Statistical Inference project Part 2

Overview

This report is for second part of the course project of the Coursera course "Statistical Inference" which is a part of specialization "Data Science". In this second part, we perform basic inferential analyses using the ToothGrowth data in the R datasets package.

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

```
#load the dataset
library(datasets)
data(ToothGrowth)

# check the dataset details
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 ...
```

```
# number of rows of dataset
nrow(ToothGrowth)
```

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

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

# look at the dataset variables after conversion
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: Factor w/ 3 levels "0.5","1","2": 1 1 1 1 1 1 1 1 1 1 ...
```

3. Provide a basic summary of the data

```
# summary statistics for all variables
summary(ToothGrowth)
```

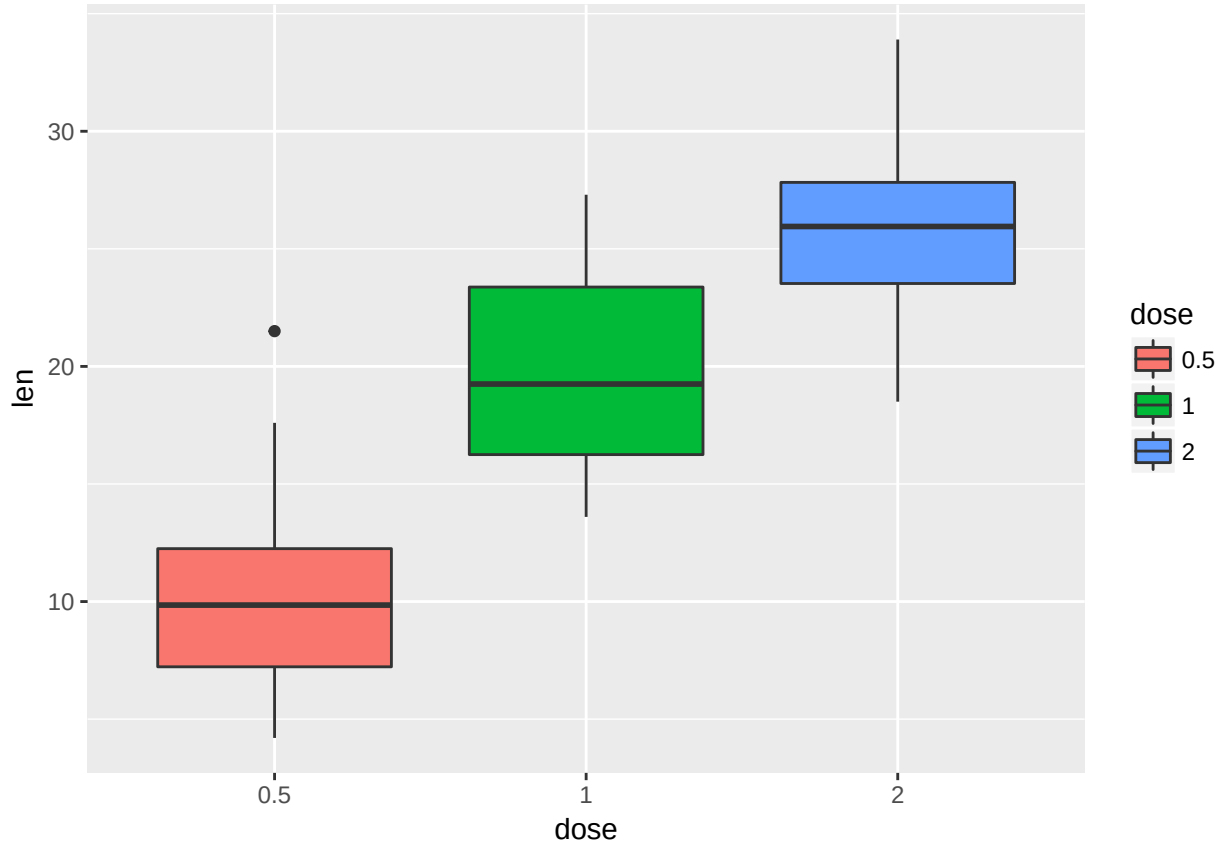
```
##       len        supp     dose
##  Min.   : 4.20   OJ:30   0.5:20
##  1st Qu.:13.07   VC:30   1  :20
##  Median :19.25           2  :20
##  Mean   :18.81
##  3rd Qu.:25.27
##  Max.   :33.90
```

```
#cases per different doses and delivery methods
table(ToothGrowth$dose, ToothGrowth$supp)
```

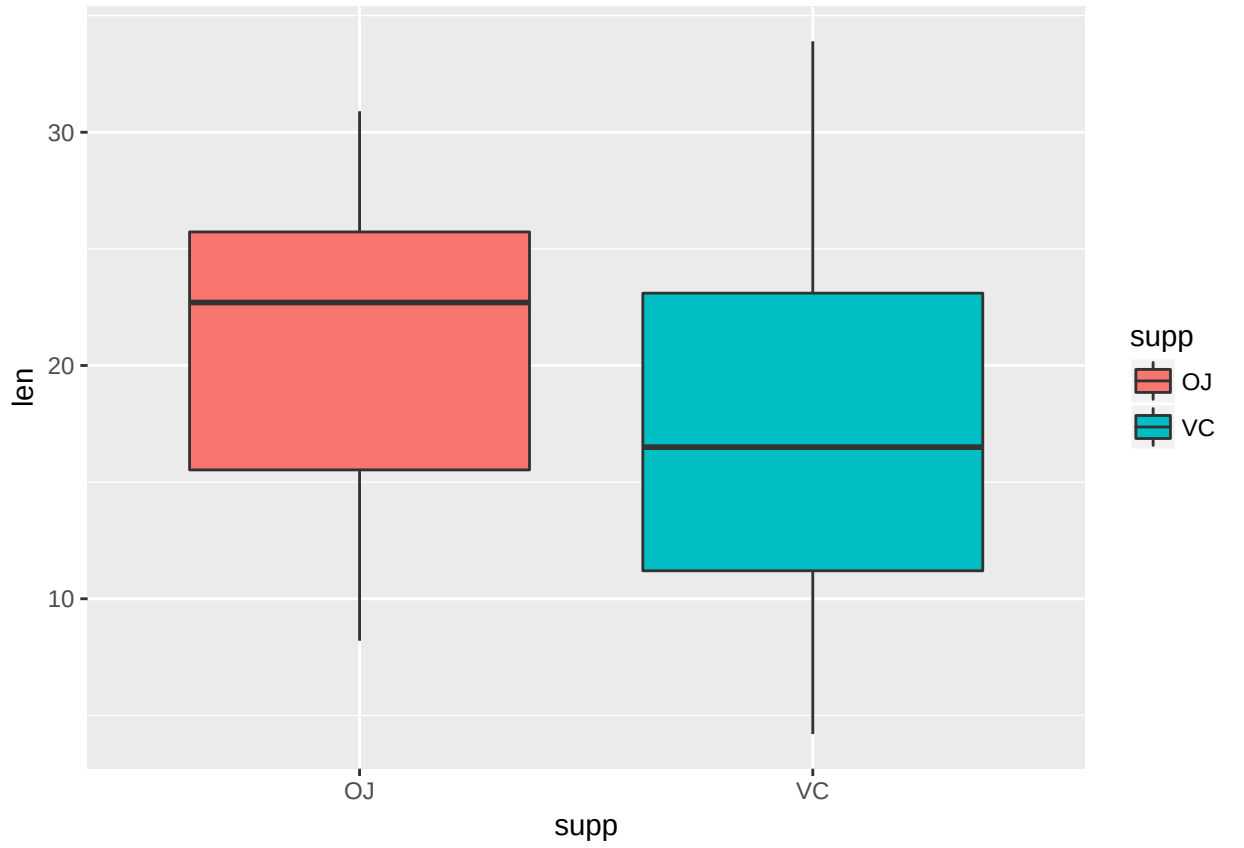
```
##
##       OJ VC
##   0.5 10 10
##   1   10 10
##   2   10 10
```

```
#loading ggplot2
library(ggplot2)
```

```
ggplot(aes(x=dose, y=len), data=ToothGrowth) + geom_boxplot(aes(fill=dose))
```




```
# visualization of tooth growth as function of supplement type
ggplot(aes(x=supp, y=len), data=ToothGrowth) + geom_boxplot(aes(fill=supp))
```



3. Use confidence intervals and/or 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 confidence interval contains zero which indicates that we can not reject the null hypothesis which states that the different supplement types have no effect on tooth length

```
# first create three sub-groups as per dose level pairs
ToothGrowth.doses_0.5_1.0 <- subset (ToothGrowth, dose %in% c(0.5, 1.0))
ToothGrowth.doses_0.5_2.0 <- subset (ToothGrowth, dose %in% c(0.5, 2.0))
ToothGrowth.doses_1.0_2.0 <- subset (ToothGrowth, dose %in% c(1.0, 2.0))
```

```r
# Check for group differences due to different dose levels (0.5, 1.0)
# assuming unequal variances between the two groups
t.test(len ~ dose, data = ToothGrowth.doses_0.5_1.0)
```

```
##
##  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
```

```r
# Check for group differences due to different dose levels (0.5, 2.0)
t.test(len ~ dose, data = ToothGrowth.doses_0.5_2.0)
```

```
##
##  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
```

```r
# Check for group differences due to different dose levels (1.0, 2.0)
t.test(len ~ dose, data = ToothGrowth.doses_1.0_2.0)
```

```
##
##  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 3 subsets, the confidence levels does not contain 0 and mean length increases with increase in dose. We can reject the null hypothesis and put forward that increasing the dose increases in tooth length

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

CONCLUSIONS: 1. supplement type has no impact on tooth growth 2. Dosage impacts tooth length, increase of which increases the tooth length

ASSUMPTIONS: 1. The sample size of guinea pigs is 60, which is used for conclusions 2. For the t-tests, the

variances are assumed to be different for the two groups being compared.