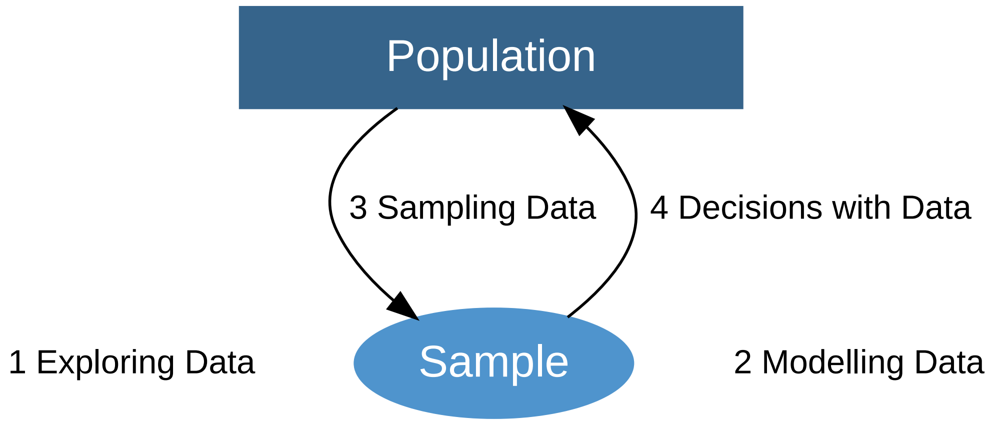
2 Sample T Test

Decisions with Data | Tests for relationships

Unit Overview



Module4 Decisions with Data

Test for a Proportion

How can we make evidence based decisions? Is an observed result due to chance or something else? How can we test whether a population has a certain proportion?

Tests for a Mean

How can we test whether a population has a certain mean? Or whether 2 populations have the same mean?

Tests for a Relationship [DATA1001/MATH1115]

How can we test whether 2 variables are linearly related? How can we test whether a categorical variable is in certain proportions?

Topic 34 2 Sample T Tests

Data Story

What is the effect of a Red Bull on heart rate?

Red Bull data

- Red Bull is an energy drink advertised to “give you wings”.
- What does research say about the medical effects of drinking a Red Bull?

Scenario1

Consider the following data on heart rates (beats per minute), for 2 independent groups of Sydney students, collected 20 minutes after the ‘RedBull’ group had drunk a 250ml cold can of Red Bull.

No_Redbull	84	76	68	80	64	62	74	84	68	96	80	64	65	66
Redbull	72	88	72	88	76	75	84	80	60	96	80	84	NA	NA

Statistical Thinking

- Write down a hypothesis you could test.
- Would you consider a test of differences like the caffeine data? Why or why not?

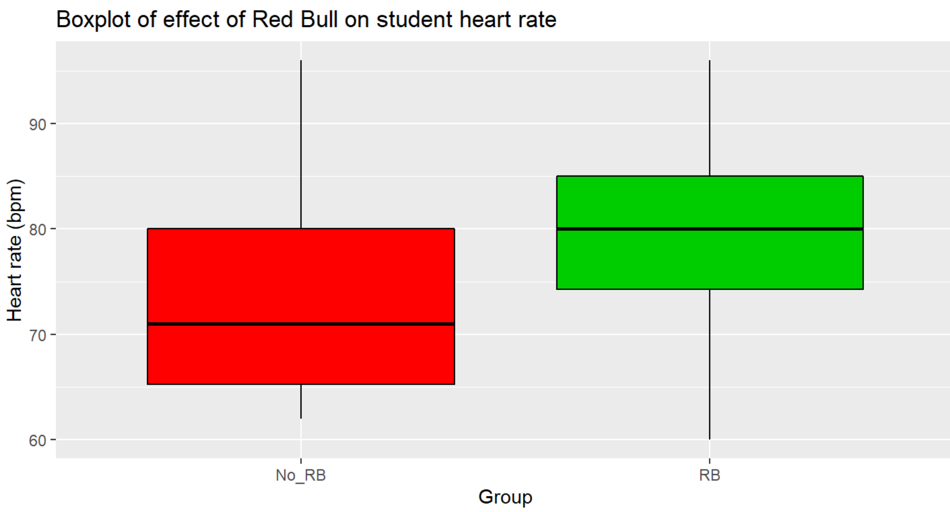
```
No_RB <- c(84, 76, 68, 80, 64, 62, 74, 84, 68, 96, 80, 64, 65, 66)
RB <- c(72, 88, 72, 88, 76, 75, 84, 80, 60, 96, 80, 84)

RB_data <- data.frame(group = c(rep("No_RB", 14), rep("RB", 12)), rate = c(No_RB,
    RB))
```

Visualising data

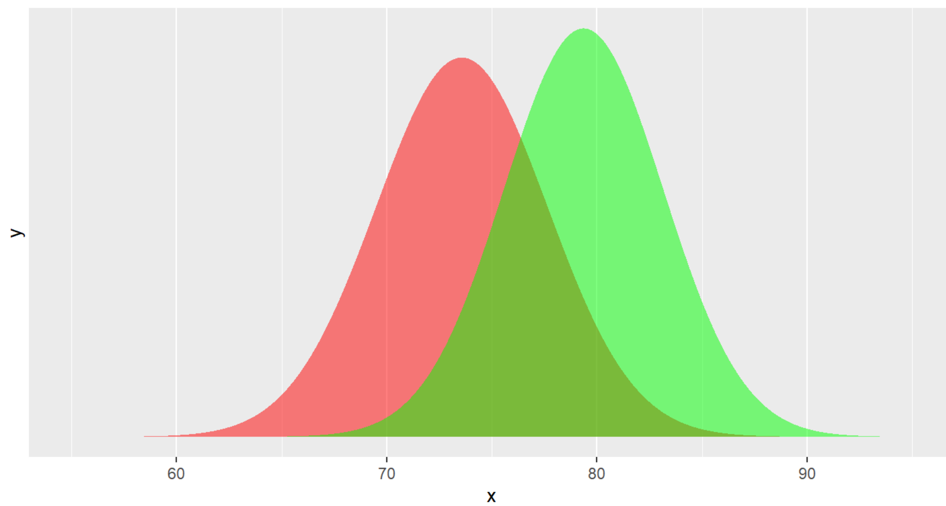
Comparative Boxplots of data

- We can compare the shape and spread of the 2 samples (as a window into the 2 populations) using boxplots.



Comparative Normal curves of the 2 means

- Assuming that the 2 populations are normally distributed, we can compare the distributions of the two populations using the sample means and SDs as estimates in the plots.



Inference: 2 Sample T Test

Inference

Based on the previous 3 plots, do you think that the heart rate of a student is impacted by drinking a Red Bull?

Inference

- While visualisation of the data gives us an initial **glimpse** at the possible relationship between the two populations (those who have drunk a Red Bull and those who have not), we often want to make a **decision** on whether the mean of the two populations is the same or different.

- **Inference** is making a decision about population parameter(s) based on a sample. How would we go about this?

3 main steps (Revision)

1. Set up research question

H: Hypothesis H_0 vs H_1

2. Weigh up evidence

A: Assumptions
T: Test Statistic
P: P-value

3. Explain conclusion

C: Conclusion

H: Hypotheses for 2 Sample T Test

Example (2-sided alternative hypothesis)

Does Red Bull change the heart rate in students?

- Let μ_1 = mean heart rate of our control (no Red Bull)
- Let μ_2 = mean heart rate of our treatment (Red Bull)
- H_0 : There is no difference: $\mu_1 = \mu_2$, or $\mu_1 - \mu_2 = 0$.
- H_1 : There is a difference: $\mu_1 \neq \mu_2$, or $\mu_1 - \mu_2 \neq 0$.

Example (1-sided alternative hypothesis)

Does Red Bull increase the heart rate in students?

- Write down the hypothesis for a 1-sided test.

How do we choose between a 1-sided or a 2-sided test?

Oracle

A: Assumptions for 2 Sample T Test

A1. The 2 samples are independent

- The two samples (Red Bull and Control) contain different people. This design differs from the caffeine one in which the same person is tested at both 0 and 13 mg level of caffeine and we consider the sample of differences from each pair.

A2. The 2 populations have equal spread (SD/variance)

- We assume that the 2 populations have the **same variation** in heart rate.

- Check: Boxplots, Histograms, Variance Test

- Note: This asumption can be relaxed by using the Welch 2 sample T Test.

A3. The 2 populations are Normal

- We assume that the 2 populations have Normally distributed heart rates.
- Check: Boxplots (expect no or few outliers), Histograms, QQ-Plots, Normality Test

T: Test Statistic for 2 Sample T Test

Test Statistic for 2 Sample T Test (equal variance)

We compare 2 populations, so the observed test statistic is

$$\text{test statistic} = \frac{\text{OV - EV}}{\text{SE}} = \frac{\bar{x}_1 \text{ - } \bar{x}_2 \text{ - } 0}{\text{SE}}$$

where

$$SE = \sqrt{\text{SD}_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} \qquad df = n_1 + n_2 - 2$$

based on the pooled SD, where $\text{SD}_p^2 = \frac{(n_1 - 1)\text{SD}_1^2 + (n_2 - 1)\text{SD}_2^2}{n_1 + n_2 - 2}$

Statistical Thinking

Think about the equation for the test statistic:

- What would cause the numerator to increase in size and how will this effect the size of the test statistic?
- What would cause the denominator to increase in size and how will this effect the size of the test statistic?

Calculating the Test Statistic (and P-value)

By formula

```
m1 = mean(No_RB)
m2 = mean(RB)
n1 = length(No_RB)
n2 = length(RB)
sdp2 = ((n1 - 1) * sd(No_RB)^2 + (n2 - 1) * sd(RB)^2)/(n1 + n2 - 2)
se = sqrt(sdp2 * (1/n1 + 1/n2))
TestStat = (m1 - m2 - 0)/se
Pvalue = 2 * pt(abs(TestStat), n1 + n2 - 2, lower.tail = F)
c(TestStat, Pvalue)
```

```
## [1] -1.5418064  0.1362041
```

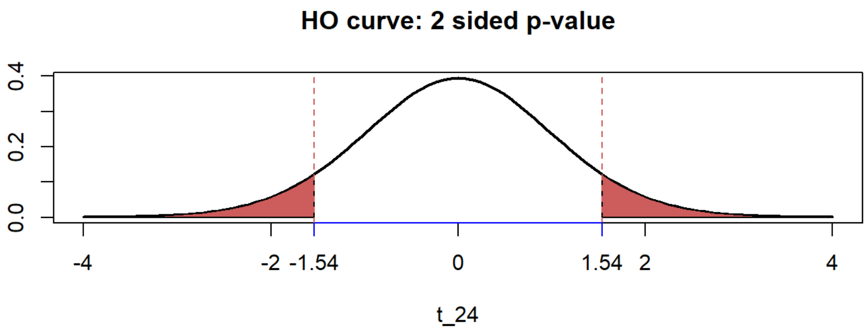
Using R function `t.test`

The test statistic and P-value can be obtained using `t.test` in one go.

```
t.test(No_RB, RB, var.equal = T)
```

```
##
##  Two Sample t-test
##
## data:  No_RB and RB
## t = -1.5418, df = 24, p-value = 0.1362
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -13.892538   2.011586
## sample estimates:
## mean of x mean of y
##  73.64286  79.58333
```

Note the argument `var.equal = T` indicates that we are doing a two sample T test assuming equal population variances.


HO curve: 2 sided p-value
0.0
0.2
0.4
-4
-2
-1.54
0
1.54
2
4
t_24

C: Conclusion

Putting everything together!

H:

- Let μ_1 = mean heart rate of our control (no Red Bull)
- Let μ_2 = mean heart rate of our treatment (Red Bull)
- H_0 : There is no difference: $\mu_1 = \mu_2$
- H_1 : There is a difference: $\mu_1 \neq \mu_2$

A: We assume that the 2 populations are Normally distributed with equal variance, and the 2 samples are independent.

T: Observed test statistic is -1.5418

P: The p-value is 0.1362

C:

Statistical conclusion:

- As the p-value > 0.05, we retain the null hypothesis. (Note also that that the 95% Confidence Interval for the mean difference is (-13.89,2.01) which contains the null hypothesis value 0.)
- Thus data is consistent with hypothesis that the mean heart rates are equal.

Scientific conclusion:

The data suggests that the consumption of Redbull by university students does not have an effect on heart rate.

Checking assumptions for 2 Sample T Test

Assumptions for 2 Sample T Test

A1. The 2 samples are independent

- Check: Context

A2. The 2 populations have equal spread (SD/variance)

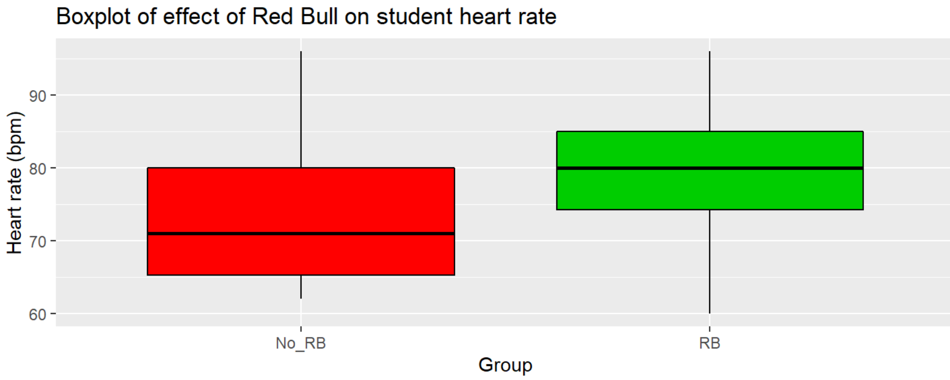
- Check: Boxplots, Histograms, Variance Test

A3. The 2 populations are Normal

- Check: Boxplots (expect no or few outliers), Histograms, QQ-Plots, Normality Test

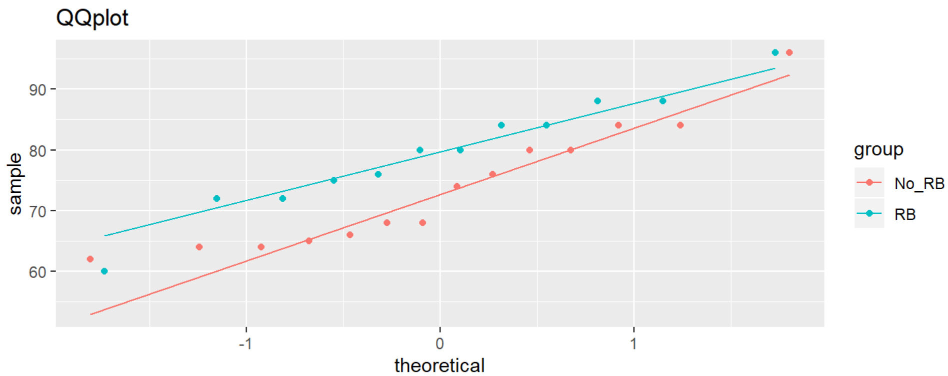
Comparative Boxplots - for normality and equality of variance assumptions

- Normality: the 2 samples look symmetrical and so are consistent with Normality, although the control one looks a bit asymmetric.
- Equality of variance: the 2 samples have similar spread and so are consistent with the equality of variance assumption.



Q-Q Plot - for Normality

- A quantile-quantile plot (Q-Q Plot) graphs the theoretical quantiles based on the normal curve against the actual quantiles. If the line formed by the points is reasonably straight, then we can safely assume that the data is normally distributed. Both plots (red = no Redbull, green = Redbull) show a reasonably straight line.



```
require(ggplot2)
p3 = ggplot(RB_data, aes(sample = rate, colour = group)) + stat_qq() + stat_qq_line() +
    ggtitle("QQplot")
p3
```

Shapiro-Wilk Test - for Normality

- The Shapiro-Wilk test tests the null hypothesis that the data is Normal.
  - H_0 : The data is consistent with Normal
  - H_1 : The data is not consistent with Normal
- While useful, this test is very sensitive to sample size (for example, small samples will almost always be retained as Normal). Therefore, it is recommended to use this test in conjunction with graphical methods.

```
shapiro.test(No_RB)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  No_RB
## W = 0.90604, p-value = 0.138
```

```
shapiro.test(RB)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  RB
## W = 0.97459, p-value = 0.9524
```

- Here, both tests give large p-values (p-value>0.05), so considering the previous plots, we conclude that the populations could be Normally distributed.

Levene’s Test (F-Test) - for equal spread

- The F-Test tests the null hypothesis that the spread is equal between the two populations.
- Let σ_1^2 = variance of the Control group (No Redbull)
- Let σ_2^2 = variance of the Treament group (Redbull)
  - H_0 : There is no difference: $\sigma_1^2 = \sigma_2^2$
  - H_1 : There is a difference: $\sigma_1^2 \neq \sigma_2^2$
- If the variation is very different between the 2 samples, we may increase the chance of falsely rejecting the H_0 (called type I error).

```
var.test(No_RB, RB)
```

```
##
##  F test to compare two variances
##
## data:  No_RB and RB
## F = 1.1357, num df = 13, denom df = 11, p-value = 0.8428
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.334832 3.631266
## sample estimates:
## ratio of variances
##           1.135659
```

The F-test gives p-value > 0.05 suggesting that data is consistent with equal variance.

Overall Conclusion about Assumptions

A2. The 2 populations have equal spread (SD)

- The boxplots suggest that the variation is similar between the 2 populations, as is confirmed by the F-Test.

A3. The 2 populations are Normal

- The 2 samples suggest the 2 populations could be Normal, from both the boxplots, the Q-Q Plot and the Shapiro-Wilk Test.

Hence we proceed with a 2 Sample T Test (assuming equal variance).

Other Tests

Assumptions not met?

- What if the assumptions for the 2 Sample T Test are not met?
- There's lots more tests, with different assumptions, to deal with the following situations:

Data	Appropriate Test(s)
2 samples have unequal variance	Welch 2 Sample T Test
2 samples suggest Non-Normality	Transformations or Non Parametric Tests
2 samples are matched pairs	Paired T-Test (using sample of differences)

Unequal variances (Welch 2 Sample T Test)

Suppose the variance was much larger in the group who have not had caffeine (Redbull).

```
set.seed(10)
No_RB <- No_RB + rnorm(14, 0, 20)  # New simulated version of No-RB with larger variance
var.test(No_RB, RB)
```

```
##
##  F test to compare two variances
##
## data:  No_RB and RB
## F = 3.9709, num df = 13, denom df = 11, p-value = 0.02813
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##   1.170767 12.697012
## sample estimates:
## ratio of variances
##           3.970923
```

Now we see that the p-value from the variance test is < 0.05. Therefore we reject the null hypothesis and question whether the variances of the 2 populations are equal.

- In this case we use the argument `var.equal = FALSE`. This performs the Welch 2 Sample T Test. The p-value is > 0.05 so our conclusion concurs with the previous 2 Sample T Test for equal variances.

```
t.test(No_RB, RB, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  No_RB and RB
## t = -1.6111, df = 19.747, p-value = 0.123
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -21.180686   2.728887
## sample estimates:
## mean of x mean of y
##  70.35743  79.58333
```

Non-Normality

- Dealing with non-Normal data is an area hotly debated in the stats world.
- One option is to transform non-Normal data (eg log or square root) before performing a T Test.
- An alternative, particularly for smaller datasets, is to use non-parametric tests such as the Mann-Whitney-Wilcoxon test: `wilcox.test`.

Non-Independent Data (Paired T Test)

- Sometimes it is desirable to analyse **dependent** data. We often design an experiment to take advantage of this dependency in order to control variation between experimental groups.

How could we design the Redbull experiment such that the 2 groups are dependent?

Scenario2

- Suppose we measure the heart rate of 12 students **before** and **after** they consume a Redbull can, like the elite cyclists example. We call this **paired data** because we have one experimental unit (a student) with two treatments (heart rate before and after RedBull), as seen in previous lecture.

- We can record the data as follows:

StudentdID	1	2	3	4	5	6	7	8	9	10	11	12
Before_Redbull	84	76	68	80	64	62	74	84	68	96	80	64
After_Redbull	72	88	72	88	76	75	84	80	60	96	80	84

Why is this experimental design useful compared to having independent samples?

Preparation: Calculate differences

- Calculate the **differences** between `Before_Redbull - After_Redbull`.

```
diff = RB - No_RB
diff
```

```
## [1] -12  12   4   8  12  13  10  -4  -8   0   0  20
```

- The Paired T Test is just a 1 Sample T Test on the differences!

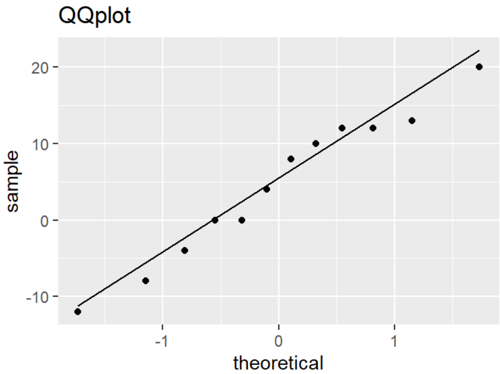
Performing the Paired T Test

H

- H_0 : The difference in means is zero : $\mu_d = 0$
- H_1 : The difference in means is not zero : $\mu_d \neq 0$

A

We assume that the population of differences is Normally distributed.



- The QQ-plot looks reasonably linear suggesting the data is reasonably Normally distributed.

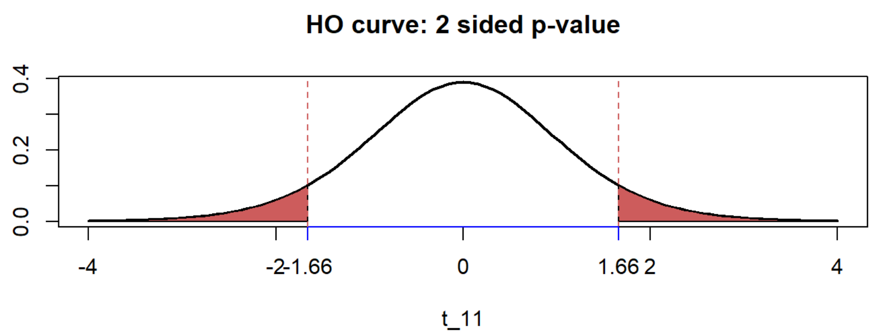
T

Use the `t.test` function with the argument `paired = T`.

```
No_RB <- c(84, 76, 68, 80, 64, 62, 74, 84, 68, 96, 80, 64)
RB <- c(72, 88, 72, 88, 76, 75, 84, 80, 60, 96, 80, 84)
t.test(No_RB, RB, paired = T)
```

P

```
##
##  Paired t-test
##
## data:  values by ind
## t = -1.6578, df = 11, p-value = 0.1256
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -10.668294   1.501628
## sample estimates:
## mean of the differences
##               -4.583333
```



C

Statistical conclusion:

- As the p-value > 0.05, we retain the null hypothesis.
- Note also that that the 95% Confidence Interval for the mean difference is (-10.668294,1.501628) which contains the null hypothesis value 0.)
- Thus the data is consistent with the hypothesis that the difference in mean heart rates is 0.

Scientific conclusion:

- The data suggests that the consumption of Redbull by university students does not have an effect on heart rate.

Summary

Key Words

Inference, groups, 2 Sample T Test, test assumptions, Normal distribution, independent, variance/variation/spread, independent, Q-Q Plot, Shapiro-Wilk Test, Levene's Test, dependent, experimental design, paired T Test, transformations, parametric tests, Welch 2 Sample T Test

Further Thinking

- Is there a disadvantage to doing consecutive tests such as the variance test followed by a T Test?

- Should you report just the p-value or include other information such as confidence intervals?

Mathematical formulae

Two sample T Test

$$t = \frac{\overline{y}_1 - \overline{y}_2}{SE}$$

Confidence intervals difference two sample means

$$95\%CI = (\overline{y}_1 - \overline{y}_2) \pm t_{crit}^{0.025} . SE$$

Pooled standard deviation (SD_p)

$$\mathrm{SD}_p^2 = \frac{(n_1 - 1)\,\mathrm{SD}_1^2 + (n_2 - 1)\,\mathrm{SD}_2^2}{n_1 + n_2 - 2}$$

Standard error of the difference (equal variance)

$$SE = \sqrt{\mathrm{SD}_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} \qquad df = n_1 + n_2 - 2$$

Standard error of the difference (unequal variance)

$$SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \qquad df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left(\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}\right)} \quad \text{round down}$$