Effect of Vitamin C on Tooth Growth in Guinea Pigs - Exploratory & Inferential Analyses

Connor Gooding

8/18/2017

Overview

This report employs exploratory and inferential data analyses on the ToothGrowth dataset in R. The given data describes the effect of Vitamin C on tooth growth on guinea pigs varied by supplement type (Vitamin C or Orange Juice).

Data Summary and Assumptions

```
data(ToothGrowth)
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  
```

Taking an initial look at the structure of the raw dataset, the data is comprised of 60 observations of 3 variables:

```
* len: Tooth length
* supp: Type of supplement used (either Vitamin C or Orange Juice)
* dose: Dosage administered to the subject, in mg/day
```

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

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

Upon further observation, it becomes apparent that only three dosage levels were administered: 0.5 mg/day, 1.0 mg/day, and 2.0 mg/day.

It is assumed throughout this report that the data is normal but of relatively small sample size, meaning that the T-test will be used for hypothesis testing of these variables.

Exploratory Analysis

An initial look at length vs. dosage suggests that there is a positive relationship between the variables for both orange juice and Vitamin C supplements, with a stronger increase in length occurring as Vitamin C increases in dosage.

From both plots, it is evident that six subsets of the data, representing each possible combination of supplement type and dosage, can be extracted and tested against each other. Their means and standard deviations are summarized below.

```r
VC <- subset(ToothGrowth, supp == "VC")
OJ <- subset(ToothGrowth, supp == "OJ")

VCdose <- split(VC, VC$dose)
OJdose <- split(OJ, OJ$dose)

sapply(VCdose, function(x) c(round(mean(x$len), 3), round(sd(x$len), 3)))
```

```
##         0.5      1      2
## [1,] 7.980 16.770 26.140
## [2,] 2.747  2.515  4.798
```

```r
sapply(OJdose, function(x) c(round(mean(x$len), 3), round(sd(x$len), 3)))
```

```
##         0.5      1      2
## [1,] 13.23 22.700 26.060
## [2,]  4.46  3.911  2.655
```

The following hypotheses can be drawn from the exploratory analysis:

```
1. Tooth length has a positive relationship with dosage level for both Vitamin C and
   ascorbic acid.
2. Orange juice has a greater impact on tooth length than Vitamin C.
```

Inferential Analysis - Hypothesis Testing

Considering the data can be neatly partitioned into six approximately normal subsets across two supplement types and three dosage levels, the hypothesis tests taken for inferential analysis will assess the significance of mean difference across supplement type (holding dosage level constant) and dosage level (holding supplement type constant), respectively. The null hypothesis for each of these tests is that the difference between the means of each subset is zero ($\mu\ 1 = \mu\ 2$).

T-Testing Variation in Dosage Levels

Comparing the Vitamin C subsets to each other via T-test returns p-values << .05. 0.5 mg was compared with 1.0 mg, and 1.0m was compared with 2.0 mg. Since both tests yielded appropriate p-values to reject the null hypotheses ($\alpha = 0.5$), it follows that a test between 0.5 mg and 2.0 mg would yield the same end result.

```r
c("0.5 vs. 1.0:", t.test(VCdose[[1]]$len, VCdose[[2]]$len, var.equal = TRUE,
                         alternative = "less")$p.value)
```

```
## [1] "0.5 vs. 1.0:"         "3.24613229907883e-07"
```

```r
c("1.0 vs 2.0:", t.test(VCdose[[2]]$len, VCdose[[3]]$len, var.equal = TRUE,
                        alternative = "less")$p.value)
```

```
## [1] "1.0 vs 2.0:"          "1.69878896276978e-05"
```

A similar conclusion is drawn with the Orange Juice subsets. While both tests allow for rejection of the null hypothesis for $\alpha = .05$, it should be noted that the T-test between 1.0 mg and 2.0 mg for Orange Juice only yields a p-value $p = .0187$, unlike the other three tests, which had p-values $<< 10^{-3}$.

```r
c("0.5 vs. 1.0:", t.test(OJdose[[1]]$len, OJdose[[2]]$len, var.equal = TRUE,
                         alternative = "less")$p.value)
```

```
## [1] "0.5 vs. 1.0:"         "4.17877964072189e-05"
```

```
c("1.0 vs 2.0:", t.test(OJdose[[2]]$len, OJdose[[3]]$len, var.equal = TRUE,
                        alternative = "less")$p.value)
```

```
## [1] "1.0 vs 2.0:"       "0.018681397928322"
```

T-Testing Variation in Supplement Type

Three T-tests are needed to assess variation in supplement type, one per dosage level. The tests reject their null hypotheses for dosages of 0.5 and 1.0 mg, but the test between Vitamin C and Orange Juice at 2.0 mg dosage fails to reject the null hypothesis ($\alpha = .05$).

```
c("0.5:", t.test(VCdose[[1]]$len, OJdose[[1]]$len, alternative = "less")$p.value)
```

```
## [1] "0.5:"               "0.0031793033820484"
```

```
c("1.0", t.test(VCdose[[2]]$len, OJdose[[2]]$len, alternative = "less")$p.value)
```

```
## [1] "1.0"                  "0.000519187936149939"
```

```
c("2.0", t.test(VCdose[[3]]$len, OJdose[[3]]$len, alternative = "less")$p.value)
```

```
## [1] "2.0"                "0.518074205638314"
```

Conclusion

Based on the results of the T-tests across both dimensions, it is reasonable to conclude that, for both concentrated ascorbic acid and Orange Juice as the supplement type, dosage of Vitamin C has a positive relationship with tooth length in guinea pigs. It is also reasonable to conclude that at smaller doses, orange juice yields more tooth growth than ascorbic acid. However, the same conclusion cannot be drawn for larger doses of either supplement.

These conclusions are based on the assumptions that with more variation in dosage level, the trends observed could be extrapolated. That is, it is assumed by these conclusions that as more dosage levels were tried (both less than 0.5 mg and greater than 2.0 mg), greater tooth growth would be observed in orange juice than ascorbic acid for all dosages less than 2.0 mg.

There is significant room for error by making those assumptions. Therefore, it is not advised to operate on these conclusions without testing a wider range of dosage levels and running further tests.

Appendix

Figure 1

```
ggplot( data = ToothGrowth, aes(x = factor(dose), y = len, fill = factor(dose))) +
geom_boxplot() +
facet_grid(.~supp) + xlab("Dosage") + ylab("Tooth Length") +
ggtitle("Tooth Length vs. Dosage (mg/day) by Supplement Type")
```