

Topic 6 – Two-sample t -tests – Part I

ENVX1002 Introduction to Statistical Methods

Januar Harianto

The University of Sydney

Dec 2024



THE UNIVERSITY OF
SYDNEY

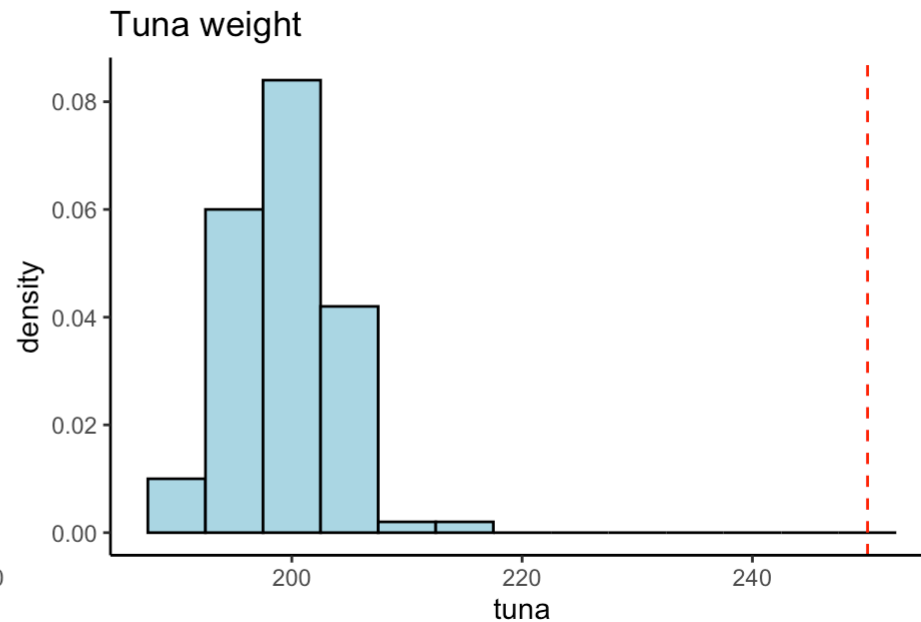
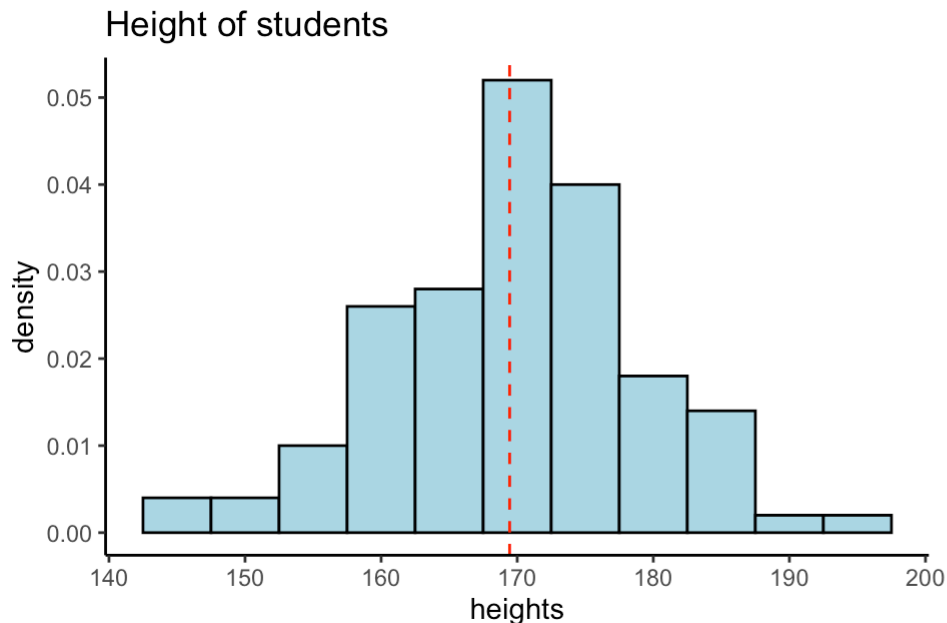
Testing differences in means

Recap: we have one sample

One-sample *t*-test: compare the sample of data to a *fixed* value of interest (e.g. a hypothesised value, or a population mean).

Examples

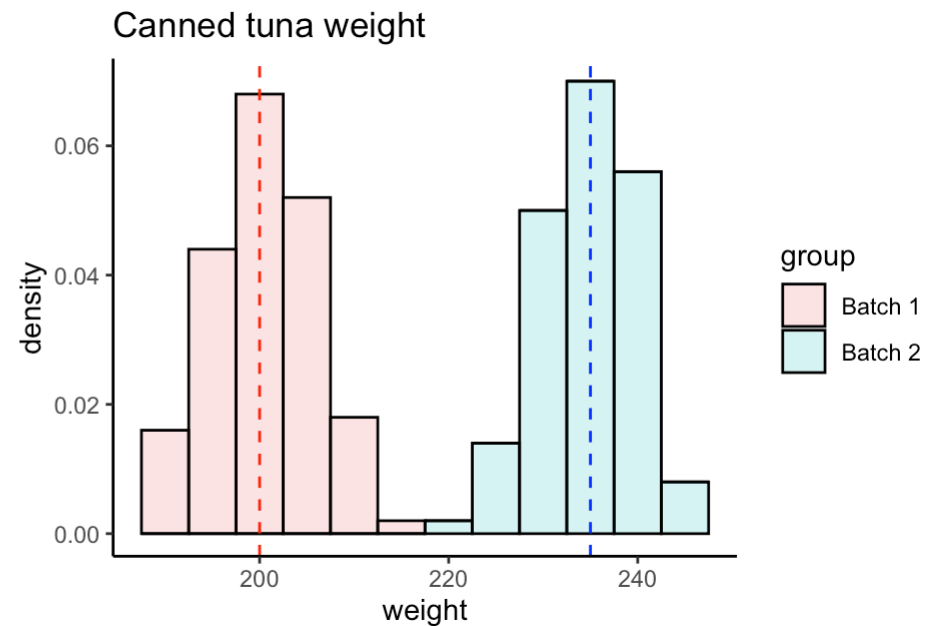
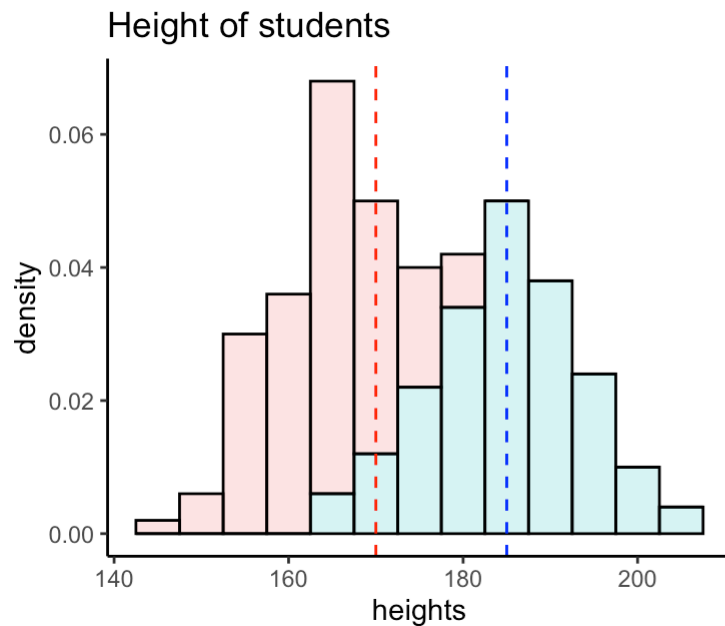
- *Is the mean height of students in this class different from the population mean of 170 cm?*
- *Is the food weight in a can of tuna different from its advertised weight of 250 g?*



What if we want to compare a sample of data to *another* sample?

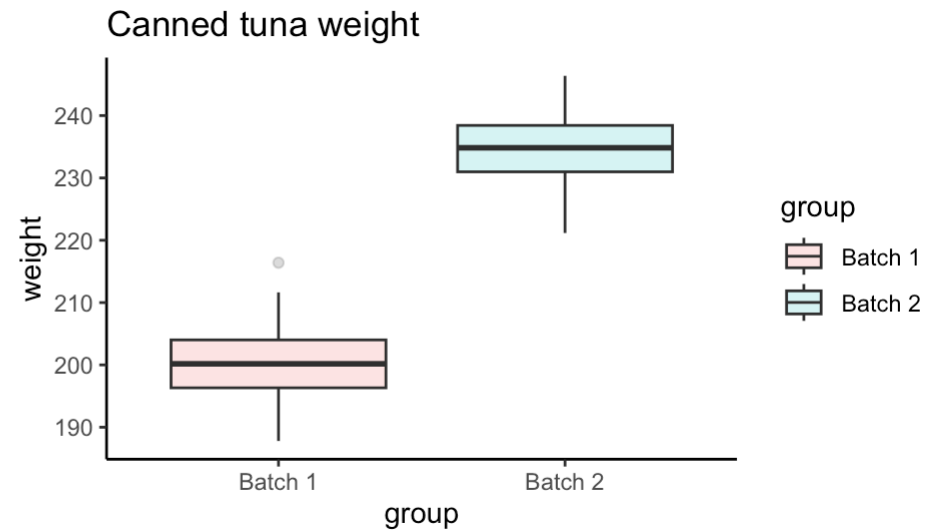
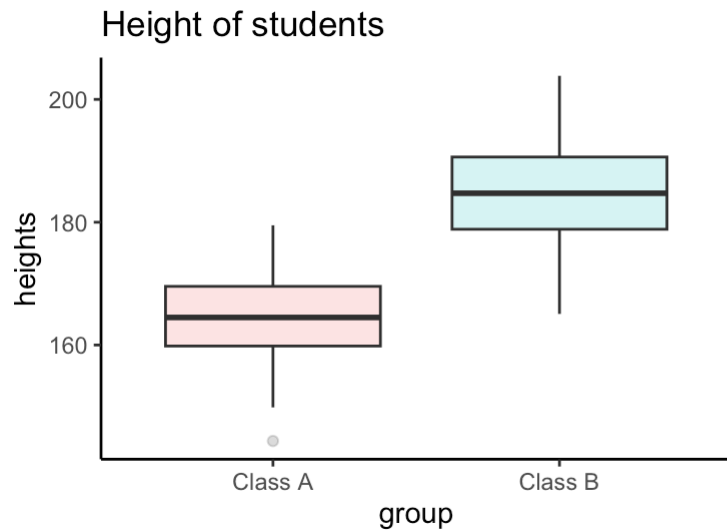
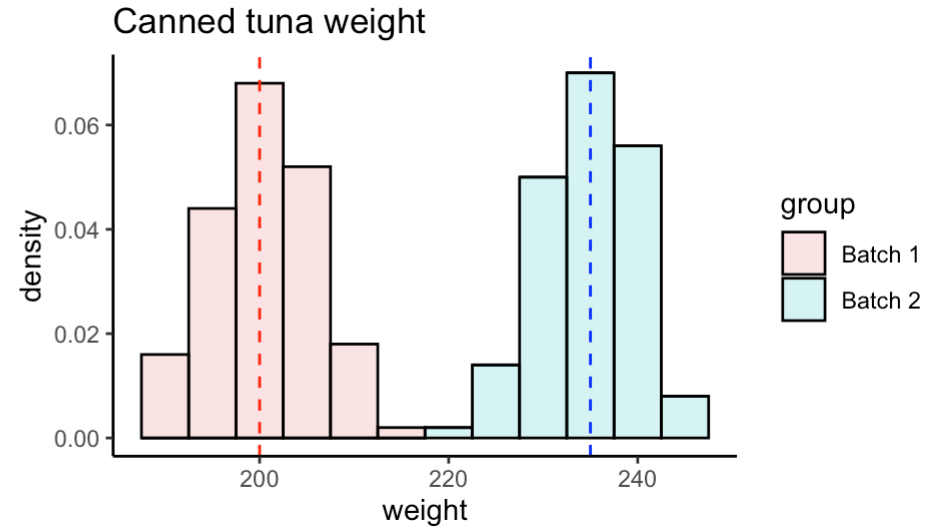
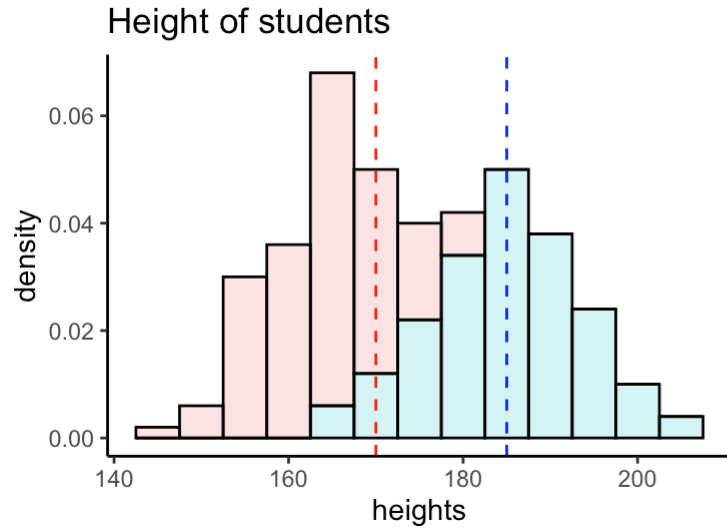
Examples

- *Is the mean height of students in this class different from the other class?*
- *Is the mean food weight in this batch of tuna different from the other batch?*

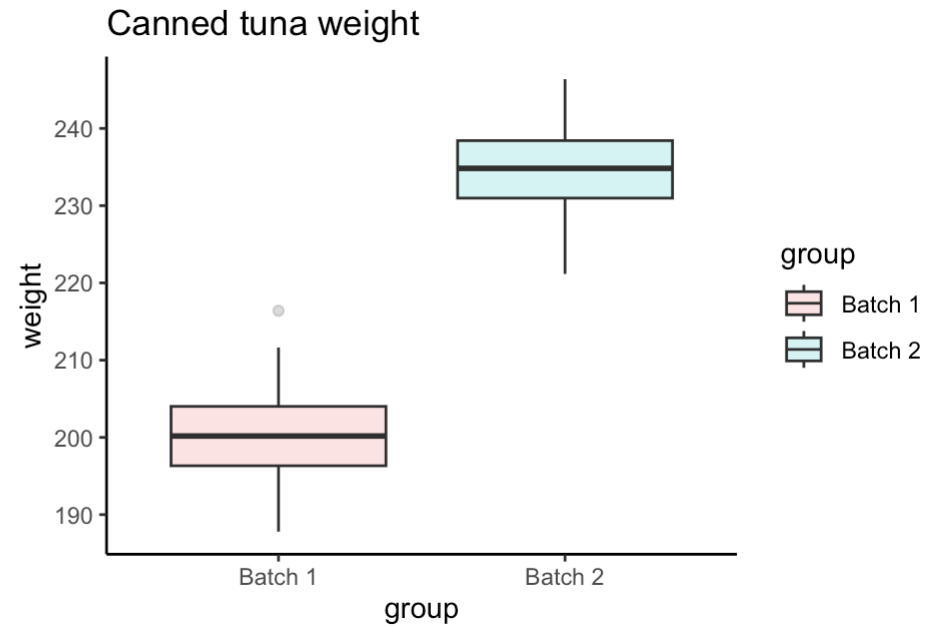
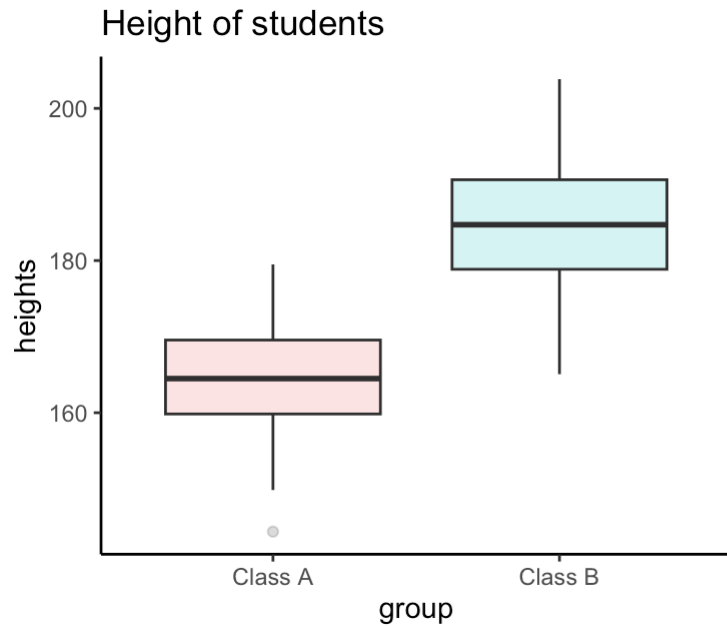


Two-sample t -test

Comparing two samples: visualisation



Some considerations: the boxplot



- Trade-off between being able to see the distribution of the data and being able to compare between groups.
- The *recommended* approach when comparing two or more groups of data in most cases.

Drinking two cans of Red Bull ‘increases risk of cardiac arrest by a fifth’ for people with an underlying heart condition

The popular drink could put people with long QT syndrome in serious danger

Sarah Young • Tuesday 18 July 2017 12:46 BST •  Comments



Does Red Bull increase the heart rate of students?

Data

A simulated example (data is not real):

Experimental design: two groups of students selected at random, *without* replacement, from the ENVX1002 cohort.

- `redbull` group: students who consumed 250 ml of Red Bull.
- `control` group: students who consumed 250 ml of water (control group).

Heart rate in beats per minute (bpm) was measured *20 minutes after consumption*.

Structure of data

```
'data.frame':  24 obs. of  2 variables:
 $ group      : chr  "redbull" "redbull" "redbull" "redbull" ...
 $ heart_rate: num  72 88 72 88 76 75 84 80 60 96 ...
```

HATPC

Hypothesis | Assumptions | Test | P-value | Conclusion

Hypothesis

For a two-sample t -test, the null hypothesis is that the means of the two groups are equal, and the alternative hypothesis is that the means are different.

$$H_0 : \mu_{\text{redbull}} = \mu_{\text{control}}$$

$$H_1 : \mu_{\text{redbull}} \neq \mu_{\text{control}}$$

Compare this to the one-sample t -test, where the null hypothesis is that the sample mean is equal to a fixed value:

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

Assumptions

The assumptions of the two-sample t -test include:

1. **Normality**: the data are normally distributed.
2. **Homogeneity of variance**: the variances of the two groups are equal.

Why are these assumptions important?

- Since the t -test compares the means of two groups, normality ensures that the means are the *best estimate* of the population means.
- Equal variances indicates that the two groups have similar “noise” influencing their means, except for the “treatment” effect.
 - ➡ In the Red Bull example, this means that the range of heart rate values in students for both groups is similar, except for the effect of consuming Red Bull.

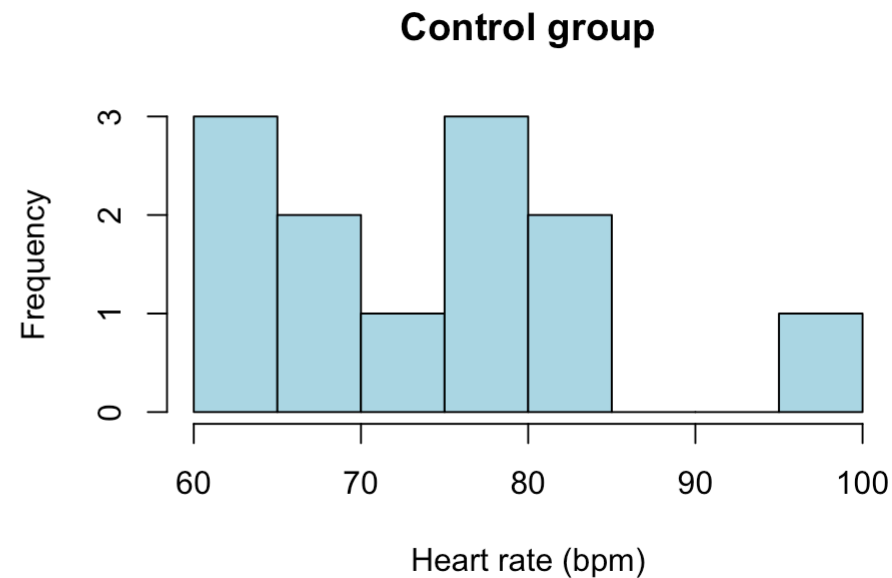
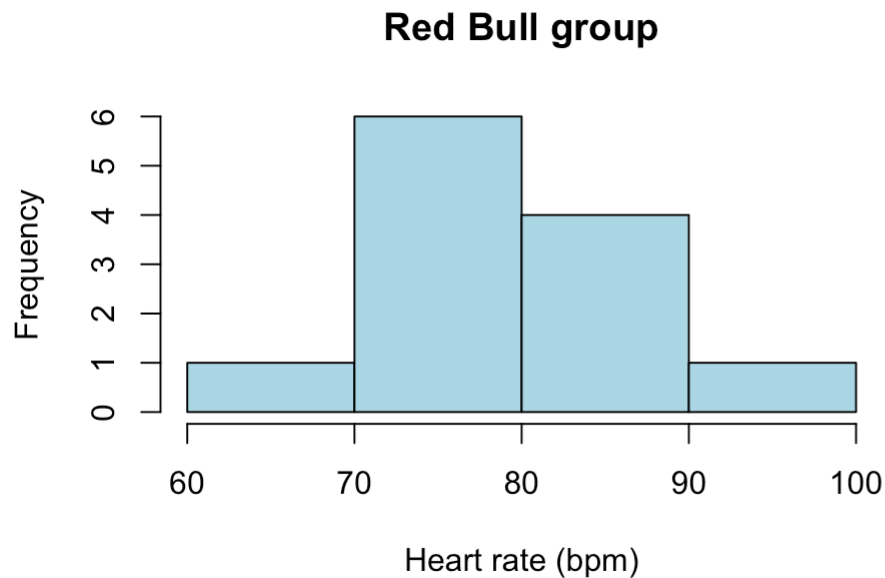
Assumption: normality

Histogram | QQ-plot | Shapiro-Wilk test

Normality: histogram

- We visually inspect the distribution of the data using histograms, generally for **each group**.
- Look out for: symmetry, skewness, and multimodality.

R base graphics ggplot2 ggpubr

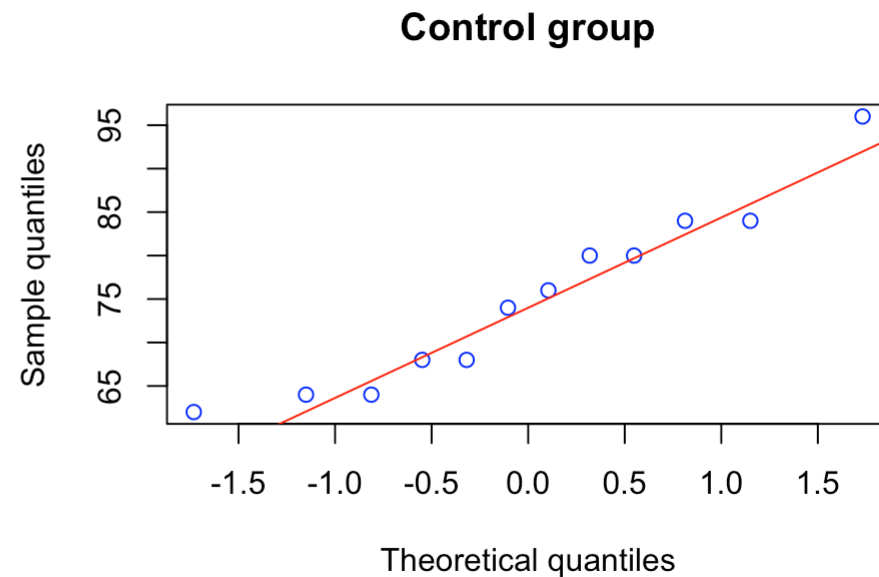
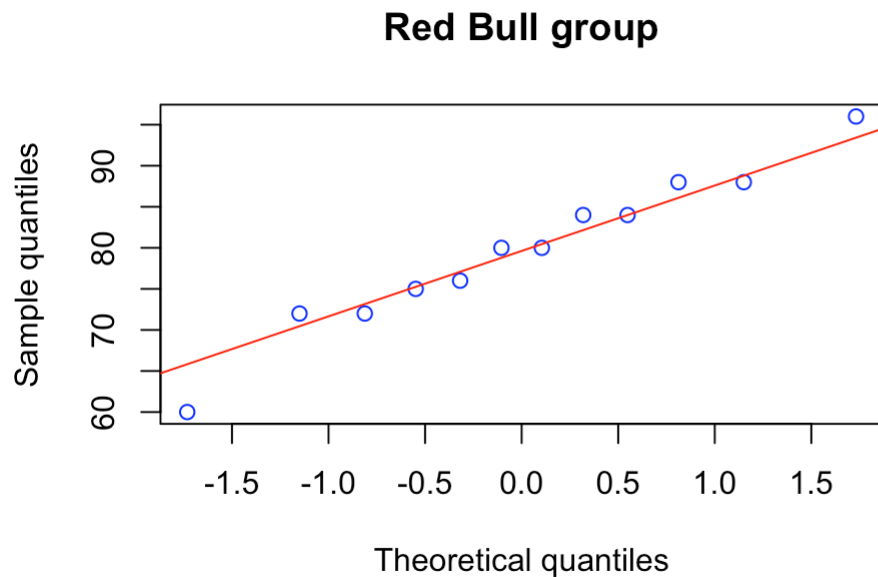


Conclusion: *The data appear to be normally distributed, but it is better to confirm this with a QQ-plot.*

Normality: QQ-plot

- The qq-plot is a graphical method to *specifically* assess the normality of the data. Again, we look at the data for **each group**.
- Look out for: deviations from the straight line.

R base graphics ggplot2 ggpubr



Conclusion: *The data appear to be normally distributed.*

Normality: formal test

- Use the Shapiro-Wilk test which tests the null hypothesis that the data are normally distributed.
- This test is sensitive to deviations from normality in the tails of the distribution, and is suitable for small sample sizes (about 5 to 50 observations).

```
Shapiro-Wilk normality test
```

```
data: redbull$heart_rate[redbull$group == "redbull"]  
W = 0.97459, p-value = 0.9524
```

```
Shapiro-Wilk normality test
```

```
data: redbull$heart_rate[redbull$group == "control"]  
W = 0.93733, p-value = 0.4643
```

Conclusion: p -values are greater than 0.05, so we do not reject the null hypothesis of normality. The data are normally distributed.

What if the normality assumption is violated?

- The t -test is robust to deviations from normality, especially for large sample sizes due to the Central Limit Theorem.
- If the sample size is small, consider using a non-parametric test (e.g. the Wilcoxon rank-sum test): **next week**
- Alternatively, transform the data: **tomorrow**

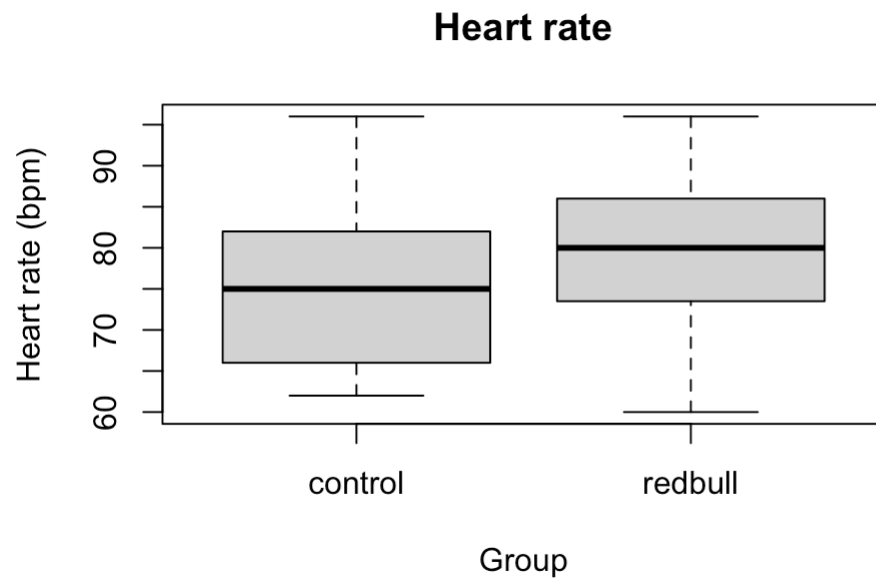
Assumption: homogeneity of variance

Boxplots | Formal tests

Equal variances: boxplot

- We visually inspect the spread of the data using boxplots, generally for **each group**.
- Look out for: differences in spread, outliers, and symmetry.

R base graphics ggplot2 ggpubr



Conclusion: *The spread of the data appears to be similar between the two groups.*

Equal variances: formal tests

- Bartlett's and Levene's tests may be used to test the null hypothesis that the variances of the groups are equal.
- These tests are sensitive to deviations from normality (Levene's test is less so compared to Bartlett's), and are suitable for small sample sizes.

Levene's test Bartlett's test

```
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  1    0.289 0.5962
      22
```

Conclusion: p -values are greater than 0.05, so we do not reject the null hypothesis of equal variances. The variances of the two groups are equal.

What if the equal variance assumption is violated?

Some debate exists on what to do, but choices include:

- Use the **Welch's t -test**, which is robust to unequal variances: *coming up next*
- **Transform** the data to stabilise the variances: **tomorrow**
- Perform a non-parametric test: **next week**

The Welch's t -test

- The Welch's t -test is a modification of the two-sample t -test that **does not assume equal variances**.
- Also applicable when the **sample sizes are unequal**.

Why not use the Welch's t -test all the time?

- Ongoing debate on whether to use the Welch's t -test or the Student's t -test when the variances are equal.
- The Welch's t -test is generally considered more robust, and **is the default in R's `t.test()` function**.
- You can still use the Student's t -test by setting `var.equal = TRUE` in the `t.test()` function.

Are the assumptions of normality and homogeneity of variance met?

When reporting in journals, it is common to simply state that the assumptions were met and what tests were used to confirm them, without showing the exact results of the tests!

Example 1

The assumptions of normality and homogeneity of variance were met for the data (Shapiro-Wilk test, $p > 0.05$; Levene's test, $p > 0.05$). Thus, we performed a two-sample t-test...

Example 2

Visual inspection of the histograms, QQ-plots, and boxplots showed that the data met the assumptions of both normality and homogeneity of variance. Thus, we performed a two-sample t-test..

For your lab reports, you should show the results of the tests (because we want to check your work!)

P-value and Conclusion

Performing the *t*-test

Two Sample *t*-test

```
data: heart_rate by group
t = -1.1365, df = 22, p-value = 0.268
alternative hypothesis: true difference in means between group control and group redbull is not equal to 0
95 percent confidence interval:
 -12.947117    3.780451
sample estimates:
mean in group control mean in group redbull
          75.00000          79.58333
```

Results indicate that the means of the two groups are **not** significantly different ($p = 0.27$).

Compare with the Welch's *t*-test

Welch Two Sample *t*-test

```
data: heart_rate by group
t = -1.1365, df = 21.845, p-value = 0.2681
alternative hypothesis: true difference in means between group control and group redbull is not equal to 0
95 percent confidence interval:
 -12.950568    3.783902
sample estimates:
mean in group control mean in group redbull
          75.00000          79.58333
```

Conclusion

Differences in heart rate were not statistically significant between the Red Bull and control groups ($t_{22} = -1.1$, $p = 0.27$) indicating that Red Bull did not significantly increase the heart rate of students sampled from ENVX1002.

The paired t -test

Are the two sample independent?

When testing if two samples are different from each other, we need to consider two possible scenarios:

- **Independent samples:** The samples are drawn from two different populations, **or** the samples are not related to each other – **independent groups**.
- **Related samples:** The samples are drawn from the same population, **and/or** the samples are related to each other – **repeated measures** or **matched pairs**.

If the samples are related, a **paired t -test** is more appropriate than a two-sample t -test as it accounts for the relationship between the samples that could confound the results.

Paired t -test

Experimental design (what if?)

Before

Two groups of students selected at random, *without* replacement, from the ENVX1002 cohort.

Paired design

The same student was used in a before/after experiment, where the heart rate was measured before and after consuming 250ml of Red Bull. Twelve (12) students were selected at random from the ENVX1002 cohort.

- Data is no longer independent, as the same student is measured twice.
- The student now confounds the results, as the heart rate of the same student is likely to be **correlated** even without consuming Red Bull.
- Total number of students is now 12, not 24.
- **Let's assume the data collected are exactly the same.**

Hypothesis

For a paired t -test, the null hypothesis is that the mean difference between the two groups is zero, and the alternative hypothesis is that the mean difference is different from zero.

$$H_0 : \mu_{\text{diff}} = 0$$

$$H_1 : \mu_{\text{diff}} \neq 0$$

Compare this to the two-sample t -test, where the null hypothesis is that the means of the two groups are equal:

$$H_0 : \mu_{\text{redbull}} = \mu_{\text{control}}$$

$$H_1 : \mu_{\text{redbull}} \neq \mu_{\text{control}}$$

Assumptions of the paired t -test

- The assumption of the paired t -test is that the differences between the two groups are normally distributed.
- There is no assumption of equal variances, as the paired t -test is a one-sample t -test on the differences.
 - ➡ Another way to think about it is that since the data are paired, the variance of the differences is the same for both groups.

Performing the paired t -test

There are two ways.

Method 1: Calculate the differences, then perform a one-sample t -test using

`t.test()`

One Sample t -test

```
data: diff
t = 1.6578, df = 11, p-value = 0.1256
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -1.501628 10.668294
sample estimates:
mean of x
 4.583333
```

Method 2: Use the `t.test()` function with the `paired = TRUE` argument

The results for both methods are identical; the mean difference is not significantly different from zero ($p = 0.13$).

What if the assumptions are violated?

We will explore this in tomorrow's lecture!

Thanks!

This presentation is based on the [SOLES Quarto reveal.js template](#) and is licensed under a [Creative Commons Attribution 4.0 International License](#).

References

- Quinn G. P. & Keough M. J. (2002) Experimental design and data analysis for biologists. Cambridge University Press, Cambridge, UK.
- Logan, M. (2010). Biostatistical design and analysis using R a practical guide. Hoboken, N.J., Wiley-Blackwell.