Statistical Inference Course Project part 2

Jure Bordon

Sunday, February 22, 2015

Load the ToothGrowth data and perform some basic exploratory data analyses

Load the standard dataset for ToothGrowth and since dose has only 3 different values, we convert it to factor (for easier plotting):

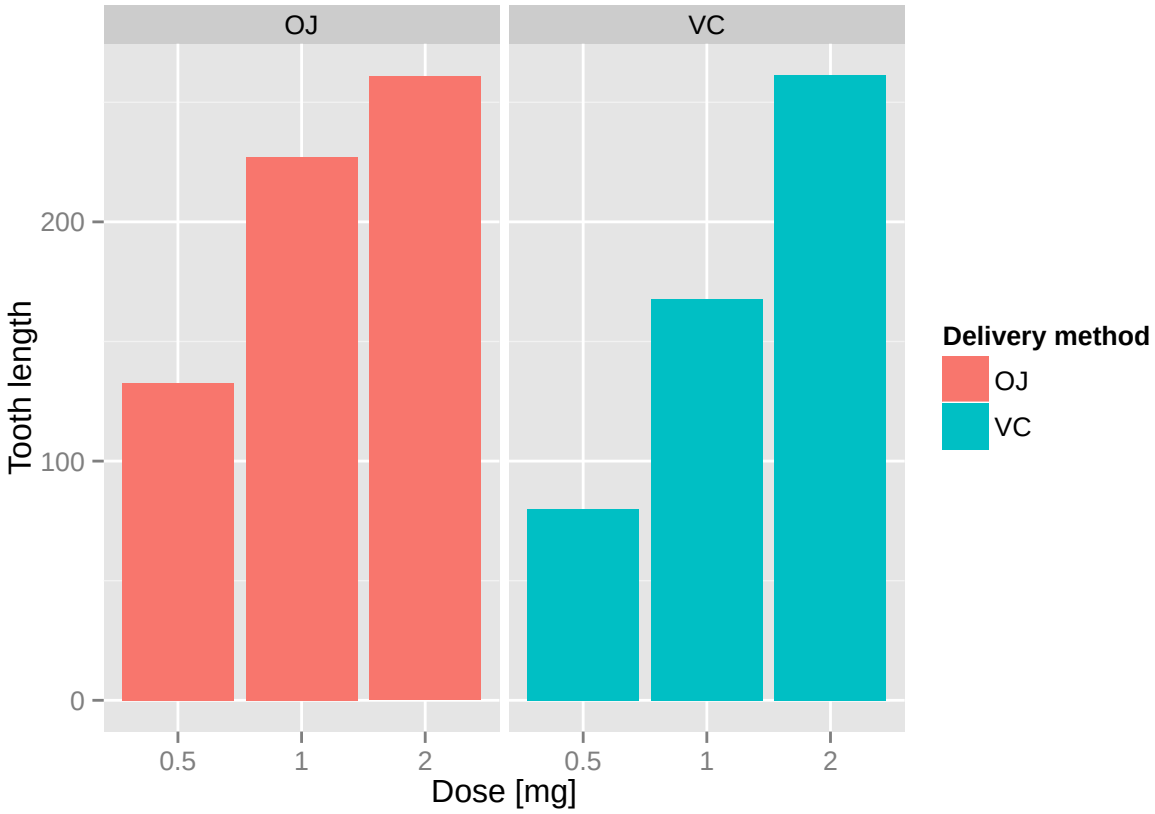
```
library(datasets)
tg <- ToothGrowth
tg$dose <- as.factor(tg$dose)
str(tg)
```

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

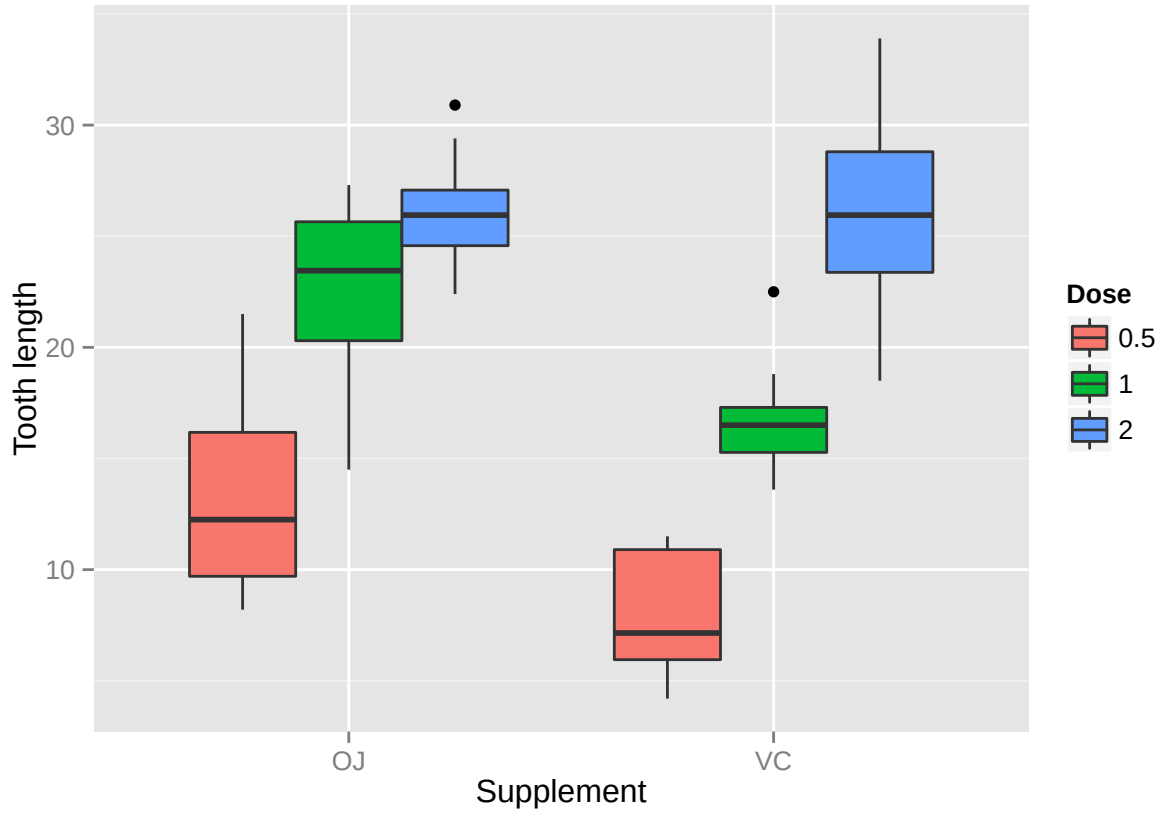
```
require(ggplot2)
```

```
## Loading required package: ggplot2
```

```
ggplot(data=tg, aes(x=dose, y=len, fill=supp)) +
    geom_bar(stat="identity") +
    facet_grid(. ~ supp) +
    xlab("Dose [mg]") +
    ylab("Tooth length") +
    guides(fill=guide_legend(title="Delivery method"))
```




```
ggplot(data=tg, aes(x=supp, y=len)) +
    geom_boxplot(aes(fill=dose)) +
    xlab("Supplement") +
    ylab("Tooth length") +
    guides(fill=guide_legend(title="Dose"))
```



Provide a basic summary of the data.

The dataset contains the length of teeth in 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1 and 2 mg). The delivery of Vitamin C was tested with two methods, orange juice or ascorbic acid.

```
summary(tg)
```

```
##       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 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 will use t.test to check if there is any difference in tooth growth when using different supplements and using different doses. We will be interested whether p-value is greater than 0.05 and if confidence interval contains 0.

Orange juice vs Vitamin C (ascorbic acid)

Dose = 0.5 mg:

```
t.test(len ~ supp,tg[tg$dose==0.5,],paired=FALSE,var.equal=TRUE)
```

```
##
##  Two Sample t-test
##
## data:  len by supp
## t = 3.1697, df = 18, p-value = 0.005304
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.770262 8.729738
## sample estimates:
## mean in group OJ mean in group VC
##            13.23             7.98
```

The mean length for OJ is 13.23 and for Vitamin C 7.98. The interval does not contain zero and p-value is less than 0.05, which means there is a statistically significant difference in effectiveness. At dose of 0.5 mg OJ is therefore more effective than Vitamin C for tooth growth.

Dose = 1 mg:

```
t.test(len ~ supp,tg[tg$dose==1,],paired=FALSE,var.equal=TRUE)
```

```
##
##  Two Sample t-test
##
## data:  len by supp
## t = 4.0328, df = 18, p-value = 0.0007807
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.840692 9.019308
## sample estimates:
## mean in group OJ mean in group VC
##            22.70            16.77
```

The mean length for OJ is 22.70 and for Vitamin C 16.77. The interval does not contain zero and p-value is less than 0.05, which means there is a statistically significant difference in effectiveness. At dose of 1 mg OJ is therefore more effective than Vitamin C for tooth growth.

Dose = 2 mg:

```
t.test(len ~ supp,tg[tg$dose==2,],paired=FALSE,var.equal=TRUE)
```

```
##
##  Two Sample t-test
##
## data:  len by supp
## t = -0.0461, df = 18, p-value = 0.9637
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
```

```
##  -3.722999  3.562999
## sample estimates:
## mean in group OJ mean in group VC
##            26.06            26.14
```

The mean length for OJ is 26.06 and for Vitamin C 26.14. The interval contains 0 and p-value is much greater than 0.05, which means we cannot assume that OJ is in any way more effective than VC at dose of 2 mg.

Dose = 0.5 mg vs Dose = 1.0 mg

```
t.test(len ~ dose,tg[tg$dose!=2,],paired=FALSE, var.equal=TRUE)
```

```
##
##  Two Sample t-test
##
## data:  len by dose
## t = -6.4766, df = 38, p-value = 1.266e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.983748  -6.276252
## sample estimates:
## mean in group 0.5   mean in group 1
##            10.605            19.735
```

The mean length for dose of 0.5 mg is 10.605 and for dose of 1 mg 19.735. The interval does not contain zero and p-value is less than 0.05, which means there is a statistically significant difference in effectiveness of the two doses. Dose of 1 mg seems to be more effective than dose of 0.5 mg.

Dose = 1.0 mg vs Dose = 2.0 mg

```
t.test(len ~ dose,tg[tg$dose!=0.5,],paired=FALSE, var.equal=TRUE)
```

```
##
##  Two Sample t-test
##
## data:  len by dose
## t = -4.9005, df = 38, p-value = 1.811e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.994387 -3.735613
## sample estimates:
## mean in group 1 mean in group 2
##          19.735          26.100
```

The mean length for dose of 1 mg is 19.735 and for dose of 2 mg 26.100. The interval does not contain zero and p-value is less than 0.05, which means there is a statistically significant difference in effectiveness of the two doses. Dose of 2 mg seems to be more effective than dose of 1 mg. From these results, we can safely assume that dose of 2 mg is also more effective than dose of 0.5 mg.

Conclusions

From these tests we can say that Orange juice seems to be more effective than Vitamin C (ascorbic acid) for tooth growth at doses 0.5 mg and 1 mg. For dose 2 mg there is no significant difference in effectivness between the two methods. We can also say that increasing the dose from 0.5 mg to 1.0 mg as well as from 1.0 mg to 2.0 mg increases the effect of tooth growth.