

# test

*Nicholas Nagle*

*January 26, 2015*

## INSTRUCTIONS:

Edit this file by replacing the instructions with text and code to answer the questions. I should be able to knit this file and reproduce the html, pdf, or word document.

A block to load in the data:

```
rm(list=ls()) # Clean the workspace to be empty
arbuthnot <- read.csv('/Users/nnagle/Dropbox/git_root/homework/homework1/data/arbuthnot.csv')
```

## Question 1

This is a question about how the axes affects how we interpret plots.

Create two plots of Male-Female Christenings, one in which the y-axis scale is set by default, and one in which the y-axis extends all the way to zero. Yes, I know that was in the tutorial. I want to see it here.

- Describe the visual appearance of the two plots. Do they “feel” like they describe the same data?
- Describe how you might be able to mislead readers by changing the scaling on graphs.
- Which plot seems more appropriate here? Why?

## Question 2

This question is designed to give you some practice with ggplot as well as describing plots in words.

Use ggplot to create a scatterplot that has Female christenings on the x-axis and Male christenings on the y-axis. Draw a 45 degree line (using `geom_abline`) to show the line where Male and Female christenings are equal. Use this figure to describe the relationship and distribution of Male and Female christenings.

## Question 3

This is a question about population, sample, representativeness and generalizability.

How do you think the christenings-based sample would compared to a births-based sample? Similar? Different? Why? Arbuthnot’s data probably included most every christening in London during this period; they probably aren’t any unreported christenings. Is this fact important? Why or why not? Would a christenings based sample be appropriate now, in the 21st century? (Hint, this last question is trickier than it might seem. Think about what causes Male/Female Births, what causes people to christen their children, and any relations or not between these)

## Question 4

What does “sex ratio at birth” mean? How does it depend on biology, culture and technology. Consider the three cases of 1) late 17th century England, 2) 21st century US, and China under the [one-child policy](#).

## Question 5

The purpose of this question is to give you a little practice using standard deviation.

- a. The average height of young women ages 18-24 is 64.5 in. The distribution of heights is approximately normal (Gaussian) with a standard deviation of  $\sigma = 2.5$ . Complete this sentence: Approximately 95% of women have a height greater than \_\_\_\_\_ in and less than \_\_\_\_\_ in.

## Question 6

The purpose of this question is to help you understand the variance.

The formula for sample variance  $\sigma^2$  of a dataset is:

$$\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N - 1}$$

Describe in words what each part of this equation is, and using the concept of “distance” describe what the sample variance measures.