Statistical Inference Course Project Part 2

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

```
library(psych)
library(ggplot2)
```

1. Load the ToothGrowth data and perform some basic exploratory data analyses.

```
##
## Attaching package: 'ggplot2'
##
## The following object is masked from 'package:psych':
##
##     %+%
```

```
library(gridExtra)
```

```
## Loading required package: grid
```

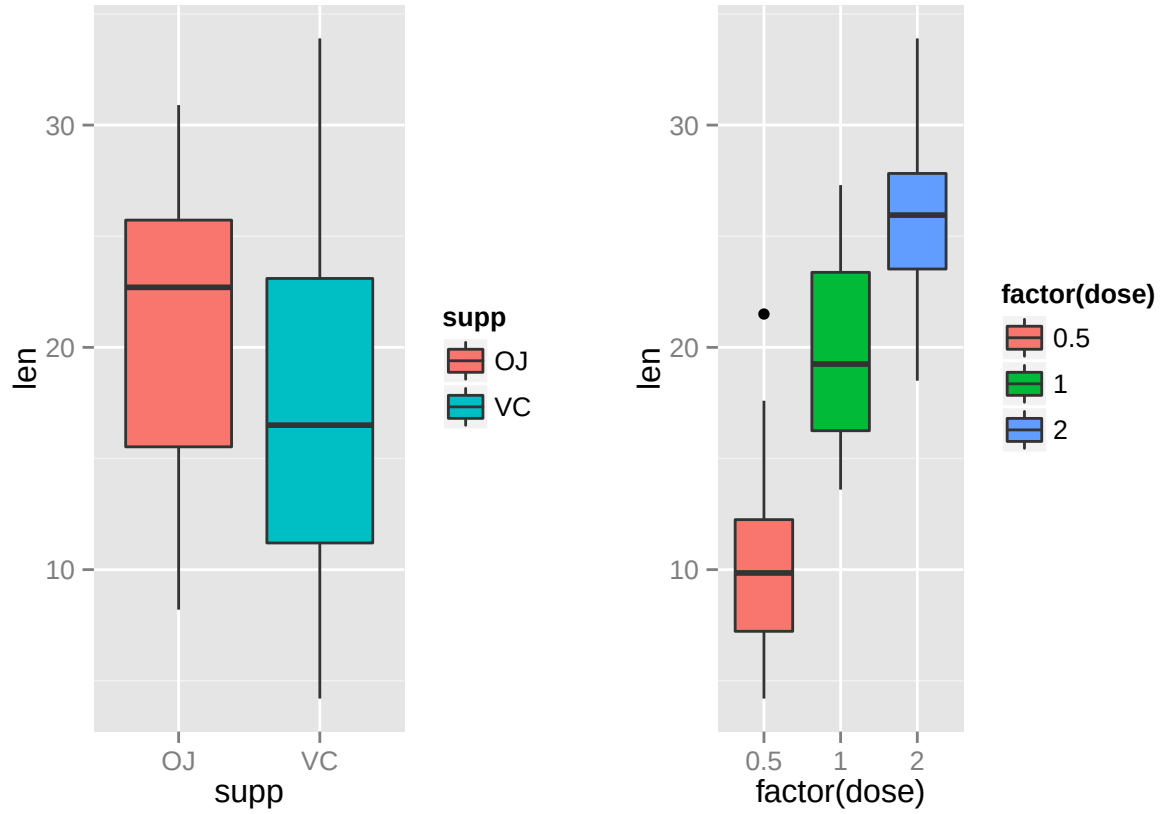
```
library(datasets)
data(ToothGrowth)
attach(ToothGrowth)
```

```
## The following object is masked _by_ .GlobalEnv:
##
##     dose
```

```
suppplot<-ggplot(aes(x = supp, y = len), data = ToothGrowth) + geom_boxplot(aes(fill = supp))

doseplot<-ggplot(aes(x = factor(dose), y =len), data = ToothGrowth) +
        geom_boxplot(aes(fill = factor(dose)))

grid.arrange(suppplot, doseplot, ncol = 2)
```




```
head(ToothGrowth)
```

2. Provide a basic summary of the data.

```
##    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
```

```
summary(ToothGrowth)
```

```
##       len        supp         dose
##  Min.   : 4.2   OJ:30   Min.   :0.50
##  1st Qu.:13.1   VC:30   1st Qu.:0.50
##  Median :19.2           Median :1.00
##  Mean   :18.8           Mean   :1.17
##  3rd Qu.:25.3           3rd Qu.:2.00
##  Max.   :33.9           Max.   :2.00
```

```
dose<-as.factor(dose)
describe(len)
```

```
##   vars  n  mean   sd median trimmed  mad min  max range  skew kurtosis
## 1    1 60 18.81 7.65  19.25   18.95 9.04 4.2 33.9  29.7 -0.14    -1.04
##     se
## 1 0.99
```

```
table(supp,dose)
```

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

```
round(with(ToothGrowth, sapply(split(len, supp), mean)), 3)
```

```
##    OJ    VC
## 20.66 16.96
```

```
aggregate(len, list(dose), mean)
```

```
##   Group.1     x
## 1     0.5 10.61
## 2       1 19.73
## 3       2 26.10
```

```
aggregate(len, list(supp, dose), mean)
```

```
##   Group.1 Group.2     x
## 1      OJ     0.5 13.23
## 2      VC     0.5  7.98
## 3      OJ       1 22.70
## 4      VC       1 16.77
## 5      OJ       2 26.06
## 6      VC       2 26.14
```

```
aggregate(len, list(supp, dose), sd)
```

```
##   Group.1 Group.2     x
## 1      OJ     0.5 4.460
## 2      VC     0.5 2.747
## 3      OJ       1 3.911
## 4      VC       1 2.515
## 5      OJ       2 2.655
## 6      VC       2 4.798
```

The ToothGrowth dataset explains the relation between the growth of teeth of 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 and ascorbic acid).

```r
# T Test by supplemant type
t.test(len ~ supp, data = ToothGrowth)
```

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

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 1.915, df = 55.31, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.171  7.571
## sample estimates:
## mean in group OJ mean in group VC
##            20.66            16.96
```

```r
# T test by dose level
Tooth.dose0.5_1.0 <- subset(ToothGrowth, dose %in% c(0.5, 1.0))
Tooth.dose0.5_2.0 <- subset(ToothGrowth, dose %in% c(0.5, 2.0))
Tooth.dose1.0_2.0 <- subset(ToothGrowth, dose %in% c(1.0, 2.0))

t.test(len ~ dose, data = Tooth.dose0.5_1.0)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -6.477, df = 37.99, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.984  -6.276
## sample estimates:
## mean in group 0.5   mean in group 1
##             10.61             19.73
```

```r
t.test(len ~ dose, data = Tooth.dose0.5_2.0)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -11.8, df = 36.88, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.16 -12.83
## sample estimates:
## mean in group 0.5   mean in group 2
##             10.61             26.10
```

```r
t.test(len ~ dose, data = Tooth.dose1.0_2.0)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -4.901, df = 37.1, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.996 -3.734
## sample estimates:
## mean in group 1 mean in group 2
##           19.73           26.10
```

```r
# T test for supplement by dose level
Tooth.dose0.5 <- subset(ToothGrowth, dose == 0.5)
Tooth.dose1.0 <- subset(ToothGrowth, dose == 1.0)
Tooth.dose2.0 <- subset(ToothGrowth, dose == 2.0)

t.test(len ~ supp, data = Tooth.dose0.5)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 3.17, df = 14.97, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.719 8.781
## sample estimates:
## mean in group OJ mean in group VC
##            13.23             7.98
```

```r
t.test(len ~ supp, data = Tooth.dose1.0)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 4.033, df = 15.36, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802 9.058
## sample estimates:
## mean in group OJ mean in group VC
##            22.70            16.77
```

```r
t.test(len ~ supp, data = Tooth.dose2.0)
```

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

```
##
## data:  len by supp
## t = -0.0461, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.798  3.638
## sample estimates:
## mean in group OJ mean in group VC
##            26.06            26.14
```

For dose 0.5, the p-value of OJ in comparison to VC is 0.0064. Since it is less than 0.05 (strong presumption against null hypothesis), it means that there is a difference between both methods.

For dose 1.0, the p-value of OJ in comparison to VC is 0.001. Since it is less than 0.05 (strong presumption against null hypothesis), it means that there is a difference between both methods.

For dose 2.0, the p-value of OJ in comparison to VC is 0.064. Since it is greater than 0.05 (low presumption against null hypothesis), it means that there is a no that much of a difference between both methods.

4. State your conclusions and the assumptions needed for your conclusions. With the values obtained it can be assumed that there is a different in the growth of the tooth while the doses are larger. By looking at the boxplot and the results, it can also be said that there is no other factor that will affect the growing process (it will depend greatly on the dose), in other words, the delivery methods are independent of the dose size.