Statistic Inference Project Part 2

April 7, 2018

Part 2: Basic Inferential Data Analysis Instructions

Now in the second portion of the project, we're going to analyze the ToothGrowth data in the R datasets package.
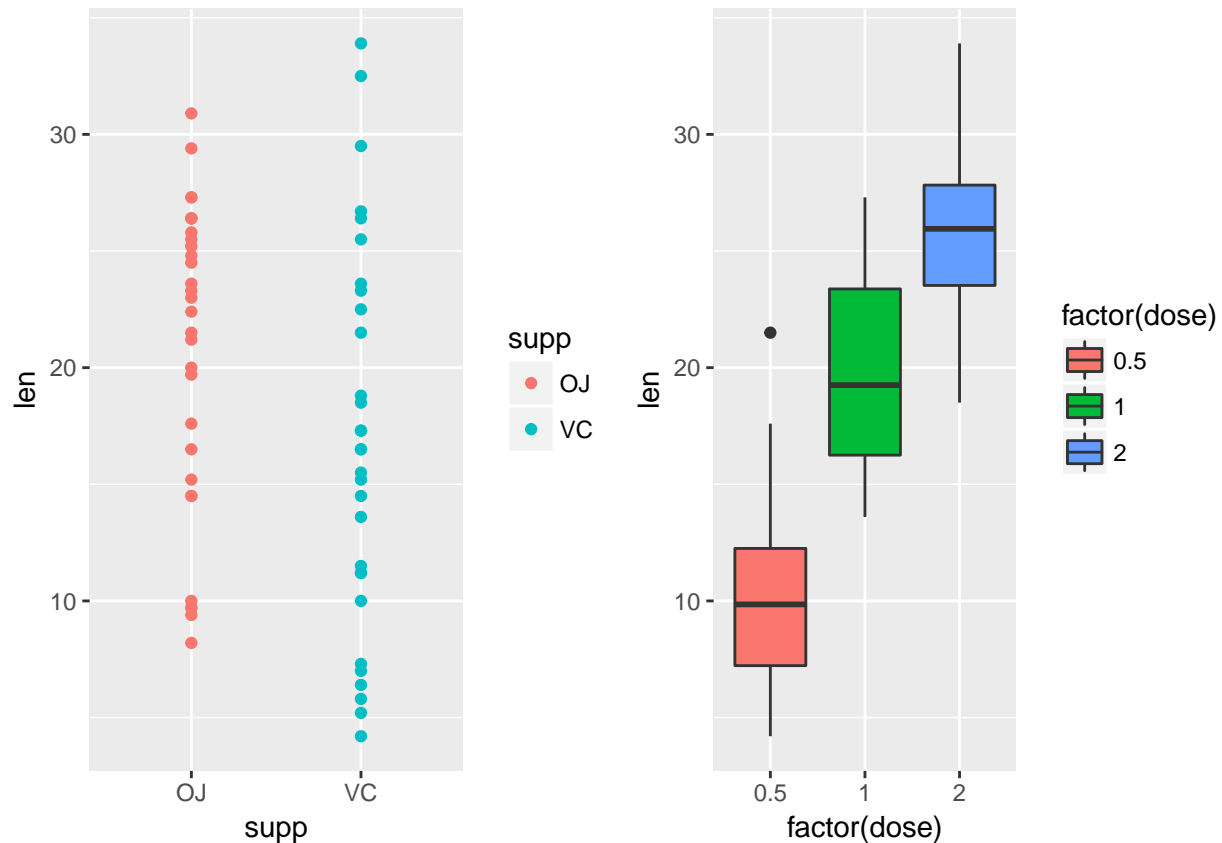
1. Load the ToothGrowth data

```
data("ToothGrowth")
data <- ToothGrowth
```

2. Provide a basic summary of the data

```
a<-ggplot( aes(x=supp, y=len), data=data) + geom_point(aes(color=supp))
b <- ggplot( aes(x=factor(dose), y=len), data=data) + geom_boxplot(aes(fill=factor(dose)))

grid.arrange(a, b, ncol=2, heights = 2)
```



```
summary(data)
```

```
##       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
```

3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. ####3.1 Compare tooth growth by supp

H0: Toothgrowth is the same from VC supp and OJ supp (the 2 delivery methods: ascorbic acid and orange juice)
H1: Toothgrowth is different

```
t.test(data=data, len~supp, paired=FALSE, var.equal=TRUE)
```

```
##
##   Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 58, p-value = 0.06039
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1670064  7.5670064
## sample estimates:
## mean in group OJ mean in group VC
##          20.66333         16.96333
```

The P value = 0.06, which is > the signifance level, 0.05
In addition, hypothesis value 0 is inside the 95% confidence interval.
Thus,we do not reject the null hypothesis. ####3b. Compare the toothgrowth by dose H0: Toothgrowth varies by dose
HA: The different in dose has no impact to Toothgrowth

```
dose05 <- data[data$dose==0.5,]
dose1 <- data[data$dose==1,]
dose2 <- data[data$dose==2,]

t.test(dose05$len,dose1$len, paired=FALSE, alternative = "two.sided")
```

```
##
##   Welch Two Sample t-test
##
## data:  dose05$len and dose1$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
```

```
t.test(dose1$len,dose2$len, paired=FALSE, alternative = "two.sided")
```

```
##
##   Welch Two Sample t-test
##
```

```
## data:  dose1$len and dose2$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
```

```r
t.test(dose05$len,dose2$len, paired=FALSE, alternative = "two.sided")
```

```
##
##  Welch Two Sample t-test
##
## data:  dose05$len and dose2$len
## 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 of x mean of y
##    10.605    26.100
```

Compare the dose samples, all p-values are close to 0.
None of the confidence interval has 0 in it.
Thus we reject H0.

4. Conclusion:

Different ways to supply the vitamin C has no impact to tooth growth.
It is possible that the difference in dose has an impact to tooth group.