Statistical Thinking

Task 1

Week 1 S2 2023

Introduction

This is the first task for ETC2420 and is worth 6 marks.
This task is due Friday Aug 11 at 4:30pm.

AIM

Essentially this task is designed to ensure that you can:
- Create an Rmarkdown document
- Adapt R code to a new task
- Write comments around R output
- Create a pdf document (via html or directly to pdf)
- Proofread and edit a document for presentation
- Submit Rmd and pdf files
- Follow instructions

Please ask for help if you get stuck

The task

We want you to analyse the present (present.csv) birth rate and compare this to the Arbuthnot example from the Tutorial 1 handout.

This task is your first attempt at analysing data and writing your analysis as a report. It is also aimed at training you to carefully consider presentation and think about what should and should not appear in a report.

The present dataset is an updated version of the historical Arbuthnot dataset. Numbers of boys and girls born in the United States between 1940 and 2002.[1]

You must create an RMarkdown document and name it with the form:

First name_Last name_ID. For example, I would name it Lachlan_Macquarie_123456. (Please use the same name as your enrolment).

[1]From OpenIntro package. Source: Mathews, T. J., and Brady E. Hamilton. "Trend analysis of the sex ratio at birth in the United States." National vital statistics reports 53.20 (2005): 1-17.

Submission

This task is due Friday Aug 11 at 4:30pm.
Submit **BOTH** your Rmd and pdf files in the Moodle link (check announcements).
You may create an html document but you **must** convert this to pdf.
Submission of incorrect files will incur a penalty (see Rubric below).

Presentation

You should clearly show which question you are answering.
- Please show the code chunks used to create your output.
- Please hide warnings.
- Make sure code is wrapped (i.e. the entire code chunk is visible on your pdf document)

Rubric

Grade	Comment
6	Well formatted; Questions answered in sentences; one or two minor errors
5.9	Good, but several minor presentation issues that need to be addressed
5.5	Difficult to read; minimal attempt to answer questions; a lot of extraneous R output (warnings etc.), incorrect file &/or incorrect file names
5	As above but file is missing
4	Major issues; can't open files
0	No attempt

Instructions

Create a new RMarkdown file and load the present.csv file. Be sure that no messages appear in your rendered document.
Pay attention to presentation/formatting. Treat this as a report to your boss.
Be sure to use proper sentences in your explanations.

Question 1

Glimpse the present dataset. Show the code chunk in your report, and the output.
How do these counts compare to Arbuthnot's? Are they of a similar magnitude?

Question 2

Generate a plot of the girls born. Use triangles instead of dots and change the colour to red. *HINT: you can set type and colour in the* `geom_point()` *command. You may have to look up the commands though*

Question 3

Briefly describe (2-3 sentences only) any interesting features of the plot. Are there any similarities to the plot from Arbuthnot's data? (Do not include the Arbuthnot plot here)

Question 4

Create new variables total, ratio, prop_girls, prop_boys. Show the R code chunk that you used to do this (Try using pipes first - if you get stuck, use the same method as in the tutorial - you should have 7 variables).

Question 5

Arbuthnot observed that boys were born in greater proportion than girls (which we could see in the tutorial graph). Make a line plot that displays the proportion of boys born over time. Add a *blue* horizontal line at 0.5. Is Arbuthnot's observation still valid between 1940 and 2002 in the US? Explain.