Statistical Inference Course Assignment Part 2

Carolinecoder

December 22, 2015

Load the ToothGrowth data and perform some basic exploratory data analyses. Provide a basic summary of the data.

Using the ?ToothGrowth function, we learn that the ToothGrowth data documents the effect of Vitamin C on tooth growth in guinea pigs. The response is the length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods (orange juice or ascorbic acid).

```
# Load the data, print the first few rows, and get information on structure of the dataframe and summary stats
library(datasets)
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
```

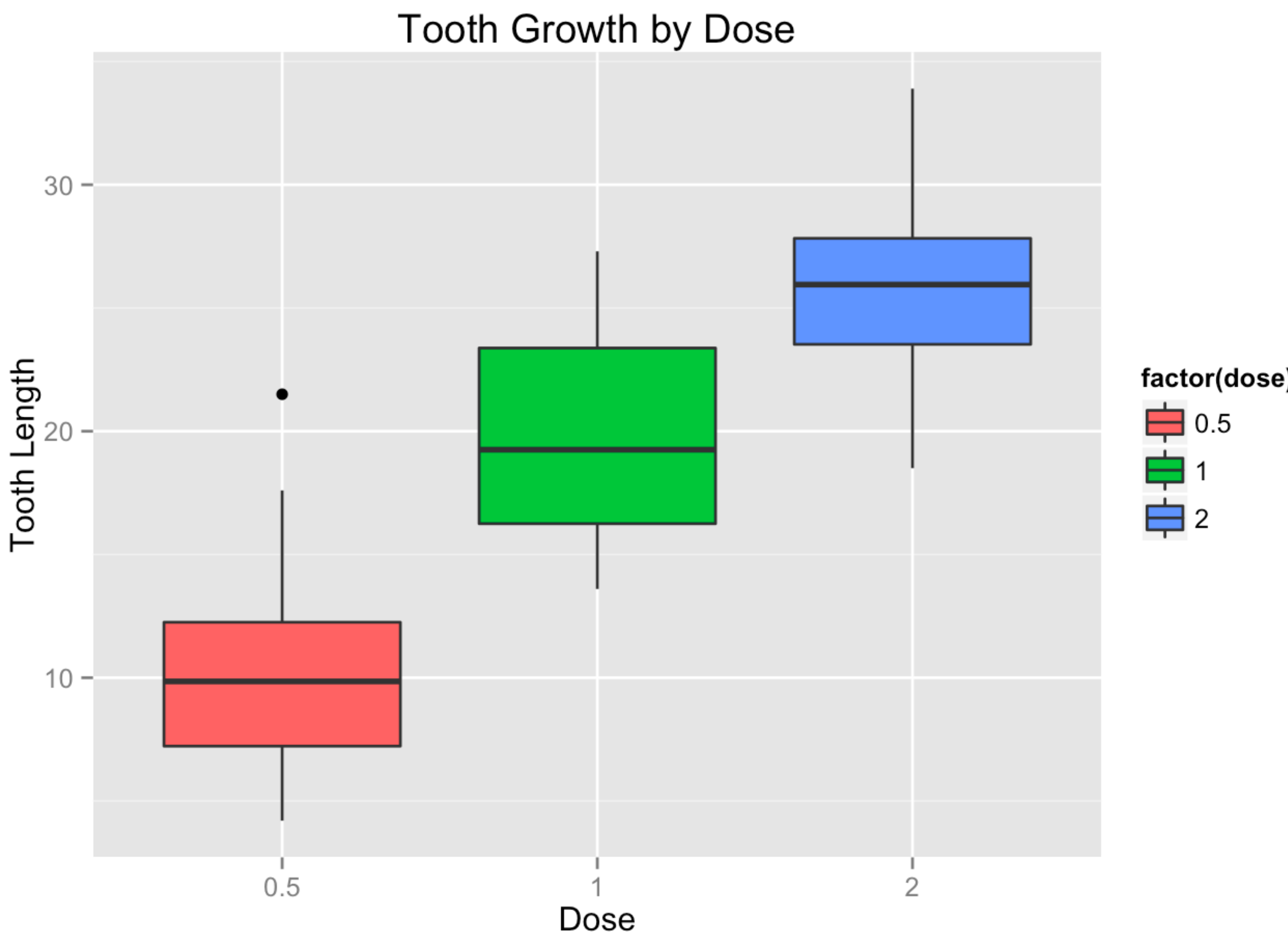
```
str(ToothGrowth)
```

```
## '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: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

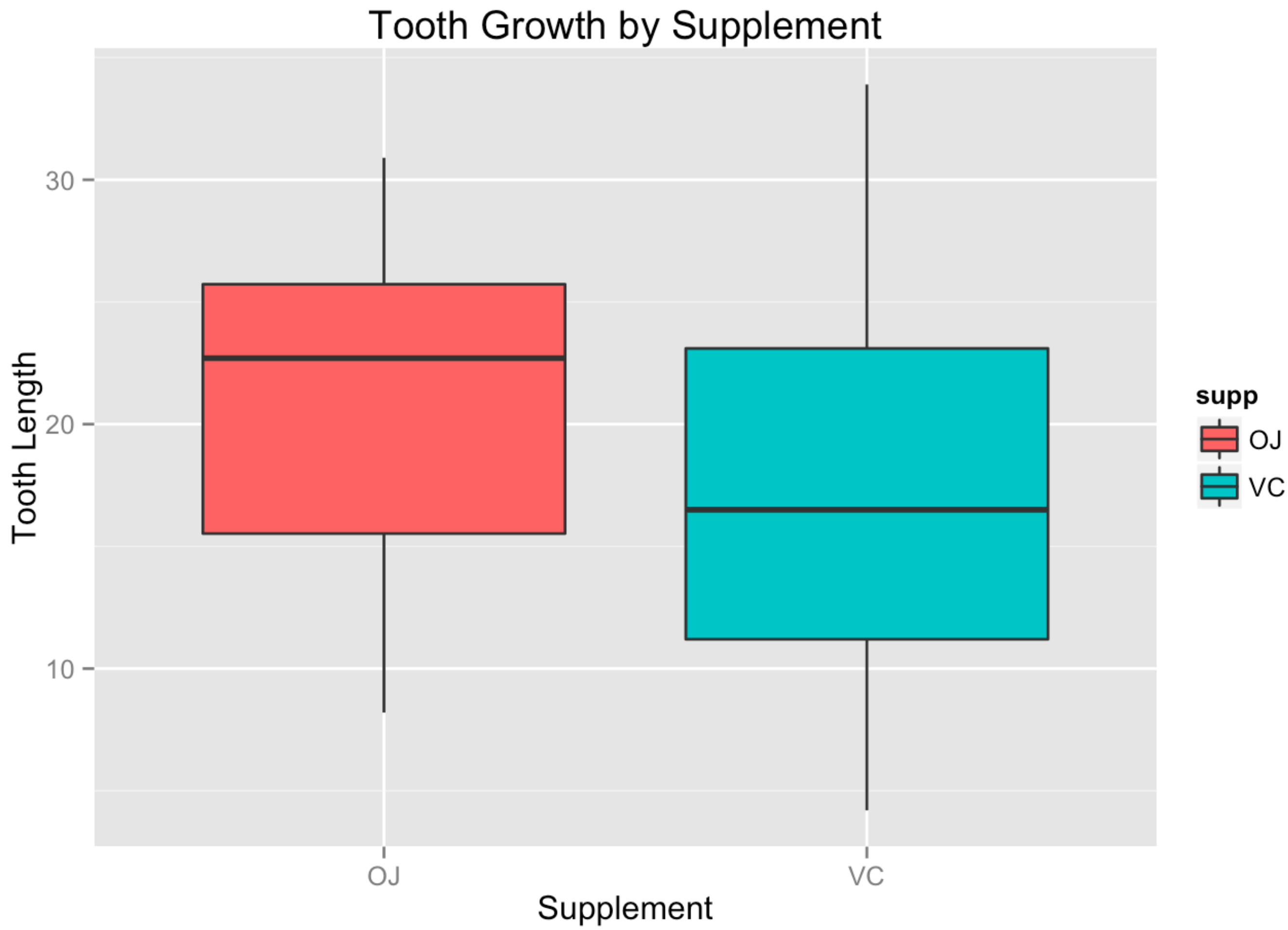
```
summary(ToothGrowth)
```

```
##       len        supp         dose      
##  Min.   : 4.20   OJ:30   Min.   :0.500  
##  1st Qu.:13.07   VC:30   1st Qu.:0.500  
##  Median :19.25           Median :1.000  
##  Mean   :18.81           Mean   :1.167  
##  3rd Qu.:25.27           3rd Qu.:2.000  
##  Max.   :33.90           Max.   :2.000
```

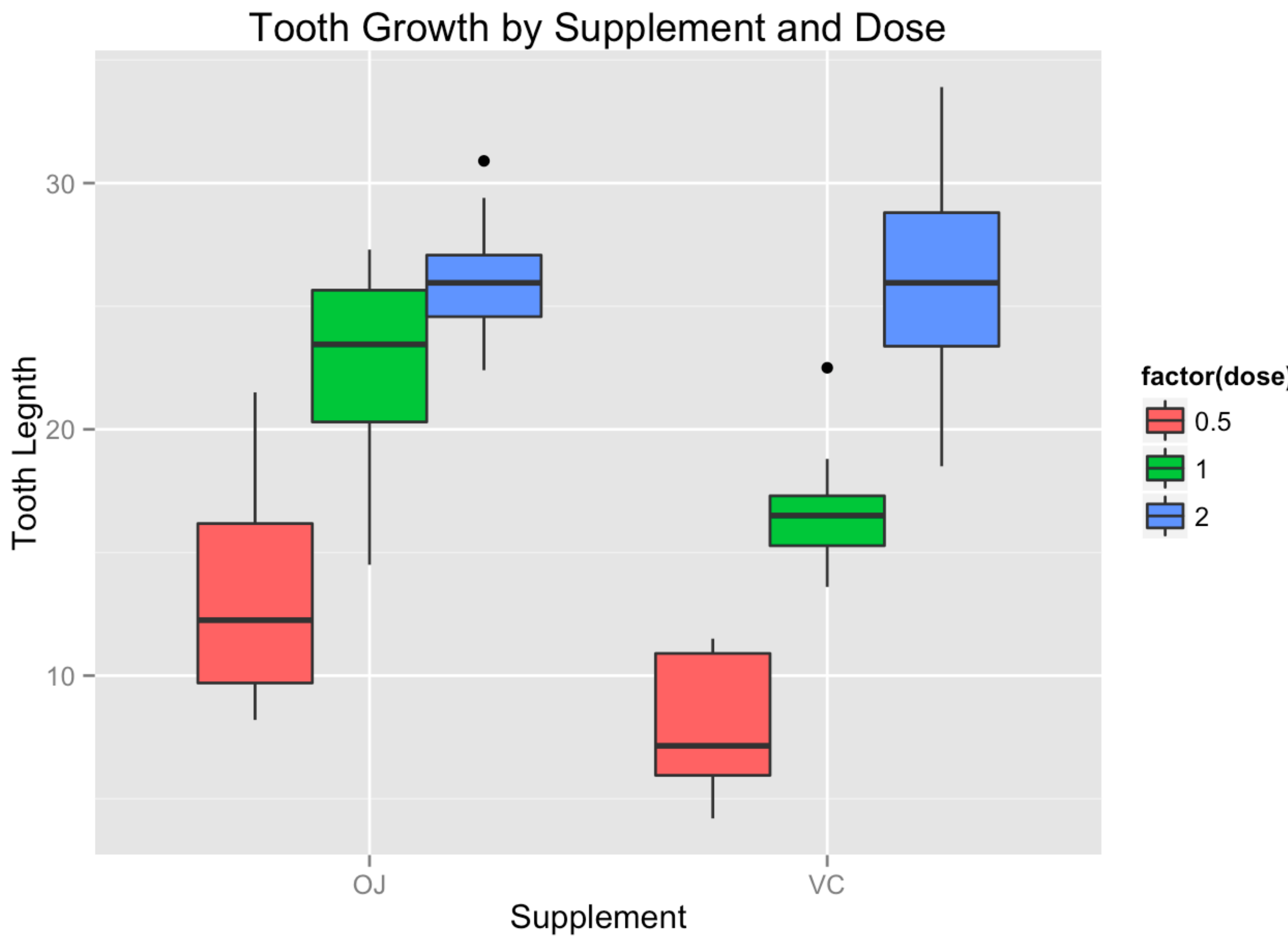
```
# Exploratory graphs by dose
library(ggplot2)
p <- ggplot(ToothGrowth, aes(factor(dose), len)) + geom_boxplot(aes(fill = facto
r(dose))) + ggtitle("Tooth Growth by Dose") + xlab("Dose") + ylab("Tooth Length")
p
```




```
# By supplement
q <- ggplot(ToothGrowth, aes(supp, len)) + geom_boxplot(aes(fill = supp)) + ggtitl
e("Tooth Growth by Supplement") + xlab("Supplement") + ylab("Tooth Length")
q
```




```
# Both at the same time
r <-  ggplot(ToothGrowth, aes(supp, len)) + geom_boxplot(aes(fill = factor(dose)))
+ ggtitle("Tooth Growth by Supplement and Dose") + xlab("Supplement") + ylab("Toot
h Legnth")
r
```



We can hypothesize, based on these exploratory graphs, that OJ may promote faster tooth growth than Vitamin C at dose levels of 0.5 and 1, though probably not at dose 2; and that higher dosage of either substance within the OJ and Vitamin C groups may also increase tooth length.

Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there's other approaches worth considering)

Supplement analysis

Perform t-tests on OJ versus Vitamin C to determine whether there the supplement really does have an effect on tooth length, or whether the higher tooth length average in the OJ group is simply a result of chance (variability in the two different samples). We are not comparing the same subjects to each other, so we will perform an independent t-test. Also, the variance within the OJ and VC groups is not the same, so we will assume unequal variance.

```
OJ_len <- ToothGrowth$len[ToothGrowth$supp=="OJ"]
var(OJ_len)
```

```
## [1] 43.63344
```

```
VC_len <- ToothGrowth$len[ToothGrowth$supp=="VC"]
var(VC_len)
```

```
## [1] 68.32723
```

```
t.test(OJ_len, VC_len, paired=FALSE, var.equal=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  OJ_len and VC_len
## 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 of x mean of y
##  20.66333  16.96333
```

Our t-test contains 0 and has a p value > 0.05, meaning that there is not a true difference between the OJ and VC means. Therefore, the difference in tooth length between the OJ and VC groups could simply be a result of chance, and we cannot draw further conclusions on the effect of OJ or VC on guinea pig tooth length.

Dosage levels analysis

Perform t-tests on 0.5 and 1 dosage levels (we can only test two samples at a time) to determine whether there the dosage level really does have an effect on tooth length, or whether the higher tooth length average in the higher dosage group is simply a result of chance (variability in the two different samples). We are not comparing the same subjects to each other, so we will perform an independent t-test. Also, while the variance within the two dose levels is similar, we will continue to assume unequal variance.

```
low_len <- ToothGrowth$len[ToothGrowth$dose==0.5]
var(low_len)
```

```
## [1] 20.24787
```

```
med_len <- ToothGrowth$len[ToothGrowth$dose==1]
var(med_len)
```

```
## [1] 19.49608
```

```
t.test(low_len, med_len, paired=FALSE, var.equal=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  low_len and med_len
## 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 of x mean of y
##    10.605    19.735
```

Our t-test does not contain 0 and has a p value < 0.05, meaning that there is a true difference between the 0.5 and 1 dosage level means. Therefore, the difference in tooth length between the 0.5 and 1 dose level is not a result of chance, and we can further explore the effect of dose on tooth length.

Perform t-tests on 1 and 2 dosage levels (we can only test two samples at a time) to determine whether there the dosage level really does have an effect on tooth length, or whether the higher tooth length average in the higher dosage group is simply a result of chance (variability in the two different samples). We are not comparing the same subjects to each other, so we will perform an independent t-test. Also, the variance within the two dose levels is quite different, so we will assume unequal variance, since this is the best approach to take when there is some doubt about population variance.

```
med_len <- ToothGrowth$len[ToothGrowth$dose==1]
var(med_len)
```

```
## [1] 19.49608
```

```
high_len <- ToothGrowth$len[ToothGrowth$dose==2]
var(high_len)
```

```
## [1] 14.24421
```

```
t.test(med_len, high_len, paired=FALSE, var.equal=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  med_len and high_len
## 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 of x mean of y
##    19.735    26.100
```

Our t-test does not contain 0 and has a p value < 0.05, meaning that there is a true difference between the 1 and 2 dosage level means. Therefore, the difference in tooth length between the 1 and 2 dose levels is not a result of chance, and we can further explore the effect of dose on tooth length.

Controlling for supplement in dosage levels analysis

Our exploratory graph that tracked dosage levels within OJ and Vitamin C suggests that increasing dosage increases tooth length across both supplement groups. This should be true, given the fact that the t-test on the supplement revealed that OJ vs Vitamin C had no effect on tooth growth. However, let us make sure this is true by conducting t-test for the 0.5 to 1, and 1 to 2 dosage increase, first for OJ and then for Vitamin C.

```
# 0.5 to 1, OJ
low_len_OJ <- ToothGrowth$len[ToothGrowth$dose==0.5 & ToothGrowth$supp=="OJ"]
med_len_OJ <- ToothGrowth$len[ToothGrowth$dose==1 & ToothGrowth$supp=="OJ"]
t.test(low_len_OJ, med_len_OJ, paired=FALSE, var.equal=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  low_len_OJ and med_len_OJ
## t = -5.0486, df = 17.698, p-value = 8.785e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -13.415634  -5.524366
## sample estimates:
## mean of x mean of y
##     13.23     22.70
```

```
# 1 to 2, OJ
med_len_OJ <- ToothGrowth$len[ToothGrowth$dose==1 & ToothGrowth$supp=="OJ"]
high_len_OJ <- ToothGrowth$len[ToothGrowth$dose==2 & ToothGrowth$supp=="OJ"]
t.test(med_len_OJ, high_len_OJ, paired=FALSE, var.equal=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  med_len_OJ and high_len_OJ
## t = -2.2478, df = 15.842, p-value = 0.0392
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -6.5314425 -0.1885575
## sample estimates:
## mean of x mean of y
##     22.70     26.06
```

```
# 0.5 to 1, VC
low_len_VC <- ToothGrowth$len[ToothGrowth$dose==0.5 & ToothGrowth$supp=="VC"]
med_len_VC <- ToothGrowth$len[ToothGrowth$dose==1 & ToothGrowth$supp=="VC"]
t.test(low_len_VC, med_len_VC, paired=FALSE, var.equal=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  low_len_VC and med_len_VC
## t = -7.4634, df = 17.862, p-value = 6.811e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.265712  -6.314288
## sample estimates:
## mean of x mean of y
##      7.98     16.77
```

```
# 1 to 2, VC
med_len_VC <- ToothGrowth$len[ToothGrowth$dose==1 & ToothGrowth$supp=="VC"]
high_len_VC <- ToothGrowth$len[ToothGrowth$dose==2 & ToothGrowth$supp=="VC"]
t.test(med_len_VC, high_len_VC, paired=FALSE, var.equal=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  med_len_VC and high_len_VC
## t = -5.4698, df = 13.6, p-value = 9.156e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -13.054267  -5.685733
## sample estimates:
## mean of x mean of y
##     16.77     26.14
```

State your conclusions and the assumptions needed for your conclusions.

Each one of the t-tests reveals that the increasing dosage may increase tooth length for all subjects, whether they are fed OJ or Vitamin C, but that the supplement does not have a measurable effect on tooth length for these guinea pigs. Assumptions include unequal variance for each t.test, that the guinea pigs tested are a random sample and do not have characteristics specific to them that could skew our results (for example, a special gene for fast tooth growth that is triggered by certain dosage levels), and that the dosage is not correlated with another, unknown variable that has a more robust effect on tooth length than dosage (for example, maybe higher dosage of a supplement triggers higher appetite, and it is the increase in food, rather than supplement dosage, that triggers tooth growth).