Statistical Inference

Basic inferential data analysis

We're going to analyze the ToothGrowth data in the R datasets package.

1. Load the ToothGrowth data and perform some basic exploratory data analyses
2. Provide a basic summary of the data.
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)
4. State your conclusions and the assumptions needed for your conclusions.

We load the data and perform some basic exploratory data analysis, plus the summary of th data:

```
library(datasets)

class(ToothGrowth)
```

```
## [1] "data.frame"
```

```
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  
```

```
dim(ToothGrowth)
```

```
## [1] 60  3
```

```
names <- names(ToothGrowth)
levels(ToothGrowth)
```

```
## NULL
```

```
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 ...
```

So, we see it is a data frame which has two dimensions: 60, 3, with no levels. The names are len, supp, dose.

Next, we go to point three where we plot what we have in Appendix as code chunk 1 with figure 1. From it we can see that the tooth increases as the dosage increases. We create a regression analysis to test in deeper the data. This model can be found in the Appendix in code chunk 2 with results. Afterwards we apply the confidence intervals function and we have it in the Appendix in code chunk 3 with results. Next, in the code chunk 4 (see Appendix), we apply the t test functions, so we get the `p` values.

From these analyses in point 3, we can go to conclusions (point 4):

1. As stated earlier from the box plots we see that: the tooth increases as the dosage increases.
2. The p values calculated earlier show that each variable proves the variability in tooth length.

Appendix

Code chunk 1 with figure 1

```
library(ggplot2)
comparePlot <- ggplot(ToothGrowth, aes(factor(dose), len, fill = factor(dose)))
comparePlot +
      geom_boxplot() +
      facet_grid(. ~ supp) +
      xlab('Factor by dose') +
      ylab('Length of tooth')
```

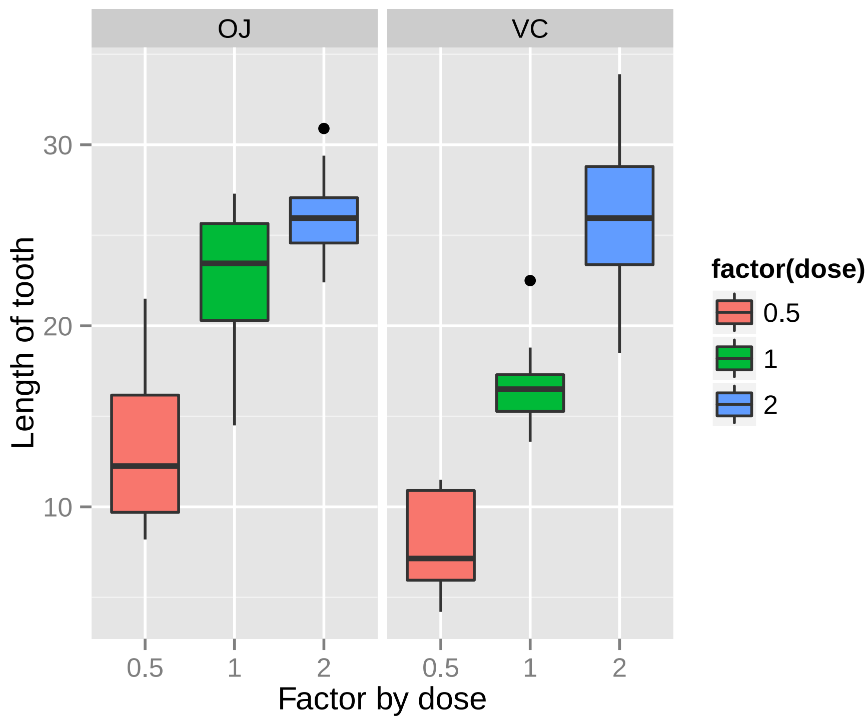

Figure 1: plot of chunk 1

Code chunk 2 with results:

```
myModel <- lm(len ~ dose + supp, ToothGrowth)
summary(myModel)
```

```
##
## Call:
## lm(formula = len ~ dose + supp, data = ToothGrowth)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -6.600 -3.700  0.373  2.116  8.800
##
```

```
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    9.273      1.282    7.23  1.3e-09 ***
## dose           9.764      0.877   11.14  6.3e-16 ***
## suppVC        -3.700      1.094   -3.38   0.0013 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.24 on 57 degrees of freedom
## Multiple R-squared:  0.704,  Adjusted R-squared:  0.693
## F-statistic: 67.7 on 2 and 57 DF,  p-value: 8.72e-16
```

Code chunk 3 with results:

```
confint(myModel)
```

```
##              2.5 % 97.5 %
## (Intercept)  6.705  11.84
## dose         8.008  11.52
## suppVC      -5.890  -1.51
```

Code chunk 4 with results:

```
firstDose <- ToothGrowth[ToothGrowth$dose == 0.5,]
testFirstDose <- t.test(len ~ supp, data = firstDose)
testFirstDose$p.value;
```

```
## [1] 0.006359
```

```
secondDose <- ToothGrowth[ToothGrowth$dose == 1.0,]
testSecondDose <- t.test(len ~ supp, data = secondDose)
testSecondDose$p.value;
```

```
## [1] 0.001038
```

```
thirdDose <- ToothGrowth[ToothGrowth$dose == 2.0,]
testThirdDose <- t.test(len ~ supp, data = thirdDose)
testThirdDose$p.value;
```

```
## [1] 0.9639
```