

# T-Test: Independent Samples

Data Analysis for Psychology in R 1

Semester 2, Week 7

**Dr Emma Waterston**

Department of Psychology  
The University of Edinburgh

# Learning Objectives

- Understand when to use an independent samples  $t$ -test
- Understand the null hypothesis for an independent sample  $t$ -test
- Understand how to calculate the test statistic
- Know how to conduct the test in [R](#)

# Topics for Today

- Conceptual background and introduction to our example
- Calculations and R-functions
- Assumptions and effect size

# Independent T-Test Purpose & Data

- The independent  $t$ -test is used when we want to test the difference in mean between two measured groups.
- The groups must be independent:
  - No person can be in both groups.
- Examples:
  - Treatment versus control group in an experimental study
  - Married versus not married
- Data Requirements:
  - A continuously measured variable
  - A binary variable denoting groups

# t-statistic

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - \delta_0}{SE_{(\bar{x}_1 - \bar{x}_2)}}$$

- Where
  - $\bar{x}_1$  and  $\bar{x}_2$  are the sample means in each group
  - $\delta_0$  is the hypothesised population difference in means in the null hypothesis ( $\mu_1 - \mu_2$ )
  - $SE_{(\bar{x}_1 - \bar{x}_2)}$  is standard error of the difference
- Sampling distribution is a  $t$ -distribution with  $n - 2$  degrees of freedom, where  $n = n_1 + n_2$ .

# Standard Error Difference

- First calculate the pooled standard deviation.

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

- Then use this to calculate the SE of the difference.

$$SE_{(\bar{x}_1 - \bar{x}_2)} = s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

# Hypotheses

- Two-tailed:

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

- One-tailed:

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 < \mu_2$$

$$H_1 : \mu_1 > \mu_2$$

- Two-tailed:

$$H_0 : \mu_1 - \mu_2 = 0$$

$$H_1 : \mu_1 - \mu_2 \neq 0$$

- One-tailed:

$$H_0 : \mu_1 - \mu_2 = 0$$

$$H_1 : \mu_1 - \mu_2 < 0$$

$$H_1 : \mu_1 - \mu_2 > 0$$

# Questions?



# Example

- Example taken from Howell, D.C. (2010). *Statistical Methods for Psychology, 7th Edition*. Belmont, CA: Wadsworth Cengage Learning.
- Data from Aronson, Lustina , Good, Keough , Steele and Brown (1998). Experiment on stereotype threat.
  - Two independent groups college students (n=12 control; n=11 threat condition).
  - Both samples excel in maths.
  - Threat group told certain students usually do better in the test

# Data

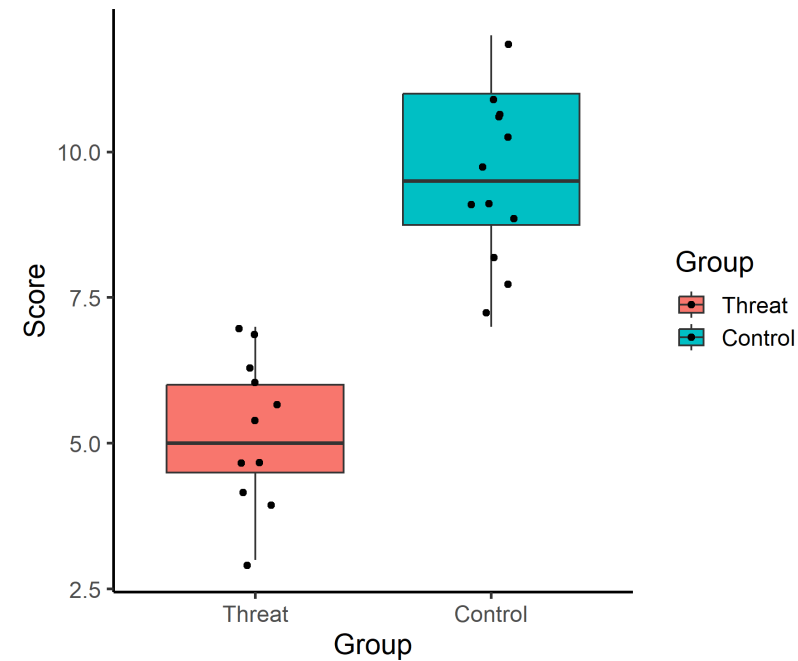
```
## # A tibble: 23 × 2
##   Group Score
##   <fct> <dbl>
## 1 Threat      7
## 2 Threat      5
## 3 Threat      6
## 4 Threat      5
## 5 Threat      6
## 6 Threat      5
## 7 Threat      4
## 8 Threat      7
## 9 Threat      4
## 10 Threat     3
## # i 13 more rows
```

# Visualizing data

- We spoke earlier in the course about the importance of visualizing our data.
- Here, we want to show the mean and distribution of scores by group.
- So we want a.....

# Visualizing data

```
ggplot(data = threat,  
       aes(x = Group, y = Score, fill = Group)) +  
  geom_boxplot() +  
  geom_jitter(width = 0.1)
```



# Hypotheses

- My hypothesis is that the threat group will perform worse than the control group.
  - This is a one-tailed hypothesis.
- And I will use an  $\alpha = .05$

# Questions?

# Calculation

- Steps in my calculations:
  - Calculate the sample mean in both groups  $\bar{x}_1$  and  $\bar{x}_2$ .
  - Calculate the pooled SD ( $s_p$ ).
  - Check I know my  $n$ .
  - Calculate the standard error ( $SE$ ).
- Use all this to calculate  $t$ .

# Calculation

```
threat %>%  
  group_by(Group) %>%  
  summarise(  
    Mean = round(mean(Score),2),  
    SD = round(sd(Score),2),  
    n = n()  
  ) %>%  
  kable(digits = 2) %>%  
  kable_styling(full_width = FALSE)
```

Group	Mean	SD	n
Threat	5.27	1.27	11
Control	9.58	1.51	12



# Calculation

Group	Mean	SD	n
Threat	5.27	1.27	11
Control	9.58	1.51	12

- Calculate pooled standard deviation:

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{(11 - 1) * 1.27^2 + (12 - 1) * 1.51^2}{11 + 12 - 2}} = \sqrt{\frac{10 * 1.27^2 + 11 * 1.51^2}{11 + 12 - 2}} = \sqrt{\frac{41.21}{21}} = 1.401$$

- Calculate the standard error:

$$SE_{(\bar{x}_1 - \bar{x}_2)} = s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 1.401 \sqrt{\frac{1}{11} + \frac{1}{12}} = 1.401 * 0.417 = 0.584$$

# Calculation

- Steps in my calculations:
  - Calculate the sample mean in both groups - Threat ( $\bar{x}_1 = 5.27$ ), Control ( $\bar{x}_2 = 9.58$ ).
  - Calculate the pooled SD ( $s_p = 1.401$ ).
  - Check I know my  $n$  - Threat ( $n_1 = 11$ ) and Control ( $n_2 = 12$ ) -  $n = 23$ .
  - Calculate the standard error ( $SE = 0.584$ ).
- Use all this to calculate  $t$ .

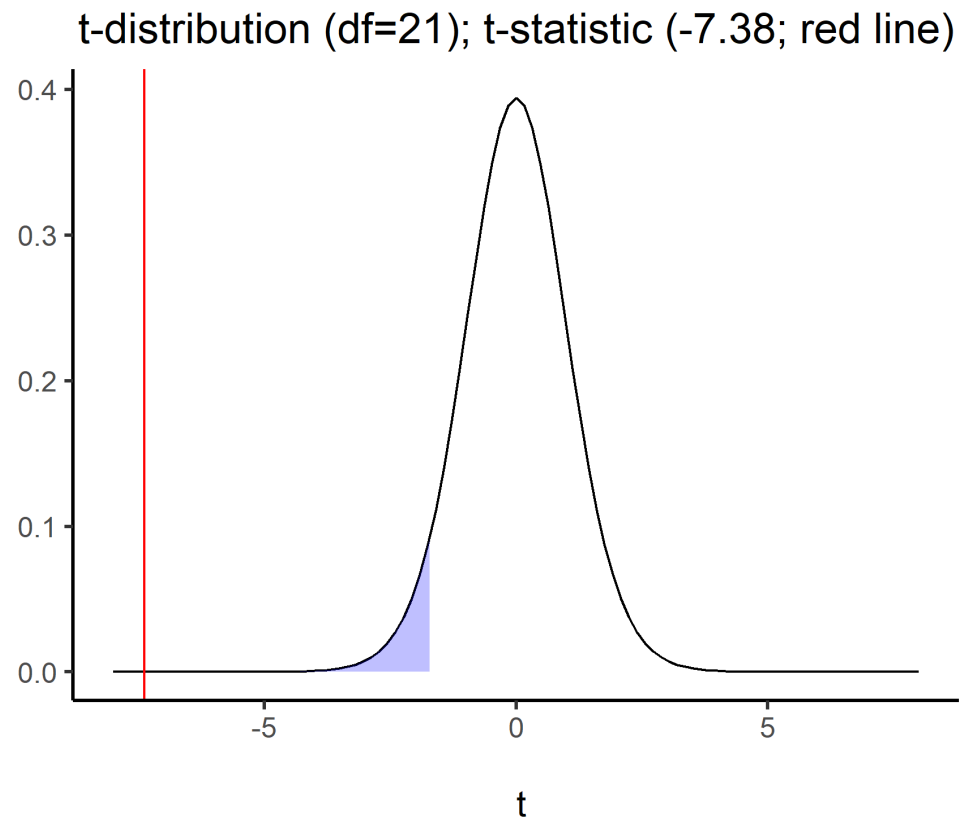
$$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{SE_{(\bar{x}_1 - \bar{x}_2)}} = \frac{5.27 - 9.58}{0.584} = -7.38$$

- So in our example  $t = -7.38$
- Note: When doing hand calculations there might be a small amount of rounding error when we compare to  $t$  calculated in [R](#).

# Is our test significant?

- We have all the pieces we need:
  - Degrees of freedom =  $n - 2 = (12 + 11) - 2 = 23 - 2 = 21$
  - We have our  $t$ -statistic (-7.38)
  - Hypothesis to test (one-tailed)
  - $\alpha$  level (.05).
- So now all we need is the critical value from the associated  $t$ -distribution in order to make our decision.

# Is our test significant?



```
tibble(  
  LowerCrit = round(qt(0.05, 21), 2),  
  Exactp = 1-pt(7.3817, 21)  
)
```

```
## # A tibble: 1 × 2  
##   LowerCrit      Exactp  
##   <dbl>        <dbl>  
## 1     -1.72 0.000000146
```

# Is my test significant?

- So our critical value is -1.72
  - Our  $t$ -statistic (-7.38) is larger than this
  - So we reject the null hypothesis

# Our Test: In R

```
res <- t.test(threat$Score ~ threat$Group,  
             alternative = "less",  
             mu = 0,  
             var.equal = TRUE,  
             conf.level = 0.95)
```

```
res
```

```
##  
##      Two Sample t-test  
##  
## data:  threat$Score by threat$Group  
## t = -7.3817, df = 21, p-value = 1.458e-07  
## alternative hypothesis: true difference in means between group Threat and group Control is less than 0  
## 95 percent confidence interval:  
##      -Inf -3.305768  
## sample estimates:  
##  mean in group Threat mean in group Control  
##      5.272727      9.583333
```

# Write up

An independent sample  $t$ -test was used to determine whether the average maths score of the stereotype threat group ( $n = 11$ ) was significantly lower ( $\alpha = .05$ ) than the control group ( $n = 12$ ). There was a significant difference in test score between the control ( $M = 9.58, SD = 1.51$ ) and threat ( $M = 5.27, SD = 1.27$ ) groups, where the scores were significantly lower in the threat group ( $t(21) = -7.38, p < .001, one - tailed$ ). Therefore, we can reject the null hypothesis. The direction of difference supports our directional hypothesis and indicates that the threat group performed more poorly than the control group.

# Questions?



# Assumption checks summary

	Description	One-Sample t-test	Independent Sample t-test	Paired Sample t-test
Normality	Continuous variable (and difference) is normally distributed.	Yes (Population)	Yes (Both groups/ Difference)	Yes (Both groups/ Difference)
Tests:	Descriptive Statistics; Shapiro-Wilks Test; QQ-plot			
Independence	Observations are sampled independently.	Yes	Yes (within and across groups)	Yes (within groups)
Tests:	None. Design issue.			
Homogeneity of variance	Population level standard deviation is the same in both groups.	NA	Yes	NA
Tests:	F-test			
Matched Pairs in data	For paired sample, each observation must have matched pair.	NA	NA	Yes
Tests:	None. Data structure issue.			

# Assumptions

- The independent sample  $t$ -test has the following assumptions:
  - Independence of observations within and across groups.
  - Continuous variable is approximately normally distribution **within both groups**.
    - Equivalently, that the difference in means is normally distributed.
  - Homogeneity of variance across groups.

# Assumption checks: Normality

- Descriptive statistics:
  - Skew:
    - Below are some rough guidelines on how to interpret skew.
    - No strict cuts for skew - these are loose guidelines.

Verbal label	Magnitude of skew in absolute value
Generally not problematic	$  \text{Skew}   < 1$
Slight concern	$1 >   \text{Skew}   < 2$
Investigate impact	$  \text{Skew}   > 2$

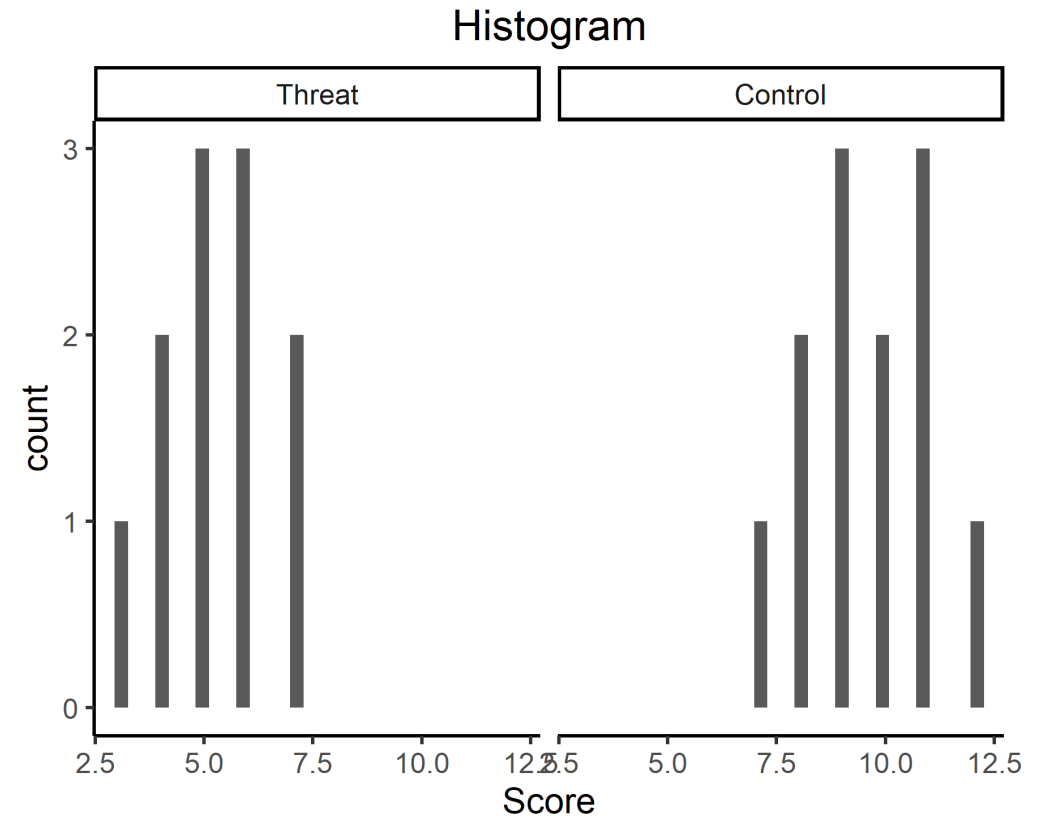
# Skew

```
library(psych)
threat %>%
  group_by(Group) %>%
  summarise(
    skew = round(skew(Score),2)
  )
```

```
## # A tibble: 2 × 2
##   Group      skew
##   <fct>    <dbl>
## 1 Threat  -0.2
## 2 Control -0.07
```

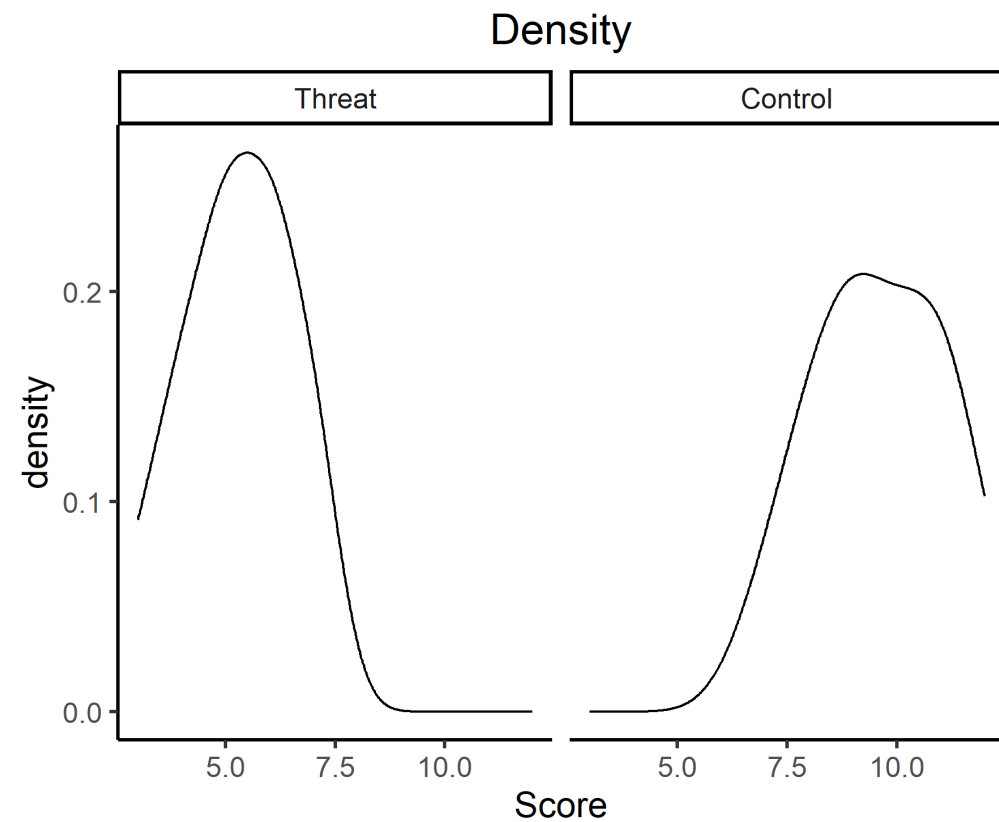
# Histograms

```
ggplot(threat, aes(x=Score)) +  
  geom_histogram() +  
  facet_wrap(~ Group) +  
  labs(title = "Histogram")
```



# Density

```
ggplot(threat, aes(x=Score)) +  
  geom_density() +  
  facet_wrap(~ Group) +  
  labs(title = "Density")
```

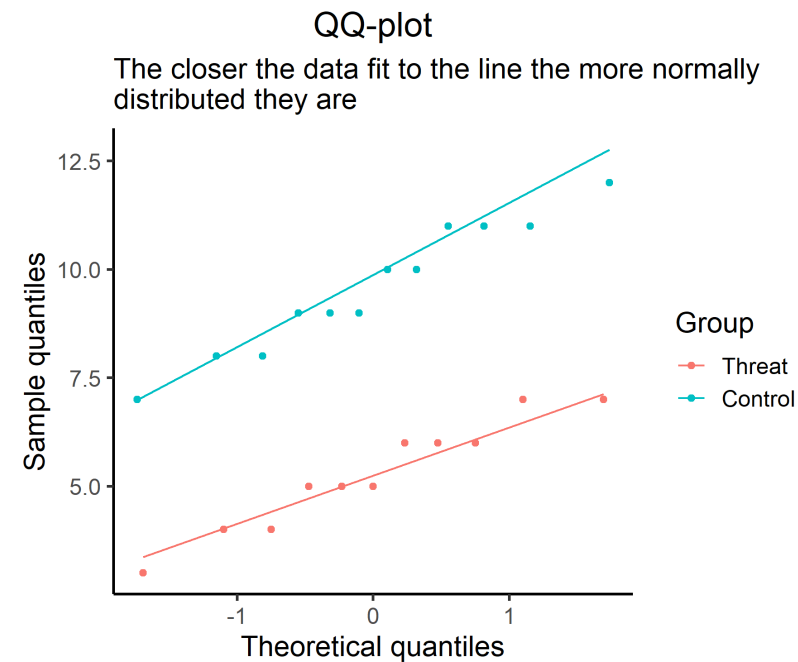


# Assumption checks: Normality

- QQ-plots:
  - Plots the sorted quantiles of one data set (distribution) against sorted quantiles of data set (distribution).
  - Quantile = the percent of points falling below a given value.
  - For a normality check, we can compare our own data to data drawn from a normal distribution

# QQ-plots

```
ggplot(data = threat,  
       aes(sample = Score, colour = Group)) +  
  geom_qq() +  
  geom_qq_line() +  
  labs(title="QQ-plot",  
       subtitle="The closer the data fit to the  
       x = \"Theoretical quantiles\",  
       y = \"Sample quantiles\"")
```



- This looks reasonable in both groups



# Assumption checks: Normality

- Shapiro-Wilks test:
  - Checks properties of the observed data against properties we would expect from normally distributed data.
  - Statistical test of normality.
  - $H_0$ : data = a normal distribution.
  - $p\text{-value} < \alpha$  = reject the null, data are not normal.
    - Sensitive to  $n$  as all  $p$ -values will be.
    - In very large  $n$ , normality should also be checked with QQ-plots alongside statistical test.

# Shapiro-Wilks in R

```
threat %>%  
  filter(Group == "Control") %>%  
  pull(Score) %>%  
  shapiro.test()
```

```
##  
##      Shapiro-Wilk normality test  
##  
## data:  .  
## W = 0.95538, p-value = 0.7164
```

$W = 0.96, p = .716$

```
thr <- threat %>%  
  filter(Group == "Threat") %>%  
  select(Score)  
shapiro.test(thr$Score)
```

```
##  
##      Shapiro-Wilk normality test  
##  
## data:  thr$Score  
## W = 0.93979, p-value = 0.518
```

$W = 0.94, p = .518$

# Assumption checks: Homogeneity of variance

- The  $F$ -test is a test that compares the variances of two groups.
  - This test is preferable for  $t$ -test.
  - $H_0$ : Population variances are equal.
  - $p\text{-value} < \alpha$  = reject the null, the variances differ across groups.

# F-test R

```
var.test(threat$Score ~ threat$Group, ratio = 1)
```

```
##  
##      F test to compare two variances  
##  
## data:  threat$Score by threat$Group  
## F = 0.71438, num df = 10, denom df = 11, p-value = 0.6038  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
##  0.2026227 2.6181459  
## sample estimates:  
## ratio of variances  
##      0.7143813
```

- Why `ratio = 1`?

- $H_0 : \sigma_1^2 = \sigma_2^2$ 
  - $H_1 : \sigma_1^2 \neq \sigma_2^2$

- $H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1$ 
  - $H_1 : \frac{\sigma_1^2}{\sigma_2^2} \neq 1$

# Violation of homogeneity of variance

- If the variances differ, we can use a Welch test.
- Conceptually very similar, but we do not use a pooled standard deviation.
  - As such our estimate of the SE of the difference changes
  - As do our degrees of freedom

# Welch test

- If the variances differ, we can use a Welch test.
- Test statistic = same:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - \delta_0}{SE_{(\bar{x}_1 - \bar{x}_2)}}$$

- SE calculation:

$$SE_{(\bar{x}_1 - \bar{x}_2)} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

- And degrees of freedom (don't worry, not tested):

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$$

# Welch: In R

```
t.test(threat$Score ~ threat$Group,  
       alternative = "less",  
       mu = 0,  
       var.equal = FALSE, #default, only here to highlight difference  
       conf.level = 0.95)
```

```
##  
##      Welch Two Sample t-test  
##  
## data:  threat$Score by threat$Group  
## t = -7.4379, df = 20.878, p-value = 1.346e-07  
## alternative hypothesis: true difference in means between group Threat and group Control is less than 0  
## 95 percent confidence interval:  
##      -Inf -3.313093  
## sample estimates:  
##  mean in group Threat mean in group Control  
##      5.272727      9.583333
```

# Cohen's D: Independent samples t-test

- Independent-sample  $t$ -test (if you do have equality of variances):

$$D = \frac{(\bar{x}_1 - \bar{x}_2) - \delta_0}{s_p}$$

- $\bar{x}_1$  = mean group 1
- $\bar{x}_2$  = mean group 2
- $\delta_0$  is the hypothesised population difference in means in the null hypothesis ( $\mu_1 - \mu_2$ )
- $s_p$  = pooled standard deviation
- Independent-sample  $t$ -test (if you do not have equality of variances):
  - Calculate via `cohens_d()` function from `effectsize` package in R - do not calculate by hand.
- Recall the common "cut-offs" for  $D$ -scores:

Verbal label	Magnitude of $D$ in absolute value
Small (or weak)	$\leq 0.20$
Medium (or moderate)	$\approx 0.50$
Large (or strong)	$> 0.80$



# Cohen's D in R

```
library(effectsize)
cohens_d(threat$Score ~ threat$Group,
         mu = 0,
         alternative = "less",
         var.equal = TRUE,
         conf.level = 0.95)
```

```
## Cohen's d |          95% CI
## -----
## -3.08      | [-Inf, -2.02]
##
## - Estimated using pooled SD.
## - One-sided CIs: lower bound fixed at [-Inf].
```

# Write up: Assumptions

The DV of our study, Score, was measured on a continuous scale, and data were independent (participants belonged to one of two groups - Control or Threat). The assumption of normality was visually assessed (via histograms, density plots, and a QQplot) as well as statistically via a Shapiro-Wilks test. The QQplots did not show much deviation from the diagonal line in either group, and the Shapiro-Wilks test for both the Control ( $W = 0.96, p = .716$ ) and Threat ( $W = 0.94, p = .518$ ) conditions suggested that the samples came from a population that was normally distributed. This was inline with the histogram and density plots for each group, which suggested that Score was normally distributed (and where  $skew < 1$ ). Based on the results of our  $F$ -test, there was no significant difference between the two population variances ( $F(10, 11) = 0.71, p = .604$ ). The size of the effect was found to be large ( $D = -3.08$ ).

# Summary

- Today we have covered:
  - Basic structure of the independent-sample  $t$ -test
  - Calculations
  - Interpretation
  - Assumption checks
  - Effect size measures

# Tasks

- Read the lab worked example and complete your lab
- Complete any assigned readings
- Go to office hours if you have questions
  - Emma's Office Hours = Tuesdays 10:30-11:30 in G15, 7 George Square
- Complete the weekly quiz
  - Opens Monday at 9am
  - Closes Sunday at 5pm