine Switzerland skiers were studied during two one-week periods in 1998. One group of 140 skiers receiving a placebo each day and the other 139 receiving one gram of ascorbic acid (known as Vitamin C) per day. 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 ascorbic acid group contracts a cold 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)

Outcome	Ascorbic Asid	Placebo
With cold	17	31
With no cold	122	109
Total	139	140

(i) Find the sample proportion of skiers developing colds in the placebo and treatment groups.

(4 marks)

(ii) Calculate the pooled proportion of the developed colds in the study.

(2 marks)

(iii) State and prove the hypothesis that the probability of contracting a cold is the different if given a placebo or Vitamin C at 5% level of significance.

(10 marks)

(iv) Find the 95% confidence interval of $p_1 - p_2$ and state the appropriate conclusion.

(5 marks)

(b) The Minitab output of the cervical dysplasia data are as follows :

```
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 of for Sample 1 and Sample 2.
(2 marks)

(ii) State the hypothesis based on the output given above.
(2 marks)

Q2 (a) A complementary subject for Bachelor student in UTHM have two tests and a final exam assessment. A linear regression model for the Final as response variable on Test 1 and Test 2 as explanatory variables for 24 students who took the subject were done as shown in the Appendix **Q2**. However, there was some argument whether the final exam assessment could reasonably be predicted from the two tests.

(i) What is the fraction of the variability in the Final is explained by the two tests ?
(1 mark)

(ii) Give the fitted regression equation.
(1 mark)

(iii) From the results, it appears that one of the tests was more useful than the other in terms of predicting the Final. Which is the test ? How do you know ?
(2 marks)

(iv) Based on the output, find the correlation coefficient between Test 1 and Final. What is the correlation value ?
(5 marks)

(v) The F value is given as 12.12. What hypothesis is being tested with this F value for this analysis ? What conclusion do you reach about this hypothesis ?
(3 marks)

(vi) Find the standard deviation of the Final. By using this value, make a conclusion as to whether the regression was useful.
(3 marks)

(b) Two groups of statisticians were surveyed regarding opinions as to the price of gold on December 31, 2012. These statisticians were either academic (major employment by college or university) or professional (major employment by bank or securities firm). The result could be seen in the Table **Q2(b)** below.

TABLE Q2(b)

Group	Number	Mean Average	Standard Deviation
Academic	24	250	40
Professional	30	210	30

The assumption made for this sample data population are equal standard deviations.

(i) Calculate the pooled standard deviation of the data above.

(4 marks)

(ii) Find a 95% confidence interval for the population mean difference between these groups.

(6 marks)

SECTION B

Q3 (a) The question based on the data on the effect of Vitamin C on tooth growth in a group of rabbits. The variables are `len`, which is tooth length as a response variable, `supp` of which the levels are `OJ` (as orange juice) and `VC` (as ascorbic acid) and `dose` of the supplement in milligrams (0.5, 1.0 and 2.0) where `R` considers `dose` to be numerical. The result could be seen in the Appendix **Q3**.

(i) What is the command `summary()` for ?

(1 mark)

(ii) What is the `object` of this analysis ?

(1 mark)

(iii) Examine the number for `object` column and row.

(2 marks)

(iv) How many variables involve in this object ?

(1 mark)

(v) What is the type of variable for `supp` ?

(1 mark)

(vi) What is the meaning of the `R command` below ?

```
ToothGrowth$dose <- factor(ToothGrowth$dose,
levels=c(0.5,1.0,2.0),
labels=c("low","med","high"))
```

(2 marks)

(vii) From the second `summary(ToothGrowth)`, what is the most obvious result differ to previous `summary(ToothGrowth)` result ?

(2 marks)

(viii) What is the meaning of `R command` below ?
`ToothGrowth[seq(1,60,5),]`

(2 marks)

(ix) Figure **Q3(a)** shows the plot of Tooth Length vs Dose. What is the `R command` should we use to plot the figure ?

FIGURE Q3(a)

(3 marks)

(b) From **Q3(a)**, the ANOVA test was run where two or more factor names separated by colons refer to the interaction between those factors.

(i) What is the obvious result could you interpret from the `R command` of `summary(aov.out)` ?

(2 marks)

(ii) What is the model formula specifies by the `R command` below ?
`aov.out <- aov(len ~ supp * dose, data=ToothGrowth)`

(3 marks)

(iii) What is the `R command` below mean ?

```
TukeyHSD(aov.out,which=c("dose"), conf.level=.99)
```

(2 marks)

(iv) Figure **Q3(b)** shows the plot of Tooth Length vs Supplement:Dose. What is the `R command` should we use to plot the figure ?

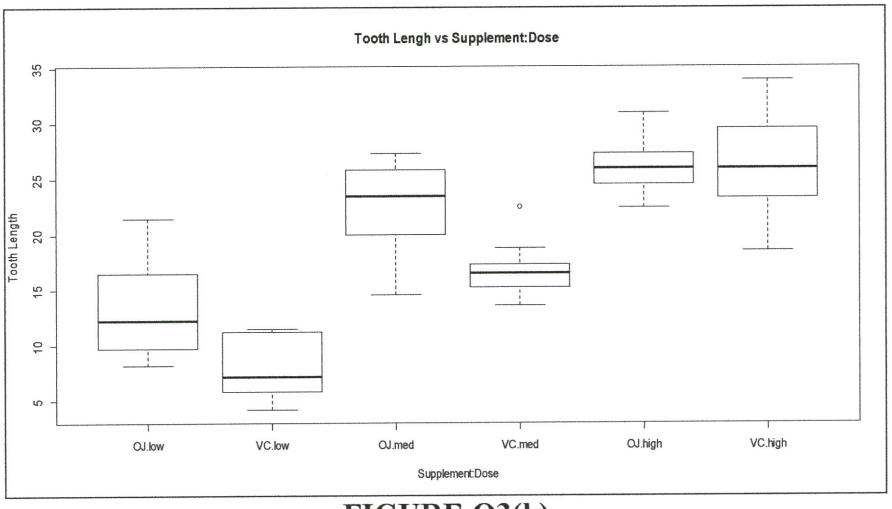

FIGURE Q3(b)

(3 marks)

Q4 (a) The fish burger slices prepared by Kyros Kebab & Burger is supplied by a local fish packing factory known as Costcutter. The fish burger was ordered to the specification that each slice have a weight of 25 grams and contain no more than eleven grams of fat. The requirement that the weight be 25 grams is easily checked and there has never been a problem occurred. However, some concern that the fish burger slices might contain more than eleven grams of fat. The decision to do a quality control inspection was arrange for the random sampling of 80 fish burger slices where the fish burger slices had an average fat content of 14 grams, with a sample standard deviation of 1.2 grams.

(i) Calculate the t – statistic of the data given above.

(4 marks)

State either **True** or **False** to each of the following questions (ii) to (vi).

(ii) This problem can be regarded as a test of the null hypothesis sample $H_0 : \mu = 11$.

(1 mark)

(iii) The parameter μ represents the mean weight of the 80 fish burger slices in the sample.

(1 mark)

(iv) The t statistic has 79 degrees of freedom.

(1 mark)

(v) Most published t tables have no line for 79 degrees of freedom.

(1 mark)

(vi) At the 5% level of significance, one would accept the hypothesis in **Q4(a)(i)**.

(1 mark)

(b) The marketing group in Alor Setar, Kedah plan to locate the regular drinkers of the Coke cans. The target population is defined for adults in the city of Alor Setar only. This is a situation of detecting events of low probability, as this is not a very popular drink.

(i) Suppose that 0.004 of the adults in Alor Setar are regular drinkers of Coke. What is the sample size should the marketing group use if they wants a probability of at least 80% of finding at least one Coke drinker in the sample ?

(7 marks)

(ii) Suppose that 0.02 of the adults in Alor Setar are regular drinkers of Coke. What is the sample size should the marketing group use if they wants a probability of at least 80% of finding at least one Coke drinker in the sample ?

(7 marks)

(iii) Why an inequality direction reversed in the calculation above ?

(2 marks)

Q5 (a) The SPSS output is about the Maintenance System hire by one selected Mobile Company in Perak. By using the result given in the Appendix **Q5**, answer the following questions carefully.

(i) What is the proportion of participants are male ?

(2 marks)

(ii) What is the frequency of respondents have lots of computer experience ?

(1 mark)

(iii) What is the different percentage of the respondents with some experience in IR experience and computer experience ?

(2 marks)

(iv) What is the missing terms refer in the SPSS Output given ?

(1 mark)

(v) What is the percentage of the respondents were satisfied with the result ?

(1 mark)

(vi) What is the proportion of the respondents were generally dissatisfied overall with the Maintenance System ?

(2 marks)

(vii) How many respondents taken at most 14 minutes using Maintenance System ?

(2 marks)

(b) The `anscombe` dataset is one of the 102 available datasets in the `R package`. The data in the datasets package are lazy loaded in `R workspace` where no need to install the datasets package. Based on the output in Appendix **Q5**, answer each of the following questions.

(i) How many variables in the object of `anscombe` ?

(1 mark)

(ii) How many respondent in the `anscombe` dataset ?

(1 mark)

(iii) Is there a linear relationship of the x's and the y's ?

(1 mark)

(iv) What is the different of `R command cor(anscombe)` and `cor(anscombe[["x3"]],anscombe[["y2"]])` ?

(2 marks)

(v) What is the different of `R command summary(anscombe)` and `summary(anscombe)[,1:6]` ?

(3 marks)

(vi) What is the obvious different of the `R command` of the last `summary(anscombe)` from the previous `summary(anscombe)` in the output ?

(2 marks)

(vii) What is the type of variables in the column four of the `anscombe` dataset ?

(1 mark)

(viii) The Pearson correlation coefficient measures the strength of the linear relationship between two variables. Based on the Figure **Q5(b)**, what is the type of relationship between two variables ? What is the `R command` to plot the figure ?

(3 marks)

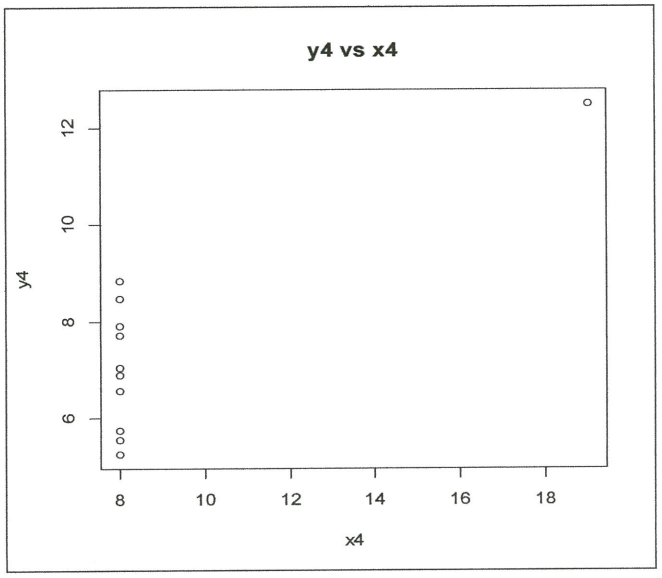

FIGURE Q5(b)

- END OF QUESTION -

FINAL EXAMINATION

SEMESTER/SESSION : SEM I/2013/2014 PROGRAMME : 2 BWQ
COURSE NAME : STATISTICAL PACKAGES COURSE CODE : BWB 20703

APPENDIX Q2

```
> data1 <- read.table("data1.txt",header=T)
> print(data1)
   Test1 Test2 Final
1    110    80   105
2    110   102   103
3     75    81    85
4    118   108   128
5     30    33    83
6    118    64    90
7     59    56   115
8     52    78    79
9     57    69    75
10   105    75   128
11    45    42   106
12   140   110   155
13    47    65    84
14   135    80   120
15    97    95   132
16   100    95   130
17   100    90   103
18    32    51   100
19    42    12    45
20   105    99   121
21    60   101   150
22    75    77    96
23   125   102   150
24    92    91   100
> names(data1)
[1] "Test1" "Test2" "Final"
> model1 <- lm(Final~Test1 + Test2,data=data1)
> summary(model1)
Call:
lm(formula = Final ~ Test1 + Test2, data = data1)
Residuals:
   Min     1Q Median     3Q    Max
-24.90 -14.83  -3.48  16.87  30.21
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  46.8453    12.9416   3.620  0.00161 **
Test1         0.1277     0.1702   0.751  0.46126
Test2         0.6463     0.2299   2.811  0.01047 *
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 18.95 on 21 degrees of freedom
Multiple R-squared:  0.5358,    Adjusted R-squared:  0.4916
F-statistic: 12.12 on 2 and 21 DF,  p-value: 0.0003162
> anova(model1)
Analysis of Variance Table
Response: Final
```

```
          Df Sum Sq Mean Sq F value    Pr(>F)
Test1      1 5867.2  5867.2 16.3420 0.0005869 ***
Test2      1 2836.7  2836.7  7.9012 0.0104705 *
Residuals 21 7539.6   359.0

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

> names(data2)[4:5] <- c("Fits", "Residuals")

> print(data2)
   Test1 Test2 Final      Fits  Residuals
1    110    80   105 112.59986  -7.599855
2    110   102   103 126.81911 -23.819109
3     75    81    85 108.77630 -23.776300
4    118   108   128 131.71878  -3.718776
5     30    33    83  72.00548  10.994523
6    118    64    90 103.28027 -13.280267
7     59    56   115  90.57468  24.425320
8     52    78    79 103.89996 -24.899958
9     57    69    75  98.72155 -23.721545
10   105    75   128 108.72965  19.270349
11    45    42   106  79.73811  26.261890
12   140   110   155 135.82108  19.178923
13    47    65    84  94.85912 -10.859116
14   135    80   120 115.79263   4.207370
15    97    95   132 120.63456  11.365442
16   100    95   130 121.01769   8.982309
17   100    90   103 117.78604 -14.786043
18    32    51   100  83.89483  16.105165
19    42    12    45  59.96508 -14.965085
20   105    99   121 124.24157  -3.241565
21    60   101   150 119.78723  30.212770
22    75    77    96 106.19098 -10.190981
23   125   102   150 128.73477  21.265226
24    92    91   100 117.41068 -17.410684

> # stem-and-leaf plot of the residuals
> stem(data2$Residuals,scale=2)

  The decimal point is 1 digit(s) to the right of the |

  -2 | 5444
  -1 | 755310
  -0 | 843
   0 | 49
   1 | 11699
   2 | 146
   3 | 0
```

FINAL EXAMINATION

SEMESTER/SESSION : SEM I/2013/2014 PROGRAMME : 2 BWQ
COURSE NAME : STATISTICAL PACKAGES COURSE CODE : BWB 20703

APPENDIX Q3

```
> data(ToothGrowth)
> str(ToothGrowth)
'data.frame':   60 obs. of  3 variables:
 $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
 $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
 $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
> dim(ToothGrowth)
[1] 60  3
> names(ToothGrowth)
[1] "len"  "supp" "dose"
> print(ToothGrowth)
    len supp dose
1   4.2   VC  0.5
2  11.5   VC  0.5
3   7.3   VC  0.5
4   5.8   VC  0.5
5   6.4   VC  0.5
6  10.0   VC  0.5
7  11.2   VC  0.5
8  11.2   VC  0.5
9   5.2   VC  0.5
10  7.0   VC  0.5
11 16.5   VC  1.0
12 16.5   VC  1.0
13 15.2   VC  1.0
14 17.3   VC  1.0
15 22.5   VC  1.0
16 17.3   VC  1.0
17 13.6   VC  1.0
18 14.5   VC  1.0
19 18.8   VC  1.0
20 15.5   VC  1.0
21 23.6   VC  2.0
22 18.5   VC  2.0
23 33.9   VC  2.0
24 25.5   VC  2.0
25 26.4   VC  2.0
26 32.5   VC  2.0
27 26.7   VC  2.0
28 21.5   VC  2.0
29 23.3   VC  2.0
30 29.5   VC  2.0
31 15.2   OJ  0.5
32 21.5   OJ  0.5
33 17.6   OJ  0.5
34  9.7   OJ  0.5
35 14.5   OJ  0.5
```

```
36 10.0   OJ  0.5
37  8.2   OJ  0.5
38  9.4   OJ  0.5
39 16.5   OJ  0.5
40  9.7   OJ  0.5
41 19.7   OJ  1.0
42 23.3   OJ  1.0
43 23.6   OJ  1.0
44 26.4   OJ  1.0
45 20.0   OJ  1.0
46 25.2   OJ  1.0
47 25.8   OJ  1.0
48 21.2   OJ  1.0
49 14.5   OJ  1.0
50 27.3   OJ  1.0
51 25.5   OJ  2.0
52 26.4   OJ  2.0
53 22.4   OJ  2.0
54 24.5   OJ  2.0
55 24.8   OJ  2.0
56 30.9   OJ  2.0
57 26.4   OJ  2.0
58 27.3   OJ  2.0
59 29.4   OJ  2.0
60 23.0   OJ  2.0
> summary(ToothGrowth)
      len        supp         dose
 Min.   : 4.20   OJ:30   Min.   :0.500
 1st Qu.:13.07   VC:30   1st Qu.:0.500
 Median :19.25           Median :1.000
 Mean   :18.81           Mean   :1.167
 3rd Qu.:25.27           3rd Qu.:2.000
 Max.   :33.90           Max.   :2.000
> str(ToothGrowth)
'data.frame':   60 obs. of  3 variables:
 $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
 $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
 $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
> dim(ToothGrowth)
[1] 60  3
> ToothGrowth$dose <-
factor(ToothGrowth$dose,levels=c(0.5,1.0,2.0),labels=c("low","med","hig
h"))
> str(ToothGrowth)
'data.frame':   60 obs. of  3 variables:
 $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
 $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
 $ dose: Factor w/ 3 levels "low","med","high": 1 1 1 1 1 1 1 1 1 1 ...
> summary(ToothGrowth)
      len        supp      dose
 Min.   : 4.20   OJ:30   low :20
 1st Qu.:13.07   VC:30   med :20
 Median :19.25           high:20
 Mean   :18.81
 3rd Qu.:25.27
 Max.   :33.90
```

```
> ToothGrowth[seq(1,60,5),]
    len supp dose
1   4.2   VC  low
6  10.0   VC  low
11 16.5   VC  med
16 17.3   VC  med
21 23.6   VC high
26 32.5   VC high
31 15.2   OJ  low
36 10.0   OJ  low
41 19.7   OJ  med
46 25.2   OJ  med
51 25.5   OJ high
56 30.9   OJ high
> replications(len~supp*dose,data=ToothGrowth)
     supp      dose supp:dose
       30        20        10
> replications(len~supp*dose,data=ToothGrowth[1:58,])
$supp
supp
OJ VC
28 30

$dose
dose
 low  med high
  20   20   18

$`supp:dose`
    dose
supp low med high
  OJ  10  10    8
  VC  10  10   10
>
> aov.out = aov(len ~ supp * dose, data=ToothGrowth)
> model.tables(aov.out, type="means", se=T)
Tables of means
Grand mean

18.81333

 supp
supp
    OJ     VC
20.663 16.963

 dose
dose
   low    med   high
10.605 19.735 26.100

 supp:dose
    dose
supp low   med   high
  OJ 13.23 22.70 26.06
  VC  7.98 16.77 26.14
```

```
Standard errors for differences of means
          supp   dose supp:dose
        0.9376 1.1484    1.6240
replic.     30     20        10
> summary(aov.out)
            Df Sum Sq Mean Sq F value   Pr(>F)
supp         1  205.3   205.3  15.572 0.000231 ***
dose         2 2426.4  1213.2  92.000  < 2e-16 ***
supp:dose    2  108.3    54.2   4.107 0.021860 *
Residuals   54  712.1    13.2
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> TukeyHSD(aov.out,which=c("dose"),conf.level=0.95)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = len ~ supp * dose, data = ToothGrowth)

$dose
           diff       lwr       upr   p adj
med-low   9.130  6.362488 11.897512 0.0e+00
high-low 15.495 12.727488 18.262512 0.0e+00
high-med  6.365  3.597488  9.132512 2.7e-06
```

FINAL EXAMINATION

SEMESTER/SESSION : SEM I/2013/2014
COURSE NAME : STATISTICAL PACKAGES

PROGRAMME : 2 BWQ
COURSE CODE : BWB 20703

APPENDIX Q5

Gender of respondent

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	male	20	37.0	37.0	37.0
	female	34	63.0	63.0	100.0
	Total	54	100.0	100.0	

IR experience

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	none	3	5.6	5.6	5.6
	some	36	66.7	66.7	72.2
	lots	15	27.8	27.8	100.0
	Total	54	100.0	100.0	

Computer experience

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	some	34	63.0	63.0	63.0
	lots	20	37.0	37.0	100.0
	Total	54	100.0	100.0	

Internet experience

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	some	32	59.3	59.3	59.3
	lots	22	40.7	40.7	100.0
	Total	54	100.0	100.0	

Satisfaction with results

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	dissatisfied	24	44.4	45.3	45.3
	neutral	15	27.8	28.3	73.6
	satisfied	14	25.9	26.4	100.0
	Total	53	98.1	100.0	
Missing	missing	1	1.9		
Total		54	100.0		

General feelings overall

		Frequenc y	Percen t	Valid Percent	Cumulative Percent
Valid	dissatisfied	19	35.2	35.8	35.8
	neutral	18	33.3	34.0	69.8
	satisfied	16	29.6	30.2	100.0
	Total	53	98.1	100.0	
Missing	missing	1	1.9		
Total		54	100.0		

Photo Search

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	used	1	1.9	2.0	2.0
	not used	50	92.6	98.0	100.0
	Total	51	94.4	100.0	
Missing	missing	3	5.6		
Total		54	100.0		

Time taken in minutes

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1-4mins	1	1.9	2.1	2.1
	5-9mins	11	20.4	22.9	25.0
	10-14mins	7	13.0	14.6	39.6
	15-19mins	13	24.1	27.1	66.7
	20-24mins	6	11.1	12.5	79.2
	25-29mins	3	5.6	6.3	85.4
	30-34mins	3	5.6	6.3	91.7
	35-39mins	1	1.9	2.1	93.8
	55-59mins	1	1.9	2.1	95.8
	60-65mins	2	3.7	4.2	100.0
	Total	48	88.9	100.0	
Missing	missing	6	11.1		
Total		54	100.0		

```
> anscombe
   x1 x2 x3 x4    y1   y2    y3    y4
1  10 10 10  8  8.04 9.14  7.46  6.58
2   8  8  8  8  6.95 8.14  6.77  5.76
3  13 13 13  8  7.58 8.74 12.74  7.71
4   9  9  9  8  8.81 8.77  7.11  8.84
5  11 11 11  8  8.33 9.26  7.81  8.47
6  14 14 14  8  9.96 8.10  8.84  7.04
7   6  6  6  8  7.24 6.13  6.08  5.25
8   4  4  4 19  4.26 3.10  5.39 12.50
9  12 12 12  8 10.84 9.13  8.15  5.56
10  7  7  7  8  4.82 7.26  6.42  7.91
11  5  5  5  8  5.68 4.74  5.73  6.89
> class(anscombe)
[1] "data.frame"
> summary(anscombe)
       x1              x2              x3              x4           y1
 Min.   : 4.0   Min.   : 4.0   Min.   : 4.0   Min.   : 8   Min.   :
4.260
 1st Qu.: 6.5   1st Qu.: 6.5   1st Qu.: 6.5   1st Qu.: 8   1st Qu.:
6.315
 Median : 9.0   Median : 9.0   Median : 9.0   Median : 8   Median :
7.580
 Mean   : 9.0   Mean   : 9.0   Mean   : 9.0   Mean   : 9   Mean   :
7.501
 3rd Qu.:11.5   3rd Qu.:11.5   3rd Qu.:11.5   3rd Qu.: 8   3rd Qu.:
8.570
 Max.   :14.0   Max.   :14.0   Max.   :14.0   Max.   :19   Max.
:10.840
```

```
      y2              y3              y4
 Min.   :3.100   Min.   : 5.39   Min.   : 5.250
 1st Qu.:6.695   1st Qu.: 6.25   1st Qu.: 6.170
 Median :8.140   Median : 7.11   Median : 7.040
 Mean   :7.501   Mean   : 7.50   Mean   : 7.501
 3rd Qu.:8.950   3rd Qu.: 7.98   3rd Qu.: 8.190
 Max.   :9.260   Max.   :12.74   Max.   :12.500
> summary(anscombe)[,1:6]
       x1              x2              x3              x4
 "Min.   : 4.0  " "Min.   : 4.0  " "Min.   : 4.0  " "Min.   : 8  "
 "1st Qu.: 6.5  " "1st Qu.: 6.5  " "1st Qu.: 6.5  " "1st Qu.: 8  "
 "Median : 9.0  " "Median : 9.0  " "Median : 9.0  " "Median : 8  "
 "Mean   : 9.0  " "Mean   : 9.0  " "Mean   : 9.0  " "Mean   : 9  "
 "3rd Qu.:11.5  " "3rd Qu.:11.5  " "3rd Qu.:11.5  " "3rd Qu.: 8  "
 "Max.   :14.0  " "Max.   :14.0  " "Max.   :14.0  " "Max.   :19  "
       y1              y2
 "Min.   : 4.260  " "Min.   :3.100  "
 "1st Qu.: 6.315  " "1st Qu.:6.695  "
 "Median : 7.580  " "Median :8.140  "
 "Mean   : 7.501  " "Mean   :7.501  "
 "3rd Qu.: 8.570  " "3rd Qu.:8.950  "
 "Max.   :10.840  " "Max.   :9.260  "
> cor(anscombe[["x1"]],anscombe[["y1"]])
[1] 0.8164205
> cor(anscombe[["x2"]],anscombe[["y2"]])
[1] 0.8162365
> cor(anscombe[["x3"]],anscombe[["y3"]])
[1] 0.8162867
> cor(anscombe[["x4"]],anscombe[["y4"]])
[1] 0.8165214
> cor(anscombe)
           x1         x2         x3         x4         y1         y2
y3
x1  1.0000000  1.0000000  1.0000000 -0.5000000  0.8164205  0.8162365
0.8162867
x2  1.0000000  1.0000000  1.0000000 -0.5000000  0.8164205  0.8162365
0.8162867
x3  1.0000000  1.0000000  1.0000000 -0.5000000  0.8164205  0.8162365
0.8162867
x4 -0.5000000 -0.5000000 -0.5000000  1.0000000 -0.5290927 -0.7184365 -
0.3446610
y1  0.8164205  0.8164205  0.8164205 -0.5290927  1.0000000  0.7500054
0.4687167
y2  0.8162365  0.8162365  0.8162365 -0.7184365  0.7500054  1.0000000
0.5879193
y3  0.8162867  0.8162867  0.8162867 -0.3446610  0.4687167  0.5879193
1.0000000
y4 -0.3140467 -0.3140467 -0.3140467  0.8165214 -0.4891162 -0.4780949 -
0.1554718
```

```
           y4
x1 -0.3140467
x2 -0.3140467
x3 -0.3140467
x4  0.8165214
y1 -0.4891162
y2 -0.4780949
y3 -0.1554718
y4  1.0000000
> summary(anscombe)
       x1              x2              x3          x4           y1
 Min.   : 4.0   Min.   : 4.0   Min.   : 4.0   yes:10   Min.   : 4.260
 1st Qu.: 6.5   1st Qu.: 6.5   1st Qu.: 6.5   no : 1   1st Qu.: 6.315
 Median : 9.0   Median : 9.0   Median : 9.0            Median : 7.580
 Mean   : 9.0   Mean   : 9.0   Mean   : 9.0            Mean   : 7.501
 3rd Qu.:11.5   3rd Qu.:11.5   3rd Qu.:11.5            3rd Qu.: 8.570
 Max.   :14.0   Max.   :14.0   Max.   :14.0            Max.   :10.840
       y2              y3              y4
 Min.   :3.100   Min.   : 5.39   Min.   : 5.250
 1st Qu.:6.695   1st Qu.: 6.25   1st Qu.: 6.170
 Median :8.140   Median : 7.11   Median : 7.040
 Mean   :7.501   Mean   : 7.50   Mean   : 7.501
 3rd Qu.:8.950   3rd Qu.: 7.98   3rd Qu.: 8.190
 Max.   :9.260   Max.   :12.74   Max.   :12.500
```