Statistical Inference Course - Project 2: Part 2

David Proksch

4/21/2017

Statistical Inference Course - Project 2, Part 2

Project Overview

The goal of this project is to load the Tooth Growth dataset and perform basic data exploration and analysis. The complete source code for the analysis can be found at https://github.com/dproksch/Stat-Inf-projet2

Load the required Data

```
library(ggplot2)
library(datasets)

data("ToothGrowth")
toothGrowth <- ToothGrowth
toothGrowth$dose <- as.factor(toothGrowth$dose)
```

Basic Summary of the Data

```
str(toothGrowth)
```

```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: Factor w/ 3 levels "0.5","1","2": 1 1 1 1 1 1 1 1 1 1 ...
```

```
summary(toothGrowth)
```

```
##       len        supp     dose
##  Min.   : 4.20   OJ:30   0.5:20
##  1st Qu.:13.07   VC:30   1  :20
##  Median :19.25           2  :20
##  Mean   :18.81
##  3rd Qu.:25.27
##  Max.   :33.90
```

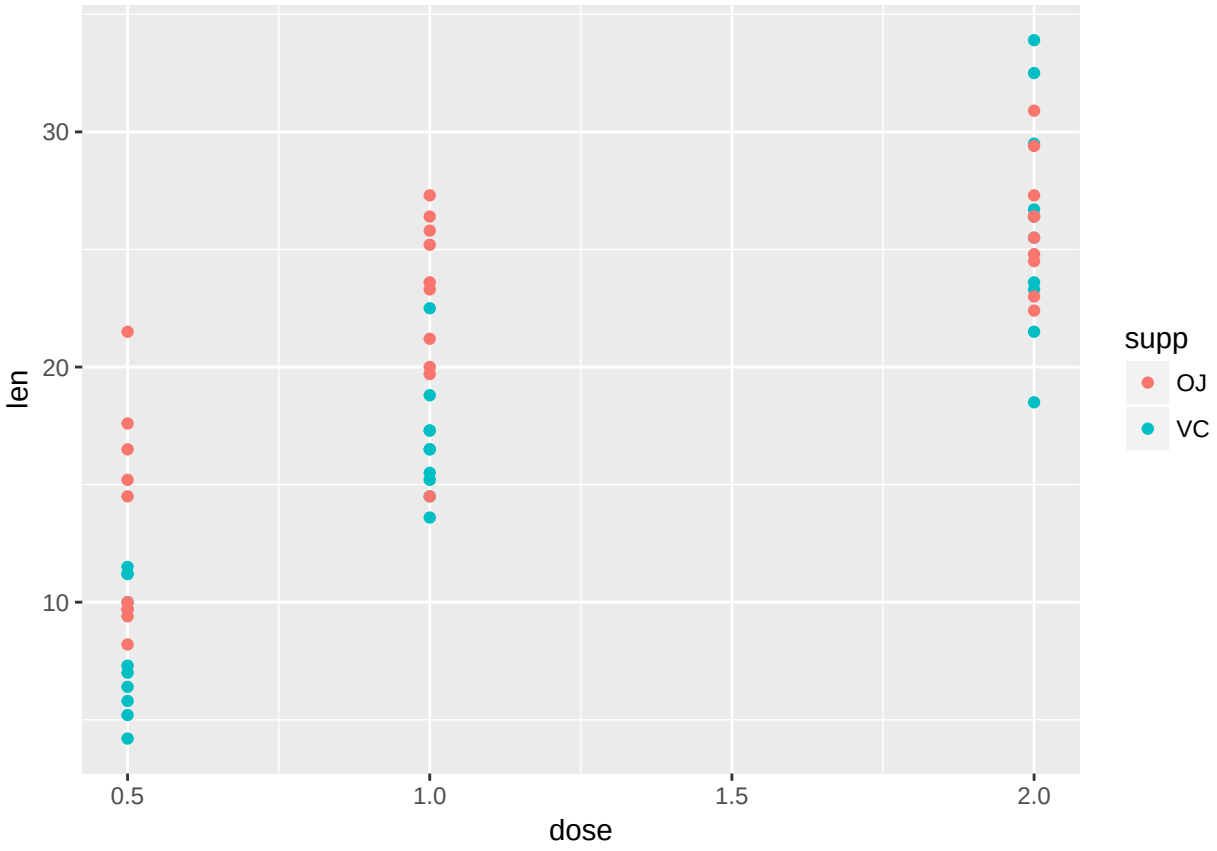
```
head(toothGrowth)
```

```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

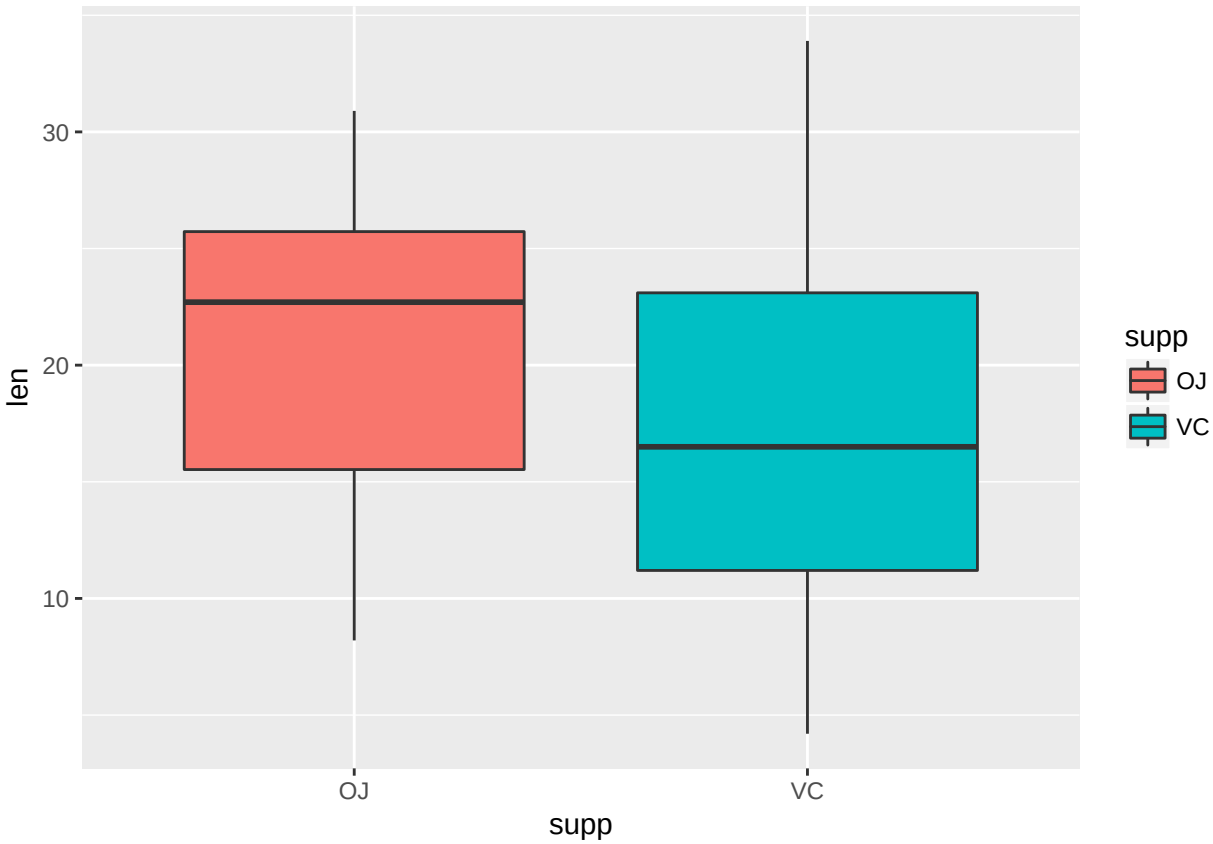
```
table(toothGrowth$supp, toothGrowth$dose)
```

```
##
##      0.5  1  2
##   OJ  10 10 10
##   VC  10 10 10
```

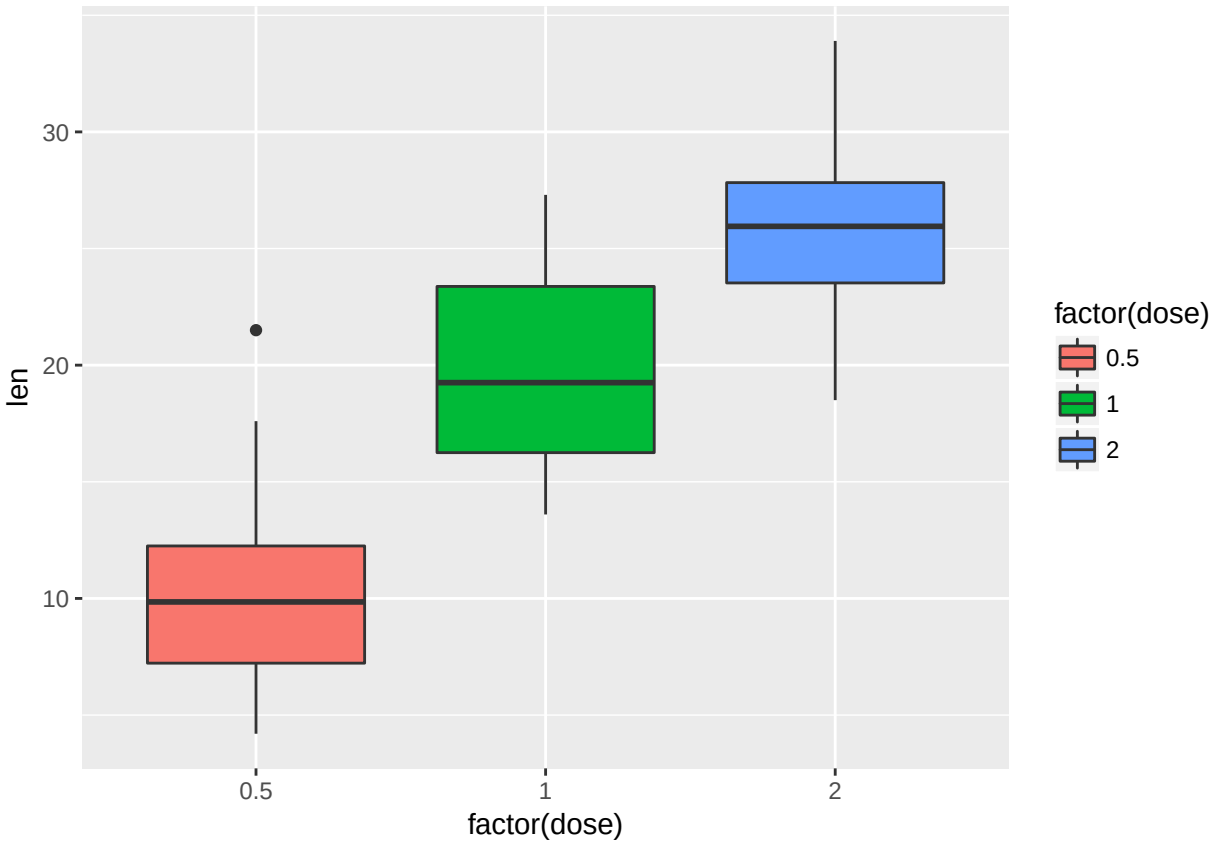
Compare tooth length and Vitamin C doses using a scatterplot -



Compare tooth length and delivery methods using a boxplot -



Compare Vitamin C dose levels and tooth length with a boxplot -



Initial data analysis suggests that the dosage appears to have a significant impact on the length of teeth. The impact of the supplement is not as obvious. The length of the teeth appears to be maximized when Vitamin C is given as a supplement.

Hypothesis Testing and Confidence Intervals

Impact of Supplements

Test for a correlation between delivery method and tooth growth. The null hypothesis is the there is no correlation between delivery method and tooth length.

```
t.test(len ~ supp, paired = F, var.equal = F, data = ToothGrowth)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1710156  7.5710156
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

A confidence interval of [-0.171, 7.571] does not allow us to reject the null hypothesis.

Impact of Dosage

Test for a correlation between the dose level and tooth growth. The null hypothesis is there is no correlation between dose level and tooth growth.

```
dose1 <- subset(ToothGrowth, dose %in% c(0.5, 1.0))
dose2 <- subset(ToothGrowth, dose %in% c(0.5, 2.0))
dose3 <- subset(ToothGrowth, dose %in% c(1.0, 2.0))
t.test(len ~ dose, paired = F, var.equal = F, data = dose1)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.983781  -6.276219
## sample estimates:
## mean in group 0.5   mean in group 1
##            10.605            19.735
```

```
t.test(len ~ dose, paired = F, var.equal = F, data = dose2)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.15617 -12.83383
## sample estimates:
## mean in group 0.5   mean in group 2
##            10.605            26.100
```

```
t.test(len ~ dose, paired = F, var.equal = F, data = dose3)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -4.9005, df = 37.101, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.996481 -3.733519
## sample estimates:
## mean in group 1 mean in group 2
##          19.735          26.100
```

The confidence intervals ([-11.98, -6.276] for doses 0.5 and 1.0, [-18.16, -12.83] for doses 0.5 and 2.0, and [-8.996, -3.734] for doses 1.0 and 2.0) allow for the rejection of the null hypothesis. This suggests there is a correlation between dose levels and tooth length.

Impact of the Supplement with Dosage Levels

Test for a correlation between the dosage levels and tooth growth with each dosage level.

```
Tooth.dose5 <- subset(ToothGrowth, dose == 0.5)
Tooth.dose1 <- subset(ToothGrowth, dose == 1.0)
Tooth.dose2 <- subset(ToothGrowth, dose == 2.0)
t.test(len ~ supp, paired = F, var.equal = F, data = Tooth.dose5)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
## mean in group OJ mean in group VC
##            13.23             7.98
```

```
t.test(len ~ supp, paired = F, var.equal = F, data = Tooth.dose1)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
## mean in group OJ mean in group VC
##            22.70            16.77
```

```
t.test(len ~ supp, paired = F, var.equal = F, data = Tooth.dose2)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = -0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean in group OJ mean in group VC
##            26.06            26.14
```

Dose levels 0.5mg and 1.0mg([1.72, 8,78] within 0.5mg, [2.80, 9.06] within 1.0mg) allow for the rejection of the null hypothesis and a confirmation that there is correlation between doseage levels and tooth length. The confidence interval for dose level 2.0[-3.80, 3.64] is not enough to reject the null hypothesis.

Conclusions and Assumptions

It must assumed that:

* these poplulations are independent,
* the variances between populations are different,
* a random population was used,
* the population was comprised of similar guinea pigs,
* measurement error was accounted for with significant digits,
* double blind research methods were used.

For the populations to be independent, 60 guinea pigs would have to be used so each combination of dose level and delivery method were not affected by the other methods.

Working under the assumption that the above conditions are true, it is possible to infer that there is a significant difference between tooth length and dose levels across both delivery methods. A higher dose level consistently led to longer teeth. Initially it appeared that the delivery method had no significant impact on tooth length, but when controlling for dose level we discovered that there was a significant difference at 0.5mg and 1.0mg, but not at 2.0mg. Based on this evidence, it appears that orange juice is a better delivery method with a larger impact on tooth length for a given dose of Vitamin C, but above a maximum dose level there is no further improvement.