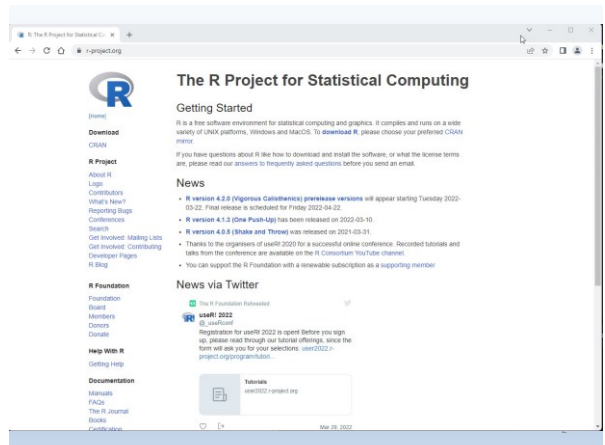


Research Methods

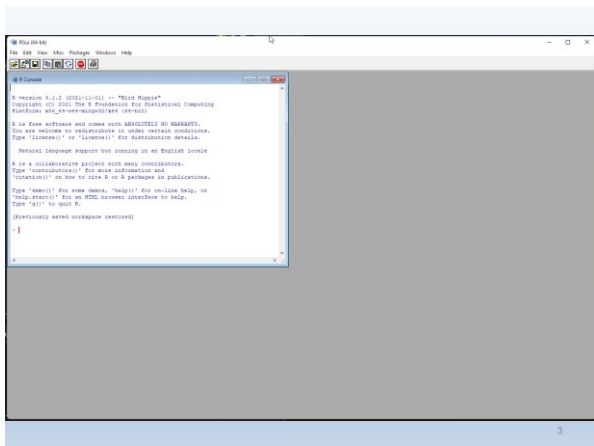
Introduction to Statistics with R

Andy Gillett

1

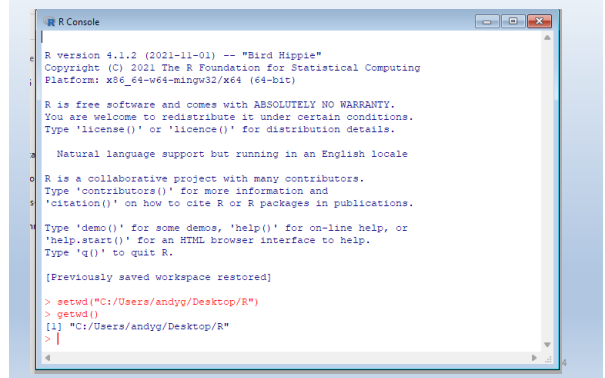


2



3

Working Directory



4

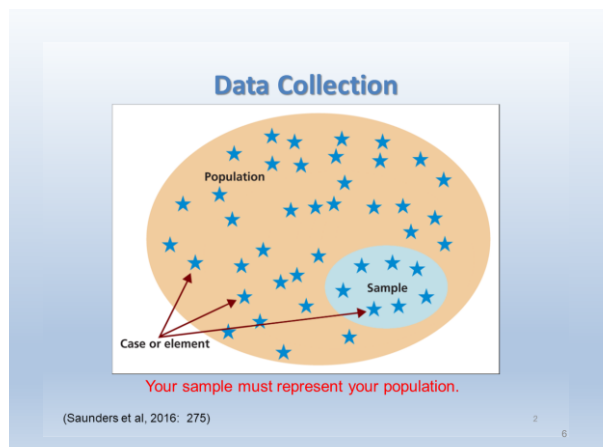
Describing Data using Statistics

Descriptive	Inferential
Description	Inference
Sample	Population

“When an individual uses descriptive statistics, he talks about the data he has; but with inferential statistics, he talks about data he does not have.”

(Popham & Sirotnik, 1973: 40)

5



6

Describing Data using Statistics

Description

- Central Tendency
- Dispersion

Inference

- Exploring Relationships
 - Correlation
 - Regression
- Comparing Groups
 - t-Test
 - ANOVA
 - One-Way
 - Two-Way
 - ANCOVA

7

Describing Data using Statistics

Description:

- The central tendency, which is the common, middle or average value.
- The dispersion, which is how the data values are spread, or dispersed, around the central tendency.

8

Gender Height (m)

Male 1.89
Female 1.75
Male 1.64
Male 1.81
Female 1.59
Male 1.94
Female 1.88
Female 1.65
Male 1.93
Male 1.74
Female 1.70
Male 1.82
Male 1.91
Female 1.69
Female 1.68
Female 1.80

9

Data

	A	B	C	D	E	F	G	H	I
1	Gender	Height							
2	Male	1.89							
3	Male	1.64							
4	Male	1.81							
5	Male	1.94							
6	Male	1.93							
7	Male	1.74							
8	Male	1.82							
9	Male	1.91							
10	Female	1.75							
11	Female	1.59							
12	Female	1.88							
13	Female	1.65							
14	Female	1.7							
15	Female	1.69							
16	Female	1.68							
17	Female	1.8							
18									
19									

10

MFheights.csv

Gender	Height
Male	1.89
Male	1.64
Male	1.81
Male	1.94
Male	1.93
Male	1.74
Male	1.82
Male	1.91
Female	1.61
Female	1.59
Female	1.52
Female	1.65
Female	1.64
Female	1.69
Female	1.68
Female	1.45

11

Import into R

```
>  
>  
>  
> MFheights <- read.csv("MFheights.csv", header = TRUE, colClasses = c("factor", "numeric"))  
> MFheights  
  Gender Height  
1 Male 1.89  
2 Male 1.64  
3 Male 1.81  
4 Male 1.94  
5 Male 1.93  
6 Male 1.74  
7 Male 1.82  
8 Male 1.91  
9 Female 1.61  
10 Female 1.59  
11 Female 1.52  
12 Female 1.65  
13 Female 1.64  
14 Female 1.69  
15 Female 1.68  
16 Female 1.45  
>  
>  
>
```

12

Summary

```
R Console
4 Male 1.94
5 Male 1.93
6 Male 1.74
7 Male 1.82
8 Male 1.91
9 Female 1.75
10 Female 1.59
11 Female 1.88
12 Female 1.65
13 Female 1.70
14 Female 1.69
15 Female 1.68
16 Female 1.90
> summary(MFheights)
 1. Gender Height
Female:0 Min. 1.590
Male :0 1st Qu.:1.600
      Median:1.775
      Mean :1.776
      3rd Qu.:1.883
      Max. :1.940
> summary(MFheights$Height)
      Min. 1st Qu. Median Mean 3rd Qu. Max.
 1.590 1.600 1.775 1.776 1.883 1.940
```

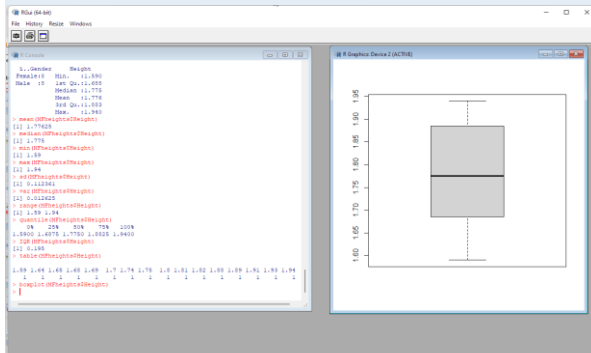
13

Various Descriptive Statistics

```
R Console
> summary(MFheights)
 1. Gender Height
Female:0 Min. 1.590
Male :0 1st Qu.:1.600
      Median:1.775
      Mean :1.776
      3rd Qu.:1.883
      Max. :1.940
> mean(MFheights$Height)
[1] 1.77625
> median(MFheights$Height)
[1] 1.775
> min(MFheights$Height)
[1] 1.59
> max(MFheights$Height)
[1] 1.94
> sd(MFheights$Height)
[1] 0.112361
> var(MFheights$Height)
[1] 0.012625
> range(MFheights$Height)
[1] 1.59 1.94
> quantile(MFheights$Height)
 0% 25% 50% 75% 100%
1.5900 1.6075 1.7750 1.8825 1.9400
> IQR(MFheights$Height)
[1] 0.195
> table(MFheights$Height)
 1.59 1.64 1.68 1.69 1.7 1.74 1.75 1.8 1.81 1.82 1.88 1.89 1.91 1.93 1.94
 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
```

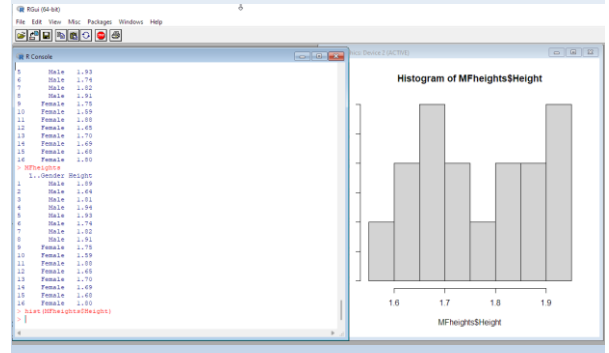
14

Boxplot



15

Histogram



16

Describing Data using Statistics

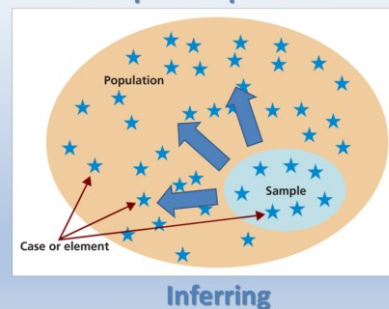
Descriptive	Inferential
Description	Inference
Sample	Population

“When an individual uses descriptive statistics, he talks about the data he has; but with inferential statistics, he talks about data he does not have.”

(Popham & Sirotnik, 1973: 40)

17

Sample – Population



(Saunders et al, 2016: 275)

18

Describing Data using Statistics

Description

- ❑ Central Tendency
 - Mean, Median, Mode
- ❑ Dispersion
 - Range, Standard Deviation, IQR

Inference

- ❑ Exploring Relationships
 - Correlation
 - Regression
- ❑ Comparing Groups
 - t-Test
 - ANOVA
 - One-Way
 - Two-Way

19

- ❑ Central Tendency
 - Mean, Median, Mode
- ❑ Dispersion
 - Range, Standard Deviation, IQR

- ❑ Exploring Relationships
 - Correlation
 - Regression
- ❑ Comparing Groups
 - t-Test
 - ANOVA
 - One-Way
 - Two-Way

21

Correlation

There is much interest in whether there is a relationship between pupils' ability at sport and their maths ability. Accordingly a study was carried out in which both sport and maths ability were rated on each of 10 pupils sampled at random from a class of first year pupils in a secondary school. Sport ability was rated on a scale of 1 to 20. Maths ability was measured by a pencil and paper test on a scale from 1 to 30.

20

22

Correlation

Pupil No.	1	2	3	4	5	6	7	8	9	10
Sports score	15	19	12	14	9	16	20	14	15	19
Maths score	24	29	20	27	16	28	28	23	21	29

21

Data

	A	B	C	D	E	F
1	ID	Sport	Maths			
2	1	15	24			
3	2	19	29			
4	3	12	20			
5	4	14	27			
6	5	9	16			
7	6	15	28			
8	7	12	28			
9	8	14	23			
10	9	15	21			
11	10	19	29			
12						
13						
14						
15						

22

Import Data

The screenshot shows an R console window titled "R Console". The prompt ">" appears multiple times. Then, the command `MathsSport <- read.csv("maths-sport.csv", header = TRUE, colClasses = c("numeric", "numeric", "numeric"))` is entered. This is followed by `> MathsSport`, which displays the contents of the imported dataset as follows:

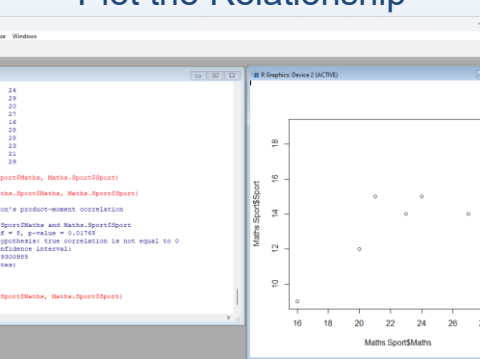
ID	Sport	Maths
1	10	26
2	19	28
3	12	20
4	14	27
5	9	16
6	17	20
7	12	28
8	14	23
9	15	21
10	19	29

The prompt ">" appears again at the bottom.

23

23

Plot the Relationship



The image shows an RStudio window with a script editor on the left and a plot window on the right. The script editor contains the following R code:

```
# Maths (84.80)
File History Plots Windows

# Example
cor <- cor(x, y)

1 1 15 24
2 2 15 19
3 3 12 20
4 4 14 27
5 5 9 16
6 6 15 28
7 7 12 20
8 8 14 23
9 9 15 21
10 10 19 29

> cor(Maths.SportsSpent, Maths.SportsSpent)
[1] 0.726103
> cor.test(Maths.SportsSpent, Maths.SportsSpent)

Pearson's product-moment correlation

data: Maths.SportsSpent and Maths.SportsSpent
t = 2.9752, df = 9, p-value = 0.01761
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.376442 0.890095
sample estimates:
cor
0.726103

# plot(Maths.SportsSpent, Maths.SportsSpent)
#
```

The plot window displays a scatter plot titled "R Graphics Device 1 (AC7X)". The x-axis is labeled "Maths SportsSpent" and the y-axis is labeled "Maths SportsSpent". The plot shows 10 data points, indicating a positive correlation between the two variables. The data points are approximately as follows:

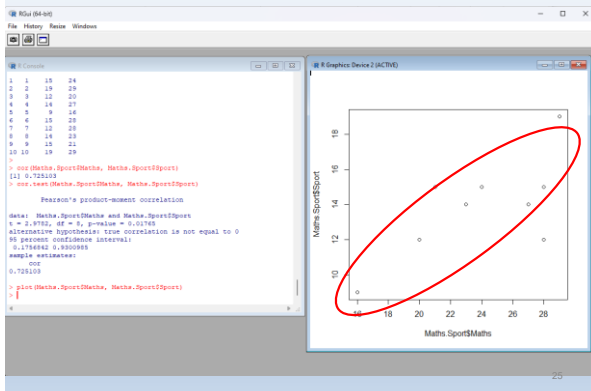
Maths SportsSpent (x)	Maths SportsSpent (y)
1	15
2	15
3	12
4	14
5	9
6	15
7	12
8	14
9	15
10	19

The plot window also shows a legend with the text "R Graphics Device 1 (AC7X)" and a small "Legend" button.

9/4

24

Correlation



25

Correlation

```

R Console

> Maths.Sport <- read.csv("maths-sport.csv", header = TRUE, colClasses=)
> Maths.Sport
  ID Sport Maths
1  1   15    24
2  2   19    29
3  3   12    20
4  4   14    27
5  5    9    16
6  6   15    28
7  7   12    28
8  8   14    23
9  9   15    21
10 10   19    29
>
> cor(Maths.Sport$Maths, Maths.Sport$Sport)
[1] 0.725103
> cor.test(Maths.Sport$Maths, Maths.Sport$Sport)

Pearson's product-moment correlation

data: Maths.Sport$Maths and Maths.Sport$Sport
t = 2.9782, df = 8, p-value = 0.01765
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.1756942 0.9300985
sample estimates:
cor
0.725103

```

26

Correlation

The correlation, r , has the value 0.725.

The p -value is the probability that the correlation found between the two variables could be a chance result. As p here is less than 0.05 the correlation is unlikely to be a chance result.

The conclusion is that there is a relationship between SPORT scores and MATHS scores in the wider population of first year secondary school pupils. The results can be generalised.

27

27

Regression

A study looked at the possibility of predicting students' future reading ability (READING) from their intelligence score (IQ).

Ten children were included in the investigation, selected at random from a primary school class. The children's IQ scores were obtained when they entered the school and the corresponding reading score was obtained at the end of the first year. The reading score was the result out of 100 on a test.

28

28

IQ

1 Which is the odd one out?

A Orange
B Apple
C Potato
D Banana

2 What is 22×22 ?

A 474
B 484
C 464
D 504

3 If $X = 5$ and $Y = 7$
 $X + (Y - X) =$

A 5
B 7
C 8
D 17

4 What shape completes the sequence?

A B C D

5 What shape completes the sequence?

A B C D

29

29

IQ : Reading

Pupil No.	1	2	3	4	5	6	7	8	9	10
IQ Score	98	115	124	97	135	105	117	101	110	127
Reading	54	74	81	59	88	70	78	73	84	86

30

30

IQ : Reading

	A	B	C	D	E	F	G
1	ID	IQ	Reading				
2		1	98	54			
3		2	115	74			
4		3	124	81			
5		4	97	59			
6		5	135	88			
7		6	105	70			
8		7	117	78			
9		8	101	73			
10		9	110	84			
11		10	127	86			
12							
13							
14							

31

Hypotheses

H_0 : READING score cannot be predicted from IQ score

H_1 : READING score can be predicted from IQ score

H_0 = **null hypothesis** - no relationship between two measured phenomena.

Rejecting or disproving the **null hypothesis** is a central task in modern scientific practice.

Statistically significant means that the null hypothesis has been rejected.

32

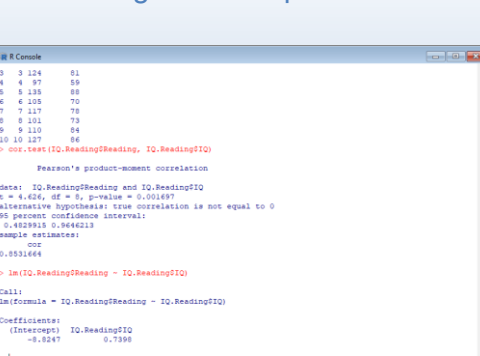
Import the Data

The screenshot shows an R console window titled "R Console". It contains several prompt symbols ">" followed by commands and their output:

```
> 
> 
> 
> 
> 
> 
> 
> 
> 
> 
> 
> 
> 
> 
> 
> 
> 
> 
> 
> IO.reading <- read.csv("IO-reading.csv", header = TRUE, asIsClasses = c("numeric", "numeric", "numeric"))
> IO.reading
      ID   IQ    Reading
1     1    90         64
2     2   115        74
3     3   124        81
4     4    97        59
5     5   135        88
6     6   105        70
7     7   117        78
8     8   101        73
9     9   110        84
10    10   127       86
> |
|_>
```

33

Regression Equation



The screenshot shows an R console window with the following content:

```

R Console
3 3 124 81
4 4 97 59
5 5 195 88
6 6 105 70
7 7 117 78
8 8 101 73
9 9 110 84
10 10 127 86
> cor.test(IQ.Reading~Reading, IQ.Reading~IQ)

Pearson's product-moment correlation

data: IQ.Reading~Reading and IQ.Reading~IQ
t = 4.426, df = 8, p-value = 0.001897
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.4029315 0.944213
sample estimates:
      cor 
0.8531668 
> lm(IQ.Reading~Reading + IQ.Reading~IQ)

Call:
lm(formula = IQ.Reading~Reading + IQ.Reading~IQ)

Coefficients:
(Intercept)  IQ.Reading~IQ
      -8.8247       0.7398

>

```

34

34

Regression Equation

$$x = a + by$$
$$a = -8.825$$
$$b = .740$$

35

Regression

The regression equation

$$\text{“READING} = -8.825 + 0.740 \times \text{IQ”}$$

can be obtained by equating the dependent variable **Reading** to the *CONSTANT* value -8.825 , plus the independent variable **IQ** multiplied by its coefficient 0.740 .

36

IQ: Reading + Availability of Books

	A	B	C	D	E	F
1	ID	IQ	Books	Reading		
2	1	98	5	54		
3	2	115	25	74		
4	3	124	37	81		
5	4	97	10	59		
6	5	135	90	88		
7	6	105	80	70		
8	7	117	88	78		
9	8	101	73	73		
10	9	110	80	84		
11	10	127	100	86		
12						
13						

37

Hypotheses

H_0 : READING score cannot be predicted from a combination of IQ score and number of Books Available

H_1 : READING score can be predicted from a combination of IQ score and number of Books Available

38

IQ/Books : Reading

	A	B	C	D	E	F
1	ID	IQ	Books	Reading		
2	1	98	5	54		
3	2	115	25	74		
4	3	124	37	81		
5	4	97	10	59		
6	5	135	90	88		
7	6	105	80	70		
8	7	117	88	78		
9	8	101	73	73		
10	9	110	80	84		
11	10	127	100	86		
12						
13						

IQ-Books-Reading.csv

39

Import the Data into R

```
IQ.Books.Reading <- read.csv("IQ-Books-Reading.csv", header = TRUE, colClasses = c("numeric", "numeric", "numeric", "numeric"))
```

```
IQ.Books.Reading <- read.csv("IQ-Books-Reading.csv", header = TRUE, colClasses = c("numeric", "numeric", "numeric", "numeric"))
```

	A	B	C	D	E	F
1	ID	IQ	Books	Reading		
2	1	98	5	54		
3	2	115	25	74		
4	3	124	37	81		
5	4	97	10	59		
6	5	135	90	88		
7	6	105	80	70		
8	7	117	88	78		
9	8	101	73	73		
10	9	110	80	84		
11	10	127	100	86		

40

Regression Equation

```
Call:
lm(formula = IQ.Books.Reading$Reading ~ IQ.Books.Reading$IQ + IQ.Books.Reading$Books)
lmfit: 29.2 on 8 and 67, p-value: 0.0000000
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	6.47837	18.58373	0.448	0.6703
IQ.Books.Reading\$IQ	0.53111	0.14093	3.769	0.0070 **
IQ.Books.Reading\$Books	0.14066	0.05118	2.749	0.0287 *

Multiple R-squared: 0.8689, Adjusted R-squared: 0.8315

41

Multiple Regression Equation

$$x = a + by + cz$$

$$a = 6.478$$

$$b = .531$$

$$c = .140$$

42

Multiple Regression

The multiple regression equation

“READING = 6.478 + 0.531 × IQ + .140 x Books Available”

can be obtained by equating the dependent variable **Reading** to the *CONSTANT* value 6.478, plus the independent variables **IQ** multiplied by its coefficient 0.531 and the number of **Books Available** multiplied by .140.

43

IQ/Books/Sex : Reading

	A	B	C	D	E	F
1	ID	IQ	Books	Sex	Reading	
2	1	98	5	1	54	
3	2	115	25	2	74	
4	3	124	37	2	81	
5	4	97	10	1	59	
6	5	135	90	1	88	
7	6	105	80	1	70	
8	7	117	88	2	78	
9	8	101	73	2	73	
10	9	110	80	1	84	
11	10	127	100	2	86	
12						
13						

IQ-Books-Sex-Reading.csv

44

Import the Data into R

```
IQ.Books.Sex.Reading <- read.csv("IQ-Books-Sex-Reading.csv", header = TRUE, colClasses = c("numeric", "numeric", "numeric", "numeric", "numeric"))
```

```
IQ.Books.Sex.Reading <- read.csv("IQ-Books-Sex-Reading.csv", header = TRUE, colClasses = c("numeric", "numeric", "numeric", "numeric", "numeric"))
```

```
1  1  98  5  1  54
2  2 115 25  2  74
3  3 124 37  2  81
4  4  97 10  1  59
5  5 135 90  1  88
6  6 105 80  1  70
7  7 117 88  2  78
8  8 101 73  2  73
9  9 110 80  1  84
10 10 127 100 2  86
```

45

Regression Equation

```
model <- lm(IQ.Books.Sex.Reading$Reading ~ IQ.Books.Sex.Reading$IQ + IQ.Books.Sex.Reading$Sex + IQ.Books.Sex.Reading$Books)
```

```
> model <- lm(IQ.Books.Sex.Reading$Reading ~ IQ.Books.Sex.Reading$IQ + IQ.Books.Sex.Reading$Sex + IQ.Books.Sex.Reading$Books)
> summary(model)

Call:
lm(formula = IQ.Books.Sex.Reading$Reading ~ IQ.Books.Sex.Reading$IQ + IQ.Books.Sex.Reading$Sex + IQ.Books.Sex.Reading$Books)

Residuals:
    Min       1Q   Median       3Q      Max
-4.6713 -2.7087 -0.6339  1.9184  6.6990

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    6.3863    15.3608   0.416  0.6921
IQ.Books.Sex.Reading$IQ  0.5079     0.1542   3.294  0.0165 *
IQ.Books.Sex.Reading$Sex  1.8091     3.2472   0.557  0.5976
IQ.Books.Sex.Reading$Books 0.1405     0.0539   2.606  0.0403 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.87 on 6 degrees of freedom
Multiple R-squared:  0.8784,    Adjusted R-squared:  0.8191
F-statistic: 14.08 on 3 and 6 DF, p-value: 0.004031
```

46

Multiple Regression Equation

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    6.3863    15.3608   0.416  0.6921
IQ.Books.Sex.Reading$IQ  0.5079     0.1542   3.294  0.0165 *
IQ.Books.Sex.Reading$Sex  1.8091     3.2472   0.557  0.5976
IQ.Books.Sex.Reading$Books 0.1405     0.0539   2.606  0.0403 *
```

Reading = 6.39 + .5 x IQ + .14 x Books + 1.8 x Sex

Note that Sex is a categorical variable with two levels (Males = 1 & Females = 2).

In this case the regression weighting (1.8) is added to the predicted Reading value. So for females 1.8 (1.8 x 1) would be added and for males 3.6 (1.8 x 2) would be added.

47

Multiple Regression: Relative Importance

Which variables are important?

Install and load the “QuantPsyc” package, and use the function “lm.beta” on your model.

```
> lm.beta(model)
            IQ.Books.Sex.Reading$IQ  IQ.Books.Sex.Reading$Sex  IQ.Books.Sex.Reading$Books
            0.58571381            0.08464255            0.44603647
```

You now have the standardised regression coefficients, which you can now compare.

IQ: 0.58571381

Sex: 0.08464255

Books Available: 0.44603647

And you can see that the most important predictor is **IQ**, followed closely by **Books Available**. **Sex** is not contributing very much.

48

47

48

Describing Data using Statistics

Description

Inference

- Exploring Relationships
- Comparing Groups

"When an individual uses descriptive statistics, he talks about the data he has; but with inferential statistics, he talks about data he does not have."

(Popham & Sirotnik, 1973: 40)

49

Comparing Groups (2 groups)

An experiment is carried out to see if announcing the items on a conveyor belt as they pass helps children remember the items better. Twenty-two children, all aged 12 years, were randomly allocated; 10 to the announcement group, and the remaining 12 to the no-announcement group. The children were shown the items on the conveyor belt with or without announcing each item as they passed. The number of items recalled out of each of the ten letters in the list was recorded for each child:

50

Data

ID	Method	Recall
1	1 Announcement	5
2	2 Announcement	7
3	3 Announcement	8
4	4 Announcement	9
5	5 Announcement	5
6	6 Announcement	4
7	7 Announcement	9
8	8 Announcement	8
9	9 Announcement	7
10	10 Announcement	4
11	11 No Announcement	3
12	12 No Announcement	8
13	13 No Announcement	5
14	14 No Announcement	6
15	15 No Announcement	3
16	16 No Announcement	5
17	17 No Announcement	9
18	18 No Announcement	3
19	19 No Announcement	3
20	20 No Announcement	4
21	21 No Announcement	7
22	22 No Announcement	4

51

Import Data

```

R Console
lm(formula = IQ.ReadingTQ ~ IQ.ReadingReading)
Coefficients:
(Intercept)  IQ.ReadingReading
39.4035      0.9539

> Announcement <- read.csv("Announcement.csv", header = TRUE, colClasses = c("fct", "integer"))
> Announcement
  1..12  Method Recall
1  1  Announcement  5
2  2  Announcement  7
3  3  Announcement  8
4  4  Announcement  9
5  5  Announcement  5
6  6  Announcement  4
7  7  Announcement  9
8  8  Announcement  8
9  9  Announcement  7
10 10  Announcement  4
11 11 No Announcement  3
12 12 No Announcement  8
13 13 No Announcement  5
14 14 No Announcement  6
15 15 No Announcement  3
16 16 No Announcement  5
17 17 No Announcement  9
18 18 No Announcement  3
19 19 No Announcement  3
20 20 No Announcement  4
21 21 No Announcement  7
22 22 No Announcement  4

```

52

Summary

```

R Console
1..12  Method Recall
1  1  Announcement  5
2  2  Announcement  7
3  3  Announcement  8
4  4  Announcement  9
5  5  Announcement  5
6  6  Announcement  4
7  7  Announcement  9
8  8  Announcement  8
9  9  Announcement  7
10 10  Announcement  4
11 11 No Announcement  3
12 12 No Announcement  8
13 13 No Announcement  5
14 14 No Announcement  6
15 15 No Announcement  3
16 16 No Announcement  5
17 17 No Announcement  9
18 18 No Announcement  3
19 19 No Announcement  3
20 20 No Announcement  4
21 21 No Announcement  7
22 22 No Announcement  4

> summary(Announcement)
      1..12  Method      Recall
Min.   1.00  Announcement 10  Min.   19.000
1st Qu. 4.25  No Announcement 12  1st Qu. 14.000
Median 11.50                                     Median 15.000
Mean   12.50                                     Mean   15.727
3rd Qu.14.75                                     3rd Qu.17.750
Max.   12.00                                     Max.   19.000

```

53

Draw a Boxplot



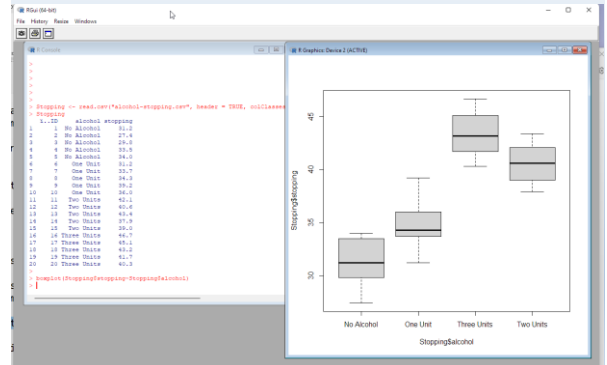
54

Compare Data

```
R Console
> Stopping <- read.csv("alcohol-stopping.csv", header=TRUE, colClasses=c("numeric", "factor", "numeric"))
> Stopping
  19 units_of_alcohol stopping
1 1 No Alcohol 31.2
2 2 No Alcohol 27.4
3 3 No Alcohol 29.8
4 4 No Alcohol 33.5
5 5 No Alcohol 34.0
6 6 One Unit 31.2
7 7 One Unit 33.7
8 8 One Unit 34.3
9 9 One Unit 39.2
10 10 One Unit 36.0
11 11 Two Units 42.1
12 12 Two Units 40.6
13 13 Two Units 43.4
14 14 Two Units 37.9
15 15 Two Units 39.0
16 16 Three Units 46.7
17 17 Three Units 45.1
18 18 Three Units 43.2
19 19 Three Units 41.7
20 20 Three Units 40.3
```

61

Compare Data



62

ANOVA

```
R Console
> Stopping
  1 1 No Alcohol 31.2
  2 2 No Alcohol 27.4
  3 3 No Alcohol 29.8
  4 4 No Alcohol 33.5
  5 5 No Alcohol 34.0
  6 6 One Unit 31.2
  7 7 One Unit 33.7
  8 8 One Unit 34.3
  9 9 One Unit 39.2
 10 10 One Unit 36.0
 11 11 Two Units 42.1
 12 12 Two Units 40.6
 13 13 Two Units 43.4
 14 14 Two Units 37.9
 15 15 Two Units 39.0
 16 16 Three Units 46.7
 17 17 Three Units 45.1
 18 18 Three Units 43.2
 19 19 Three Units 41.7
 20 20 Three Units 40.3

> boxplot(Stopping$stopping~Stopping$alcohol)
> anova <- aov(Stopping$stopping ~ Stopping$alcohol)
> summary(anova)
              Df Sum Sq Mean Sq F value    Pr(>F)    
Stopping$alcohol  3  456.1  152.04   21.92 6.52e-06 ***
Residuals       16  111.0    6.94                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> |
```

63

Analysis

```
              Df Sum Sq Mean Sq F value    Pr(>F)    
Stopping$alcohol  3  456.1  152.04   21.92 6.52e-06 ***
Residuals       16  111.0    6.94                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$p < 0.05$
Therefore significant

The conclusion is that there is a significant difference between the distance it takes to stop a car simulator according to the amount of alcohol consumed.

64

Multiple Comparisons

```
R Console
 14 14 Two Units 37.9
 15 15 Two Units 39.0
 16 16 Three Units 46.7
 17 17 Three Units 45.1
 18 18 Three Units 43.2
 19 19 Three Units 41.7
 20 20 Three Units 40.3

> boxplot(Stopping$stopping~Stopping$alcohol)
> anova <- aov(Stopping$stopping ~ Stopping$alcohol)
> summary(anova)
              Df Sum Sq Mean Sq F value    Pr(>F)    
Stopping$alcohol  3  456.1  152.04   21.92 6.52e-06 ***
Residuals       16  111.0    6.94                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> TukeyHSD(anova)
Tukey multiple comparisons of means
50% family-wise confidence level

Fit: aov(formula = Stopping$stopping ~ Stopping$alcohol)
$Stopping$alcohol
      diff      lwr      upr    p adj
One Unit-No Alcohol  3.70 -1.0654663  8.465466 0.1596059
Three Units-No Alcohol 12.22  7.4545337 16.985466 0.0000091
Two Units-No Alcohol  9.42  4.6545337 14.185466 0.0001901
Three Units-One Unit  8.52  3.7545337 13.285466 0.0005430
Two Units-One Unit  5.72  0.9545337 10.485466 0.0161514
Two Units-Three Units -2.80 -7.5654663  1.965466 0.3649771
```

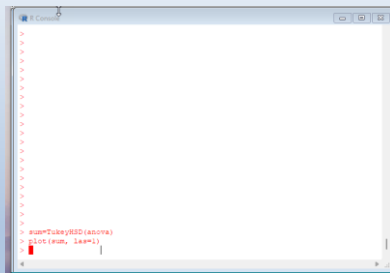
65

Multiple Comparisons

```
      diff      lwr      upr    p adj
One Unit-No Alcohol  3.70 -1.0654663  8.465466 0.1596059
Three Units-No Alcohol 12.22  7.4545337 16.985466 0.0000091
Two Units-No Alcohol  9.42  4.6545337 14.185466 0.0001901
Three Units-One Unit  8.52  3.7545337 13.285466 0.0005430
Two Units-One Unit  5.72  0.9545337 10.485466 0.0161514
Two Units-Three Units -2.80 -7.5654663  1.965466 0.3649771
```

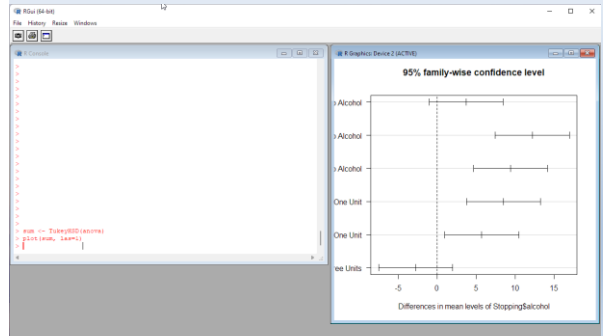
66

Tukey Plot



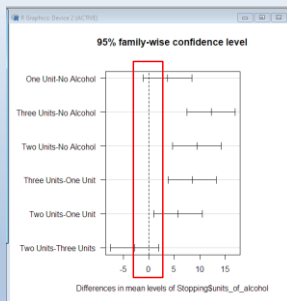
67

Tukey Plot



68

Tukey Plot



69

Comparing Groups

An experiment was carried out in which a sample of adults was selected and randomly allocated to four groups. All were asked to consume amounts of alcohol as follows:

- Group 1: NO ALCOHOL
- Group 2: ONE UNIT OF ALCOHOL
- Group 3: TWO UNITS OF ALCOHOL
- Group 4: THREE UNITS OF ALCOHOL

after which they were asked drive a car simulator and perform an emergency stop at 40 miles per hour. The distance to stopping the simulator in metres by each group member is shown below.

Is this the same for both males and females?

70

Comparing Groups (4 groups and M/F)

The results were as follows.

M A L E S	NO ALCOHOL	31.2	27.4	29.8	33.5	34.0
	ONE UNIT	31.2	33.7	34.3	39.2	36.0
	TWO UNITS	42.1	40.6	43.4	37.9	39.0
	THREE UNITS	46.7	45.1	43.2	41.7	40.3
F E M A L E S	NO ALCOHOL	35.1	32.3	35.4	38.1	35.9
	ONE UNIT	32.8	35.9	36.1	40.9	38.2
	TWO UNITS	38.1	35.9	40.2	39.8	37.6
	THREE UNITS	39.8	31.7	41.2	39.1	37.9

Stopping Distance in Metres

71

Comparing Groups (4 groups and M/F)

id	alcohol	gender	stopping
1	1 0 Units	Male	31.2
2	2 0 Units	Male	27.4
3	3 0 Units	Male	29.8
4	4 0 Units	Male	33.5
5	5 0 Units	Male	34
6	6 1 Unit	Male	31.2
7	7 1 Unit	Male	33.7
8	8 1 Unit	Male	34.3
9	9 1 Unit	Male	39.2

alcohol-gender-stopping.csv

13	14 2 Units	Male	37.9
16	15 2 Units	Male	39
17	16 3 Units	Male	46.7
18	17 3 Units	Male	45.1
19	18 3 Units	Male	43.2
20	19 3 Units	Male	41.7
21	20 3 Units	Male	40.3
22	21 0 Units	Female	35.1
23	22 0 Units	Female	32.3
24	23 0 Units	Female	35.4
25	24 0 Units	Female	38.1
26	25 0 Units	Female	35.9
27	26 1 Unit	Female	32.8
28	27 1 Unit	Female	35.9
29	28 1 Unit	Female	36.1
30	29 1 Unit	Female	40.9
31	30 1 Unit	Female	38.2

72

71

72

Three Questions

1. Is there a difference in the stopping distance according to the amount of alcohol consumed?
 2. Is there a difference in the stopping distance between male and females?
 3. Is the effect of amount of alcohol consumed on stopping distance different for males and females?
- 1 & 2 = main effect
3 = interaction

73

Import Data

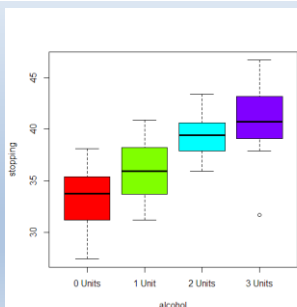
```
> stopping <- read.csv("alcohol-gender-stopping.csv",
header = TRUE, colClasses = c("numeric", "factor", "factor",
"numeric"))
```

```
R Console
> setwd("C:/Users/andy/Desktop/3")
> stopping <- read.csv("alcohol-gender-stopping.csv", header = TRUE, colClasses =
> stopping
1. add alcohol gender stopping
1 1 0 Unite Male 31.2
2 2 0 Unite Male 27.4
3 3 0 Unite Male 25.0
4 4 0 Unite Male 33.5
5 5 0 Unite Male 34.0
6 6 1 Unite Male 31.2
7 7 1 Unite Male 33.7
8 8 1 Unite Male 34.3
9 9 1 Unite Male 39.2
10 10 1 Unite Male 36.0
11 11 2 Unite Male 42.1
12 12 2 Unite Male 40.6
13 13 2 Unite Male 43.4
14 14 2 Unite Male 37.9
15 15 2 Unite Male 39.0
16 16 3 Unite Male 46.7
17 17 3 Unite Male 48.1
18 18 3 Unite Male 43.2
19 19 3 Unite Male 41.7
20 20 3 Unite Male 40.3
21 21 0 Unite Female 35.1
```

74

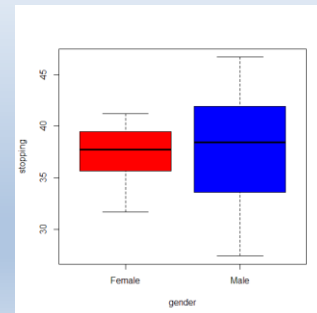
```
> boxplot(stopping~alcohol, data=stopping)
```

```
> boxplot(stopping~alcohol, data=stopping,
col=rainbow(4))
```



75

```
> boxplot(stopping~gender, data=stopping,
col=c("red", "blue"))
```



76

```
> tapply(stopping$stopping, stopping$gender, mean)
```

```
R Console
25 25 0 Unite Female 35.9
26 26 1 Unite Female 32.8
27 27 1 Unite Female 35.9
28 28 1 Unite Female 36.1
29 29 1 Unite Female 40.9
30 30 1 Unite Female 39.2
31 31 2 Unite Female 38.1
32 32 2 Unite Female 38.9
33 33 2 Unite Female 40.2
34 34 2 Unite Female 39.0
35 35 2 Unite Female 37.6
36 36 3 Unite Female 38.9
37 37 3 Unite Female 31.7
38 38 3 Unite Female 41.2
39 39 3 Unite Female 39.1
40 40 3 Unite Female 37.9
> boxplot(stopping~alcohol, data=stopping)
> boxplot(stopping~gender, data=stopping)
> boxplot(stopping~gender, data=stopping, col=c("blue", "red"))
> boxplot(stopping~gender, data=stopping, col=c("red", "blue"))
> boxplot(stopping~alcohol, data=stopping, col=rainbow(4))
> tapply(stopping$stopping, stopping$gender, mean)
Female Male
37.100 37.815
> |
```

77

```
> tapply(stopping$stopping, stopping$alcohol, mean)
```

```
R Console
28 28 1 Unite Female 36.1
29 29 1 Unite Female 40.9
30 30 1 Unite Female 39.2
31 31 2 Unite Female 38.1
32 32 2 Unite Female 38.9
33 33 2 Unite Female 40.2
34 34 2 Unite Female 39.0
35 35 2 Unite Female 37.6
36 36 3 Unite Female 38.9
37 37 3 Unite Female 31.7
38 38 3 Unite Female 41.2
39 39 3 Unite Female 39.1
40 40 3 Unite Female 37.9
> boxplot(stopping~alcohol, data=stopping)
> boxplot(stopping~gender, data=stopping, col=c("blue", "red"))
> boxplot(stopping~gender, data=stopping, col=c("red", "blue"))
> boxplot(stopping~alcohol, data=stopping, col=rainbow(4))
> tapply(stopping$stopping, stopping$gender, mean)
Female Male
37.100 37.815
> tapply(stopping$stopping, stopping$alcohol, mean)
0 Units 1 Unit 2 Units 3 Units
33.27 35.03 39.46 40.67
> |
```

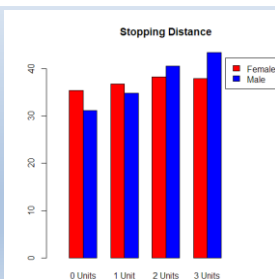
78

```
> tapply(stopping$stopping, list(stopping$gender,
stopping$alcohol), mean)
```

```
R Console
32 32 2 0 Units Female 35.9
33 33 2 0 Units Female 40.2
34 34 2 0 Units Female 39.8
35 35 2 0 Units Female 37.6
36 36 2 0 Units Female 39.8
37 37 2 0 Units Female 31.7
38 38 2 0 Units Female 41.2
39 39 2 0 Units Female 39.1
40 40 2 0 Units Female 37.9
> boxplot(stopping$alcohol, data=stopping)
> boxplot(stopping$gender, data=stopping)
> boxplot(stopping$gender, data=stopping, col=c("blue", "red"))
> boxplot(stopping$gender, data=stopping, col=c("red", "blue"))
> boxplot(stopping$alcohol, data=stopping, col=rainbow(4))
> tapply(stopping$stopping, stopping$gender, mean)
Female Male
37.100 37.015
> tapply(stopping$stopping, stopping$alcohol, mean)
0 Units 1 Unit 2 Units 3 Units
33.27 35.83 39.46 40.67
> tapply(stopping$stopping, list(stopping$gender, stopping$alcohol), mean)
Female 0 Units 1 Unit 2 Units 3 Units
35.36 36.78 39.32 37.94
Male 31.17 34.88 40.60 42.40
> |
```

79

```
> count <- tapply(stopping$stopping,
list(stopping$gender, stopping$alcohol), mean)
> barplot(count, main="Stopping Distance",
beside=TRUE, xlim=c(0,16), legend=TRUE,
col=c("red", "blue"))
```



80

ANOVA

One-Way Between-Groups ANOVA
Single independent variable.

Two-Way Between-Groups ANOVA
Factorial/Multifactorial ANOVA

2 independent variables - two factors

81

2-Way ANOVA

```
> anova2 = aov(stopping ~ alcohol + gender +
alcohol:gender, data = stopping)
> summary(anova2)
```

82

2-Way ANOVA

```
> anova2 = aov(stopping ~ alcohol + gender +
alcohol:gender, data = stopping)
> summary(anova2)
```

```
R Console
[Previously saved workspace restored]
> setwd("C:/Users/andy/Desktop/RData")
> stopping <- read.csv("alcohol-gender-stopping.csv", header = TRUE, colClasses=)
> aov(stopping ~ alcohol + gender + alcohol:gender, data = stopping)
Call:
aov(formula = stopping ~ alcohol + gender + alcohol:gender, data = stopping)

Terms:
alcohol gender alcohol:gender Residuals
Sum of Squares 344.2407 1.7223 138.5087 230.7960
Deg. of Freedom 3 1 3 32

Residual standard error: 2.68587
Estimated effects may be unbalanced
> anova2 <- aov(stopping ~ alcohol + gender + alcohol:gender, data = stopping)
> summary(anova2)
Df Sum Sq Mean Sq F value Pr(>F)
alcohol 3 344.2 114.75 15.910 1.65e-06 ***
gender 1 1.7 1.72 0.239 0.6284
alcohol:gender 3 138.5 46.17 6.401 0.0016 **
Residuals 32 230.8 7.21
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

83

2-Way ANOVA

```
R Console
[Previously saved workspace restored]
> setwd("C:/Users/andy/Desktop/RData")
> stopping <- read.csv("alcohol-gender-stopping.csv", header = TRUE, colClasses=)
> aov(stopping ~ alcohol + gender + alcohol:gender, data = stopping)
Call:
aov(formula = stopping ~ alcohol + gender + alcohol:gender, data = stopping)

Terms:
alcohol gender alcohol:gender Residuals
Sum of Squares 344.2407 1.7223 138.5087 230.7960
Deg. of Freedom 3 1 3 32

Residual standard error: 2.68587
Estimated effects may be unbalanced
> anova2 <- aov(stopping ~ alcohol + gender + alcohol:gender, data = stopping)
> summary(anova2)
Df Sum Sq Mean Sq F value Pr(>F)
alcohol 3 344.2 114.75 15.910 1.65e-06 ***
gender 1 1.7 1.72 0.239 0.6284
alcohol:gender 3 138.5 46.17 6.401 0.0016 **
Residuals 32 230.8 7.21
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> |
```

84

2-Way ANOVA

```
R Console
37.100 37.515
> tapply(stopping$stopping, stopping$alcohol, mean)
0 Units 1 Unit 2 Units 3 Units
33.27 35.83 39.46 40.67
> tapply(stopping$stopping, list(stopping$gender, stopping$alcohol), mean)
0 Units 1 Unit 2 Units 3 Units
Female 35.36 36.78 38.32 37.94
Male 31.18 34.88 40.60 43.40
> count <- tapply(stopping$stopping, list(stopping$gender, stopping$alcohol), m)
>
> barplot(count, main="Stopping Distance", beside=TRUE, , xlim=c(0,15), legend=2
> barplot(count, main="Stopping Distance", beside=TRUE, , xlim=c(0,16), legend=2
> anova2 <- aov(stopping ~ alcohol + gender + alcohol:gender, data = stopping)
> summary(anova2)
Analysis of Variance Table

Response: stopping
          Df Sum Sq Mean Sq F value    Pr(>F)
alcohol    3 344.24 114.747 15.9097 1.652e-06 ***
gender      1  1.72   1.722  0.2388  0.628413
alcohol:gender 3 138.51 46.170  6.4014  0.001603 **
Residuals 32 230.80   7.212
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> |
```

85

```
R Console
37.100 37.515
> tapply(stopping$stopping, stopping$alcohol, mean)
0 Units 1 Unit 2 Units 3 Units
33.27 35.83 39.46 40.67
> tapply(stopping$stopping, list(stopping$gender, stopping$alcohol), mean)
0 Units 1 Unit 2 Units 3 Units
Female 35.36 36.78 38.32 37.94
Male 31.18 34.88 40.60 43.40
> count <- tapply(stopping$stopping, list(stopping$gender, stopping$alcohol), m)
>
> barplot(count, main="Stopping Distance", beside=TRUE, , xlim=c(0,15), legend=2
> barplot(count, main="Stopping Distance", beside=TRUE, , xlim=c(0,16), legend=2
> anova2 <- aov(stopping ~ alcohol + gender + alcohol:gender, data = stopping)
> summary(anova2)
Analysis of Variance Table

Response: stopping
          Df Sum Sq Mean Sq F value    Pr(>F)
alcohol    3 344.24 114.747 15.9097 1.652e-06 ***
gender      1  1.72   1.722  0.2388  0.628413
alcohol:gender 3 138.51 46.170  6.4014  0.001603 **
Residuals 32 230.80   7.212
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> |
```

alcohol = Amount of Alcohol Consumed – significant ($p=0.0000 < 0.05$)

86

```
R Console
37.100 37.515
> tapply(stopping$stopping, stopping$alcohol, mean)
0 Units 1 Unit 2 Units 3 Units
33.27 35.83 39.46 40.67
> tapply(stopping$stopping, list(stopping$gender, stopping$alcohol), mean)
0 Units 1 Unit 2 Units 3 Units
Female 35.36 36.78 38.32 37.94
Male 31.18 34.88 40.60 43.40
> count <- tapply(stopping$stopping, list(stopping$gender, stopping$alcohol), m)
>
> barplot(count, main="Stopping Distance", beside=TRUE, , xlim=c(0,15), legend=2
> barplot(count, main="Stopping Distance", beside=TRUE, , xlim=c(0,16), legend=2
> anova2 <- aov(stopping ~ alcohol + gender + alcohol:gender, data = stopping)
> summary(anova2)
Analysis of Variance Table

Response: stopping
          Df Sum Sq Mean Sq F value    Pr(>F)
alcohol    3 344.24 114.747 15.9097 1.652e-06 ***
gender      1  1.72   1.722  0.2388  0.628413
alcohol:gender 3 138.51 46.170  6.4014  0.001603 **
Residuals 32 230.80   7.212
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> |
```

alcohol = Amount of Alcohol Consumed – significant ($p=0.0000 < 0.05$)
gender = M/F – not significant ($p=0.63$)

87

```
R Console
37.100 37.515
> tapply(stopping$stopping, stopping$alcohol, mean)
0 Units 1 Unit 2 Units 3 Units
33.27 35.83 39.46 40.67
> tapply(stopping$stopping, list(stopping$gender, stopping$alcohol), mean)
0 Units 1 Unit 2 Units 3 Units
Female 35.36 36.78 38.32 37.94
Male 31.18 34.88 40.60 43.40
> count <- tapply(stopping$stopping, list(stopping$gender, stopping$alcohol), m)
>
> barplot(count, main="Stopping Distance", beside=TRUE, , xlim=c(0,15), legend=2
> barplot(count, main="Stopping Distance", beside=TRUE, , xlim=c(0,16), legend=2
> anova2 <- aov(stopping ~ alcohol + gender + alcohol:gender, data = stopping)
> summary(anova2)
Analysis of Variance Table

Response: stopping
          Df Sum Sq Mean Sq F value    Pr(>F)
alcohol    3 344.24 114.747 15.9097 1.652e-06 ***
gender      1  1.72   1.722  0.2388  0.628413
alcohol:gender 3 138.51 46.170  6.4014  0.001603 **
Residuals 32 230.80   7.212
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> |
```

88

```
R Console
37.100 37.515
> tapply(stopping$stopping, stopping$alcohol, mean)
0 Units 1 Unit 2 Units 3 Units
33.27 35.83 39.46 40.67
> tapply(stopping$stopping, list(stopping$gender, stopping$alcohol), mean)
0 Units 1 Unit 2 Units 3 Units
Female 35.36 36.78 38.32 37.94
Male 31.18 34.88 40.60 43.40
> count <- tapply(stopping$stopping, list(stopping$gender, stopping$alcohol), m)
>
> barplot(count, main="Stopping Distance", beside=TRUE, , xlim=c(0,15), legend=2
> barplot(count, main="Stopping Distance", beside=TRUE, , xlim=c(0,16), legend=2
> model <- list(stopping ~ alcohol + gender + alcohol:gender, data = stopping)
> anova(model)
Analysis of Variance Table

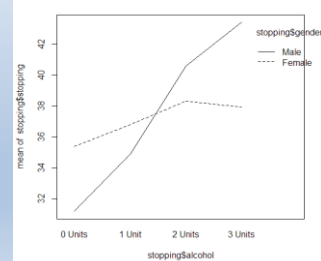
Response: stopping
          Df Sum Sq Mean Sq F value    Pr(>F)
alcohol    3 344.24 114.747 15.9097 1.652e-06 ***
gender      1  1.72   1.722  0.2388  0.628413
alcohol:gender 3 138.51 46.170  6.4014  0.001603 **
Residuals 32 230.80   7.212
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> |
```

alcohol:gender - Interaction effect = whether the influence of alcohol on stopping distance depends on whether you are male or female – significant ($p=0.0016$)

89

Interaction Plot

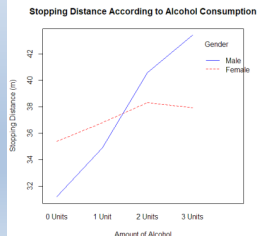
> interaction.plot(stopping\$alcohol, stopping\$gender, stopping\$stopping)



90

Interaction Plot

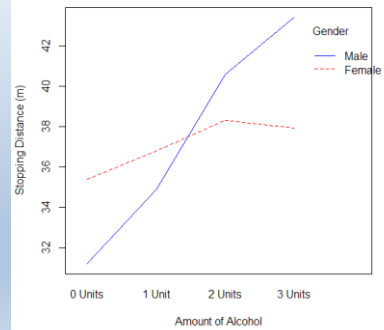
```
> interaction.plot(stopping$alcohol, stopping$gender,
stopping$stopping, col=c("red", "blue"),
trace.label="Gender", xlab="Amount of Alcohol",
ylab="Stopping Distance (m)", main="Stopping Distance
According to Alcohol Consumption")
```



91

Interaction Plot

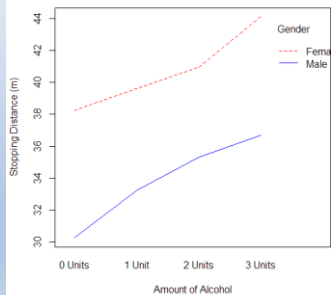
Stopping Distance According to Alcohol Consumption



92

No Interaction

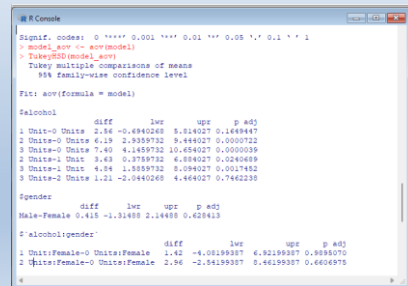
Stopping Distance According to Alcohol Consumption



93

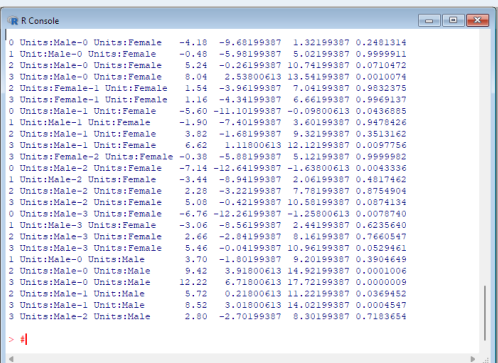
Tukey HSD

```
> model_aov <- aov(model)
> TukeyHSD(model_aov)
```



94

Tukey HSD



95

Tukey HSD

```

3 Units:Male-1 Unit:Female 6.62 0.0097756
0 Units:Male-2 Units:Female -7.14 0.0043336
0 Units:Male-3 Units:Female -6.76 0.0078740
2 Units:Male-0 Units:Male 9.42 0.0001006
3 Units:Male-0 Units:Male 12.22 0.0000009
2 Units:Male-1 Unit:Male 5.72 0.0369452
3 Units:Male-1 Unit:Male 8.52 0.0004547

```

96

Interaction

The conclusion is that the effect of alcohol on car stopping distance was different for male subjects than for females.

A two-way between-groups analysis of variance was conducted to explore the impact of gender and amount of alcohol consumed on a car stopping distance. There was a significant interaction effect between the amount of alcohol consumed and the gender of the person driving the car on the distance needed to stop the car $F(3, 32) = 6.4, p < 0.05$.

97

97

Comparing Groups

An experiment was carried out in which a sample of adults was selected and randomly allocated to four groups. All were asked to consume amounts of alcohol as follows:

Group 1: NO ALCOHOL
Group 2: ONE UNIT OF ALCOHOL
Group 3: TWO UNITS OF ALCOHOL
Group 4: THREE UNITS OF ALCOHOL

after which they were asked drive a car simulator and perform an emergency stop at 40 miles per hour. The distance to stopping the simulator in metres by each group member is shown below.

Does age make a difference?

98

98

Analysis of Covariance

If three or more distinct groups – need to control for 3rd variable:

Use one-way analysis of covariance or ANCOVA.

99

99

Data for Analysis of Covariance

An experiment was carried out in which a sample of adults was selected and randomly allocated to four groups. All were asked to consume amounts of alcohol as follows:

Group 1: NO ALCOHOL
Group 2: ONE UNIT OF ALCOHOL
Group 3: TWO UNITS OF ALCOHOL
Group 4: THREE UNITS OF ALCOHOL

after which they were asked drive a car simulator and perform an emergency stop at 40 miles per hour.

However, we need to control for age.

100

100

Three Questions

1. Is there a difference in the stopping distance according to the amount of alcohol drunk?
2. Is there a difference in the stopping distance depending on age?
3. Does age affect the stopping distance according to the amount of alcohol drunk?

101

101

Data for Analysis of Covariance

The stopping distances (m) obtained by each group member is shown below.
The results were as follows:

No Alcohol	31.2	27.4	29.8	33.5	34.0
One Unit	31.2	33.7	34.3	39.2	36.0
Two Units	42.1	40.6	43.4	37.9	39.0
Three Units	46.7	45.1	43.2	41.7	40.3

102

102

	Age	Stopping Distance (m)
No Alcohol	25.00	31.20
No Alcohol	24.00	27.40
No Alcohol	23.50	29.80
No Alcohol	24.70	33.50
No Alcohol	24.20	34.00
One Unit	24.50	31.20
One Unit	26.00	33.70
One Unit	27.00	34.30
One Unit	22.00	39.20
One Unit	25.00	36.00
Two Units	24.30	42.10
Two Units	24.60	40.60
Two Units	28.00	43.40
Two Units	27.00	37.90
Three Units	26.00	46.70
Three Units	25.00	45.10
Three Units	23.00	43.20
Three Units	28.00	41.70
Three Units	31.00	40.30

103

Data for Analysis of Covariance

id	alcohol	age	stopping
1	0	25	31.2
2	0	24	27.4
3	0	23.5	29.8
4	0	24.7	33.5
5	0	24.2	34
6	1	24.5	31.2
7	1	26	33.7
8	1	27	34.3
9	1	22	39.2
10	1	25	36
11	2	24.3	42.1
12	2	24.6	40.6
13	2	28	43.4
14	2	27	37.9
15	3	26	39
16	3	23	46.7
17	3	25	45.1
18	3	23	43.2
19	3	28	41.7
20	3	31	40.3

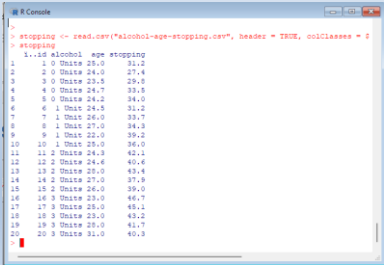
alcohol-age-stopping.csv

104

104

Import Data for Analysis of Covariance

```
> stopping <- read.csv("alcohol-age-stopping.csv",
header = TRUE, colClasses =
c("numeric","factor","numeric","numeric"))
```

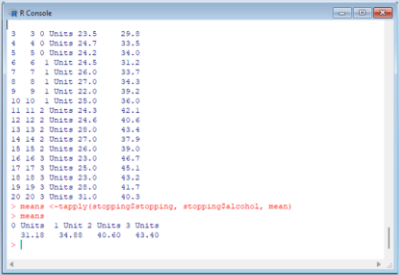


105

105

Data: Means

```
> means <- tapply(stopping$stopping,
stopping$alcohol, mean)
> means
```

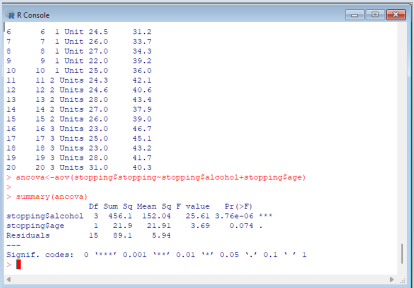


88

106

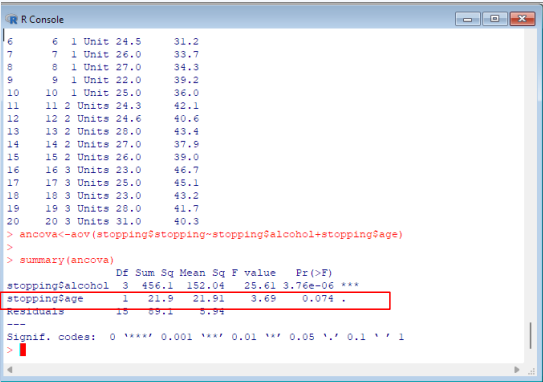
Analysis of Covariance

```
> ancova<-aov(stopping$stopping ~
stopping$alcohol+stopping$age)
> summary(ancova)
```



107

107



108

108

No Significant Effect For Age (p=.074)

```

R Console
6 6 1 Unit 24.5 31.2
7 7 1 Unit 26.0 33.7
8 8 1 Unit 27.0 34.3
9 9 1 Unit 22.0 39.2
10 10 1 Unit 25.0 36.0
11 11 2 Units 24.3 42.1
12 12 2 Units 24.6 40.6
13 13 2 Units 28.0 43.4
14 14 2 Units 27.0 37.9
15 15 2 Units 26.0 39.0
16 16 3 Units 23.0 46.7
17 17 3 Units 25.0 45.1
18 18 3 Units 23.0 43.2
19 19 3 Units 28.0 41.7
20 20 3 Units 31.0 40.3
> ancova<-aov(stopping$stopping~stopping$alcohol+stopping$age)
> summary(ancova)
              Df Sum Sq Mean Sq F value    Pr(>F)    
stopping$alcohol 3  456.1  152.04    25.61 3.76e-06 ***
stopping$age      1    0.0    0.00      0.00 0.9754 .    
Residuals       15   89.1    5.94                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Significant Effect For Amount of Alcohol (= <0.05)

109

TukeyHSD

> TukeyHSD(ancova,"stopping\$alcohol")

```

R Console
> ancova<-aov(stopping$stopping ~ stopping$alcohol+stopping$age)
> TukeyHSD(ancova,"stopping$alcohol")
Tukey multiple comparisons of means
 95% family-wise confidence level

Fit: aov(formula = stopping$stopping ~ stopping$alcohol + stopping$age)

$`stopping$alcohol`
      diff      lwr      upr    p adj
1 Unit-0 Units  3.70 -0.7417801  8.14178 0.1196368
2 Units-0 Units  9.42  4.9782199 13.86178 0.0001083
3 Units-0 Units 12.22  7.7782199 16.66178 0.0000052
2 Units-1 Unit   5.72  1.2782199 10.16178 0.0100402
3 Units-1 Unit   8.52  4.0782199 12.96178 0.0003028
3 Units-2 Units  2.80 -1.6417801  7.24178 0.3039307

Warning message:
In replications(paste("=", xx), data = mf) :
non-factors ignored: stopping$age

```

110

Estimated Marginal Means

```

> library('emmeans')
> emmeans(ancova, 'alcohol')

```

```

R Console
      diff      lwr      upr    p adj
1 Unit-0 Units  3.70 -0.7417801  8.14178 0.1196368
2 Units-0 Units  9.42  4.9782199 13.86178 0.0001083
3 Units-0 Units 12.22  7.7782199 16.66178 0.0000052
2 Units-1 Unit   5.72  1.2782199 10.16178 0.0100402
3 Units-1 Unit   8.52  4.0782199 12.96178 0.0003028
3 Units-2 Units  2.80 -1.6417801  7.24178 0.3039307

Warning message:
In replications(paste("=", xx), data = mf) :
non-factors ignored: stopping$age
Error in emmeans(ancova, "alcohol") : could not find function "emmeans"
> library("emmeans")
> emmeans(ancova, "alcohol")
alcohol emmean    SE df lower.CL upper.CL
0 Units  30.6  1.13 15  28.2    33.0
1 Unit   34.7  1.10 15  32.3    37.0
2 Units  41.0  1.11 15  38.6    43.3
3 Units  43.8  1.11 15  41.4    46.2

Confidence level used: 0.95

```

111

Presenting the Results from One-Way ANCOVA

A one-way between-groups analysis of covariance was conducted to compare the effect of amount of alcohol drunk on stopping distance on a group of adult drivers. The independent variable was the amount of alcohol drunk (none, 1 unit, 2 units, 3 units) and the dependent variable consisted of the stopping distance after an emergency stop in the car at 40 miles per hour. Participants' age was used as the covariate in this analysis. After adjusting for age, a significant difference ($F(3, 15) = 25.61, p < .001$) and was found between the amount of alcohol drunk and the stopping distance.

112

Comparison of Groups

A sample of male and female adults was selected and randomly allocated to four teaching groups using different teaching methods, after which they were given a test.

100

Two-Way ANCOVA

Method	Gender	Pre-Test Score	Final Test Score
Method 1	Male	25.00	31.20
Method 1	Male	24.00	27.40
Method 1	Male	23.50	29.80
Method 1	Male	24.70	33.50
Method 1	Male	24.20	34.00
Method 2	Male	24.50	31.20
Method 2	Male	26.00	33.70
Method 2	Male	27.00	34.30
Method 2	Male	22.00	39.20

114

Method	Gender	Pre-Test Score	Post-Test Score
Method 1	Male	25.00	28.00
Method 1	Male	24.00	27.00
Method 1	Male	23.50	26.50
Method 1	Male	24.70	27.10
Method 1	Male	24.20	26.00
Method 2	Male	24.50	27.20
Method 2	Male	23.00	26.20
Method 2	Male	23.00	26.20
Method 2	Male	24.30	27.50
Method 2	Male	24.80	26.00
Method 3	Male	26.00	28.50
Method 3	Male	27.00	27.00
Method 3	Male	26.00	29.00
Method 3	Male	25.00	26.70
Method 3	Male	23.00	25.10
Method 4	Male	23.00	25.10
Method 4	Male	26.00	27.20
Method 4	Male	26.00	25.20
Method 4	Male	23.00	25.10
Method 4	Male	23.00	25.10
Method 1	Female	25.00	28.00
Method 1	Female	24.00	27.00
Method 1	Female	23.50	26.50
Method 1	Female	24.70	27.10
Method 1	Female	24.20	26.00
Method 2	Female	24.50	27.20
Method 2	Female	23.00	26.20
Method 2	Female	23.00	26.20
Method 2	Female	24.30	27.50
Method 2	Female	24.80	26.00
Method 3	Female	26.00	28.50
Method 3	Female	27.00	27.00
Method 3	Female	26.00	29.00
Method 3	Female	25.00	26.70
Method 3	Female	23.00	25.10
Method 4	Female	23.00	25.10
Method 4	Female	26.00	27.20
Method 4	Female	26.00	25.20
Method 4	Female	23.00	25.10
Method 4	Female	23.00	25.10

115

Comparison of Groups (t-test)

A sample of adults was selected and randomly allocated to four teaching groups using different teaching methods, after which they were given a test.

Males/Females

One continuous variable – Final Test Score

103

Comparison of Groups (t-test)

```
> test <- read.csv("method-pretest-gender-test-score.csv", header = TRUE, colClasses = c("numeric", "factor", "factor", "numeric", "numeric"))
```

104

Comparison of Groups (t-test)

```
> test <- read.csv("method-pretest-gender-test-score.csv", header = TRUE, colClasses = c("numeric", "factor", "factor", "numeric", "numeric"))
```

```
R Console
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'station()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.test()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Previously saved workspace restored]

> method <- read.csv("method-pretest-gender-test-score.csv", header = TRUE, colClasses = c("numeric", "factor", "factor", "numeric", "numeric"))
> test <- read.csv("method-pretest-gender-test-score.csv", header = TRUE, colClasses = c("numeric", "factor", "factor", "numeric", "numeric"))
> test
  Method Gender Pre-Test Score Post-Test Score
1  Method 1 Male    25.00    28.00
2  Method 1 Male    24.00    27.00
3  Method 1 Male    23.50    26.50
4  Method 1 Male    24.70    27.10
5  Method 1 Male    24.20    26.00
6  Method 2 Male    24.50    27.20
7  Method 2 Male    23.00    26.20
8  Method 2 Male    23.00    26.20
9  Method 2 Male    24.30    27.50
10 Method 2 Male    24.80    26.00
11 Method 3 Male    26.00    28.50
12 Method 3 Male    27.00    27.00
13 Method 3 Male    26.00    29.00
14 Method 3 Male    25.00    26.70
15 Method 3 Male    23.00    25.10
16 Method 4 Male    23.00    25.10
17 Method 4 Male    26.00    27.20
18 Method 4 Male    26.00    25.20
19 Method 4 Male    23.00    25.10
20 Method 4 Male    23.00    25.10
21 Method 1 Female    25.00    28.00
22 Method 1 Female    24.00    27.00
23 Method 1 Female    23.50    26.50
24 Method 1 Female    24.70    27.10
25 Method 1 Female    24.20    26.00
26 Method 2 Female    24.50    27.20
27 Method 2 Female    23.00    26.20
28 Method 2 Female    23.00    26.20
29 Method 2 Female    24.30    27.50
30 Method 2 Female    24.80    26.00
31 Method 3 Female    26.00    28.50
32 Method 3 Female    27.00    27.00
33 Method 3 Female    26.00    29.00
34 Method 3 Female    25.00    26.70
35 Method 3 Female    23.00    25.10
36 Method 4 Female    23.00    25.10
37 Method 4 Female    26.00    27.20
38 Method 4 Female    26.00    25.20
39 Method 4 Female    23.00    25.10
40 Method 4 Female    23.00    25.10
```

104

Comparison of Groups (t-test)

```
> t.test(test$score ~ test$gender)
```

105

Comparison of Groups (t-test)

```
> t.test(test$score ~ test$gender)
```

```
R Console
30  Method 2 Female    24.2    29.2
31  Method 3 Female    24.5    29.1
32  Method 3 Female    24.0    28.9
33  Method 3 Female    27.0    29.2
34  Method 3 Female    22.0    29.0
35  Method 3 Female    25.0    27.6
36  Method 4 Female    24.3    29.0
37  Method 4 Female    24.4    27.7
38  Method 4 Female    28.0    27.2
39  Method 4 Female    27.0    29.1
40  Method 4 Female    24.0    27.9
> t.test(test$score ~ test$gender)

Welch Two Sample t-test

data: test$score by test$gender
t = -0.30285, df = 38.199, p-value = 0.7642
alternative hypothesis: true difference in means between group Female and group
95 percent confidence interval:
 -3.221062  2.391062
sample estimates:
mean in group Female    mean in group Male
      27.100           27.515
```

105

115

116

117

118

119

120

Data for Analysis of Variance (One-Way ANOVA)

A sample of adults was selected and randomly allocated to four teaching groups using different teaching methods, after which they were given a test.

Group 1: Method 1

Group 2: Method 2

Group 3: Method 3

Group 4: Method 4

One categorical variable – Method

One continuous variable – Final Test Score

121

Analysis of Variance (One-Way ANOVA)

```
> anova1<-aov(test$score ~ test$method)
> summary(anova1)
```

```
R Console
> anova1<-aov(test$score ~ test$method)
> summary(anova1)
Data: test$score by test$method
t = -0.30205, df = 20.159, p-value = 0.7642
alternative hypothesis: true difference in means between group Female and group0
95 percent confidence interval:
-0.221062 2.991062
sample estimates:
mean in group Female mean in group Male
37.100 37.515

> anova1<-aov(test$score ~ test$method)
> summary(anova1)
Df Sum Sq Mean Sq F value Pr(>F)
test$method 3 344.2 114.75 11.13 2.89e-05 ***
Residuals 36 371.0 10.31
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> TukeyHSD(anova1)
Tukey multiple comparisons of means
95% family-wise confidence level

Fit: aov(formula = test$score ~ test$method)
$`test$method`
      diff      lwr      upr      p adj
Method 2-Method 1 2.56 -1.3066912 6.426691 0.2978604
Method 3-Method 1 6.19 2.3233088 10.056691 0.0006686
Method 4-Method 1 7.40 3.5333088 11.266691 0.0000537
Method 3-Method 2 3.63 -0.3366912 7.496691 0.0733277
Method 4-Method 2 4.84 0.9733088 8.706691 0.0093145
Method 4-Method 3 1.21 -2.6566912 5.076691 0.8336921
```

107

121

122

Analysis of Variance (One-Way ANOVA)

TukeyHSD(anova1)

```
R Console
> anova1<-aov(test$score ~ test$method)
> summary(anova1)
Df Sum Sq Mean Sq F value Pr(>F)
test$method 3 344.2 114.75 11.13 2.89e-05 ***
Residuals 36 371.0 10.31
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> TukeyHSD(anova1)
Tukey multiple comparisons of means
95% family-wise confidence level

Fit: aov(formula = test$score ~ test$method)
$`test$method`
      diff      lwr      upr      p adj
Method 2-Method 1 2.56 -1.3066912 6.426691 0.2978604
Method 3-Method 1 6.19 2.3233088 10.056691 0.0006686
Method 4-Method 1 7.40 3.5333088 11.266691 0.0000537
Method 3-Method 2 3.63 -0.3366912 7.496691 0.0733277
Method 4-Method 2 4.84 0.9733088 8.706691 0.0093145
Method 4-Method 3 1.21 -2.6566912 5.076691 0.8336921
```

123

Data for Analysis of Variance (Two-Way ANOVA)

A sample of adults was selected and randomly allocated to four teaching groups using different teaching methods, after which they were given a test.

Group 1: Method 1

Group 2: Method 2

Group 3: Method 3

Group 4: Method 4

Males/Females

Two categorical variables – Method & Gender

One continuous variable – Final Test Score

124

124

Analysis of Variance (Two-Way ANOVA)

```
> anova2<-aov(test$score ~ test$method +
test$gender + test$method:test$gender)
> summary(anova2)
```

```
R Console
> TukeyHSD(anova1)
Tukey multiple comparisons of means
95% family-wise confidence level

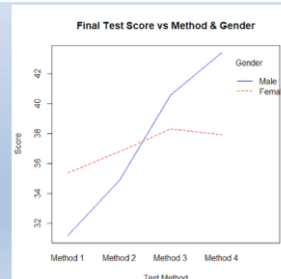
Fit: aov(formula = test$score ~ test$method)
$`test$method`
      diff      lwr      upr      p adj
Method 2-Method 1 2.56 -1.3066912 6.426691 0.2978604
Method 3-Method 1 6.19 2.3233088 10.056691 0.0006686
Method 4-Method 1 7.40 3.5333088 11.266691 0.0000537
Method 3-Method 2 3.63 -0.3366912 7.496691 0.0733277
Method 4-Method 2 4.84 0.9733088 8.706691 0.0093145
Method 4-Method 3 1.21 -2.6566912 5.076691 0.8336921

> anova2<-aov(test$score ~ test$method + test$gender + test$method:test$gender)
> summary(anova2)
Df Sum Sq Mean Sq F value Pr(>F)
test$method 3 344.2 114.75 15.910 1.65e-06 ***
test$gender 1 1.7 1.72 0.239 0.6204
test$method:test$gender 3 135.5 45.17 6.401 0.0016 **
Residuals 32 230.6 7.21
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

125

Analysis of Variance (Two-Way ANOVA)

```
> interaction.plot(test$method, test$gender, test$score,
col=c("red", "blue"), trace.label="Gender", xlab="Test
Method", ylab="Score", main= "Final Test Score vs Method &
Gender")
```



111

126

Analysis of Variance (Two-Way ANOVA)

> TukeyHSD(anova2)

```
R Console
> test$gender
      diff      lwr      upr      p adj
Male-Female 0.415 -1.31488 2.14488 0.628413

> test$method:test$gender
      diff      lwr      upr      p adj
Method 2:Female-Method 1:Female 1.42 -4.08199387 6.92199387 0.9895070
Method 3:Female-Method 1:Female 2.94 -5.142199387 8.462199387 0.4604973
Method 4:Female-Method 1:Female 2.59 -2.92199387 8.08199387 0.7917676
Method 1:Male-Method 1:Female -4.18 -9.68199387 1.32199387 0.2481314
Method 2:Male-Method 1:Female -0.45 -5.95199387 5.02199387 0.9995911
Method 3:Male-Method 1:Female 8.24 -0.26199387 10.74199387 0.0710472
Method 4:Male-Method 1:Female 8.04 2.83000613 13.84199387 0.0010074
Method 3:Female-Method 2:Female 1.54 -3.94199387 7.04199387 0.9323775
Method 4:Female-Method 2:Female 1.16 -4.34199387 6.64199387 0.9949137
Method 1:Male-Method 2:Female -5.60 -11.10199387 -0.09500613 0.0436885
Method 2:Male-Method 2:Female -1.90 -7.40199387 3.60199387 0.9478426
Method 3:Male-Method 2:Female 3.82 -1.68199387 9.32199387 0.3513162
Method 4:Male-Method 2:Female 6.62 1.11800613 12.12199387 0.0097786
Method 1:Female-Method 3:Female -0.38 -5.88199387 5.12199387 0.9998982
Method 2:Female-Method 3:Female -7.14 -12.64199387 -1.63800613 0.0043336
Method 3:Male-Method 3:Female -3.44 -8.94199387 2.06199387 0.4817462
Method 4:Male-Method 3:Female 2.28 -3.22199387 7.78199387 0.8784904
Method 4:Male-Method 3:Female 8.08 -0.42199387 10.88199387 0.0874139
```

127

Data for Analysis of Variance (One-Way ANCOVA)

A sample of adults was selected and randomly allocated to four teaching groups using different teaching methods, after which they were given a test.

Group 1: Method 1

Group 2: Method 2

Group 3: Method 3

Group 4: Method 4

Pre-Test/Final Test

One categorical variable – Method

Two continuous variables – Pre-Test Score & Final Test Score

128

Analysis of Variance (One-Way ANCOVA)

> ancova1<-aov(test\$score ~ test\$method + test\$pretest)

```
R Console
> ancova1<-aov(test$score ~ test$method + test$pretest)
> summary(ancova1)
Df Sum Sq Mean Sq F value Pr(>F)
test$method 3 344.2 114.75 10.828 3.52e-05 ***
test$pretest 1 0.1 0.13 0.012 0.914
Residuals 35 370.9 10.60
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> ancova1<-aov(test$score ~ test$method + test$pretest)
> summary(ancova1)
Df Sum Sq Mean Sq F value Pr(>F)
test$method 3 344.2 114.75 10.828 3.52e-05 ***
test$pretest 1 0.1 0.13 0.012 0.914
Residuals 35 370.9 10.60
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

129

Analysis of Variance (One-Way ANCOVA)

> TukeyHSD(anova1)

```
R Console
> ancova1<-aov(test$score ~ test$method + test$pretest)
> summary(ancova1)
Df Sum Sq Mean Sq F value Pr(>F)
test$method 3 344.2 114.75 10.828 3.52e-05 ***
test$pretest 1 0.1 0.13 0.012 0.914
Residuals 35 370.9 10.60
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> TukeyHSD(ancova1)
Tukey multiple comparisons of means
95% family-wise confidence level

Fit: aov(formula = test$score ~ test$method)

> test$method
      diff      lwr      upr      p adj
Method 2-Method 1 2.56 -1.30464912 6.4264912 0.2976894
Method 3-Method 1 4.19 2.3233088 10.0566912 0.0004686
Method 4-Method 1 7.40 3.5333088 11.2666912 0.0000337
Method 3-Method 2 3.63 -0.2364912 7.4964912 0.0723277
Method 4-Method 2 4.84 0.8733088 7.7066912 0.0093148
Method 4-Method 3 1.21 -2.4564912 5.0766912 0.8364921
```

130

Data for Analysis of Variance (Two-Way ANCOVA)

A sample of adults was selected and randomly allocated to four teaching groups using different teaching methods, after which they were given a test.

Group 1: Method 1

Group 2: Method 2

Group 3: Method 3

Group 4: Method 4

Pre-Test/Final Test

Male/Female

Two categorical variables – Method & Gender

Two continuous variables – Pre-Test Score & Final Test Score

131

Analysis of Variance (Two-Way ANOVA)

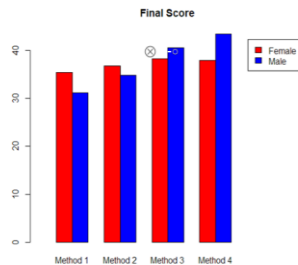
> tapply(test\$score, list(test\$gender, test\$method), mean)

```
R Console
> tapply(test$score, list(test$gender, test$method), mean)
Female 35.36 36.78 38.32 37.94
Male 31.18 34.08 40.60 43.40
```

132

Analysis of Variance (Two-Way ANOVA)

```
> plot <- tapply(test$score, list(test$gender, test$method),
mean)
> barplot(plot, main="Final Score", beside=TRUE, xlim=c(0,16),
legend=TRUE, col=c("red", "blue"))
```



133

Analysis of Variance (Two-Way ANCOVA)

```
> test <- read.csv("method-pretest-gender-test-
score.csv", header = TRUE, colClasses =
c("numeric", "factor", "factor", "numeric", "numeric"))
```

```
R Console
Natural language support but running in an English locale
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Previously saved workspace restored]

> test <- read.csv("method-pretest-gender-test-score.csv", header = TRUE, colClasses =
c("numeric", "factor", "factor", "numeric", "numeric"))
> anova2 <- aov(test$score ~ test$method + test$gender + test$pretest + test$method:test$gender)
> summary(anova2)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
test\$method	3	346.2	115.4	25.642	2.24e-06 ***
test\$gender	1	1.7	1.72	0.235	0.63141
test\$pretest	1	0.2	0.20	0.022	0.88067
test\$method:test\$gender	3	161.7	53.9	4.439	0.00462 **
Residuals	31	227.4	7.34		

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

134

Analysis of Variance (Two-Way ANCOVA)

```
> anova2 <- aov(test$score ~
test$method + test$gender + test$pretest +
test$method:test$gender)
> summary(anova2)
```

```
R Console
Natural language support but running in an English locale
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Previously saved workspace restored]

> test <- read.csv("method-pretest-gender-test-score.csv", header = TRUE, colClasses =
c("numeric", "factor", "factor", "numeric", "numeric"))
> anova2 <- aov(test$score ~ test$method + test$gender + test$pretest + test$method:test$gender)
> summary(anova2)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
test\$method	3	346.2	115.4	25.642	2.24e-06 ***
test\$gender	1	1.7	1.72	0.235	0.63141
test\$pretest	1	0.2	0.20	0.022	0.88067
test\$method:test\$gender	3	161.7	53.9	4.439	0.00462 **
Residuals	31	227.4	7.34		

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

135

Analysis of Variance (Two-Way ANCOVA)

```
> TukeyHSD(anova2, "test$method")
```

```
R Console
1: In replications(paste("~", xx), data = mf) :
non-factors ignored: test$pretest
2: In TukeyHSD.anova2(1) :
'which' specified some non-factors which will be dropped
> TukeyHSD(anova2, "test$method")
Error in TukeyHSD(anova2, "test$method") : object 'anova2' not found
> TukeyHSD(anova2, "test$method")
99% family-wise confidence level

Fit: aov(formula = test$score ~ test$method + test$gender + test$pretest + test$method:test$gender, data = test)
Df testMethod      diff      lwr      upr      p adj
Method 2-Method 1  2.54 -0.7274332  5.874233 0.1712114
Method 3-Method 1  4.19  2.9025768  5.477423 0.0000086
Method 4-Method 1  7.40  4.1125768  10.697423 0.0000052
Method 3-Method 2  1.63  0.3425768  2.917423 0.0239983
Method 4-Method 2  4.84  1.5525768  8.127423 0.0019930
Method 4-Method 3  1.21 -2.0774232  4.497423 0.7510044

Warning message:
In replications(paste("~", xx), data = mf) :
non-factors ignored: test$pretest
```

136

Analysis of Variance (Two-Way ANCOVA)

```
> TukeyHSD(anova2)
```

```
R Console
$ test$gender:      diff      lwr      upr      p adj
Male-Female 0.415 -1.331011 2.161011 0.6314094

$ test$method:test$gender:      diff      lwr      upr      p adj
Method 2:Female-Method 1:Female  1.4904050 -4.0027199  7.0415294 0.3072040
Method 3:Female-Method 1:Female  3.0297913 -2.5093943  8.5952210 0.4444416
Method 4:Female-Method 1:Female  2.6208764 -2.9401541  8.1813072 0.7875594
Method 1:Male-Method 1:Female  -4.1948465 -8.6531753  0.4644800 0.2009005
Method 2:Male-Method 1:Female  -0.4550455 -4.9157730  4.0056820 0.9999941
Method 3:Male-Method 1:Female  5.2586219 -0.3021087  10.8183264 0.0747149
Method 4:Male-Method 1:Female  8.056385  3.5241061  12.5886745 0.0011221
Method 1:Female-Method 2:Female  1.1493326 -4.0119520  6.3106172 0.9333953
Method 4:Female-Method 2:Female  1.1401176 -4.4206312  6.7008593 0.9973566
Method 1:Male-Method 2:Female  -5.8750455 -11.1377730 -0.6123170 0.0490448
Method 2:Male-Method 2:Female  -1.3550403 -7.4962350  4.7861543 0.9495200
Method 3:Male-Method 2:Female  3.751631 -1.7854764  9.3335039 0.1704770
Method 4:Male-Method 2:Female  6.404372  1.0936479  11.7150971 0.012511
Method 1:Female-Method 3:Female  -0.0092180 -5.9699474  5.9607114 0.9999972
Method 1:Male-Method 3:Female  -7.1247701 -12.6511087 -1.5983674 0.0050053
Method 2:Male-Method 3:Female  -3.4848369 -9.0456764 -2.0759389 0.4766485
Method 3:Male-Method 3:Female  2.2259304 -3.2490020  7.7468619 0.8921988
Method 4:Male-Method 3:Female  5.0550455 -0.5050522  10.6157730 0.0942714
Method 1:Male-Method 4:Female  -6.7151631 -12.2759939 -1.1543326 0.0095227
```

137

Analysis of Variance (Two-Way ANCOVA)

```
library("effects")
result <- aov(score ~ pretest + method*gender, data=test)
effect_method = effect("method*gender", result)
ata.frame(effect_method)
```

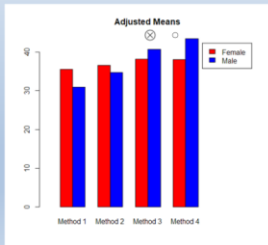
```
R Console
>
> tapply(test$score, list(test$gender, test$method), mean)
      Female      Male
Method 1 35.34 36.70 38.32 37.94
Method 2 31.15 34.00 40.40 43.40

> result <- aov(score ~ pretest + method*gender, data=test)
> effect_method = effect("method*gender", result)
> data.frame(effect_method)
  method gender  F15    se lower upper
1 Method 1 Female 35.9523 1.259494 33.02406 38.14440
2 Method 2 Female 34.9587 1.241004 32.46119 37.12605
3 Method 3 Female 38.2485 1.216322 35.76306 40.72527
4 Method 4 Female 38.0579 1.223494 35.54204 40.55279
5 Method 1 Male 30.9887 1.241004 28.46119 33.52605
6 Method 2 Male 34.9045 1.216322 32.32396 37.22527
7 Method 3 Male 40.7179 1.223494 37.22204 43.21279
8 Method 4 Male 43.4932 1.217460 40.97762 45.90864
```

138

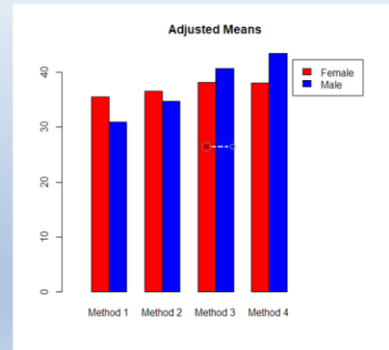
Analysis of Variance (Two-Way ANCOVA)

```
> means <- data.frame(effect_method)
> plot <- tapply(means$fit, list(means$gender,
means$method), mean)
> barplot(plot, main="Final Score, controlling for Pre-Test",
beside=TRUE, xlim=c(0,20), legend=TRUE)
```



139

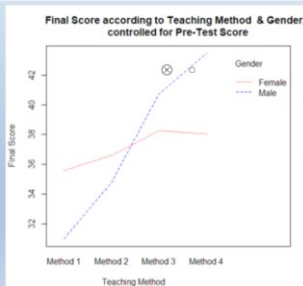
Analysis of Variance (Two-Way ANCOVA)



140

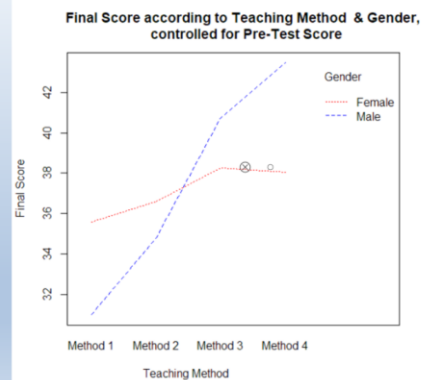
Analysis of Variance (Two-Way ANCOVA)

```
> interaction.plot(means$method, means$gender, means$fit,
col=c("red", "blue"), trace.label="Gender", xlab="Teaching
Method", ylab="Final Score", main= "Final Score according to
Teaching Method & Gender, n controlled for Pre-Test Score")
```



141

Analysis of Variance (Two-Way ANCOVA)



142

Presenting the Results from Two-Way ANCOVA

A 2 by 2 between-groups analysis of covariance was conducted to assess the effectiveness of four methods of teaching English for male and female participants. The independent variables were the teaching method (method 1, 2, 3 or 4) and gender. The dependent variable was scores on the final test, administered following completion of the intervention programs. Scores on the pre-test administered prior to the commencement of the programs were used as a covariate to control for individual differences.

After adjusting for pre-test scores, there was a significant interaction effect: $F(1, 31) = 6.439, p < .001$, with a large effect size (partial eta squared = .384). These results suggest that males and females respond differently to the four methods. Males showed a more substantial increase in the final test score after participation the class using methods 1 and 2 than methods 3 & 4, whereas females did not differ so much.

143

$\chi^2 - \chi$ Squared

```
R Console
> smokers <- matrix(c(18,35,24,14), ncol=2)
> chisq.test(smokers)

Pearson's Chi-squared test with Yates' continuity correction

data: smokers
X-squared = 6.4615, df = 1, p-value = 0.01102

> smokers
     [,1] [,2]
[1,] 18  24
[2,] 35  14

> smokers
     [,1] [,2]
[1,] 18  24
[2,] 35  14

> smokers <- matrix(c(18,35,24,14), ncol=2)
> rownames(smokers) <- c("Male", "Female")
> colnames(smokers) <- c("Smoke", "Not Smoke")

> smokers
      Smoke Not Smoke
Male      18       24
Female   35       14
```

57

143

144

$\chi^2 - \chi$ Squared

```
R Console
X-squared = 6.4615, df = 1, p-value = 0.01102

> smokers
      [,1] [,2]
[1,] 18  24
[2,] 35  14

> smokers
      [,1] [,2]
[1,] 18  24
[2,] 35  14
> smokers <- matrix(c(18,35,24,14), ncol=2)
> rownames(smokers) <- c("Male", "Female")
> colnames(smokers) <- c("Smoke", "Not Smoke")
> smokers
      Smoke Not Smoke
Male    18      24
Female  35      14
> chisq.test(smokers)

Pearson's Chi-squared test with Yates' continuity correction

data: smokers
X-squared = 6.4615, df = 1, p-value = 0.01102
```

145

```
[1,] 18 24
[2,] 35 14
> smokers
      [,1] [,2]
[1,] 18  24
[2,] 35  14
> smoker
> rownames
> colnames
> smoker
      S
Male
Female
> chisq.
```

Pearson's Chi-squared test with Yates' continuity corr

```
data: smokers
X-squared = 6.4615, df = 1, p-value = 0.01102
> |
```

146

$\chi^2 - \chi$ Squared

A B C D				A B C D E				A B C D E			
ID	Sex	Smoke		34	33 Male	Not Smoke		64	63 Female	Smoke	
1	1 Male	Smoke		35	34 Male	Not Smoke		65	64 Female	Smoke	
2	2 Male	Smoke		36	35 Male	Not Smoke		66	65 Female	Smoke	
3	3 Male	Smoke		37	36 Male	Not Smoke		67	66 Female	Smoke	
4	4 Male	Smoke		38	37 Male	Not Smoke		68	67 Female	Smoke	
5	5 Male	Smoke		39	38 Male	Not Smoke		69	68 Female	Smoke	
6	6 Male	Smoke		40	39 Male	Not Smoke		70	69 Female	Smoke	
7	7 Male	Smoke		41	40 Male	Not Smoke		71	70 Female	Smoke	
8	8 Male	Smoke		42	41 Male	Not Smoke		72	71 Female	Smoke	
9	9 Male	Smoke		43	42 Male	Not Smoke		73	72 Female	Smoke	
10	10 Male	Smoke		44	43 Female	Smoke		74	73 Female	Smoke	
11	11 Male	Smoke		45	44 Female	Smoke		75	74 Female	Smoke	
12	12 Male	Smoke		46	45 Female	Smoke		76	75 Female	Smoke	
13	13 Male	Smoke		47	46 Female	Smoke		77	76 Female	Smoke	
14	14 Male	Smoke		48	47 Female	Smoke		78	77 Female	Smoke	
15	15 Male	Smoke		49	48 Female	Smoke		79	78 Female	Not Smoke	
16	16 Male	Smoke		50	49 Female	Smoke		80	79 Female	Not Smoke	
17	17 Male	Smoke		51	50 Female	Smoke		81	80 Female	Not Smoke	
18	18 Male	Smoke		52	51 Female	Smoke		82	81 Female	Not Smoke	
19	19 Male	Not Smoke		53	52 Female	Smoke		83	82 Female	Not Smoke	
20	20 Male	Not Smoke		54	53 Female	Smoke		84	83 Female	Not Smoke	
21	21 Male	Not Smoke		55	54 Female	Smoke		85	84 Female	Not Smoke	
22	22 Male	Not Smoke		56	55 Female	Smoke		86	85 Female	Not Smoke	
23	23 Male	Not Smoke		57	56 Female	Smoke		87	86 Female	Not Smoke	
24	24 Male	Not Smoke		58	57 Female	Smoke		88	87 Female	Not Smoke	
25	25 Male	Not Smoke		59	58 Female	Smoke		89	88 Female	Not Smoke	
26	26 Male	Not Smoke		60	59 Female	Smoke		90	89 Female	Not Smoke	
27	27 Male	Not Smoke		61	60 Female	Smoke		91	90 Female	Not Smoke	
28	28 Male	Not Smoke		62	61 Female	Smoke		92	91 Female	Not Smoke	

147

$\chi^2 - \chi$ Squared

```
R Console
```

```
Type 'q()' to quit R.
```

```
> [Previously saved workspace restored]
```

```
> smoke <- read.csv("smoke.csv", header = TRUE, colClasses = c("numeric", "factor"))
```

```
> smoke
```

	1: ID	Sex	Smoke
1	1	Male	Smoke
2	2	Male	Smoke
3	3	Male	Smoke
4	4	Male	Smoke
5	5	Male	Smoke
6	6	Male	Smoke
7	7	Male	Smoke
8	8	Male	Smoke
9	9	Male	Smoke
10	10	Male	Smoke
11	11	Male	Smoke
12	12	Male	Smoke
13	13	Male	Smoke
14	14	Male	Smoke
15	15	Male	Smoke
16	16	Male	Smoke
17	17	Male	Smoke
18	18	Male	Smoke
19	19	Male	Hot Smoke
20	20	Male	Hot Smoke
21	21	Male	Hot Smoke

148

$\chi^2 - \chi$ Squared



R Console

```

> table(smoke$sex, smoke$Smoke)

```

	Not Smoke	Smoke
Female	14	35
Male	24	18

149

$\chi^2 - \chi$ Squared

```
R Console
```

```
74 74 Female Smoke
75 75 Female Smoke
76 76 Female Smoke
77 77 Female Smoke
78 78 Female Not Smoke
79 79 Female Not Smoke
80 80 Female Not Smoke
81 81 Female Not Smoke
82 82 Female Not Smoke
83 83 Female Not Smoke
84 84 Female Not Smoke
85 85 Female Not Smoke
86 86 Female Not Smoke
87 87 Female Not Smoke
88 88 Female Not Smoke
89 89 Female Not Smoke
90 90 Female Not Smoke
91 91 Female Not Smoke
> HfSmoke <- table(smoke$sex, smoke$Smoke)
> HfSmoke

      Not Smoke Smoke
Female      14     35
Male       24     10
>
```

150

$\chi^2 - \chi$ Squared

```
R Console
#1 #1 Female Not Smoke
#2 #2 Female Not Smoke
#3 #3 Female Not Smoke
#4 #4 Female Not Smoke
#5 #5 Female Not Smoke
#6 #6 Female Not Smoke
#7 #7 Female Not Smoke
#8 #8 Female Not Smoke
#9 #9 Female Not Smoke
#0 #0 Female Not Smoke
#1 #1 Female Not Smoke
> MFsmoke <- table(smoke$sex, smoke$smoke)
> MFsmoke

      Not Smoke Smoke
Female      14      35
Male       24      10
> chisq.test(MFsmoke)

Pearson's Chi-squared test with Yates' continuity correction

data: MFsmoke
X-squared = 6.4615, df = 1, p-value = 0.01102
> |
4
```

151

$\chi^2 - \chi$ Squared

```
R Console
#1 #1 Female Not Smoke
#2 #2 Female Not Smoke
#3 #3 Female Not Smoke
#4 #4 Female Not Smoke
#5 #5 Female Not Smoke
#6 #6 Female Not Smoke
#7 #7 Female Not Smoke
#8 #8 Female Not Smoke
#9 #9 Female Not Smoke
#0 #0 Female Not Smoke
#1 #1 Female Not Smoke
> MFsmoke <- table(smoke$sex, smoke$smoke)
> MFsmoke

      Not Smoke Smoke
Female      14      35
Male       24      10
> chisq.test(MFsmoke)

Pearson's Chi-squared test with Yates' continuity correction

data: MFsmoke
X-squared = 6.4615, df = 1, p-value = 0.01102
> |
4
```

152

Describing Data using Statistics

Description

- Central Tendency
- Dispersion

Inference

- Exploring Relationships
 - Correlation
 - Regression
- Comparing Groups
 - t-Test
 - ANOVA
 - One-Way
 - Two-Way
 - ANCOVA

153

153

References & Further Reading

- Brown, J. D. & Rogers, T. S. (2002). *Doing second language research*. Oxford: Oxford University Press.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York: Routledge.
- Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford: Oxford University Press.
- De Vries, A. & Meys, J. (2015). *R for dummies* (2nd ed.). London: Wiley.
- De Winter, J. C. F. & Dodou, D. (2012). Five point Likert items: t-test versus Mann-Whitney-Wilcoxon. *Practical Assessment, Research and Evaluation*, 15(11).
- Field, A., Miles, J. & Field, Z. (2013). *Discovering statistics using R*. London: Sage.
- Hand, D. J. (2008). *Statistics: A very short introduction*. Oxford: Oxford University Press.
- Hatch, E. & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics*. Boston, MA: Heinle & Heinle.
- Kabacoff, R. I. (2015). *R in action* (2nd ed.). Shelter Island, NY: Manning.
- Larson-Hall, J. (2016). *A guide to doing statistics in second language research using SPSS and R* (2nd ed.). London: Routledge.
- Lowie, W. & Seton, B. (2013). *Essential statistics for applied linguistics*. London: Palgrave Macmillan.
- Popham, W. J. & Sirotnik, K. A. (1973). *Educational statistics: Use and interpretation* (2nd ed.). New York: Harper & Row.
- Salkind, N. J. (2011). *Statistics for people who (think they) hate statistics* (4th ed.). London: Sage.
- Saunders, M. & Lewis, P. (2012). *Doing research in business and management*. London: FT/Prentice Hall.
- Saunders, M., Lewis, P. & Thornhill, A. (2016). *Research methods for business students* (7th ed.). London: FT/Prentice Hall.
- Shavelson, R. (1988). *Statistical reasoning for the behavioral sciences*. Boston: Allyn and Bacon.
- Speigelhalter, D. (2020). *The art of statistics: Learning from data*. London: Pelican.

154

154