Lab 03A: Week 8

Contents

The **specific aims** of this lab are:

- understand the issues around multiple testing and how to overcome these issues
- practice performing (one-way) ANOVA and interpreting the output
- check ANOVA assumptions

The unit **learning outcomes** addressed are:

- LO1 Formulate domain/context specific questions and identify appropriate statistical analysis.
- LO2 Extract and combine data from multiple data resources.
- LO3 Construct, interpret and compare numerical and graphical summaries of different data types including large and/or complex data sets.
- LO5 Identify, justify and implement appropriate parametric or non-parametric statistical tests.
- LO6 Formulate, evaluate and interpret appropriate linear models to describe the relationships between multiple factors.
- LO8 Create a reproducible report to communicate outcomes using a programming language.

1 Quick quiz

1.1 Multiple testing

1.1 Multiple testing

I generate 4 samples of size $n = 20$ from a $N(3, 1)$ distribution and for each of these 4 samples, I test the null hypothesis $H_0 : \mu = 3$ against the alternative $H_1 : \mu \neq 3$ using $\alpha = 0.1$ as my level of significance.

What's the probability that I falsely reject **at least one** of the four null hypotheses?

a. 0.1

b. 0.3439

c. 0.4

d. 0.0001

e. 0

1.2 Checking for normality

Which of these is the best (graphical) method to determine whether a set of data come from a normal distribution?

a. histogram

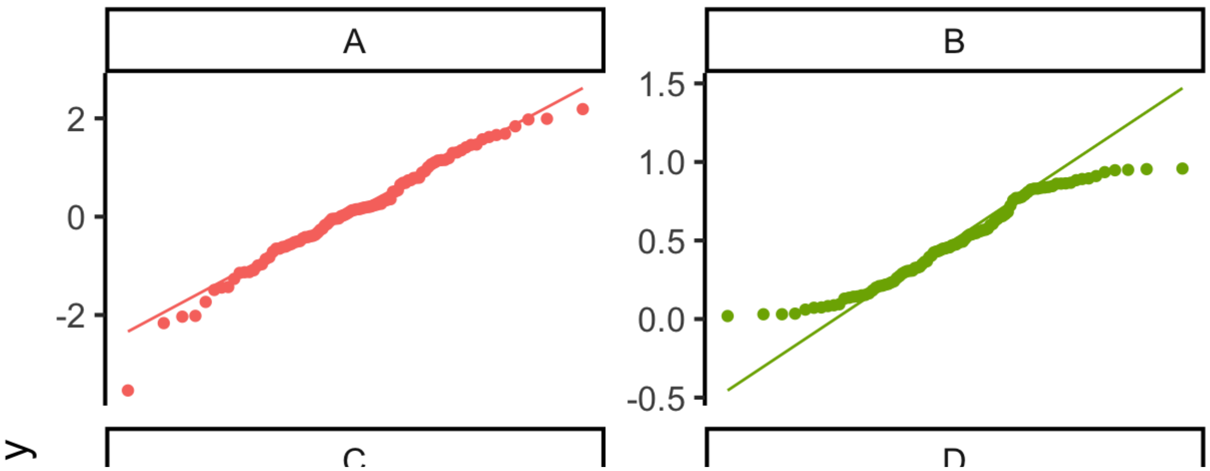
b. boxplot

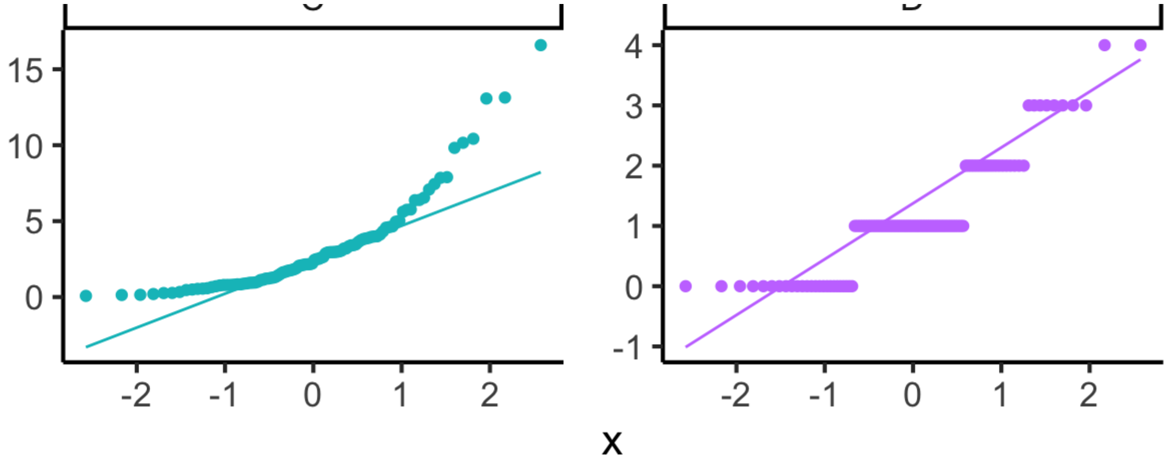
c. normal quantile (QQ) plot

d. all of the above

1.3 Identifying normal

Which of these normal QQ plots indicate that the data come from a normal population?





1.4 Which test?

1. Which method of inference is most useful for testing whether average adult body temperature is 37 degrees Celsius?

a. 1-sample t-test

b. 2-sample t-test

c. paired t-test

d. ANOVA

2. Which method of inference is most useful for testing for a difference in the average scores of 2 judges who judge the same set of dives?

a. 1-sample t-test

b. 2-sample t-test

c. paired t-test

d. ANOVA

3. Which method of inference is most useful for testing whether house prices in Sydney are more expensive than in Melbourne?

a. 1-sample t-test

b. 2-sample t-test

c. paired t-test

d. ANOVA

4. To compare the average leaf thickness of 3 different species of conifers, a scientist randomly samples 10 leaves from each species. What hypothesis test is appropriate to determine if there is a

samples 10 leaves from each species. What hypothesis test is appropriate to determine if there is a difference among the mean leaf thicknesses of the 3 species?

a. 1-sample t-test

b. 2-sample t-test

c. paired t-test

d. ANOVA

5. To compare the average leaf thickness of 3 different species of conifers, a scientist randomly samples 10 leaves from each species. What is the distribution of the ANOVA test statistic T, if H_0 is true?

a. $T \sim F_{3,\, 30}$

b. $T \sim F_{2,\, 30}$

c. $T \sim F_{3,\, 27}$

d. $T \sim F_{2,\, 27}$

e. $T \sim F_{3,\, 10}$

f. $T \sim F_{2,\, 10}$

6. To compare the average leaf thickness of 3 different species of conifers, a scientist randomly samples 10 leaves from each species. An ANOVA yields a p-value of 0.007. What is the appropriate conclusion?

a. There is no evidence of a difference among the mean leaf thicknesses of the 3 conifer species.

b. There is strong evidence that at least 2 of the mean leaf thicknesses are different.

c. There is strong evidence that all 3 mean leaf thicknesses are different.

d. There is some evidence that the mean leaf thicknesses are different, but not enough to reject H_0.

2 Group question

The lifetimes of 10 Brand A batteries, 12 Brand B batteries, and 9 Brand C batteries were recorded. A partially filled in ANOVA table is below. What is the value of the test statistic F?

```
             Df Sum Sq Mean Sq F value   Pr(>F)
Brand        __  238.1  ______   _____   ______
Residuals    __  _____  ______
Total        __ 1524.1
```

a. 1.666

b. 2.348

c. 2.592

d. 2.777

3 Questions

3.1 Critical Flicker Frequency

If a light is flickering but at a very high frequency, it appears to not be flickering at all. Thus there exists a “critical flicker frequency” where the flickering changes from “detectable” to “not detectable” and this varies from person to person.

The critical flicker frequency and iris colour for 19 people were obtained as part of a study into the relationship between critical frequency flicker and eye colour. They are given in the file `flicker.txt`.

```r
library(tidyverse)
flicker = read_tsv("https://raw.githubusercontent.com/DATA2002/data/master/flicker.txt")
glimpse(flicker)
```

```
Rows: 19
Columns: 2
$ Colour  <chr> "Brown", "Brown", "Brown", "Brown", "Brown", "Brown"…
$ Flicker <dbl> 26.8, 27.9, 23.7, 25.0, 26.3, 24.8, 25.7, 24.5, 26.4…
```

a. Generate side by side boxplots as well as normal QQ plots for the flicker data. Do your plots support the assumptions required for an ANOVA test to be valid? Explain.

b. Use the `aov()` function to perform an ANOVA test for the equality of means flicker level across each of the three eye colours.

c. Using the output, write out the hypothesis test in full. Be sure to state the null and alternative hypothesis, assumptions, test statistic (with distribution), observed test statistic, p-value and an appropriate conclusion.

3.2 Blonds

In an investigation into the relationship between tolerance to pain and hair colour, men and women of various ages were divided into 4 groups based on hair colour and given a pain sensitivity test. Each person’s “pain threshold score” (higher score means higher pain tolerance) is recorded in the file `blonds.txt`.

```r
pain = read_tsv("https://raw.githubusercontent.com/DATA2002/data/master/blonds.txt")
glimpse(pain)
```

```
Rows: 19
Columns: 2
$ HairColour <chr> "LightBlond", "LightBlond", "LightBlond", "LightB…
$ Pain       <dbl> 62, 60, 71, 55, 48, 63, 57, 52, 41, 43, 42, 50, 4…
```

1. Change `HairColour` so that the ordering is preserved from lightest to darkest. Hint use: `factor()`
2. Generate boxplots and QQ plots to check the ANOVA assumptions.
3. What do the boxplots suggest about the null hypothesis that pain thresholds are the same regardless of hair colour?
4. Test this hypothesis formally using ANOVA. Does there seem to be a relationship between hair colour and pain threshold?!

3.3 Hedenfalk data

The package **sgof** has a data set `Hedenfalk` (Conde and Una Alvarez 2020). You may need to `install.packages("sgof")`. Use `?Hedenfalk` to find out more about this data set. (If you can't install the package, the data can be found here and the help page can be found here.)

a. How many p-values are in the data set?

b. Generate a histogram of the (unadjusted) Hedenfalk p-values.

c. How many (unadjusted) p-values are significant at the 5% level of significance? What proportion of all p-values in the data set is this?

d. Why is it a good idea to consider adjusted p-values?

e. Using `p.adjust()` find the Bonferroni and BH p-values. Plot histograms of each and find the number of significant results after adjustment for both.

f. Comment on the difference between the Bonferroni method and the BH method.

4 For practice after the computer lab

From Larsen and Marx (2012) you could work through Case Study 12.2.1 and then consider these questions: 12.2.1, 12.2.2, 12.2.3, 12.2.4, 12.2.5, 12.2.6, 12.2.7, 12.2.8, 12.2.11, 12.2.12 and 12.2.13.

You can also attempt the DataCamp chapter on comparing many means.

References

Conde, Irene Castro, and Jacobo de Una Alvarez. 2020. *Sgof: Multiple Hypothesis Testing*. https://CRAN.R-project.org/package=sgof.

Larsen, Richard J., and Morris L. Marx. 2012. *An Introduction to Mathematical Statistics and Its Applications*. 5th ed. Boston, MA: Prentice Hall.