

Statistical Inference Project

Matthew Henderson

14 February 2021

Inferential Data Analysis of ToothGrowth Dataset

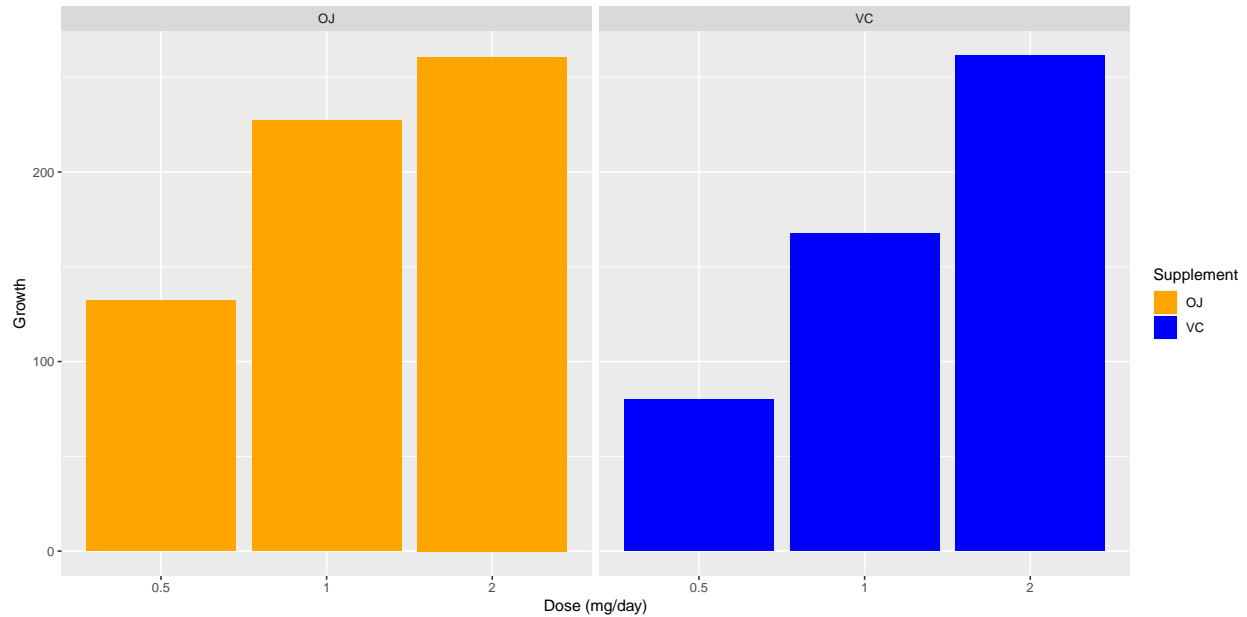
The ToothGrowth dataset from the datasets library contains data on the effect of Vitamin C on tooth growth in Guinea Pigs. The output is the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. There were two different delivery methods for the vitamin C, orange juice or ascorbic acid and three different doses (0.5, 1 or 2mg/day). Each Guinea pig recieved vitamin C by one of the two delivery methods and one of the three doses.

```
library(datasets)
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 ...
```

The question is does the method of delivery or dose of Vitamin C effect the length of growth?

```
library(dplyr)
library(ggplot2)
sumTG <- ToothGrowth %>% group_by(supp, dose) %>% summarise(Growth = sum(len))
g5 <- ggplot(sumTG, aes(x = as.factor(dose), y = Growth, fill = supp)) + geom_bar(stat = "identity")
g5 <- g5 + facet_grid(. ~ supp) +
  scale_fill_manual(values = c("orange", "blue")) +
  labs(x = "Dose (mg/day)", y = "Growth", fill = "Supplement")
g5
```



It looks like the Orange juice is a more effective method of delivery for doses of 0.5 and 1mg/day however there does not seem to be much of a difference for 2mg/day.

Hypothesis Testing

We want to see if there is any statistical evidence of a difference in lengths between the two delivery methods for Vitamin C. To do this we split the data into two groups for the guinea pigs that had orange juice and those that had ascorbic acid. The null hypothesis will be that the mean difference between these two groups is 0 and the alternative hypothesis is that the mean difference is not equal to zero. Hence:

$$H_0 : \mu = 0$$

versus:

$$H_a : \mu > 0$$

where μ is the mean difference between the two groups. We will assume this sample mean follows a t-distribution with $n_1 + n_2 - 2 = 58$ degrees of freedom ($\alpha = 58$). The two groups are independent of each other so this is not a paired t-test.

```
g1 <- filter(ToothGrowth, supp == "OJ")$len
g2 <- filter(ToothGrowth, supp == "VC")$len
t.test(g1, g2, paired = FALSE, var.equal = FALSE)$p.value
```

```
## [1] 0.06063451
```

```
t.test(g1, g2, paired = FALSE, var.equal = FALSE)$conf.int
```

```
## [1] -0.1710156 7.5710156
## attr("conf.level")
## [1] 0.95
```

The output of the test tells us that we cannot reject the null hypothesis with 95% confidence since the 95% confidence value for the mean contains 0.

Our exploratory analysis, seems to show that the 2mg/day dose seems to be an outlier in the dataset so we might be interested to see what happens if we split the 2mg/day group from the 0.5mg and 1mg/day groups.

```
# split the 0.5mg and 1mg doses from the 2mg dose
lowDose1 <- g1[1:20]
lowDose2 <- g2[1:20]
t.test(lowDose1, lowDose2, paired = FALSE, var.equal = FALSE)$p.value
```

```
## [1] 0.00423861
```

```
t.test(lowDose1, lowDose2, paired = FALSE, var.equal = FALSE)$conf
```

```
## [1] 1.875234 9.304766
## attr("conf.level")
## [1] 0.95
```

```
# test the 2mg dose group
highDose1 <- g1[21:30]
highDose2 <- g2[21:30]
t.test(highDose1, highDose2, paired = FALSE, var.equal = FALSE)$p.value
```

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

```
t.test(highDose1, highDose2, paired = FALSE, var.equal = FALSE)$conf
```

```
## [1] -3.79807 3.63807
## attr("conf.level")
## [1] 0.95
```

With a p-value of ($p = 0.01$) we can reject the null hypothesis ($H_0 : \mu = 0$) and accept the alternative hypothesis ($H_a : \mu > 0$) that the orange juice delivery method creates a larger tooth growth for doses 0.5mg and 1mg/day. However, we cannot reject the null hypothesis when the dose is 2mg/day.

Conclusion

The tests show that we can be confident that the orange juice delivery method is a more effective supplement of Vitamin C than the ascorbic acid for smaller doses of 0.5mg and 1mg/day. However, when the dose is increased to 2mg/day there is no statistical evidence of a difference between the two delivery methods.