Problem Set 7

JP

10/8/2021

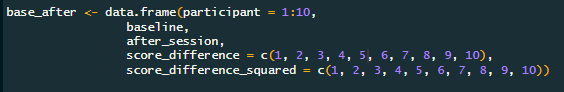
# Paired-Samples t-test

1. Get the difference between each pair of scores
2. Square each difference score
3. Get the mean difference
4. Get the variance/standard deviation of differences
5. Get the standard error of the mean differences
6. Get the t-obtained value
7. Get the degrees of freedom
8. Get the effect size
9. Find out if the t-obtained value is statistically significant
10. Report the t-test statistic finding. Remember to include the means and standard deviations for the two time points.

# Formulas

You are conducting an experiment where you are interested in how many words a one year old toddler can say. Your experimental manipulation is that every child will be given two two-hour long sessions to try and build a stronger vocabulary. You want to know if these lessons are enough for children to have a stronger stronger vocabulary than when they first arrived. You decided to get baseline data by having them say all the words they know and then you asked a second time after the vocabulary sessions. What test are you running and is the difference statistically significant?

**NOTE** once you fill in the values for score\_difference and score\_difference\_squared you will run everything from base\_after <- to the double parantheses )) like the picture below. The values in the screenshot are not the correct answers.



set.seed(093021)  
  
baseline <- rnorm(10, mean = 1.5, sd =.5)  
after\_session <- rnorm(10, mean = 4.1, sd = 2)  
  
baseline <- round(baseline, 2)  
after\_session <- round(after\_session, 2)  
  
base\_after <- data.frame(participant = 1:10,  
 baseline,  
 after\_session,  
 score\_difference = c("\_\_\_\_\_\_\_\_\_\_\_\_\_", "\_\_\_\_\_\_\_\_\_\_\_\_\_", "\_\_\_\_\_\_\_\_\_\_\_\_\_", "\_\_\_\_\_\_\_\_\_\_\_\_\_", "\_\_\_\_\_\_\_\_\_\_\_\_\_",  
 "\_\_\_\_\_\_\_\_\_\_\_\_\_", "\_\_\_\_\_\_\_\_\_\_\_\_\_", "\_\_\_\_\_\_\_\_\_\_\_\_\_", "\_\_\_\_\_\_\_\_\_\_\_\_\_", "\_\_\_\_\_\_\_\_\_\_\_\_\_"),  
 score\_difference\_squared = c("\_\_\_\_\_\_\_\_\_\_\_\_\_", "\_\_\_\_\_\_\_\_\_\_\_\_\_", "\_\_\_\_\_\_\_\_\_\_\_\_\_", "\_\_\_\_\_\_\_\_\_\_\_\_\_", "\_\_\_\_\_\_\_\_\_\_\_\_\_",  
 "\_\_\_\_\_\_\_\_\_\_\_\_\_", "\_\_\_\_\_\_\_\_\_\_\_\_\_", "\_\_\_\_\_\_\_\_\_\_\_\_\_", "\_\_\_\_\_\_\_\_\_\_\_\_\_", "\_\_\_\_\_\_\_\_\_\_\_\_\_"))  
  
base\_after

## participant baseline after\_session score\_difference score\_difference\_squared  
## 1 1 2.05 6.60 \_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_  
## 2 2 1.25 3.82 \_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_  
## 3 3 1.75 6.05 \_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_  
## 4 4 1.11 3.87 \_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_  
## 5 5 1.46 5.79 \_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_  
## 6 6 1.75 1.87 \_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_  
## 7 7 1.61 5.91 \_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_  
## 8 8 1.85 6.71 \_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_  
## 9 9 2.09 2.29 \_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_  
## 10 10 1.35 9.36 \_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_

base\_after$baseline

## [1] 2.05 1.25 1.75 1.11 1.46 1.75 1.61 1.85 2.09 1.35

base\_after$after\_session

## [1] 6.60 3.82 6.05 3.87 5.79 1.87 5.91 6.71 2.29 9.36

1. Get the difference between each pair of scores

6.60 - 2.05

## [1] 4.55

3.82 - 1.25

## [1] 2.57

6.05 - 1.75

## [1] 4.3

3.87 - 1.11

## [1] 2.76

5.79 - 1.46

## [1] 4.33

1.87 - 1.75

## [1] 0.12

5.99 - 1.61

## [1] 4.38

6.71 - 1.85

## [1] 4.86

2.29 - 2.09

## [1] 0.2

9.36 - 1.35

## [1] 8.01

1. Square each difference score

4.55^2

## [1] 20.7025

2.57^2

## [1] 6.6049

4.3^2

## [1] 18.49

2.76^2

## [1] 7.6176

4.33^2

## [1] 18.7489

.12^2

## [1] 0.0144

4.38^2

## [1] 19.1844

4.86^2

## [1] 23.6196

.2^2

## [1] 0.04

8.01^2

## [1] 64.1601

base\_after

## participant baseline after\_session score\_difference score\_difference\_squared  
## 1 1 2.05 6.60 4.55 20.7025  
## 2 2 1.25 3.82 2.57 6.6049  
## 3 3 1.75 6.05 4.30 18.4900  
## 4 4 1.11 3.87 2.76 7.6176  
## 5 5 1.46 5.79 4.33 18.7489  
## 6 6 1.75 1.87 0.12 0.0144  
## 7 7 1.61 5.91 4.30 18.4900  
## 8 8 1.85 6.71 4.86 23.6196  
## 9 9 2.09 2.29 0.20 0.0400  
## 10 10 1.35 9.36 8.01 64.1601

1. Get the mean difference

(4.55 + 2.57 + 4.3 + 2.76 + 4.33 + .12 + 4.38 + 4.86 + .2 + 8.01)/10

## [1] 3.608

1. Get the variance/standard deviation of differences

36.08^2

## [1] 1301.766

20.70 + 6.60 + 18.49 + 7.62 + 18.75 + .01 + 18.49 + 23.62 + .04 + 64.16

## [1] 178.48

1301.77/10

## [1] 130.177

178.49 - 130.18

## [1] 48.31

10 - 1

## [1] 9

48.31/9

## [1] 5.367778

48.31/9

## [1] 5.367778

1. Get the standard error of the mean differences

5.37/10

## [1] 0.537

sqrt(.54)

## [1] 0.7348469

1. Get the t-obtained value

3.6 - 0

## [1] 3.6

3.6/.73

## [1] 4.931507

1. Get the degrees of freedom

10-1

## [1] 9

1. Get the effect size

sqrt(5.37)

## [1] 2.317326

3.6/2.32

## [1] 1.551724

1. Find out if the t-obtained value is statistically significant

4.93 > 2.26

Therefore, our finding is statistically significant.

1. Report the t-test statistic finding. Remember to include the means and standard deviations for the two time points.

Note. Get the baseline and after means and standard deviations

2.05 + 1.25 + 1.75 + 1.11 + 1.46 + 1.75 + 1.61 + 1.85 + 2.09 + 1.35

## [1] 16.27

16.27/10

## [1] 1.627

# mean = 1.63  
  
2.05^2 + 1.25^2 + 1.75^2 + 1.11^2 + 1.46^2 + 1.75^2 + 1.61^2 + 1.85^2 + 2.09^2 + 1.35^2

## [1] 27.4589

16.27^2

## [1] 264.7129

264.71/10

## [1] 26.471

27.46 - 26.47

## [1] 0.99

.99/9

## [1] 0.11

sqrt(.11)

## [1] 0.3316625

# sd = 0.33

6.6 + 3.82 + 6.05 + 3.87 + 5.79 + 1.87 + 5.91 + 6.71 + 2.29 + 9.36

## [1] 52.27

52.27/10

## [1] 5.227

# mean = 5.23  
  
6.6^2 + 3.82^2 + 6.05^2 + 3.87^2 + 5.79^2 + 1.87^2 + 5.91^2 + 6.71^2 + 2.29^2 + 9.36^2

## [1] 319.5587

52.27^2

## [1] 2732.153

2732.15/10

## [1] 273.215

319.56 - 273.22

## [1] 46.34

46.34/9

## [1] 5.148889

sqrt(5.15)

## [1] 2.269361

# sd = 2.27

A paired-samples t-test was conducted looking at before and after vocabulary sessions for a sample of 10 toddlers. The difference in number of words spoken before the vocabulary sessions (*M* = 1.63, *SD* = 0.33) to after the sessions (*M* = 5.23, *SD* = 2.27) was found to result in a statistically significant difference; *t*(9) = 4.93, *p* < .05. The toddlers in this sample spoke significantly more words after the vocabulary sessions than before the sessions. The magnitude of the association was representative of having a small sample size, as the vocabulary sessions had a strong effect on the amount of words spoken (d = 1.55).

You are conducting an experiment where you are interested in getting smokers to stop smoking can help build lung capacity. You are going to test this by conducting an smoking cessation program for a year and having your sample run a mile. Before your program begins, you have every participant run a mile and get the times for each participant. After the year of the program to stop participants from smoking, you decide to have them run a mile again and track the times. What test are you running and is the difference statistically significant?

set.seed(093021)  
  
smoking <- rnorm(7, mean = 20, sd = 8.1)  
no\_smoking <- rnorm(7, mean = 8, sd = 5.3)  
  
smoking <- round(smoking, 2)  
no\_smoking <- round(no\_smoking, 2)  
  
smoking

## [1] 28.97 15.91 24.05 13.70 19.34 24.07 21.83

no\_smoking

## [1] 11.75 14.22 6.37 14.61 7.27 13.16 7.39

smoke\_df <- data.frame(participant = 1:7,  
 smoking,  
 no\_smoking,  
 score\_difference = c("\_\_\_\_\_\_\_\_\_\_\_\_\_", "\_\_\_\_\_\_\_\_\_\_\_\_\_", "\_\_\_\_\_\_\_\_\_\_\_\_\_", "\_\_\_\_\_\_\_\_\_\_\_\_\_", "\_\_\_\_\_\_\_\_\_\_\_\_\_",  
 "\_\_\_\_\_\_\_\_\_\_\_\_\_", "\_\_\_\_\_\_\_\_\_\_\_\_\_"),  
 score\_difference\_squared = c("\_\_\_\_\_\_\_\_\_\_\_\_\_", "\_\_\_\_\_\_\_\_\_\_\_\_\_", "\_\_\_\_\_\_\_\_\_\_\_\_\_", "\_\_\_\_\_\_\_\_\_\_\_\_\_", "\_\_\_\_\_\_\_\_\_\_\_\_\_",  
 "\_\_\_\_\_\_\_\_\_\_\_\_\_", "\_\_\_\_\_\_\_\_\_\_\_\_\_"))  
  
smoke\_df

## participant smoking no\_smoking score\_difference score\_difference\_squared  
## 1 1 28.97 11.75 \_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_  
## 2 2 15.91 14.22 \_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_  
## 3 3 24.05 6.37 \_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_  
## 4 4 13.70 14.61 \_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_  
## 5 5 19.34 7.27 \_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_  
## 6 6 24.07 13.16 \_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_  
## 7 7 21.83 7.39 \_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_

smoke\_df$smoking

## [1] 28.97 15.91 24.05 13.70 19.34 24.07 21.83

smoke\_df$no\_smoking

## [1] 11.75 14.22 6.37 14.61 7.27 13.16 7.39

1. Get the difference between each pair of scores

28.97 - 11.75

## [1] 17.22

15.91 - 14.22

## [1] 1.69

24.05 - 6.37

## [1] 17.68

13.7 - 14.61

## [1] -0.91

19.34 - 7.27

## [1] 12.07

24.07 - 13.16

## [1] 10.91

21.83 - 7.39

## [1] 14.44

1. Square each difference score

17.22^2

## [1] 296.5284

1.69^2

## [1] 2.8561

17.68^2

## [1] 312.5824

(-.91)^2

## [1] 0.8281

12.07^2

## [1] 145.6849

10.91^2

## [1] 119.0281

14.44^2

## [1] 208.5136

smoke\_df

## participant smoking no\_smoking score\_difference score\_difference\_squared  
## 1 1 28.97 11.75 17.22 296.5284  
## 2 2 15.91 14.22 1.69 2.8561  
## 3 3 24.05 6.37 17.68 312.5824  
## 4 4 13.70 14.61 -0.91 0.8281  
## 5 5 19.34 7.27 12.07 145.6849  
## 6 6 24.07 13.16 10.91 119.0281  
## 7 7 21.83 7.39 14.44 208.5136

1. Get the mean difference

(17.22 + 1.69 + 17.68 + (-.91) + 12.07 + 10.91 + 14.44)/7

## [1] 10.44286

1. Get the variance/standard deviation of differences

73.1^2

## [1] 5343.61

296.53 + 2.86 + 312.58 + .83 + 145.68 + 119.03 + 208.51

## [1] 1086.02

5343.61/7

## [1] 763.3729

1086.02 - 763.37

## [1] 322.65

7 - 1

## [1] 6

322.65/6

## [1] 53.775

1. Get the standard error of the mean differences

53.78/7

## [1] 7.682857

sqrt(7.68)

## [1] 2.771281

1. Get the t-obtained value

10.44 - 0

## [1] 10.44

10.44/2.77

## [1] 3.768953

1. Get the degrees of freedom

7 - 1

## [1] 6

1. Get the effect size

sqrt(53.78)

## [1] 7.333485

10.44/7.33

## [1] 1.424284

1. Find out if the t-obtained value is statistically significant

3.77 > 2.447

Therefore, our finding is statistically significant.

1. Report the t-test statistic finding. Remember to include the means and standard deviations for the two time points.

Note. Get the smoking and non-smoking means and standard deviations

28.97 + 15.91 + 24.05 + 13.70 + 19.34 + 24.07 + 21.83

## [1] 147.87

147.87/7

## [1] 21.12429

# mean = 21.12  
  
28.97^2 + 15.91^2 + 24.05^2 + 13.70^2 + 19.34^2 + 24.07^2 + 21.83^2

## [1] 3288.431

147.87^2

## [1] 21865.54

21865.54/7

## [1] 3123.649

3288.43 - 3123.65

## [1] 164.78

164.78/6

## [1] 27.46333

sqrt(27.46)

## [1] 5.240229

# sd = 5.24

11.75 + 14.22 + 6.37 + 14.61 + 7.27 + 13.16 + 7.39

## [1] 74.77

74.77/7

## [1] 10.68143

# mean = 10.68  
  
11.75^2 + 14.22^2 + 6.37^2 + 14.61^2 + 7.27^2 + 13.16^2 + 7.39^2

## [1] 874.9505

74.77^2

## [1] 5590.553

5590.553/7

## [1] 798.6504

874.95 - 798.65

## [1] 76.3

76.3/6

## [1] 12.71667

sqrt(12.72)

## [1] 3.566511

# sd = 3.57

A paired-samples t-test was conducted looking at the amount of time it took to run a mile in a sample of smokers. Participants were assessed before and after a smoking cessation program to reduce the frequency of smoking in the sample. There was a statistically significant difference in the time needed to complete the mile for participants; *t*(6) = 3.77, *p* < .05. Participants spent on average 21.12 minutes (*SD* = 5.24) to complete the mile before the cessation program. After the program’s completion, participants took on average 10.68 (*SD* = 3.57) to complete the mile. The strength of the association was strong between the reduction in smoking and amount of time to run a mile (d = 1.42).