

# Group Project: Queensland Flying Fox Monitoring

300700 Statistical Decision Making, Autumn 2017

---

Due: Friday of Week 13 (19 May)

This Group Project is about data on flying fox populations obtained from the Queensland Government website<sup>1</sup>.

The data set `FlyingFoxMonitoring` provided with this project in csv format contains data on the population sizes of three species of flying foxes at various locations.

## 1 Load and clean up data

[1 mark] Some of the records (rows) in the data set are inconsistent, in that either the total species count does not equal the sum of the counts for the three species, or the variable `FlyingFoxesAbsent` does not match the total count being zero or not.

Remove inconsistent records from the data set.

## 2 Camp sizes

For this part, we only consider records for camps that are not empty, that is, where there are flying foxes present.

- (a) [2 marks] Analyse the distribution of the total camp size (`TotalSpecies.Count`) and the distribution of the *natural logarithm*<sup>2</sup> of the total camp size for non-empty camps.

Explain why the logarithm of the camp size is the more meaningful variable to consider.

- (b) [2 marks] Use bootstrapping to compute a 95% confidence interval for the mean logarithm of the total camp size for non-empty camps.

Interpret your result.

## 3 Species proportions

- (a) [2 marks] Compute the proportions of each species at each camp site and visualise them. Then compute the overall proportions of each species.

Discuss your results.

- (b) [2 marks] Analyse whether the total camp size and the proportion of Little Red Flying Foxes are associated by testing at a significance level of 1% whether there is evidence that the correlation between the total size of

a camp and the proportion of Little Red Flying Foxes in that camp is non-zero.

Interpret your results.

## 4 Camp size at Indooroopilly Island over time

[2 marks] Use concepts related to linear regression to test at a significance level of 1% whether there is evidence that the number of flying foxes at the Indooroopilly Island camp (EHP CampId 68) has **decreased** over time.

Use the function `as.Date` to convert the date strings in the variable `SurveyDate` into dates. (Check the manual page and make sure that you use the correct format string for the dd/mm/yy format used in the csv file!) If needed, you can use the function `as.numeric` to convert the resulting date objects to a numeric value.

## 5 Species proportions at Laidley Plainlands Road camp

We want to test whether the Species proportions at the Laidley Plainlands Road camp (EHP CampID 76) have changed between 18/10/2013 and 13/11/2013.

- (a) [1 mark] Use a  $\chi^2$ -test for goodness of fit to test whether the species counts on 13/11/2013 differ significantly from the species proportions on 18/10/2013.
- (b) [1 mark] Use a  $\chi^2$ -test for independence to test whether the species counts are associated to the categorical variable with the categories 18/10/2013 and 13/11/2013.
- (c) [1 mark] Compare the  $p$ -values obtained in (a) and (b). Discuss the difference between the tests (a) and (b), and explain which is more appropriate and why.

## Submission

One report is to be submitted by each group by the due date, containing the description of a solution to the tasks above, including any R code, and the results obtained.

**The report is to be written in R Markdown, using the template available on the unit's vUWS site.**

After editing the R Markdown file, *knit* it to Word (not HTML!) in R Studio (see <http://rmarkdown.rstudio.com>

---

<sup>1</sup><https://data.qld.gov.au>

<sup>2</sup>The natural logarithm is the inverse of the exponential function; it satisfies  $\log(a \cdot b) = \log a + \log b$  for all  $a, b > 0$ . The R function to compute the natural logarithm is called `log`.

By including this statement, all authors of this work declare that:

- We hold a copy of this assignment if the original is lost or damaged.
- We hereby certify that no part of this assignment has been copied from any other student's work or from any other source except where due acknowledgement is made in the assignment.
- No part of the assignment has been written for us by any other person except where collaboration has been authorised by the unit coordinator.
- We are aware that this work may be reproduced and submitted to plagiarism detection software programs for the purpose of detecting possible plagiarism; this software may retain a copy on its database for future plagiarism checking.
- We hereby certify that no part of this assignment or product has been submitted by any of us in another (previous or current) assessment, except where appropriately referenced, and with prior permission from the unit coordinator for this unit.
- We hereby certify that we have read and understand what the University considers to be academic misconduct, and that we are aware of the penalties that may be imposed for academic misconduct.

Name	Student Number	Contribution (%)

Figure 1: Statement to be included on the first page of each submission.

for full details on R Studio and R Markdown) and then convert the resulting .docx file to PDF using MS Word. Alternatively, you can *knit* to HTML and convert the resulting .html file to PDF using a web browser (with a suitable plugin or a virtual printer). If you have L<sup>A</sup>T<sub>E</sub>X installed (<https://www.latex-project.org>), you can also *knit* your R Markdown file directly to PDF.)

After checking that the PDF file is formatted correctly, submit it using the link *Group Project* in the *Group Project* tab on the unit's vUWS site. **Do not submit a file in any format other than PDF.** If you submit a file in another format, it may not be possible to mark your report at all, or you may lose marks due to bad formatting of code, plots or text.

The first page of your report must contain the declaration shown in Figure 1; you make this declaration by submitting your report. **Do not remove this declaration! A marker has the right not to mark your report if the above declaration is not included in the report.**

## Marking Criteria and Standards

The Group Project will contribute a maximum of 15 marks towards your final mark.

The value of each of the tasks is indicated. Marks are awarded according to the following criteria:

- choice of correct method for sampling, bootstrapping, randomisation and / or analysis;
- correctness and clarity of R code; and
- correctness and clarity of analysis or interpretation.

In addition, 1 mark is awarded for the overall quality and presentation of the report.

Groups should discuss all aspects of the project: Working in a team on a statistical project is an assessable learning outcome of the unit, and the report will be treated as a team submission. The contributions by each team member must be indicated on the cover sheet.

Remember that the marker will only see what you have written, therefore, comment your R code and clearly explain all decisions made, as well as the analysis and your interpretation of the results. (Don't expect the marker to spend ages trying to figure out what you might have meant to say!)

The formatting of your report may affect its readability and the clarity of your explanations, and hence contributes to your mark.