Statistical Inference Assessment: Basic Inferential Data Analysis Exercise

Jason Collins

Overview

In this report I examine the `ToothGrowth` dataset, which contains measurements of the cells responsible for tooth growth in 60 guinea pigs. These guinea pigs were supplemented with vitamin C by orange juice or ascorbic acid (a form of vitamin C).

There is a positive relationship between dose and tooth length, with increasing dose generally increasing length regardless of supplement type. However, the evidence for increasing length for doses above 1.0 delivered by orange juice is weak.

Data Summary

First I load ggplot2, which is used in the data exploration.

```r
#load relevant libraries
library(ggplot2)
```

I then loaded the data and viewed some summaries of the data.

```r
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 ...
```

```r
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  
```

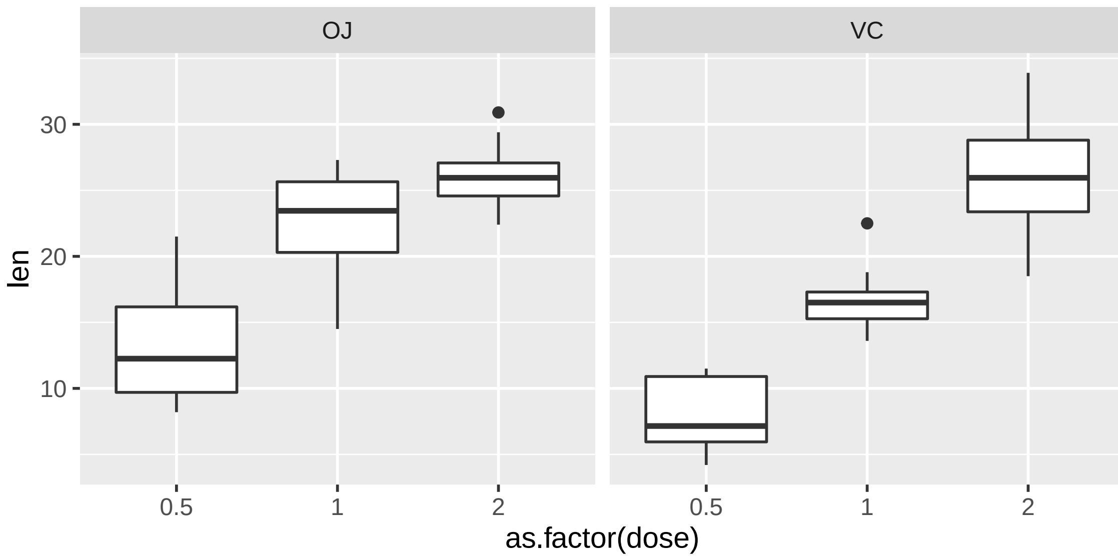
```r
unique(ToothGrowth$dose)
```

```
## [1] 0.5 1.0 2.0
```

There are two methods by which the supplement is applied - orange juice and vitamin C. It can also be seen that there are only three levels of dose.

A boxplot comparing length to dose for each or orange juice and vitamin C shows considerable difference in length with variation in dose, with a less clear effect of supplement type.

```r
g <- ggplot(ToothGrowth, aes(as.factor(dose), len)) + geom_boxplot()
g + facet_grid(. ~ supp)
```



Growth v Dose

For the first set of analysis, I examine the relationship between dose length within the two supplement types. I ran two tests within each supplement type, the first comparing a dose of 0.5 to a dose of 1.0, and the second a dose of 1.0 to a dose of 2.0. Comparison of a dose of 0.5 to 2.0 is redundant given the results of the other tests. Full results of this and later tests are contained in the Appendix.

Varying dose of orange juice supplement

```
#Select orange juice entries
OJ <- subset(ToothGrowth, supp == "OJ")

#t-test for doses of 0.5 and 1.0
OJt1 <- t.test(OJ$len[OJ$dose == 1.0], OJ$len[OJ$dose == 0.5], paired = FALSE)
OJt1$p.value; OJt1$conf.int[1:2]
```

```
## [1] 8.784919e-05
```

```
## [1]  5.524366 13.415634
```

```
#t-test for doses of 1.0 and 2.0
OJt2 <- t.test(OJ$len[OJ$dose == 2.0], OJ$len[OJ$dose == 1.0], paired = FALSE)
OJt2$p.value; OJt2$conf.int[1:2]
```

```
## [1] 0.03919514
```

```
## [1] 0.1885575 6.5314425
```

There is clearly a significant difference in length between the dose of 0.5 and 1.0, with a p-value of 8.78e-05, with a higher dose resulting in increased length. The relationship is less clear for a dose between 1.0 and 2.0, with a p-value of 0.0392. While below 0.05, in the context of the multiple comparisons, a Bonferroni correction would make it no longer significant.

Varying dose of vitamin C supplement

```
#Select vitamin C entries
VC <- subset(ToothGrowth, supp == "VC")
```

```
#t-test for doses of 0.5 and 1.0
VCt1 <- t.test(VC$len[VC$dose == 1.0], VC$len[VC$dose == 0.5], paired = FALSE)
VCt1$p.value; VCt1$conf.int[1:2]
```

```
## [1] 6.811018e-07
```

```
## [1]  6.314288 11.265712
```

```
#t-test for doses of 1.0 and 2.0
VCt2 <- t.test(VC$len[VC$dose == 2.0], VC$len[VC$dose == 1.0], paired = FALSE)
VCt2$p.value; VCt2$conf.int[1:2]
```

```
## [1] 9.155603e-05
```

```
## [1]  5.685733 13.054267
```

Both of the comparisons for the vitamin C supplement dosage demonstrate significant differences, with p-values of 6.81e-07 and 9.16e-05. A higher dose results in longer tooth length.

Growth v Supplement type

I then examined differences in the supplement type. There is little value in comparing across all doses aggregated together for the two supplement types, so I restricted the analysis to the highest dose levels.

```
OJVC1 <- t.test(VC$len[VC$dose == 2.0], OJ$len[OJ$dose == 2.0],
                paired = FALSE, var.equal = FALSE)
OJVC1$p.value; OJVC1$conf.int[1:2]
```

```
## [1] 0.9638516
```

```
## [1] -3.63807  3.79807
```

The comparison between the dose of 2.0 for each supplement type shows no significant difference, with a p-value of 0.964. I also compared the dose of 2.0 using vitamin C with a dose of 1.0.

```
OJVC2 <- t.test(VC$len[VC$dose == 2.0], OJ$len[OJ$dose == 1.0],
                paired = FALSE, var.equal = FALSE)
OJVC2$p.value; OJVC2$conf.int[1:2]
```

```
## [1] 0.09652612
```

```
## [1] -0.6843336  7.5643336
```

If there are cost or other negative consequences to higher dose, the ability to deliver a lower dose might be of value. The result of the comparison is that there is not a significant difference between the two, with a p-value of 0.097. However, this comparison could be underpowered.

Conclusion and assumptions

There is a positive relationship between dose and tooth length, with increasing dose generally increasing length regardless of supplement type. However, the evidence for increasing length for doses above 1.0 delivered by orange juice is weak.

This analysis is based on the assumption of random assignment of treatment and doses to the guinea pigs, a normal distribution of tooth lengths. I made no assumption about equality of variance between supplement and dose types.

Appendix

Varying dose of orange juice supplement

```
OJt1
```

```
##
##  Welch Two Sample t-test
##
## data:  OJ$len[OJ$dose == 1] and OJ$len[OJ$dose == 0.5]
## t = 5.0486, df = 17.698, p-value = 8.785e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   5.524366 13.415634
## sample estimates:
## mean of x mean of y
##     22.70     13.23
```

```
OJt2
```

```
##
##  Welch Two Sample t-test
##
## data:  OJ$len[OJ$dose == 2] and OJ$len[OJ$dose == 1]
## t = 2.2478, df = 15.842, p-value = 0.0392
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.1885575 6.5314425
## sample estimates:
## mean of x mean of y
##     26.06     22.70
```

Varying dose of vitamin C supplement

```
VCt1
```

```
##
##  Welch Two Sample t-test
##
## data:  VC$len[VC$dose == 1] and VC$len[VC$dose == 0.5]
## t = 7.4634, df = 17.862, p-value = 6.811e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   6.314288 11.265712
## sample estimates:
## mean of x mean of y
##     16.77      7.98
```

```
VCt2
```

```
##
##  Welch Two Sample t-test
##
## data:  VC$len[VC$dose == 2] and VC$len[VC$dose == 1]
## t = 5.4698, df = 13.6, p-value = 9.156e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
```

```
##   5.685733 13.054267
## sample estimates:
## mean of x mean of y
##     26.14     16.77
```

Growth v Supplement type

```
OJVC1
```

```
##
##  Welch Two Sample t-test
##
## data:  VC$len[VC$dose == 2] and OJ$len[OJ$dose == 2]
## 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:
##  -3.63807  3.79807
## sample estimates:
## mean of x mean of y
##     26.14     26.06
```

```
OJVC2
```

```
##
##  Welch Two Sample t-test
##
## data:  VC$len[VC$dose == 2] and OJ$len[OJ$dose == 1]
## t = 1.7574, df = 17.297, p-value = 0.09653
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.6843336  7.5643336
## sample estimates:
## mean of x mean of y
##     26.14     22.70
```

Session Info

```
sessionInfo()
```

```
## R version 3.3.2 (2016-10-31)
## Platform: x86_64-apple-darwin15.6.0 (64-bit)
## Running under: macOS Sierra 10.12.3
##
## locale:
## [1] en_AU.UTF-8/en_AU.UTF-8/en_AU.UTF-8/C/en_AU.UTF-8/en_AU.UTF-8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] ggplot2_2.1.0
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_0.12.7      digest_0.6.10    rprojroot_1.2    plyr_1.8.4
##  [5] grid_3.3.2       gtable_0.2.0     backports_1.0.5  formatR_1.4
```

```
##  [9] magrittr_1.5     scales_0.4.0     evaluate_0.10    stringi_1.1.2
## [13] reshape2_1.4.2   rmarkdown_1.3    labeling_0.3     RevoUtils_10.0.2
## [17] tools_3.3.2      stringr_1.1.0    munsell_0.4.3    yaml_2.1.13
## [21] colorspace_1.2-7 htmltools_0.3.5  knitr_1.14
```