

Statistical Inference Project 2

Samir Yelne

This is the part 2 of the course project of the Coursera course Statistical Inference. In this project we perform basic inferential analyses using the ToothGrowth data in R datasets package.

The response variable is the length of odontoblasts (teeth) in each of 10 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 or ascorbic acid).

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

```
library(datasets)
data(ToothGrowth)
# Look at the first 6 rows
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
```

```
# look at the data variables
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 ...
## $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

```
# convert variable dose to factor
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
```

2. Provide the summary of the data.

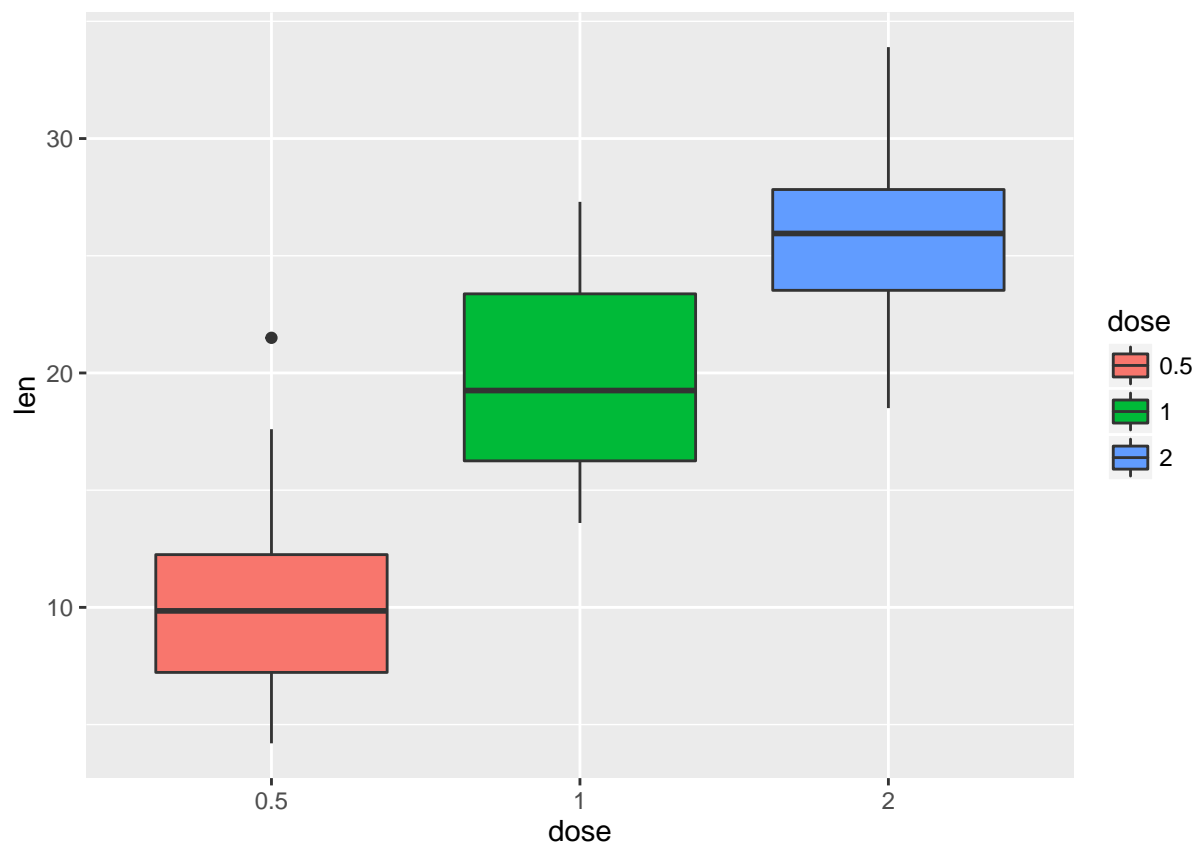
```
summary(ToothGrowth)
```

```
##      len      supp      dose
## Min.   : 4.20   OJ:30   0.5:20
## 1st Qu.:13.07   VC:30   1  :20
## Median :19.25           2  :20
## Mean   :18.81
## 3rd Qu.:25.27
## Max.   :33.90
```

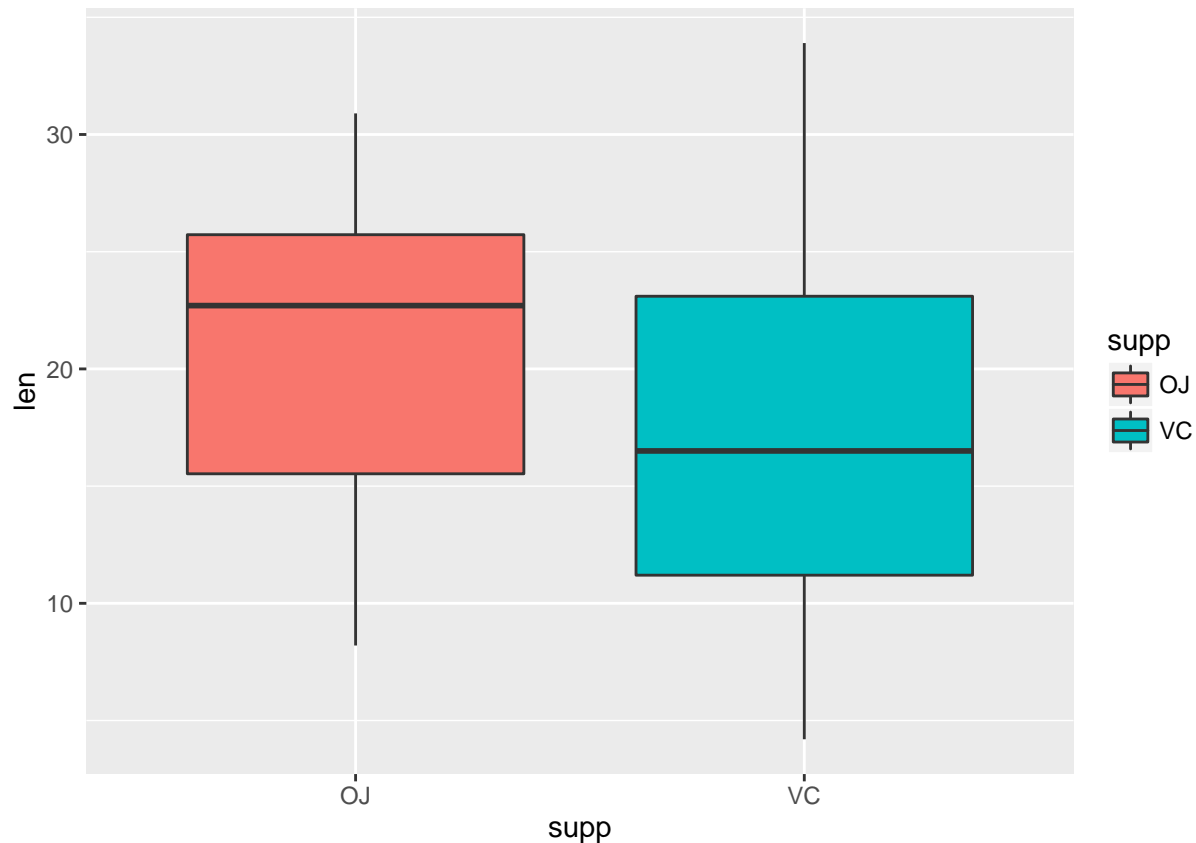
```
# use the table function to see different dose levels and delivery methods
table(ToothGrowth$dose,ToothGrowth$supp)
```

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

```
# Visualization of Tooth length as per the dose
library(ggplot2)
ggplot(ToothGrowth,aes(x=dose,y=len))+geom_boxplot(aes(fill=dose))
```



```
# Visualization of Tooth length as per supplement type
library(ggplot2)
ggplot(ToothGrowth,aes(x=supp,y=len))+geom_boxplot(aes(fill=supp))
```



3. Use confidence intervals and hypothesis tests to compare tooth growth by supp and dose

```
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 0.06063 which is greater than 0.05 and the confidence interval contains 0, that indicates we can't reject the null hypothesis that different supplement types have no effect on tooth length.

```
# create three dose level pair groups
t1 <- subset(ToothGrowth, dose %in% c(0.5,1.0))
t2 <- subset(ToothGrowth, dose %in% c(0.5,2.0))
t3 <- subset(ToothGrowth, dose %in% c(1.0,2.0))
```

```
t.test(len~dose,data=t1)
```

```
##
## 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,data=t2)
```

```
##
## 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
```

```
t.test(len~dose,data=t3)
```

```
##
## 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
```

For all the three dose level pair groups, the p value is less than 0.05 and the confidence interval does not contain zero. This indicates that we can reject the null hypothesis, and say that increase in dose levels increases tooth length.

4 Assumptions and Conclusions :

Conclusions

1. Different supplement types have effect on tooth growth.
2. Increasing the dose level leads to increased tooth growth.

Assumptions

1. Members of the sample population are representative of the entire population of guinea pigs. This assumption allows us to generalize the results.
2. The experiment was done with random assignment of guinea pigs to different dose level categories and supplement type to control for confounders affecting the outcome.