Lab 05 – Hypothesis testing

ENVX1002 Handbook

Semester 1, 2025

Welcome

Learning outcomes

- Learn to use R to calculate a 1-sample t-test
- Apply the steps for hypothesis testing from lectures
- Learn how to interpret statistical output

Before you begin

You can download the data

1) From module 5 in Canvas
2) ENVX1002_Data5.xlsx if you are viewing the html file from Github https://Github.com/envx-resources

Create a new project

Reminder (skip to step 2 if you are going to use the directory you created in your tutorial)

Step 1: Create a new project file for the practical put in your ENVX1002 Folder. *File > New Project > New Directory > New Project*.

Step 2: Download the data files from canvas or using above link and copy into your project directory.

I recommend that you make a data folder in your project directory to keep things tidy! If you make a data folder in your project directory you will need to indicate this path before the file name.

Step 3: Open a new Quarto file.

i.e. *File > New File > Quarto Document* and save it immediately i.e. File > Save.

Problems with your personal computer and R

NOTE: If you are having problems with R on your personal computer that cannot easily be solved by a demonstrator, please use the Lab PCs.

Installing packages

Remember All of the functions and data sets in R are organised into packages. There are the standard (or base) packages which are part of the source code - the functions and data sets that make up these packages are automatically available when R is opened. There are also many contributed packages. These have been written by many different authors, often to implement methods that are not available in the base packages. If you are unable to find a method in the base packages, you might be able to find it in a contributed package. The Comprehensive R Archive Network (CRAN) site (http://cran.r-project.org/) is where many contributed packages can be downloaded. Click on packages on the left hand side. We will download two packages in this class using the `install.packages` command and we then load the package into R using the library command.

Alternatively, in RStudio click on the *Packages tab > Install > type in package name > click install.*

Exercise 1: 1-sample t-test Milk Yield - Walk through

This exercise will walk you through how to test a hypothesis, check assumptions and eventually draw a conclusion on your initial hypothesis. 100 cows have their milk yield measured. Suppose we wish to test whether these milk yields (units unknown) differ significantly from the economic threshold of 11 units. (The units may possibly be litres of milk produced on a particular day).

> Fact
>
> The average Australian drinks about 100 litres of milk per year. The average cow produces between 12 and 30 litres of milk per day.

The data is in the `Milk` sheet found in the `ENVX1002_Data5.xlsx` file. You will follow the steps as outlined in the lectures:

1. Choose level of significance (α)
2. Write null and alternate hypotheses
3. Check assumptions (normal)
4. Calculate test statistic
5. Obtain P-value or critical value
6. Make statistical conclusion
7. Write a scientific (biological) conclusion

Remember you can remember the above using **HATPC**

Lets go:

1. Normally you choose *0.05* as a level of significance:

This value is generally accepted in the scientific community and is also linked to type 2 errors where choosing a lower significance increases the likelihood of a type 2 error occurring.

2. Write null and alternative hypotheses:

Question: Write down the null hypothesis and alternative hypotheses:
H_0: < Type your answer here >
H_1: < Type your answer here >

---

3. Check assumptions (normality):

a. load data:

Make sure you set your working directory first

```
# Type your R code here
```

It is always good practice to look at the data first to make sure you have the correct data, it loaded in correctly and know what the names of the columns are. This can be done by typing the name of the data `Milk` or for large datasets, use `head()` to show the first 6 lines. You can also use `str()` to look at the strucutre of the data.

```
# Type your R code here
```

b. Tests for normality:

qqplots:

```
# Type your R code here
```

Histogram and boxplots:

```
# Type your R code here
```

Question: Do the plots indicate the data are normally distributed?
Answer: < Type your answer here >

Shapiro-Wilk test of normality:

```
# Type your R code here
```

Question: Does the Shapiro-Wilk test indicate the data are normally distributed? Explain your answer.
Answer: < Type your answer here >

---

4. Calculate the test statistic

In R we achieve this via the command `t.test(milk$Yield, mu = …)` The R output first gives us the calculated t value, the degrees of freedom, and then the p-value, it then provides the 95% CI and the mean of the sample. Were `mu = …` is written enter in the hypothesised mean.

```
# write your R code here
```

5. Obtain P-value or critical value

Question: Does the hypothesised economic threshold lie within the confidence intervals?
Answer: < Type your answer here >

---

6. Make statistical conclusion

Question:: Based on the P-value, do we accept or reject the null hypothesis?
Answer: < Type your answer here >

---

7. Write a scientific (biological) conclusion

Question:: Now write a scientific (biological) conclusion based on the outcome in 6.
Answer: < Type your answer here >

Exercise 2: Stinging trees (individual or in pairs)

Data file: Stinging.csv

A forest ecologist, studying regeneration of rainforest communities in gaps caused by large trees falling during storms, read that stinging tree, **Dendrocnide excelsa**, seedlings will grow 1.5m/year in direct sunlight such as gaps. In the gaps in her study plot, she identified 9 specimens of this species and measure them in 1998 and again 1 year later.

Does her data support the published contention that seedlings of this species will average 1.5m of growth per year in direct sunlight? Also, calculate a 95% CI for the true mean. Analyse the data in R. Due to the small sample size we have to assume the data is normal.

Fact

It was found that researchers wearing welding gloves and a full body suit were still stung by the tree. The sting is extremely painful and can last for months. The pain is caused by a neurotoxin that is injected into the skin. The tree is found in the rainforests of north-eastern Australia.

Work through the steps below individually or in pairs. Add more code chunks if required (click insert -> R on above toolbar)

---

1. Choose level of significance (α)
   Answer:

---

2. Write null and alternate hypotheses
   H_0:
   H_1:

---

3. Check assumptions (normal)

Read in the data:

```
library(readxl)
sting <- read_excel("data/ENVX1002_Data5.xlsx", sheet = "Stinging")
sting
```

Plot your data:

```
# Type your R code here
```

Normality tests:

```
# Type your R code here
```

Question: Are data are normally distributed? Explain your answer.
Answer: < Type your answer here >

4. Calculate test statistic and
5. Obtain P-value or critical value

```
# Type your R code here
```

6. Make statistical conclusion
   Answer:

7. Write a scientific (biological) conclusion
   Answer:

Check you answers with teaching staff

Thanks!

Bonus take home exercices

For each of these exercises, follow the steps outlined in the lectures (and this lab!) to test your hypotheses:

1. Choose level of significance (α)
2. Write null and alternate hypotheses
3. Check assumptions (normal)
4. Calculate test statistic
5. Obtain P-value or critical value
6. Make statistical conclusion
7. Write a scientific (biological) conclusion

Exercise 1: Carrots

A farmer is growing carrots for a restaurant. The restaraunt wants their carrots to be 10 cm long, so the farmer wants to check if the carrots in their field differ significantly from the needed length.

```
#Read in data

carrots <- c(7, 7, 13, 5, 13, 10, 11, 12, 10,  9)
```

Exercise 2: Penguins

Rey has just landed on earth and notived that penguins look really similar to porgs. Using weight as the point of comparison, she wants to know if two different penguin species weigh the same as her pet Porg Stevie, who weighs 4000g.

We will be using the Palmer penguin dataset to test if chinstrap and gentoo penguins weigh the same as Stevie.

```
#install.packages("palmerpenguins")
library(palmerpenguins)
```

2.1 Chinstrap

```
#Data cleaning and subsetting for exercise
#Copy as is!

library(tidyverse)
chinstrap <-  penguins%>%
  filter(species == "Chinstrap")%>% #subset to only include chinstrap penguins
  na.omit() #exclude missing data
```

2.2 Gentoo

```
#Data cleaning and subsetting for exercise
#Copy as is!

gentoo <-penguins%>%
  filter(species == "Gentoo")%>% # Subset to only include this species
  na.omit() # exclude missing data
```

Attribution

This lab was developed using resources that are available under a Creative Commons Attribution 4.0 International license, made available on the SOLES Open Educational Resources repository.