Part 2. Basic inferential data analysis.

RE

Statistical Inference. Course Project.

Part 2. Basic inferential data analysis.

Problems.

Now in the second portion of the class, we're going to analyze the ToothGrowth data in the R datasets package.

1. Load the ToothGrowth data and perform some basic exploratory data analyses
2. Provide a basic summary of the data.
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)
4. State your conclusions and the assumptions needed for your conclusions.

Solution

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

The dataset ToothGrowth presents the effect of vitamin C on tooth growth in guinea pigs.

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 (orange juice or ascorbic acid).

The dataset is a data frame with 60 observations on 3 variables.

[,1] len *numeric* Tooth length

[,2] supp *factor* Supplement type (VC or OJ).

[,3] dose *numeric* Dose in milligrams.

Source

C. I. Bliss (1952) The Statistics of Bioassay. Academic Press.

```
data<-ToothGrowth
levels(data$supp)<-c("Orange juice", "Ascorbic acid")
head(data)
```

```
##    len          supp dose
## 1  4.2 Ascorbic acid  0.5
## 2 11.5 Ascorbic acid  0.5
```

```
## 3  7.3 Ascorbic acid  0.5
## 4  5.8 Ascorbic acid  0.5
## 5  6.4 Ascorbic acid  0.5
## 6 10.0 Ascorbic acid  0.5
```

```
str(data)
```

```
## '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 "Orange juice",..: 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(data)
```

```
##       len                  supp         dose
##  Min.   : 4.20   Orange juice :30   Min.   :0.500
##  1st Qu.:13.07   Ascorbic acid: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
```

2. Provide a basic summary of the data.

60 observations are combinations for Supplement type and dose for 10 pigs.

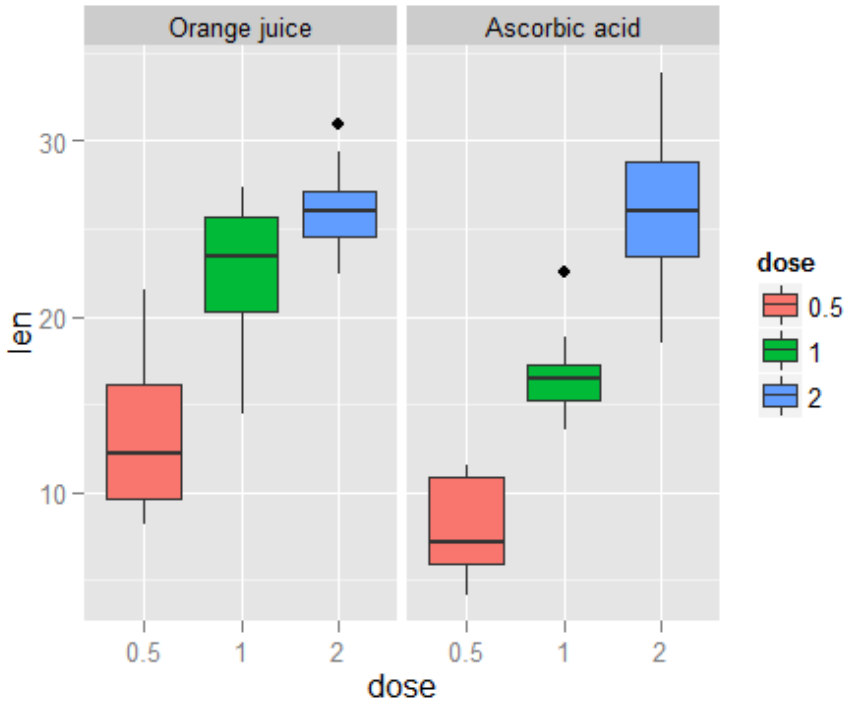
Lets transform the variable *dose* to factor.

```
data$dose<-factor(data$dose)
```

Lets perform some exploratory plot.

```
library(ggplot2)

g<-ggplot(data, aes(x=dose, y=len))+
geom_boxplot(aes(fill=dose))+facet_wrap(~supp)
g
```



It seems that 1. Vitamin C impacts on teeth groth (higher dose corresponds to longer teeths) 2. Orange juice seems more effective on groth 3. These effects are decreases on large doses.

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)

We haven't assuptions about population variance, so use appropriate t-test.

Lets compare tooth growth of dose of vitamine C.

Null H0 hypothesis: the means tooth length corresponding to different dosages are the same

For samples with 0.5 and 1 doses we obtaine:

```
data_dose_05_1<-subset(data, dose %in% c(0.5, 1))
t.test(len~dose, paired=F, var.equal=F, data=data_dose_05_1)
```

```
##
##  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
```

H0 has to be rejected with *p*-value quite small. Also confidence interval for difference for means is not include 0.

Just the same we see for samples with 1 and 2 doses

```
data_dose_1_2<-subset(data, dose %in% c(1, 2))
t.test(len~dose, paired=F, var.equal=F, data=data_dose_1_2)
```

```
##
##  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
```

Lets compare tooth growth of delivery methods under each dose.

Null H0 hypothesis: the means tooth length corresponding to different delivery methods are the same

For dose=0.5

```
t.test(len~supp, paired=F, var.equal=F, data=data[data$dose==0.5,])
```

```
##
##  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 Orange juice mean in group Ascorbic acid
##                       13.23                       7.98
```

For dose=1

```
t.test(len~supp, paired=F, var.equal=F, data=data[data$dose==1,])
```

```
##
##  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 Orange juice mean in group Ascorbic acid
##                       22.70                       16.77
```

For dose=2

```
t.test(len~supp, paired=F, var.equal=F, data=data[data$dose==2,])
```

```
##
##  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 Orange juice mean in group Ascorbic acid
##                       26.06                       26.14
```

So we have to reject null hypothesis for doses 0.5 and 1 mg becouse of small p-value.

And we haven't reasons to reject H0 under dose 2 mg.

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

So we can conclude that

1. Dosage influence on the tooth lengh (direct).
2. Delinery methods influence on the tooth lengh under low doses becouse of saturation effect.