Lab 2: Arbuthnot

STAT 20 Spring 2022

Please answer these questions in a Google Doc, with plots copied and pasted from RStudio where appropriate. Download that Google Doc as a pdf using File > Download > PDF Document and upload it to Gradescope.

Preparation: R Workshop

Please provide the R code that will generate a data frame (in `tidyverse`, these are known as "tibbles") with the following characteristics: - Three variables: Name, Favorite Color, Favorite Number - 5 observations

The code should save this data frame in an object called `my_classmates`.

Part I: Understanding the Context of the Data

1. What question did John Arbuthnot set out to answer in collecting this data?
2. What is the unit of observation in the original christening records? What are the possible variables that may have been recorded?
3. What do you think the probability is that a newborn child is recorded as a boy? What form of evidence or reasoning did you use to come to that determination?

Part II: Computing on the Data

4. What is the unit of observation in the data frame used by John Arbuthnot? What are the names of the variables and what is the type of each one?
5. What is the time frame covered by Arbuthnot's data?
6. Which year saw the greatest number of children christened?
7. What is the proportion of girls christened in 1700?
8. What is the trend over time in the total number of children christened? Please answer with a plot and written interpretation.
9. What is the trend over time in the proportion of girls christened? Please answer with a plot and written interpretation.

Part III: Extensions

10. What is the time frame covered by the present-day data?
11. In terms of general magnitude (size), how do the counts in Arbuthnot's data compare to the counts in the present-day data?
12. What is the trend over time in the proportion of girls christened? Please answer with a plot and written interpretation.
13. Based on these two data sets, what claim are you prepared to make about John Arbuthnot's original question? What reservations, if any, do you have about using this data to make the claim?