Statistical Inference CP2

Dan K. Hansen

31 May 2016

Overview:

We're going to analyze the ToothGrowth data in the R datasets package.

Tasks to do

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.

Data Processing

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

Loading the dataset:

```
library(datasets)
ds <- ToothGrowth
```

Basic exploration of the ds:

```
head(ds)
```

```
##    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
```

```
tail(ds)
```

```
##     len supp dose
## 55 24.8   OJ    2
## 56 30.9   OJ    2
## 57 26.4   OJ    2
## 58 27.3   OJ    2
## 59 29.4   OJ    2
## 60 23.0   OJ    2
```

```r
str(ds)
```

```
## '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 ...
```

Question 2: Provide a basic summary of the data

Basic summary of the dataset:

```r
summary(ds)
```

```
##       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
```

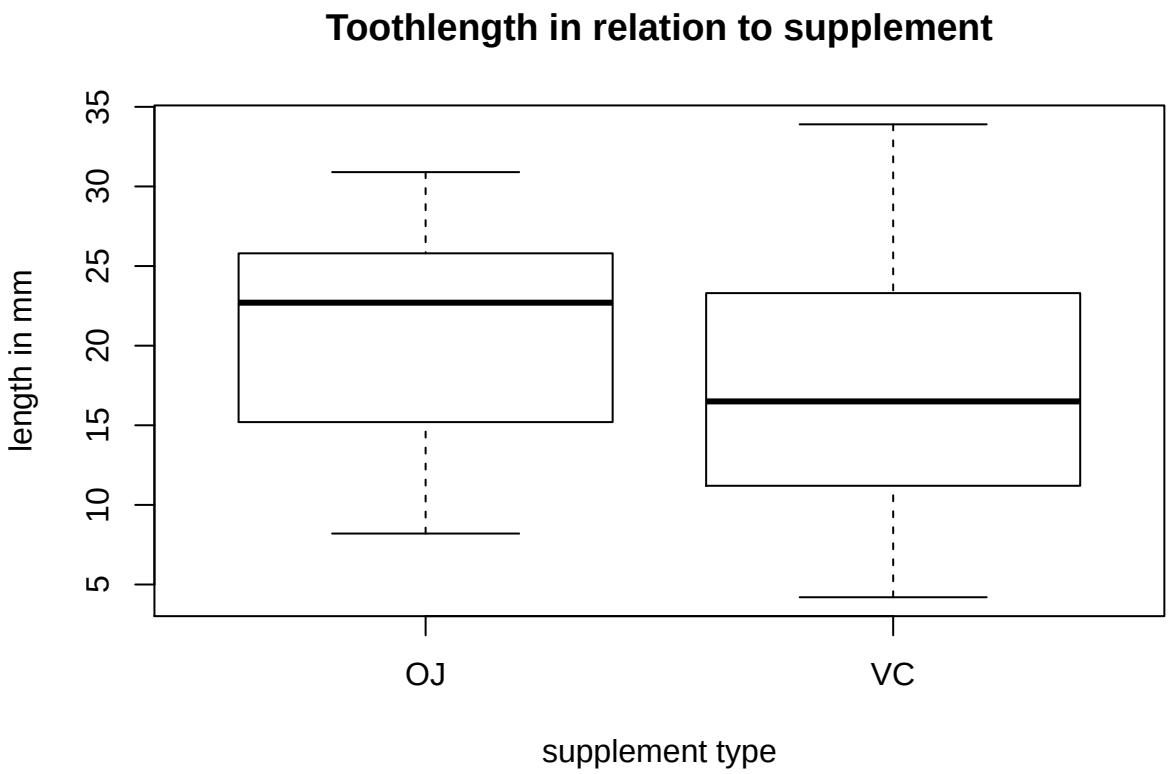
Question 3: Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.

The null hypothesis *(an indication of Status Quo and always considered to be True)* is rejected when $p < 0.05$ and not rejected when $p > 0.05$.

Confidence intervals of difference parameters not containing 0 imply that there is a statistically significant difference between the populations.

A couple of plots of toothlength in relation to supplement type and dose respectively:

```r
boxplot(ds$len~ds$supp, main="Toothlength in relation to supplement",
              ylab="length in mm", xlab="supplement type")
```



This plot seem to show that supplement OJ is slightly more efficient than supplement VC on toothgrowth.

Check H_0 that the length of teeth growth is related to supplement type:

```
t.test(ds$len ~ ds$supp)
```

```
##
##  Welch Two Sample t-test
##
## data:  ds$len by ds$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
```

$p > 0.05$ and CI contains 0 => H_0 can not be rejected

I.e. we can't say that the type of supplement have any impact on the average toothgrowth.

```
boxplot(ds$len~ds$dose, main="Toothlength in relation to dose",
              ylab="length in mm", xlab="dose in mg/day")
```

This plot shows that an increase of dose levels have an impact on toothgrowth.

Check H_0 that the length of teeth growth is related to dose level:

```
t.test(ds$len,ds$dose)
```

```
##
##  Welch Two Sample t-test
##
## data:  ds$len and ds$dose
## t = 17.81, df = 59.798, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  15.66453 19.62881
## sample estimates:
## mean of x mean of y
## 18.813333  1.166667
```

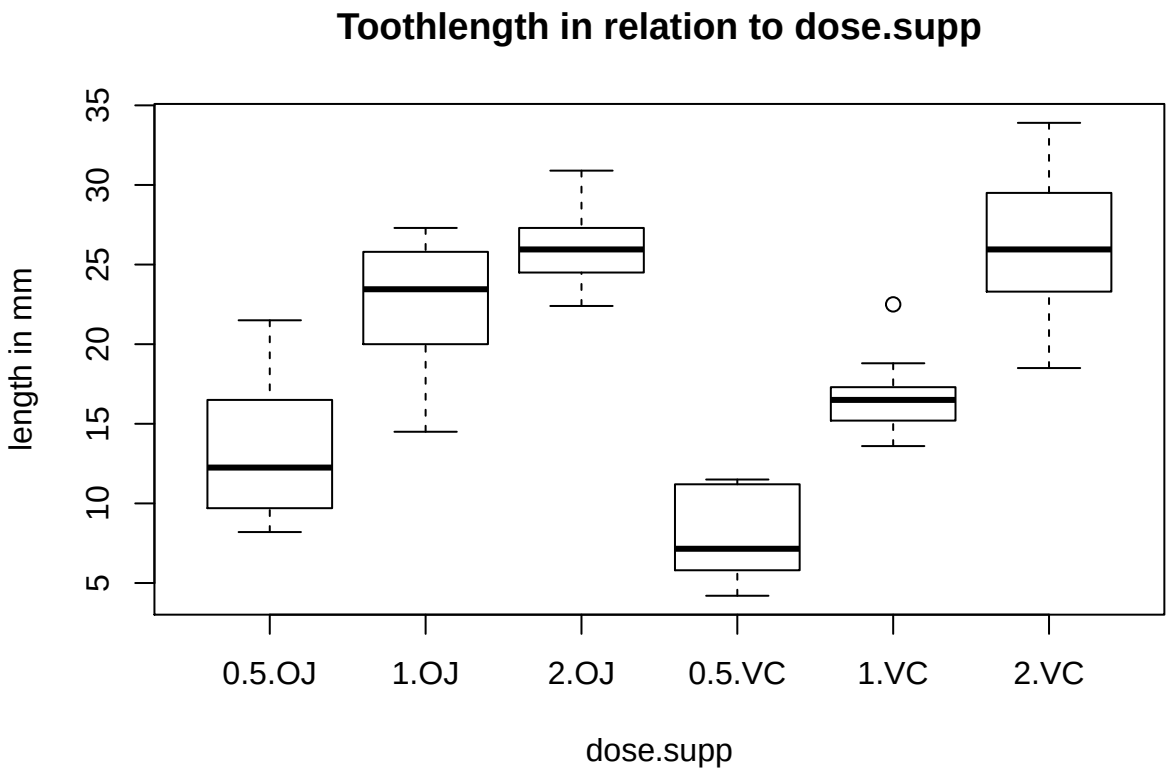
$p < 0.05$ and CI does not contain 0 => H_0 can be rejected

I.e. we can say that the dose level seem to have an impact on the average toothgrowth.

Further investigations and tests:

Plot of length in relation to dose and supplement:

```
boxplot(ds$len ~ interaction(ds$dose,ds$supp), main="Toothlength in relation to dose.supp",
              ylab="length in mm", xlab="dose.supp")
```



In this plot we can see that toothgrowth is apparently dependent on dose level, whether or not the supplement type is OJ or VC. There may be indications that OJ is slightly more effective than VC, at least at levels 0.5 and 1 mg/day. At 2 mg/day the difference is not clear. Now let's dig into some tests....

How about the teeth growth in relation to dose and supplement type?

prepare for t.test:

```
dose05_OJ <- subset(ds,ds$dose==.5 & ds$supp=="OJ")
dose05_VC <- subset(ds,ds$dose==.5 & ds$supp=="VC")
dose1_OJ <- subset(ds,ds$dose==1 & ds$supp=="OJ")
dose1_VC <- subset(ds,ds$dose==1 & ds$supp=="VC")
dose2_OJ <- subset(ds,ds$dose==2 & ds$supp=="OJ")
dose2_VC <- subset(ds,ds$dose==2 & ds$supp=="VC")
```

Or length in relation to increase of dose alone, regardless of supplement type?

prepare for t.test:

```
dose05 <- subset(ds,ds$dose==.5)
dose1 <- subset(ds,ds$dose==1)
dose2 <- subset(ds,ds$dose==2)
```

t.tests of length in relation to supplement type and dosis:

```
t.test(dose05_OJ$len,dose05_VC$len)
```

```
##
##  Welch Two Sample t-test
##
## data:  dose05_OJ$len and dose05_VC$len
## 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 of x mean of y
##     13.23      7.98
```

$p < 0.05$ and CI doesn't contain 0 => H_0 can be rejected

I.e. If dose is 0.5 mg/day there seems to be difference in toothgrowth, depending on supplement type where OJ seem to be a better catalyst for toothgrowth than VC

```
t.test(dose1_OJ$len,dose1_VC$len)
```

```
##
##  Welch Two Sample t-test
##
## data:  dose1_OJ$len and dose1_VC$len
## 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 of x mean of y
##     22.70     16.77
```

$p < 0.05$ and CI doesn't contain 0 => H_0 can be rejected

I.e. If dose is 1 mg/day there seems to be difference in toothgrowth, depending on supplement type where OJ seem to be a better catalyst for toothgrowth than VC

```
t.test(dose2_OJ$len,dose2_VC$len)
```

```
##
##  Welch Two Sample t-test
##
## data:  dose2_OJ$len and dose2_VC$len
## 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 of x mean of y
##     26.06     26.14
```

p > 0.05 and CI contain 0 => H_0 can not be rejected

I.e. If dose is 2 mg/day there don't seem to be difference in toothgrowth, depending on supplement type. Both OJ and VC seem to be equally good catalyst for toothgrowth

t.tests of length in relation to dosis increase regardless of supplement type

```
t.test(dose05$len,dose1$len)
```

```
##
##  Welch Two Sample t-test
##
## data:  dose05$len and dose1$len
## 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 of x mean of y
##    10.605    19.735
```

p < 0.05 and CI doesn't contain 0 => H_0 can be rejected

I.e. regardless of supplement a dose-increase from 0.5 to 1.0 mg/day seem to increase toothgrowth

```
t.test(dose1$len,dose2$len)
```

```
##
##  Welch Two Sample t-test
##
## data:  dose1$len and dose2$len
## 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 of x mean of y
##    19.735    26.100
```

p < 0.05 and CI doesn't contain 0 => H_0 can be rejected

I.e. regardless of supplement a dose-increase from 1 to 2.0 mg/day seem to increase toothgrowth

Conclusions:

1. Giving additives to guinea pigs either in the form of OJ or VC is resulting in an increase of toothgrowth.
2. OJ seems to be slightly better than VC using doses of 0.5 or 1 mg/day. At 2 mg/day, there seem to be no significant difference in which supplement you use.
3. Using VC, an increase in daily dose seem to increase toothgrowth in a linear fashion, whereas for OJ, the effect seems to drop off the higher the dosis.
4. More data would improve the results, and hence the conclusions... :-)