Statistical Analysis: Effect of Vitamin C on Tooth Growth

Thomas Bell

7 October 2017

Overview

This analysis looks at the effect of vitamin C on tooth growth in Guinea pigs.

The response is the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, (orange juice or ascorbic acid (a form of vitamin C and coded as VC).

Exploratory Analysis

The ToothGrowth data set is loaded into the runtime memory of R.

```
data("ToothGrowth")
dim(ToothGrowth)
```

```
## [1] 60  3
```

```
summary(ToothGrowth)
```

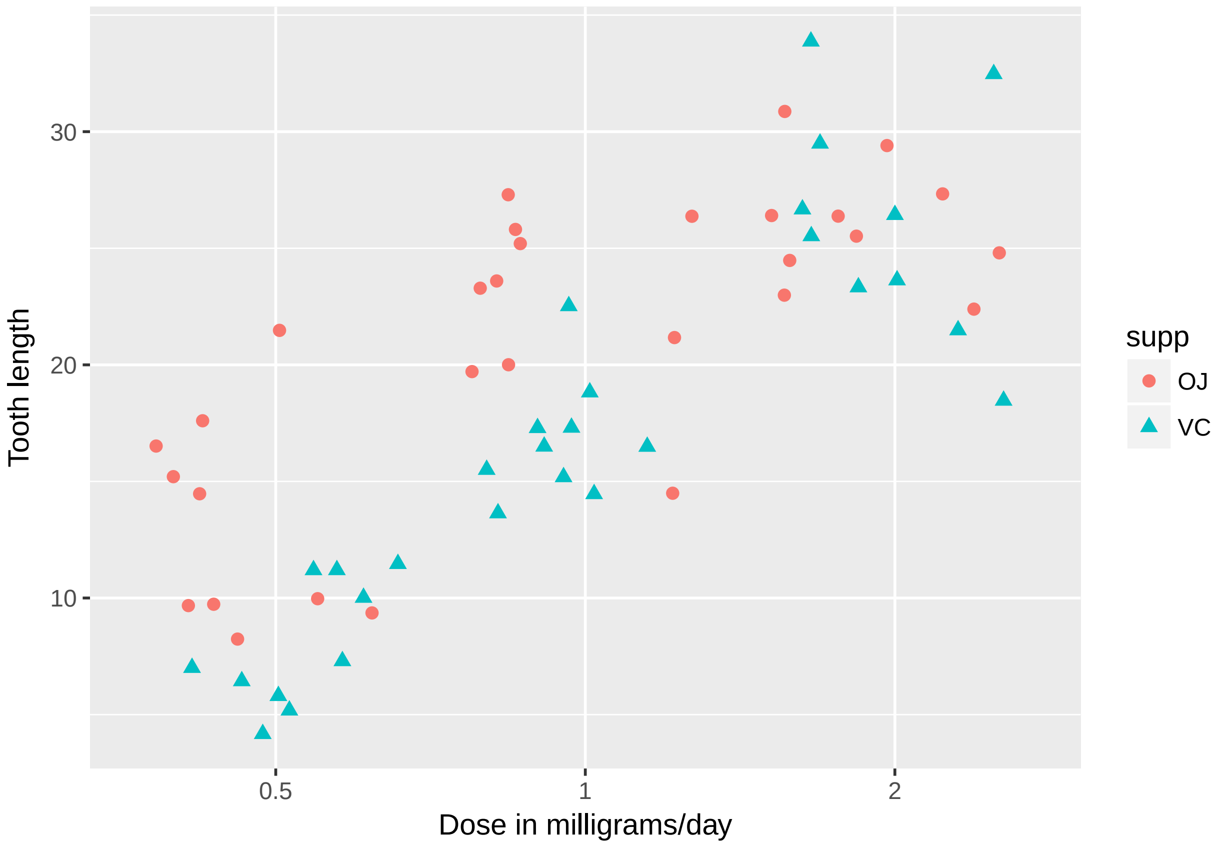
```
##       len        supp         dose      
##  Min.   : 4.20   OJ:30   Min.   :0.500  
##  1st Qu.:13.07   VC:30   1st Qu.:0.500  
##  Median :19.25           Median :1.000  
##  Mean   :18.81           Mean   :1.167  
##  3rd Qu.:25.27           3rd Qu.:2.000  
##  Max.   :33.90           Max.   :2.000  
```

Descriptions of the variables in the data set:

- len: Tooth length
- supp: Supplement type (VC or OJ).
- dose: numeric Dose in milligrams/day

Exploratory analysis; plotting the tooth length against the various doses of orange juice (OJ) and asorbic acid (VC) shows that tooth lenght increases with higher doses of supplements containing vitamin C. The red dots have the upper hand in the plot, which hints that OJ has a greater effect on tooth growth than VC. This will be further analysed through hypothesis testing.

```
ggplot(ToothGrowth, aes(x = as.factor(dose), y = len, col = supp, shape = supp)) +
  geom_jitter(stat="identity", size = 2) +
  labs(x = "Dose in milligrams/day", y = "Tooth length")
```



Hypothesis Tests

In this section we will validate the exploratory analysis results by calculating a t-confidence interval for independent groups. Independent because the Guinea pigs were divided in 6 groups, 3 for each dose, times two for OJ and VC.

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

The first test compares the average length of all doses together. Parameters for the tests:

$\mu OJ = 20.663$, $\sigma OJ = 6.606$

$\mu VC = 16.963$, $\sigma VC = 8.266$

$\alpha = 0.05$

$n = 30$

We assume the status quo that the average tooth growth is the same for the OJ as for the VC group.

$H_0 : \mu OJ = \mu VC$

our alternative hypothesis is that OJ has a greater effect on tooth growth than VC

$H_a : \mu OJ > \mu VC$

The exploratory analysis showed that the OJ data points are above VC, therefore we indicate a direction in the alternative hypothesis (greater than) and perform a single sided t-test. The sigmas of both groups indicate that the variances are not equal.

```r
t.test(OJ$len, VC$len, paired = FALSE, var.equal = FALSE, alternative = "greater")
```

```
##
##  Welch Two Sample t-test
##
## data:  OJ$len and VC$len
## t = 1.9153, df = 55.309, p-value = 0.03032
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.4682687       Inf
## sample estimates:
## mean of x mean of y
##  20.66333  16.96333
```

This test shows that H_0 should be rejected. The p-value of 0.03 is smaller than the confidence level of 5%. This means there is a 3% chance that the status quo is true. This is considered insignificant. The alternative hypothesis is accepted: OJ has a greater effect on tooth growth than VC.

In the next series of tests we repeat the same t-test but then for the different doses (2.0, 1.0, 0.5 mg/day). The results are displayed in a summary table.

```r
subset(VC, VC$dose == 2.0) -> VC_high
subset(VC, VC$dose == 1.0) -> VC_med
subset(VC, VC$dose == 0.5) -> VC_low

subset(OJ, OJ$dose == 2.0) -> OJ_high
subset(OJ, OJ$dose == 1.0) -> OJ_med
subset(OJ, OJ$dose == 0.5) -> OJ_low

t.test(OJ_high$len, VC_high$len,
       paired = FALSE,
       var.equal = FALSE,
       alternative = "greater") -> t_high
t.test(OJ_med$len, VC_med$len,
       paired = FALSE,
       var.equal = FALSE,
       alternative = "greater") -> t_med
t.test(OJ_low$len, VC_low$len,
       paired = FALSE, var.equal = FALSE,
       alternative = "greater") -> t_low

Dose <- c("2.0 mg/day","1.0 mg/day", "0.5 mg/day")
p_value <- c(round(t_high$p.value,3),round(t_med$p.value,3),round(t_low$p.value,3))
alpha <- 0.05
Conclusion <- c("Accept $H_0$ : p-value greater than confidence level; 5,18%",
                "Reject $H_0$ : p-value smaller than confidence level; accept $H_a$",
                "same as above")
stat_table <- data.frame(Dose,p_value,alpha,Conclusion)
kable(stat_table)
```

Dose	p_value	alpha	Conclusion
2.0 mg/day	0.518	0.05	Accept H_0 : p-value greater than confidence level; 5,18%
1.0 mg/day	0.001	0.05	Reject H_0 : p-value smaller than confidence level; accept H_a
0.5 mg/day	0.003	0.05	same as above

Conclusions

The scatter plot in the exploratory analysis section gave the indication that orange juice (OJ) on average has a greater effect on tooth growth of Guinea pigs than asorbic acid (VC).

When observing the treatment doses individually, it seems that OJ is greater than VC for 0.5 and 1.0 mg/day. For 2.0 mg/day it cannot be seen clearly as VC seems to have a few outliers in both directions.

Through hypothesis testing and confidence intervals the following conclusions were drawn:

- Overall OJ has greater effect on tooth growth than VC
- With a dose of 2.0 mg/day there is no significant evidence that OJ has a greater effect than VC, therefore the status quo remains valid for this dose
- With a dose of 1.0 mg/day it is very likely that OJ has a greater effect on tooth growth than VC
- With a dose of 0.5 mg/day it is very likely that OJ has a greater effect on tooth growth than VC