Statistical Inference - Basic inferential data analysis

VN

9 September 2017

Overview

This document has been prepared for the submission of the assignment (part 2) of Statistical Interfernce course. This involves an exercise around Inferential data analysis of the ToothGrowth data.

Assignment

Dataset : The response 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 (supp - orange juice or ascorbic acid)
Using the above data set :
- Provide a basic summary of the data.
- Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.
- State your conclusions and the assumptions needed for your conclusions.

Global settings

Set working directory, load dataset and ggplot2

Data Summary

Data summarized using - summary, str functions

```
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 ...
```

```
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
```

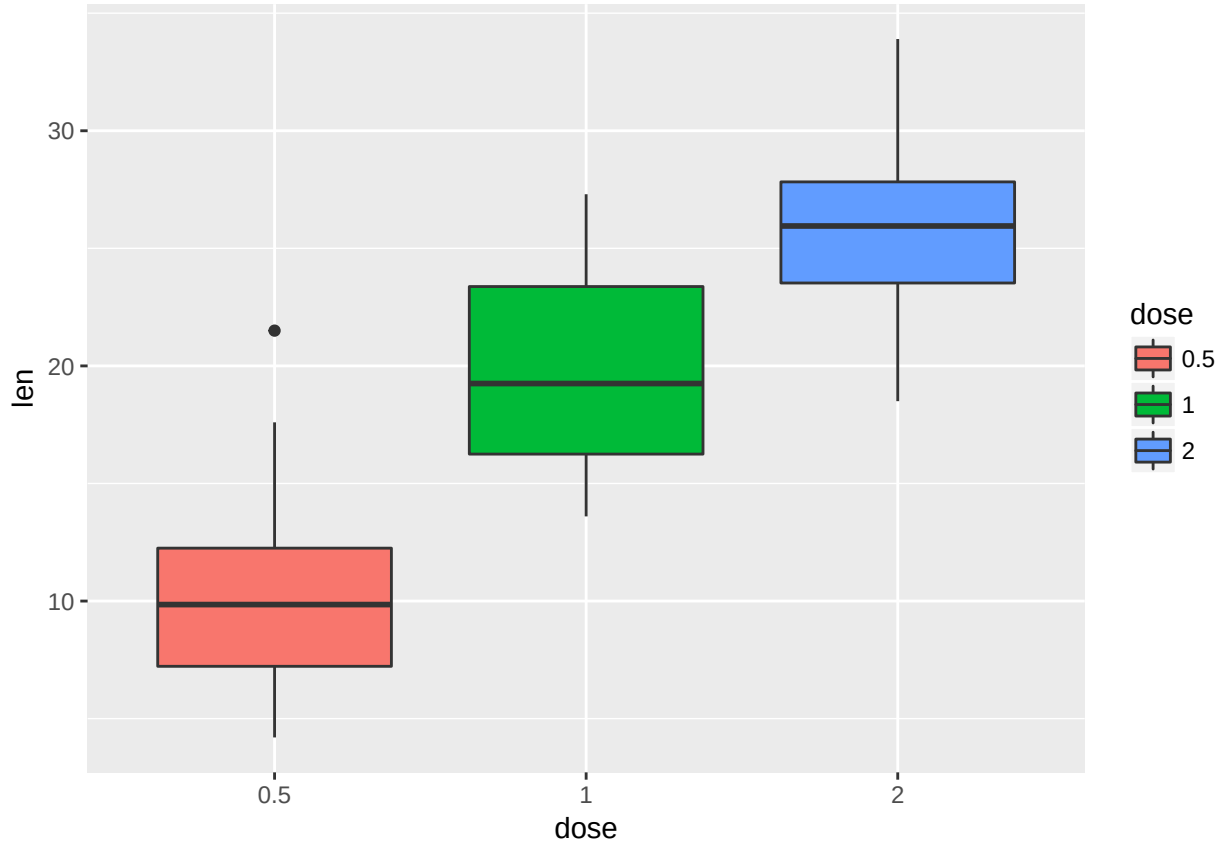
Variable dose as factor

```
ToothGrowth$dose<-as.factor(ToothGrowth$dose)
```

Visualize Data - Dose & Delivery method

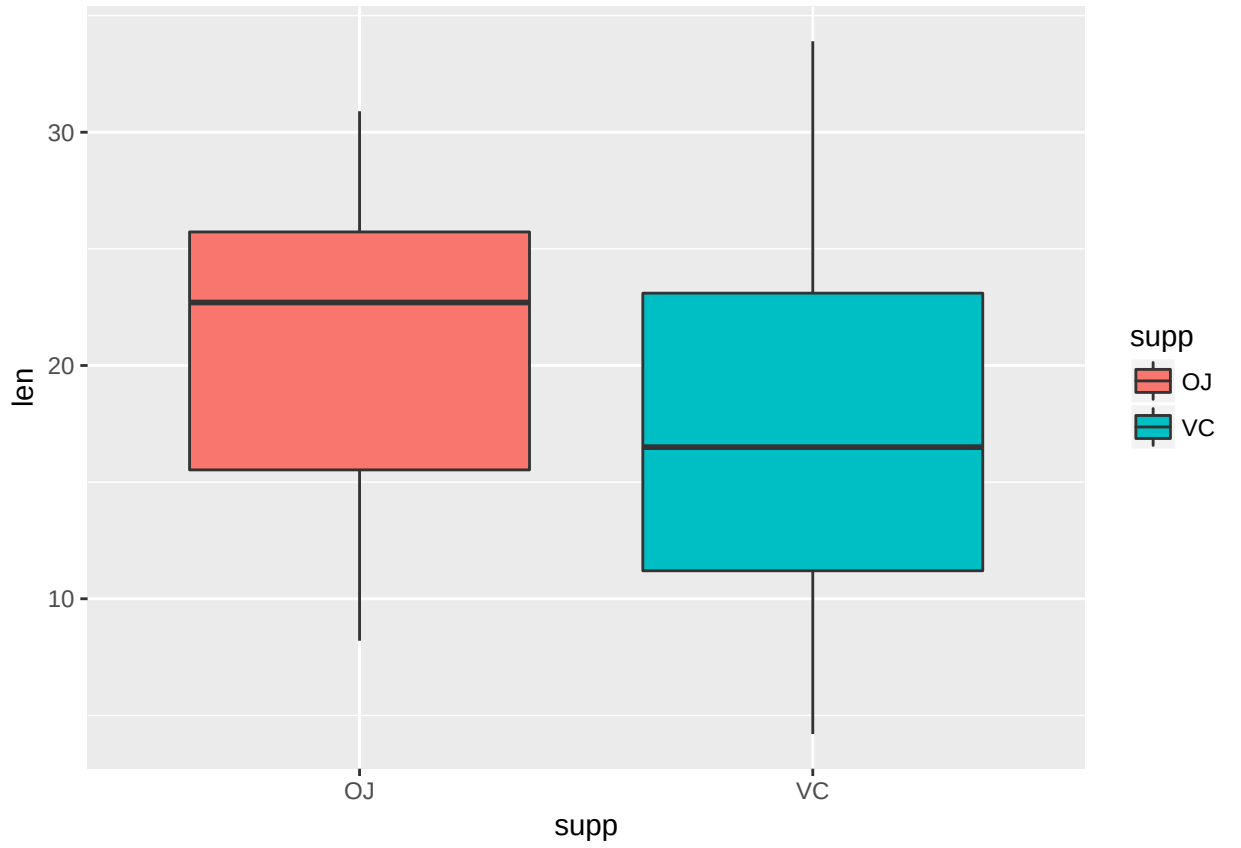
Boxplots to show response in length for the dose as factor

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



Boxplots to show response in length for the supplement as factor

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



Confidence intervals and Hypothesis tests to compare tooth growth by supp and dose

Suppliment Type To check the tooth growth by supp

```
# check for group differences due to different supplement type
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, and the confidence interval contains zero. This indicates that we can not reject the null hypothesis. That is the different supplement types have no effect on tooth length *Dosage* To check the tooth growth by dosage. For this the data is being paired amongst the values of 0.5,1.0 & 2.0

```
# 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)
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 the three dose level pairs, the p-value is less than 0.05, and the confidence interval does not contain zero in all cases. Also, the mean tooth length increases on raising the dose level. This indicates that we can reject the null hypothesis, and establish that increasing the dose level leads to an increase in tooth length

Conclusion

The confidence intervals and Hyposthesis test included above can be used to conclude :
- *1* Supplement types have no effect on tooth length
- *2* As the dosage increases tooth length increases