Coursera Data Science Project: Statistical Inference (Part 2)

Rob Rolleston

August 23, 2015

Introduction

This is the project for the statistical inference class. In it, I will use simulation to explore inference and do some simple inferential data analysis. The project consists of two parts:

1. A simulation exercise.
2. Basic inferential data analysis (this report)

Basic Inferential Data Analysis

tbd

Load and explore data

```
library(datasets)
data("ToothGrowth")
ToothGrowth_tbl <- tbl_df(ToothGrowth)
glimpse(ToothGrowth_tbl)
```

```
## Observations: 60
## Variables:
## $ len  (dbl) 4.2, 11.5, 7.3, 5.8, 6.4, 10.0, 11.2, 11.2, 5.2, 7.0, 16....
## $ supp (fctr) VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, ...
## $ dose (dbl) 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 1.0, 1....
```

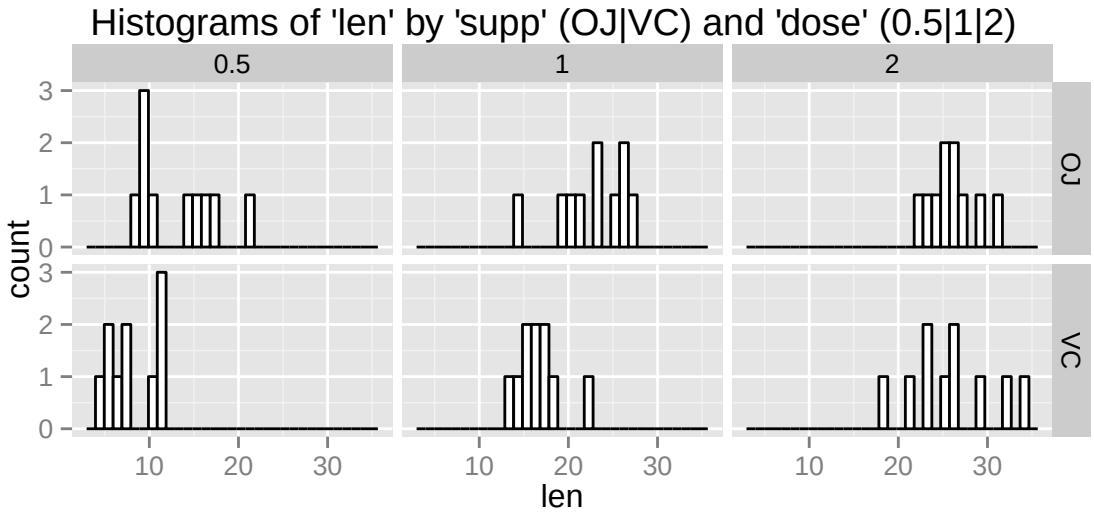
Some inspection of the ToothGrowth data indicates it has 60 observations of 3 values: len, supp, dose. The 'supp' value is a factor with only 2 levels: OJ, VC. The 'dose' value, although a number, actualy has only 3 values: 0.5, 1, 2. For processing, this value was converted to a factor.

```
ToothGrowth_tbl$dose <- as.factor(ToothGrowth_tbl$dose)
```

Data Summary

A faceted set of histograms

```
ggplot(ToothGrowth_tbl, aes(x=len)) +
  geom_histogram(color="black", fill="white") +
  facet_grid(supp ~ dose) +
  ggtitle("Histograms of 'len' by 'supp' (OJ|VC) and 'dose' (0.5|1|2) ")
```



The basic question is: **"Is ToothGrowth 'len' afected by 'supp' or 'dose'?"**
The mean, sd, and count of values of 'len' are:

```
summaryLen <- ToothGrowth_tbl %>% group_by(supp, dose) %>%
              summarize (mean = mean(len), sd=sd(len), n=n())
print(summaryLen)
```

```
## Source: local data frame [6 x 5]
## Groups: supp
##
##   supp dose  mean       sd  n
## 1   OJ  0.5 13.23 4.459709 10
## 2   OJ    1 22.70 3.910953 10
## 3   OJ    2 26.06 2.655058 10
## 4   VC  0.5  7.98 2.746634 10
## 5   VC    1 16.77 2.515309 10
## 6   VC    2 26.14 4.797731 10
```

Compare tooth growth by supp and dose

tbd

Conclusions

tbd