What t-test are we running?

JP

10/6/2021

Steps for one-sample t-test

1. Get the sample mean
2. Get the population mean
3. Get the estimated variance/standard deviation
4. Get the standard error
5. Calculate the t-obtained value
6. Get the degrees of freedom
7. Get the effect size

Steps for independent samples t-test

1. Get the means of both groups/samples
2. Get the variances of both groups/samples
3. Get the group/sample sizes (n)
4. Get the pooled variance by getting the groups'/samples' variances averaged
5. Get the standard error of the differences
6. Calculate the t-obtained value
7. Get the degrees of freedom
8. Calculate the confidence intervals
9. Get the effect size

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

Examples of t-distribution

Alpha = .05, df = 27

Alpha = .01, df = 27

Alpha = .05, df = 80

Alpha = .05, df = 71

Alpha = .01, df = 42

Alpha = .01, df = 20

Alpha = .01, df = 19

Alpha = .01, df = 1

Alpha = .01, df = 1073

Alpha = .05, df = 14

Alpha = .05, df = 350

Alpha = .05, df = 100

Alpha = .05, df = 52

Alpha = .05, df = 7

Alpha = .05, df = 2

Formulas

$$S_x^2 = \frac{\Sigma(X - \overline{X})^2}{N - 1}$$

$$S_x = \sqrt{\frac{\Sigma(X - \overline{X})^2}{N - 1}}$$

$$S_x^2 = \frac{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}{N - 1}$$

$$S_x = \sqrt{\frac{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}{N-1}}$$

$$s_D^2 = \frac{\Sigma D^2 - \frac{(\Sigma D)^2}{N}}{N-1}$$

$$S_D^2 = \sqrt{\frac{\Sigma D^2 - \frac{(\Sigma D)^2}{N}}{N-1}}$$

$$S_{pool}^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{(n_1-1)+(n_2-1)}$$

$$S_{\overline{X}} = \sqrt{\frac{S_x^2}{N}}$$

$$S_{\overline{D}} = \sqrt{\frac{S_D^2}{N}}$$

$$S_{\overline{X_1}-\overline{X_2}} = \sqrt{(S_{pool}^2)(\frac{1}{n_1}+\frac{1}{n_2})}$$

$$df = N-1$$

$$df = (n_1-1)+(n_2-1)$$

$$t_{obt} = \frac{\overline{D} - \mu_D}{S_{\overline{D}}}$$

$$t_{obt} = \frac{\overline{X} - \mu}{S_{\overline{X}}}$$

$$t_{obt} = \frac{(\overline{X_1} - \overline{X_2}) - (\mu_1 - \mu_2)}{S_{\overline{X_1}-\overline{X_2}}}$$

$$d = \frac{(\overline{X_1} - \overline{X_2})}{\sqrt{S_2}}$$

$$d = \frac{\overline{X}}{\sqrt{S_X^2}}$$

$$d = \frac{\overline{D}}{\sqrt{S_D^2}}$$

Word Problems

Problem 1

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?

```
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
```

Problem 2

You are conducting an experiment where you are interested in typing errors in adults. You are specifically interested in whether using a child's typing game will help reduce the number of typing errors. You decide to split your participants into two groups, with one group getting the child's typing game and the other getting no help typing. You decide to test these groups to see which group had less typing errors at the end of the experiment. What test are you running and is the difference statistically significant?

```
set.seed(093021)

child_game <- rnorm(9, mean = 10, sd = 4.7)
no_help <- rnorm(12, mean = 30, sd = .99)

child_game <- round(child_game, 2)
no_help <- round(no_help, 2)

child_game
```

```
## [1] 15.20  7.63 12.35  6.34  9.61 12.36 11.06 13.32 15.52
```

```
no_help
```

```
##  [1] 29.70 31.24 29.86 30.96 29.89 30.84 28.90 30.90 31.29 29.10 32.60 29.04
```

```
mean(child_game)
```

```
## [1] 11.48778
```

```
mean(no_help)
```

```
## [1] 30.36
```

```
sd(child_game)
```

```
## [1] 3.162115
```

```
sd(no_help)
```

```
## [1] 1.124399
```

```
sum(child_game)
```

```
## [1] 103.39
```

```
# child game
15.20^2+  7.63^2+ 12.35^2+  6.34^2+  9.61^2+ 12.36^2+ 11.06^2+ 13.32^2+ 15.52^2
```

```
## [1] 1267.713
```

```r
# 1267.71


sum(no_help)
```

```
## [1] 364.32
```

```r
# no help
29.70^2+ 31.24^2+ 29.86^2+ 30.96^2+ 29.89^2+ 30.84^2+ 28.90^2+ 30.90^2+ 31.29^2+ 29.10^2+ 32.60^2+ 29.0
```

```
## [1] 11074.66
```

```r
# 11074.66

sd(child_game)^2
```

```
## [1] 9.998969
```

```r
sd(no_help)^2
```

```
## [1] 1.264273
```

Problem 3

You are conducting an experiment where you are interested in how many useless Disney facts participants can state. You are interested in whether going to Disneyland for a day will help with Disney knowledge. You decide one group gets to go to Disneyland and the other can look online for an hour. You decide to test these groups at the end of the day to see who knows more about Disney. What test are you running and is the difference statistically significant?

```r
set.seed(093021)

land <- rnorm(6, mean = 16, sd = 2.2)
google <- rnorm(9, mean = 18, sd = 5)

land <- round(land, 2)
google <- round(google, 2)

land
```

```
## [1] 18.44 14.89 17.10 14.29 15.82 17.11
```

```r
google
```

```
## [1] 19.13 21.53 23.87 16.46 24.24 17.31 22.86 17.43 22.23
```

Problem 4

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?

```r
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
```

```r
no_smoking
```

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

```r
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    _____________            _____________
```

```r
smoke_df$smoking
```

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

```r
smoke_df$no_smoking
```

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

Problem 5

You are interested in the amount driving infractions a student driver makes during their driving test. You have a sample of 15 individuals from a DMV and you know how many errors they have all made during their test. You know that in order to pass, you have to have no more than 5 errors during your driving test, as has been proven throughout the state to be an acceptable amount of errors. You are now interested in whether or not your sample made less or more errors than the state allowed amount. What test are you running and is the difference statistically significant?

```r
set.seed(093021)

f1 <- rnorm(15, mean = 2, sd = 5.3)

f1 <- round(f1, 2)

f1
```

```
##  [1]  7.87 -0.68  4.65 -2.12  1.57  4.66  3.20  5.75  8.22  0.37  8.61  1.27
## [13]  7.16  1.39  6.48
```