test

Nicholas Nagle

January 26, 2015

# INSTRUCTIONS:

Edit this file by replacing the instructions with text and code to anser 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:

library(ggplot2)  
rm(list=ls()) # Clean the workspace to be empty  
arbuthnot <- read.csv('/Users/sarahbleakney/Documents/geog415/sbleakne/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.

ggplot(data=arbuthnot) +   
 geom\_line(aes(x=Year, y=Males, color='Male')) +   
 geom\_line(aes(x=Year, y=Females, color='Females')) +   
 ylab('Christenings') + labs(color='Sex')



ggplot(data=arbuthnot) +   
 geom\_line(aes(x=Year, y=Males, color='Male')) +   
 geom\_line(aes(x=Year, y=Females, color='Females')) +   
 scale\_y\_continuous(limits=c(0, 10000)) +   
 ylab('Christenings') + labs(color='Sex')



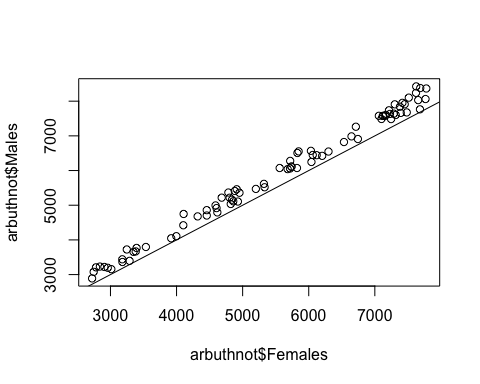
* Describe the visual appearance of the two plots. Do they "feel" like they describe the same data?  
   ---The plots are not that much different. The y-axis that begins at 0 is merely less stretched, but the trends are the same. The differences may look more dramatic since the scale is stretched.
* Describe how you might be able to mislead readers by changing the scaling on graphs. ---It is very simple to mislead readers. For instance, making the axis span a large range of values (0 to 100,000 as an extreme example) would make the two lines seem even more similar. It depends on the point you're wanting to make.
* Which plot seems more appropriate here? Why? ---In my opinion, the graph that extends down to 0 on the y-axis seems more reasonable and less skewed or stretched.

# 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.

plot(x=arbuthnot$Females, y=arbuthnot$Males) +abline(0,1)



## numeric(0)

For almost the majority of the data, there are more male christening than female christenings. This is not surprising, as there are more male births than female births. However, the ratio evens out since males have a higher infant mortality rate.

# 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) ---I would think the actual births and christenings-based birth count would be fairly similar in that time since christianity was so widespread. Yes, it is important that he collected most every christening in London during that time. If he only got data from certain churches in certain areas the results could be skewed. If the majority of data came from poorer areas it may result in a higher amount of boys. Since it was taken in all parts of the city, I would think it has a better chance of being fully representative of the entire population. If this sample were taken today, I would say it is not appropriate. Religion is often related to socio-economic status as well as culture, and while christened babies would in fact represent a sample of the population, it would likely be skewed due to these relations.

# Question 4

What does "sex ratio at birth" mean? How does it depend on biology, cuture and technology. Consider the three cases of 1) late 17th century England, 2) 21st century US, and China under the [one-child policy](http://en.wikipedia.org/wiki/One-child_policy).

The sex ratio at birth is the ratio of male births to female births, or 107:100. We discussed in class that, biologically, the male sperm swims faster and therefore the child is slightly more likely to be a boy. As for cultural factors, sex-selective abortion and infanticide are common in some cultures and will absolutely lead to a skewing of this reported ratio. Also, younger fathers lead to more male births, and societal trends influence the age of fathers, therefore affecting the sex ratio at birth. In the late 17th century England, cultural factors such as religion and age of childbirth affect family demographics and therefore sex ratios at birth. Also, the economy affects this ratio, and the ratios in the lower classes may have been affected. In the 21st century USA, the number of immigrants is apparently affecting the sex ratio at birth. In California, the large number of immigrants is a factor. Since race and ethic groups have different sex ratios at birth, the ratio is changing along with the influx of immigrants and their different cultural factors. In China under the one-child policy, only one child is allowed per family, and male babies are more valued than female babies. Because of this, families may perform an abortion if they discover they are having a girl and may try again for a boy.This is extremely likely to skew the sex ratio at birth, resulting in far more boys than girls.

# Question 5

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

1. The average height of young women ages 18-24 is 64.5 in. The distribution of heights is approximately normal (Gaussian) with a standardard deviation of . Complete this sentence: Approximately 95% of women have a height greater than **59.5** in and less than *69.5*\_\_ in.

# Question 6

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

The formula for sample variance of a dataset is:

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

The sample variance is found by taking a data point, subtracting the sample mean, and squaring the value of this difference. This is done for every data point, and these squared values are all added together. These are then divided by the number of samples, "N." It is technically divided by "N-1" but that is not too much different from N. The resulting variance will be in units squared, so data measured in meters will have a sample variance of meters squared. This number represrents the average distance squared between any data point and the mean.