Research Methods

Confidence Intervals with R

Andy Gillett

1

1

Statistical Significance

For many statistically oriented social scientists, quantitative analysis is virtually synonymous with significance testing. The whole point and purpose of the exercise is taken to be 'have we got a significant result?' Is $p < 0.05$?' This refers to *statistical significance.*

(Robson, 2011, p. 446)

2

2

Statistical Significance

If a particular result is marked as significant, we can utter a sigh of relief as this means that the observed phenomenon represents a significant departure from what might be expected by chance alone. That is, it can be assumed to be real and we can report it. However, if a particular result is not significant, we must not report it (for example, if the mean scores of two groups are not significantly different, we *must not* say that one is larger than the other even if the descriptive statistics do show some difference!).

(Dörnyei, 2007, p. 210)

3

3

Statistical Significance

It is important to stress here that statistical significance is *not* the final answer because even though a result may be statistically significant (i.e. reliable and true to the larger population), it may not be valuable. 'Significance' in the statistical sense only means 'probably true' rather than 'important'.

Dörnyei (2007, p. 210)

4

4

Effect Size

One way that you can assess the importance of your finding is to calculate the 'effect size' (also known as 'strength of association'). This is a set of statistics that indicates the relative magnitude of the differences between means, or the amount of the total variance in the dependent variable that is predictable from knowledge of the levels of the independent variable (Tabachnick & Fidell, 2013, p. 54).

(Pallant, 2016, p. 212)

5

5

Confidence Interval

While 'statistical significance' only gives us a yes/no answer about a result, the 'confidence interval' is more informative because it provides a range of scores (between two boundaries) that can be considered significant for the population. In the social sciences the usual confidence level is 95 per cent, and a 95 per cent confidence interval of a mean score can be interpreted as a range of scores that will contain the population mean with a 0.95 probability.

Dörnyei (2007, p. 210)

6

6

Confidence Interval

The Publication Manual of the American Psychological Association (APA, 2020, p. 84), which is generally seen as the most important source of guidelines on how to present our findings in applied linguistics, for example states:

> The reporting of confidence intervals ... can be an extremely effective way of reporting results. Because confidence intervals combine information on location and precision and can often be directly used to infer significance levels, they are, in general, the best reporting strategy. The use of confidence intervals is therefore strongly recommended.

7

7

Confidence Interval

A confidence interval is a range of plausible values for the index or parameter being estimated.

Ellis (2010, p. 16)

8

8

Example: Means

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

9

9

Means

Save as “MFheights.csv”

MFheights.csv - Notepad

File Edit View

```
Gender,Height
1,1.89
1,1.64
1,1.81
1,1.94
1,1.93
1,1.74
1,1.82
1,1.91
2,1.75
2,1.59
2,1.88
2,1.65
2,1.7
2,1.69
2,1.68
2,1.8
```

10

10

Means

Import into R

> MFheights <- read.csv("MFheights.csv", header = TRUE, colClasses = c(“numeric”, "factor", "numeric"))
> MFheights

```
> MFheights
   ID Gender Height
1   1   Male   1.89
2   2   Male   1.64
3   3   Male   1.81
4   4   Male   1.94
5   5   Male   1.93
6   6   Male   1.74
7   7   Male   1.82
8   8   Male   1.91
9   9 Female   1.61
10 10 Female   1.59
11 11 Female   1.52
12 12 Female   1.65
13 13 Female   1.64
14 14 Female   1.69
15 15 Female   1.68
16 16 Female   1.45
>
```

11

11

Example: Means

Look at data

> summary(MFheights)

```
> summary(MFheights)
       ID           Gender       Height
 Min.   : 1.00   Female:8   Min.   :1.450
 1st Qu.: 4.75   Male  :8   1st Qu.:1.633
 Median : 8.50              Median :1.685
 Mean   : 8.50              Mean   :1.719
 3rd Qu.:12.25              3rd Qu.:1.837
 Max.   :16.00              Max.   :1.940
```

> mean(MFheights$Height)

```
> mean(MFheights$Height)
[1] 1.719375
```

12

12

Example: Means

Using linear regression

```
> l.model <- lm(Height ~ 1, MFheights)
> l.model
```

```
> l.model <- lm(Height ~ 1, MFheights)
> l.model

Call:
lm(formula = Height ~ 1, data = MFheights)

Coefficients:
(Intercept)
      1.719
```

13

Example: Means

```
> confint(l.model, level=0.95)
```

```
> confint(l.model, level=0.95)
               2.5 %   97.5 %
(Intercept) 1.639375 1.799375
```

So the mean height of the population is 1.719, CI 95%(1.64, 1.80)

14

14

Example: Means

```
> confint(l.model, level=0.90)
```

```
> confint(l.model, level=0.90)
                  5 %     95 %
(Intercept) 1.653578 1.785172
>
```

So the mean height of the population is 1.719, CI 90%(1.65, 1.78)

15

15

Example: Means

```
> confint(l.model, level=0.99)
```

```
>
> confint(l.model, level=0.99)
               0.5 %   99.5 %
(Intercept) 1.608776 1.829974
>
>
```

So the mean height of the population is 1.719, CI 99%(1.61, 1.83)

16

16

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.

17

17

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

18

18

Correlation

Relationship between ability in maths and sports.

	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						

Save as "maths-sport.csv"

19

19

Correlation

> Maths.Sport <- read.csv("maths-sport.csv", header = TRUE, colClasses = c("numeric", "numeric", "numeric"))
> Maths.Sport

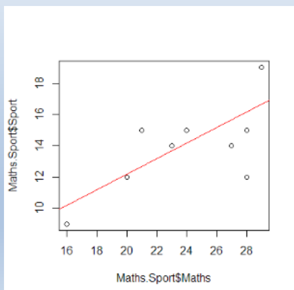
```
> Maths.Sport
   ID Sport Maths
1   1    15    24
2   2    19    29
3   3    12    20
4   4    14    27
5   5     9    16
6   6    15    28
7   7    12    28
8   8    14    23
9   9    15    21
10 10    19    29
```

20

20

Correlation

> plot(Maths.Sport$Maths,Maths.Sport$Sport)
> abline(lm(Maths.Sport$Sport ~ Maths.Sport$Maths), col='red')



21

21

Correlation

> cor(Maths.Sport$Maths,Maths.Sport$Sport)
> cor.test(Maths.Sport$Maths,Maths.Sport$Sport)

```
> 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:
 0.1756842 0.9300985
sample estimates:
     cor
0.725103
```

Yes, there is a strong significant relationship (r = 0.725, p <0.05).

22

22

Confidence Interval: Correlation

A **confidence interval for a correlation coefficient** is a range of values that is likely to contain a population correlation coefficient with a certain level of confidence.

23

23

Correlation

```
> 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:
 0.1756842 0.9300985
sample estimates:
     cor
0.725103
```

Yes, there is a strong significant relationship (r = 0.725, p <0.05, CI 95%(0.18, 0.93).

24

24

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:

Data

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

t-Test

Is there a difference between oral announcement of objects and memory?

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

Save as "Announcement.csv"

t-Test

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

```
> summary(Announcement)
       ID                      Method        Recall
 Min.   : 1.00   Announcement   :10   Min.   :3.000
 1st Qu.: 6.25   No Announcement:12   1st Qu.:4.000
 Median :11.50                        Median :5.000
 Mean   :11.50                        Mean   :5.727
 3rd Qu.:16.75                        3rd Qu.:7.750
 Max.   :22.00                        Max.   :9.000
>
```

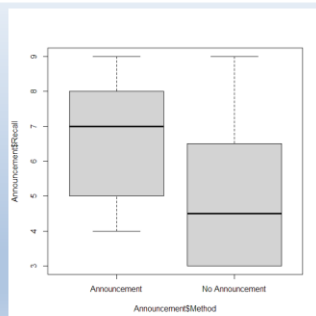

t-Test

```
> tapply(Announcement$Recall,
Announcement$Method, mean)
```

```
>
>
> tapply(Announcement$Recall, Announcement$Method, mean)
   Announcement No Announcement
            6.6             5.0
>
```


t-Test

```
> boxplot(Announcement$Recall ~ Announcement$Method)
```



t-Test

> t.test(Announcement$Recall ~ Announcement$Method)

```
> t.test(Announcement$Recall~Announcement$Method)

        Welch Two Sample t-test

data:  Announcement$Recall by Announcement$Method
t = 1.8526, df = 19.69, p-value = 0.07898
alternative hypothesis: true difference in means between group Announcement and $
95 percent confidence interval:
 -0.2033249  3.4033249
sample estimates:
   mean in group Announcement mean in group No Announcement
                          6.6                           5.0
```

No, there is no significant difference between oral announcement of objects and memory (p = 0.08).

No significant difference when CI crosses zero.

ANOVA: Comparing Groups (3 groups or more)

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.

Comparing Groups (3 groups or more)

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

Stopping Distance in Metres

ANOVA – One-Way

Is there a difference between amount of alcohol consumed and stopping distance?

ID	alcohol	stopping
1	No Alcohol	31.2
2	No Alcohol	27.4
3	No Alcohol	29.8
4	No Alcohol	33.5
5	No Alcohol	34
6	One Unit	31.2
7	One Unit	33.7
8	One Unit	34.3
9	One Unit	39.2
10	One Unit	36
11	Two Units	42.1
12	Two Units	40.6
13	Two Units	43.4
14	Two Units	37.9
15	Two Units	39
16	Three Units	46.7
17	Three Units	45.1
18	Three Units	43.2
19	Three Units	41.7
20	Three Units	40.3

Stopping distance in metres.
Save as"Stopping.csv"

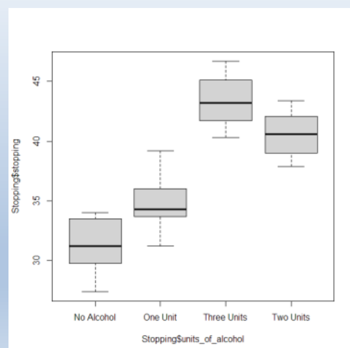
ANOVA – One-Way

Is there a difference between amount of alcohol consumed and stopping distance?

```
> Stopping
   ID 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
>
```


ANOVA – One-Way

> boxplot(Stopping$stopping ~ Stopping$alcohol)



ANOVA – One-Way

```
> anova <- aov(Stopping$stopping ~ Stopping$alcohol)
> summary(anova)
```

```
> 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
>
```

Yes ($p < 0.05$). Therefore there is a significant difference in stopping distance according to amount of alcohol drunk.

37

37

ANOVA – One-Way

Where are the differences?

```
> sum <-TukeyHSD(anova)
> sum
```

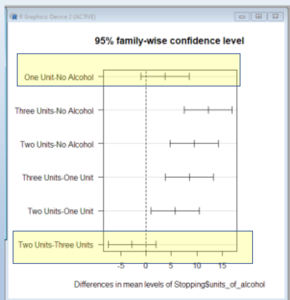
```
> TukeyHSD(anova)
  Tukey multiple comparisons of means
    95% family-wise confidence level

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
```

38

38

ANOVA – One-Way

```
> plot(sum, las=1)
```



No significant difference when lines cross zero.

39

39

References & Further Reading

American Psychological Association (2020). *Publication manual of the American Psychological Association* (7th ed.). Washington, DC: American Psychological Association.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* New York: Routledge

Crawley, M. J. (2005). *Statistics: An introduction using R.* London: Wiley.

Crawley, M. J. (2012). *The R book* (2nd ed.). London: Wiley.

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

Ellis, P. D. (2010). *The essential guide to effect sizes.* Cambridge: Cambridfe 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*. Shelter Island, NY: Manning.

Kline, R. (2004). *Beyond significance testing: Reforming data analysis methods in behavioral research.* Washington, DC: American Psychological Association.

Larson-Hall, J. (2012). How to run statistical analyses. In A. Mackey & S. M. Gass (Eds.), *Research methods in second language acquisition* (pp. 245-274). London: Wiley-Blackwell.

Larson-Hall, J. (2015). *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

Robson, C. (2011). *Real world research* (3rd ed.). London: Wiley.

Shavelson, R. (1988). *Statistical reasoning for the behavioral sciences*. Boston: Allyn and Bacon.

Tabachnick, B. G. & Fidell, L. S. (2007). *Using multivariate statistics.* (5th ed.). Boston: Allyn & Bacon.

40

40