Statistical Inference Course Project - Part 2

pvv88

Overview

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).

In the second part of the project, we're going to analyze the ToothGrowth data in the R datasets package by doing the following:

1. Load the ToothGrowth data
2. Provide a basic summary of the data
3. Perform some basic exploratory data analysis
4. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.
5. State conclusions and the assumptions needed to make them

Load the ToothGrowth data

```
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 ...
```

```
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
```

Basic summary of the data

```
summary(ToothGrowth)
```

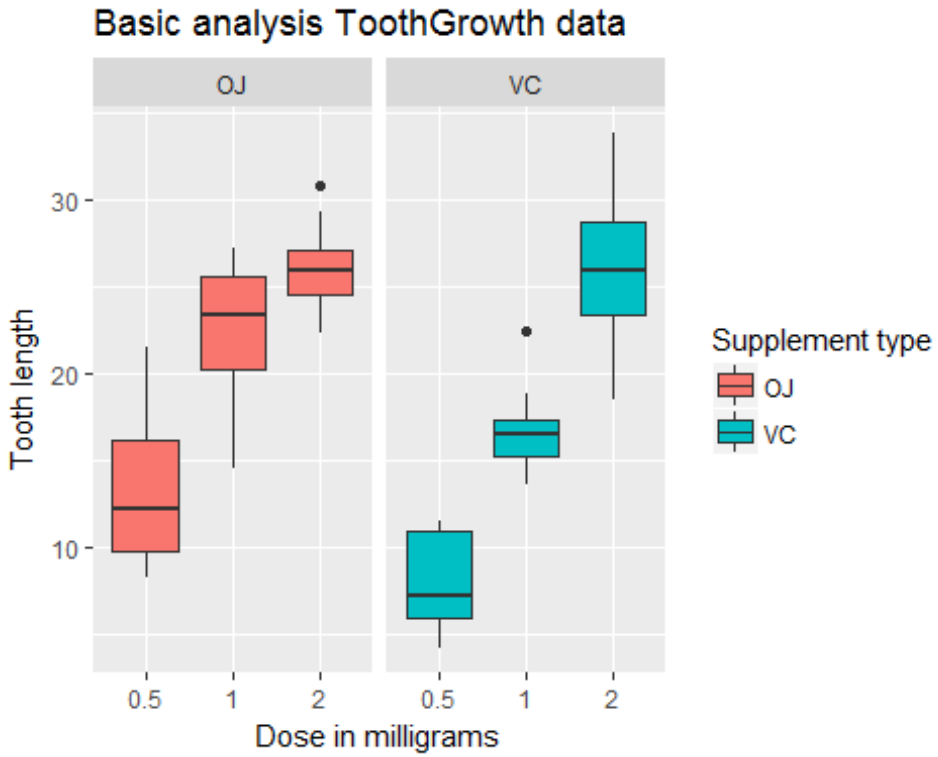
```
##       len        supp         dose
##  Min.   : 4.20   OJ:30   Min.   :0.500
##  1st Qu.:13.07   VC:30   1st Qu.:0.500
##  Median :19.25           Median :1.000
##  Mean   :18.81           Mean   :1.167
```

```
##  3rd Qu.:25.27           3rd Qu.:2.000
##  Max.   :33.90           Max.   :2.000
```

As presented above, there are 60 observations, 2 types of supplements (OJ & VC), and 3 dosage sizes (0.5, 1.0, & 2mg).

Exploratory Data Analysis

```
library(ggplot2)
ggplot(aes(x=factor(dose), y=len), data = ToothGrowth) +
  geom_boxplot(aes(fill = supp))+
  facet_grid(~supp) +
  xlab("Dose in milligrams") +
  ylab("Tooth length") +
  guides(fill=guide_legend(title="Supplement type")) +
  ggtitle("Basic analysis ToothGrowth data")
```



For doses of 0.5mg and 1.0mg, orange juice appears to help tooth growth more than ascorbic acid. However in the case of a dose of 2.0mg, both have the same impact to tooth growth. It was also observed that tooth growth increases with dose for either supplement type.

Confidence Intervals and Hypothesis Tests

Comparing tooth growth by supplement using t-test

```
t.test(len ~ supp, 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 OJ mean in group VC
##         20.66333         16.96333
```

The p-value is 0.06.This indicates that supplement types have no impact on Tooth growth.

Now we compare tooth growth by dose

```
# run t-test for 0.5 and 1.0 dosages
ToothGrowth_sub <- subset(ToothGrowth, ToothGrowth$dose %in% c(0.5,1.0))
t.test(len~dose,data=ToothGrowth_sub)

##
##  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

# run t-test for 0.5 and 2.0 dosages
ToothGrowth_sub <- subset(ToothGrowth, ToothGrowth$dose %in% c(0.5,2.0))
t.test(len~dose,data=ToothGrowth_sub)

##
##  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

# run t-test for 1.0 and 2.0 dosages
ToothGrowth_sub <- subset(ToothGrowth, ToothGrowth$dose %in% c(1.0,2.0))
t.test(len~dose,data=ToothGrowth_sub)
```

```
##
##  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
```

Since each test's p-value is less than 0.05 and its confidence interval is less than zero, we can reject the null hypothesis and can conclude that increasing the dose level leads to tooth growth.

Conclusions and the Assumptions

Given the following assumptions:

Following results can be concluded based on supporting data above:

1. From the exploratory data analysis, inreasing the dose level leads to increased tooth growth.
2. From the T-test analysis, supplement type has no impact to tooth growth.
3. From the T-test analysis, orange juice is more effective than abscorbic acid for dosages 0.5mg and 1mg. Both orange juice and abscorbic acid provides the same amount of tooth growth for .0mg dosage. However, we cannot conclude that orange juice is more effective than abscorbic acid for the entire dataset.