One-sample *t*-test in R

Cheatsheet

2024-08-10

> **ℹ Assumed knowledge**
>
> - You know how to install and load packages in R.
> - You know how to import data into R.
> - You recognise data frames and vectors.

> **❗ Data structure**
>
> The data should be in a **long format** (also known as tidy data), where each row is an observation and each column is a variable (Figure 1). If your data is not already structured this way, reshape it manually in a spreadsheet program or in R using the `pivot_longer()` function from the `tidyr` package.

Sex	BW
F	2.15
M	2.55
F	2.95
F	2.70
M	2.20
F	1.85
M	2.55
M	2.60

F	M
2.15	2.55
2.95	2.20
2.70	2.55
1.85	2.60

Figure 1: Data should be in long format (left) where each row is an observation and each column is a variable. This is the preferred format for most statistical software. Wide format (right) is also common, but may require additional steps to analyse or visualise in some instances.

> **! Data**
>
> For this cheatsheet we will use the entire possums dataset used in BIOL2022 labs.

About

The one-sample t-test is used to determine whether the mean of a single sample y is significantly different from a known or hypothesised population mean (μ). **Examples**:

- Is the mean weight of canned tuna significantly different from what was stated on the label (400 g)?
- Is the mean height of a sample of male students significantly different from the national average height (175.6 cm)?
- Is the mean number of kittens in a litter significantly different from 4?

Modelling

Is the mean **body weight** of possums (BW) significantly different from 3.5 kg?

The **simplified model** for the mathematically-adverse individual is

$$\text{body weight} \sim 3.5$$

which translates to "the body weight of possums is *around* 3.5 kg". The **statistical model** is

$$\text{body weight} = \beta_0 + \epsilon$$

where β_0 is the hypothesised population mean and ϵ is the error term.

Preparing the data

Extract **only** the variable of interest from the dataset using `select()` from the `dplyr` package – `BW`. Assign the variable to a new object – `bw` in this case.

```r
library(dplyr)
library(readxl)
possums <- read_excel("possums.xlsx", sheet = 2) # import
bw <- select(possums, BW) # select variable
```

Your own data should be in a similar format.

Analytical approaches

The traditional approach to the one-sample *t*-test is to use the `t.test()` function in R, while the modern approach is to use a general linear model (GLM) with the `lm()` or `glm()` functions.

`t.test()` function

Methods reporting

> A one-sample *t*-test was used to determine whether the mean body weight of possums was significantly different from 3.5 kg. This was computed using the `t.test()` function in R version 4.4.0 (R Core Team, 2024).

Perform the analysis

```r
t.test(bw, mu = 3.5)
```

Check assumption(s)

Normality

Any combination of one or more of the following checks can be used to assess normality:

- **Histogram**: `hist(bw$BW)`
- **Q-Q plot**: `qqnorm(bw$BW)`
- **Shapiro-Wilk test**: `shapiro.test(bw$BW)`

Include the appropriate description in your methods section.

> The normality of body weight was assessed using [insert method(s)].

How to report results

> The mean body weight of possums was significantly different from 3.5 kg (t_{19} = -10.3, 95% CI [2.3, 2.7], p < 0.001).

`lm()` function

Methods reporting

> A general linear model was used to determine whether the mean body weight of possums was significantly different from 3.5 kg. This was computed using the `lm()` function in R version 4.4.0 (R Core Team, 2024).

Perform the analysis

For a one-sample *t*-test, the formula needs to be specified as `y - μ ~ 1` where `y` is the variable of interest and μ is the hypothesised value that is being tested. The `1` indicates that the model has an intercept only i.e. we are testing whether the mean difference is significantly different from 0.

```
fit <- lm((BW - 3.5) ~ 1, data = bw)
summary(fit)
```

Check assumption(s)

Normality

With a GLM, normality can be assessed using the residuals of the model. The following checks can be used:

- **Histogram**: `hist(residuals(fit))`
- **Q-Q plot**: `qqnorm(residuals(fit))`
- **Shapiro-Wilk test**: `shapiro.test(residuals(fit))`

How to report results

There is evidence to suggest that the mean body weight of possums was significantly different from 3.5 kg (GLM, t_{19} = -10.3, p < 0.001).

Exercise(s)

Download the penguins dataset (from below if you are reading this in HTML), or load the dataset from the **palmerpenguins** package. Perform a one-sample t-test to determine whether the mean flipper length of penguins is significantly different from 200 mm.