Basic Inferential data Analysis

I. Goikoetxea

Overview

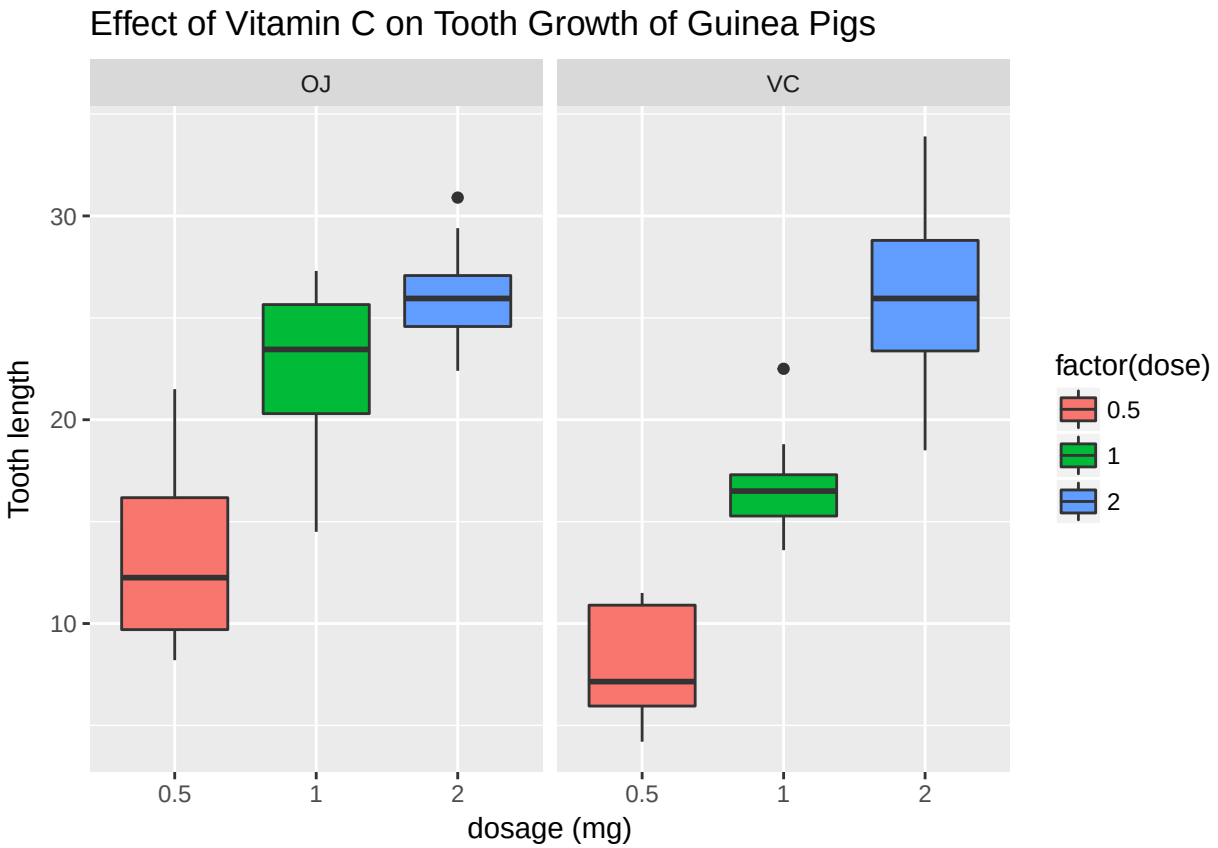
The goal is to analyse the ToothGrowth data in the R datasets package. In this dataset, the effect of Vitamin C on tooth growth in Guinea pigs is provided. The response is the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, (orange juice or ascorbic acid (a form of vitamin C and coded as VC). By means of the techniques from the Statistical Inference class, confidence intervals and/or hypothesis tests will be used to compare tooth growth by supplement and dose.

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

```
library(datasets)
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.3.2
```

```
data(ToothGrowth)
plot2 <- ggplot(data=ToothGrowth, aes(x=factor(dose), y=len, fill=factor(dose))) +
        geom_boxplot() + facet_grid(.~supp) +
        ggtitle("Effect of Vitamin C on Tooth Growth of Guinea Pigs") +
        ylab("Tooth length") + xlab("dosage (mg)")
plot2
```



The relation between tooth length for each dose and supplement method is shown in the boxplot below. The plot indicates that for larger doses of both supplements, tooth grow faster.

2. Provide a basic summary of the data.

```
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
```

3. 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)

3.1 As a factor of dosage:

```
dose_0.5 <- subset(ToothGrowth[ToothGrowth$dose==0.5,])
dose_1.0 <- subset(ToothGrowth[ToothGrowth$dose==1.0,])
```

```
dose_2.0 <- subset(ToothGrowth[ToothGrowth$dose==2.0,])
t.test(len~supp,data=dose_0.5, paired=F)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
## mean in group OJ mean in group VC
##            13.23             7.98
```

```
t.test(len~supp,data=dose_1.0, paired=F)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
## mean in group OJ mean in group VC
##            22.70            16.77
```

```
t.test(len~supp,data=dose_2.0, paired=F)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = -0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean in group OJ mean in group VC
##            26.06            26.14
```

For a dosage = 0.5 mg the confidence interval is [1.719, 8.781].The 95% confidence interval does not contain zero, therefore, we can reject the null hypothesis, and confirm that there is a correlation between tooth length and supplement at this dosage.

For a dosage = 1.0 mg the confidence interval is [2.802, 9.058]. The 95% confidence interval does not contain zero, therefore, we can reject the null hypothesis, and confirm that there is a correlation between tooth length and supplement at this dosage.

For a dosage = 2.0 mg the confidence interval is [-3.798, 3.638]. The 95% confidence interval contains zero, therefore, we can not reject the null hypothesis (it is possible that there is no correlation between tooth length and supplement at this dosage.)

3.2 As a factor of supplement

```
t.test(len~supp,data=ToothGrowth, paired=F)
```

```
##
##  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 OJ mean in group VC
##         20.66333         16.96333
```

The confidence interval [-0.171, 7.571] does not allow us to reject the null hypothesis (that there is no correlation between supplement type and tooth length).

4. State your conclusions and the assumptions needed for your conclusions.

Conclusions

Guinea pigs with OJ supplement at 0.5 and 1.0 dosages have more pronounced tooth growth than guinea pigs given VC. At a dose of 2.0 guinea pigs given OJ or VC do not have significantly different tooth growth. Increasing dosage significantly increase tooth growth for both supplements.

Assumptions

The data is not paired and the variances between populations are not equal.