Coursera Data Scientist
Statistical Inference: Week 3 Assignment Part 2
Author: Michelle Gitgit

Overview

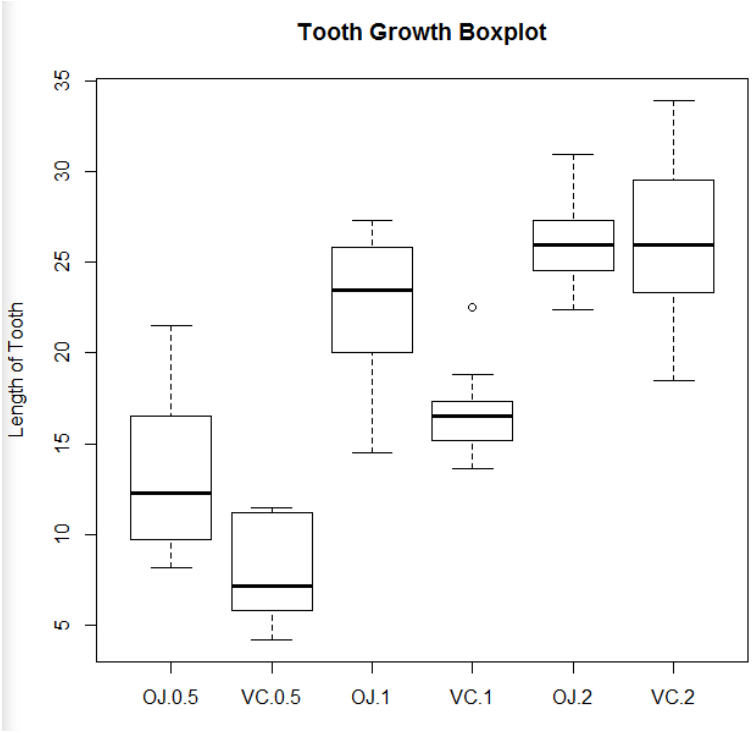
The information in this document represents the beginning analysis skills of the author. There are two parts in this report. In part 1, a simulation exercise is completed and reported on and was submitted separately. Part 2 is a basic inferential data analysis and is located below. The code for Part 2 is located in the Appendix.

Part 2: Basic Inferential Data

After loading ToothGrowth and creating a data frame, a basic summary of the data was obtained by running head(ToothGrowth). This data displayed the column headings and the first rows of the set. Next, summary(ToothGrowth). The code and output for the data are listed in the Appendix.
The minimum length was 4.2, and the maximum was 33.90. The median length (19.25) was greater than the mean (18.81). The minimum dosage given was 0.5, and the maximum dosage was 2.0. Contrary to the lengths, the median (1.0) was less than the median (1.167).

When comparing length by supplement, the t value of 1.9153 suggests it is inconclusive if there is a difference between tooth lengths of the two supplements.

When comparing a dosage of 0.5mg to 1.0mg of supplement, the t value (-6.4766, $p<0.001$) indicates significance between dosages. The same is true between 1.0mg to 2.0mg ($t=-4.9005$, $p<0.001$) and between 0.5mg and 2.0mg ($t=-11.799$, $p<0.001$). One assumption is continued taking of the supplement throughout the study.



Appendix 1: Part 2 R Code and Outputs

```
> ##This is part 2 of the week 3 statistical inference assignment by Michelle Gitgit
> ##Load ToothGrowth data
> data("ToothGrowth")
> data<-as.data.frame(ToothGrowth)
> str(data)
'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 ...
>
> ##Basic analysis and basic summary
> 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
> 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
> boxplot(len~supp * dose, data=ToothGrowth, ylab="Length of Tooth", main="Tooth Growth Boxplot")
> ToothGrowth$len <-as.numeric(ToothGrowth$len)
> ToothGrowth$dose <-as.numeric(ToothGrowth$dose)
> ToothGrowth$supp <- as.numeric(ToothGrowth$supp)
>
> table (ToothGrowth$supp, ToothGrowth$len)

   4.2 5.2 5.8 6.4 7 7.3 8.2 9.4 9.7 10 11.2 11.5 13.6 14.5 15.2 15.5 16.5
 1   0   0   0   0 0   0   1   1   2  1    0    0    0    2    1    0    1
 2   1   1   1   1 1   1   0   0   0  1    2    1    1    1    1    1    2

   17.3 17.6 18.5 18.8 19.7 20 21.2 21.5 22.4 22.5 23 23.3 23.6 24.5 24.8 25.2
 1    0    1    0    0    1 1    1    1    1    0 1    1    1    1    1    1
 2    2    0    1    1    0 0    0    1    0    1 0    1    1    0    0    0

   25.5 25.8 26.4 26.7 27.3 29.4 29.5 30.9 32.5 33.9
 1    1    1    3    0    2    1    0    1    0    0
 2    1    0    1    1    0    0    1    0    1    1
>
```

```
> 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 1 mean in group 2
       20.66333        16.96333

>
> t.test(len ~ dose, paired = F, var.equal = F, data = subset(ToothGrowth,ToothGrowth$dose!="2"))
      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 = subset(ToothGrowth,ToothGrowth$dose!="0.5"))

        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

>

paired = F, var.equal = F, data = subset(ToothGrowth,ToothGrowth$dose!="1"))

        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