Statistical Inference Project Part 2

William L.

Saturday, October 25, 2014

The second part of this project involves an analysis of the ToothGrowth dataset in the R Datasets Package. The ToothGrowth data explores the effect of vitamin C on tooth growth in guinea pigs.

Load libraries and datasets:

```
library("dplyr")
```

```
##
## Attaching package: 'dplyr'
##
## The following object is masked from 'package:stats':
##
##     filter
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library("ggplot2")
library("datasets")
```

Question 1

Load the ToothGrowth data and perform some basic exploratory data analyses

The ToothGrowth data is loaded and an examination of the structure conducted:

```
data(ToothGrowth)
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 ...
```

This dataset consists of a Data Frame with 60 observations across three variables:

- **len** is a numeric column representing tooth length
- **supp** is a factor with two levels OJ (Orange Juice) and VC (Vitamin C Supplement) representing the supplement type
- **dose** is a numeric column representing the dosage provided for each supplement type

Summarize the distribution of values:

```
TGgroups <- group_by(ToothGrowth,supp,dose)
summarise(TGgroups, count=n())
```

```
## Source: local data frame [6 x 3]
## Groups: supp
##
##   supp dose count
## 1   OJ  0.5    10
## 2   OJ  1.0    10
## 3   OJ  2.0    10
## 4   VC  0.5    10
## 5   VC  1.0    10
## 6   VC  2.0    10
```

There are 30 observations each for OJ and VC broken across 3 dosage levels. Each of the dosage levels (0.5, 1.0, 2.0) have 10 observations.

Check for any NA values which may need to be adjusted for:

```
length(which(is.na(ToothGrowth$len)))
```

```
## [1] 0
```

```
length(which(is.na(ToothGrowth$supp)))
```

```
## [1] 0
```

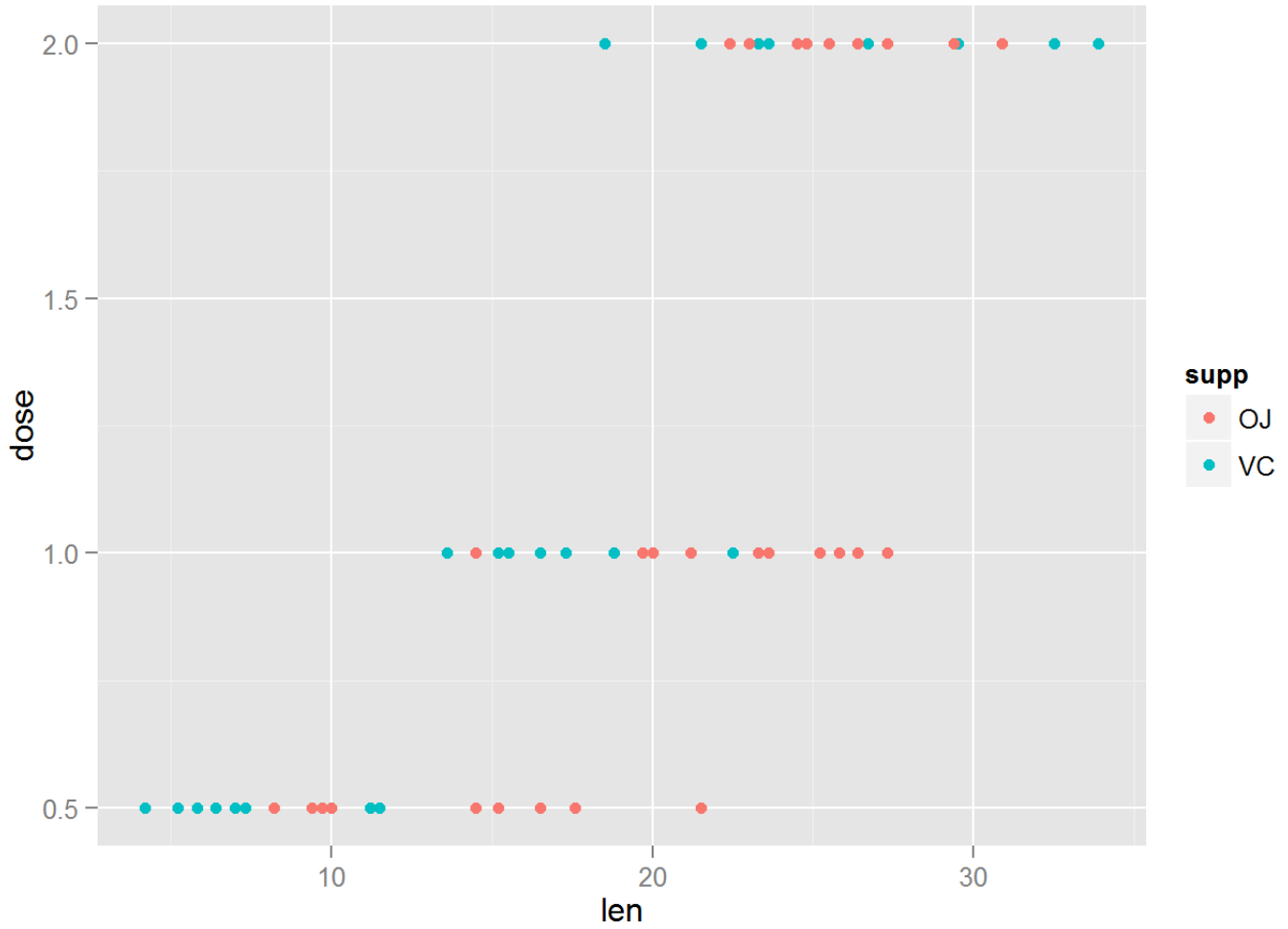
```
length(which(is.na(ToothGrowth$dose)))
```

```
## [1] 0
```

No NA values are present.

Create an initial plot of the dataset to explore potential relationships:

```
ggplot(ToothGrowth, aes(x = len, y = dose)) +
    geom_point(aes(color = supp))
```



It does appear that an increase in the dosage of Vitamin C increases tooth length. When the dose is 0.5 or 1.0 there also appears to be a better outcome of longer teeth with the OJ vs the VC supplement type, but this does not appear to hold for the 2.0 dose.

Question 2

Provide a basic summary of the data

Explore an initial summary of the full dataset:

```
summary(ToothGrowth)
```

```
##       len       supp         dose
##  Min.   : 4.2   OJ:30   Min.   :0.50
##  1st Qu.:13.1   VC:30   1st Qu.:0.50
##  Median :19.2           Median :1.00
##  Mean   :18.8           Mean   :1.17
##  3rd Qu.:25.3           3rd Qu.:2.00
##  Max.   :33.9           Max.   :2.00
```

View a summary of the mean and standard deviation for length grouped by supp and dose:

```
summarise(TGgroups, mean_len=mean(len), sd_len=sd(len))
```

```
## Source: local data frame [6 x 4]
## Groups: supp
##
##   supp dose mean_len sd_len
## 1   OJ  0.5    13.23  4.460
## 2   OJ  1.0    22.70  3.911
## 3   OJ  2.0    26.06  2.655
## 4   VC  0.5     7.98  2.747
## 5   VC  1.0    16.77  2.515
## 6   VC  2.0    26.14  4.798
```

Based on the mean for each grouping it would appear that increasing the dose of Vitamin C definately increases tooth length regardless of the delivery mechanism.

Question 3

Use confidence intervals and hypothesis tests to compare tooth growth by supp and dose

The first set of hypothesis tests will examine the relationship between the two supplement types at each dose level:

- Ho = The diffence in tooth length is not significant between the supplement types
- H1 = The diffence in tooth length is significant

```
t.test(len ~ supp, data = filter(ToothGrowth, dose==0.5))
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 3.17, df = 14.97, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.719 8.781
## sample estimates:
## mean in group OJ mean in group VC
##            13.23             7.98
```

With the p-value of **0.006** is less than 0.05 and a 95% confidence interval of **[1.71, 8.78]** the null hypothesis can be rejected. There is a significant difference between the supplement types at the 0.5 dose level.

```
t.test(len ~ supp, data = filter(ToothGrowth, dose==1.0))
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 4.033, df = 15.36, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802 9.058
## sample estimates:
## mean in group OJ mean in group VC
##            22.70            16.77
```

With the p-value of **0.001** is less than 0.05 and a 95% confidence interval of **[2.8, 9.1]** the null hypothesis can be rejected. There is a significant difference between the supplement types at the 1.0 dose level.

```
t.test(len ~ supp, data = filter(ToothGrowth, dose==2.0))
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = -0.0461, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.798  3.638
## sample estimates:
## mean in group OJ mean in group VC
##            26.06            26.14
```

With the p-value of **0.96** is greater than 0.05 and a 95% confidence interval of **[-3.8, 3.6]** the null hypothesis can not be rejected. There is no significant difference between the supplement types at the 2.0 dose level.

Though there is not a significant difference between the two supplement types at the 2.0 dosage level there is a significant difference at the lower doses (0.5 and 1.0).

The second set of hypothesis tests will examine the relationship between the dosage levels:

- Ho = The diffence in tooth length is not significant between the compared dosage levels
- H1 = The diffence in tooth length is significant

```
t.test(len ~ dose, data = filter(ToothGrowth, dose==0.5 | dose==1.0))
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -6.477, df = 37.99, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.984  -6.276
## sample estimates:
## mean in group 0.5   mean in group 1
##             10.61             19.73
```

With the p-value of **1.268e-07** is less than 0.05 and a 95% confidence interval of **[-11.98, -6.27]** the null hypothesis can be rejected. There is a highly significant difference between the tooth lengths at the 0.5 and 1.0 dose levels.

```
t.test(len ~ dose, data = filter(ToothGrowth, dose==1.0 | dose==2.0))
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -4.901, df = 37.1, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.996 -3.734
## sample estimates:
## mean in group 1 mean in group 2
##           19.73           26.10
```

With the p-value of **1.906e-05** is less than 0.05 and a 95% confidence interval of **[-8.99, -3.73]** the null hypothesis can be rejected. There is a highly significant difference between the tooth lengths at the 1.0 and 2.0 dose levels.

There is definately a significant difference between tooth length at each dosage level.

Question 4

State your conclusions and the assumptions needed for your conclusions

Based on the results of hyposthesis testing it can be concluded that a 1.0 dose of Vitamin C will increase tooth length more than a 0.5 dose and a 2.0 dose will be more effective at increasing tooth length than the 1.0 and 0.5 dosage levels. Additional testing would be required to see if tooth length continues to increase with even higher doses of Vitamin C.

With Orange Juice being significantly more effective than a Vitamin C supplement at the lower dosage levels I will make the conclusion that Orange Juice is more effective. However, this link is not as clear at the higher dosage level and more testing would be preferred. My assumption is that there are additional nutrients

provided by the Orange Juice that also assist in tooth growth.