Linear Regression

Andresa de Andrade

August 15, 2014

Contents

1 Summary of the data

We'll be working with 60 observations of 10 guinea pigs. In this project we are interested to know if there's difference in length of the teeth for different levels of vitamin C and 2 deliver methods: orange juice and ascorbic acid.

2 Exploratory Analysis

We need to check the 6 combinations of data (by supplement and by dose)

Please consider Ascorbic Acid as VC and Orange Juice as OJ

Table 1: Summary Table

Statistic	N	Mean	St. Dev.	Min	Max
Toothlength	60	18.813	7.649	4.200	33.900
Orange Juice	30	20.660	6.605	8.200	30.900
Ascorbic Acid	30	16.960	8.266	4.200	33.900
dose = 0.5	20	10.605	4.499	4.200	21.500
dose = 1	20	19.730	4.415	13.60	27.300
dose = 2	20	26.100	3.774	18.50	33.900
VC-dose = 0.5	10	07.980	2.746	4.200	11.500
VC-dose = 1	10	16.770	2.515	13.60	22.500
VC-dose = 2	10	26.140	4.797	18.50	33.900
OJ-dose = 0.5	10	13.230	4.459	8.200	21.500
OJ-dose = 1	10	22.700	3.911	14.50	27.300
OJ-dose = 2	10	26.060	2.655	22.40	30.900

- From the table above and the Figure 2 we can infer that the average of teeth length between guinea pigs that were fed with orange juice is slightly higher than the ones fed with ascorbic acid, also the variability is lower for orange juice. But we can't really say that there's a significant difference.
- When we look at different doses of vitamin C we have a a increase of teeth length as the doses are higher and a decrease in variance as the doses get bigger.
- And when check the combination between type of supplement and vitamin C doses, we have, again, a better performance of Orange Juice. For the first dose orange juice means is 60% higher than VC and Sd is 60% lower than VC (the lower the sd the better). Even though for dose =2 we have a higher average for VC, the variance is almost twice as for the orange juice, what suggests outliers or the data is more disperse around the mean as we can see in the boxplot.
- Normality Assumption : from the Figure 1, we can't really infer that we have a normal distribution because the presence of outliers is very strong in this model. Thus for tests purposes, we'll be using the T-student distribution.

3 Statistics Tests

Now that we have an idea of how data is distributed, lets check if we can confirm with a statistic test. This is the test that we'll be applying:

$H_0 : \mu_{orangejuice} = \mu_{ascorbicacid}$ versus $H_1 : \mu_{orangejuice} \neq \mu_{ascorbicacid}$

Note: we don't have a paired test because we have different guinea pigs tooth.

The p-value for difference in types of supplement(vc and oj) is greater than 0.05, which suggests that there's no difference between orange juice and ascorbic acid.

Let's test if there's difference between the doses of vitamin C:

According to the Figure 5, there's difference between treatment at 0.5 mg and 1mg. And the CI for 95% of confidence is [-11.984, -6.276]. So if the difference between means was within this interval (considering absolute numbers), we could have inferred that there's no difference between treatments.

According to the Figure 6, we also have the difference between treatments from doses at 0.5 and at 1 not significant (p-value <0.01).

Now to check if there's difference between treatment at dose = 1 and at dose = 2 we go to Figure 7 and observe that the p-value is also very low, so we can reject the null hypothesis of no difference between the the treatments.

4 #Appendix

Figure 1

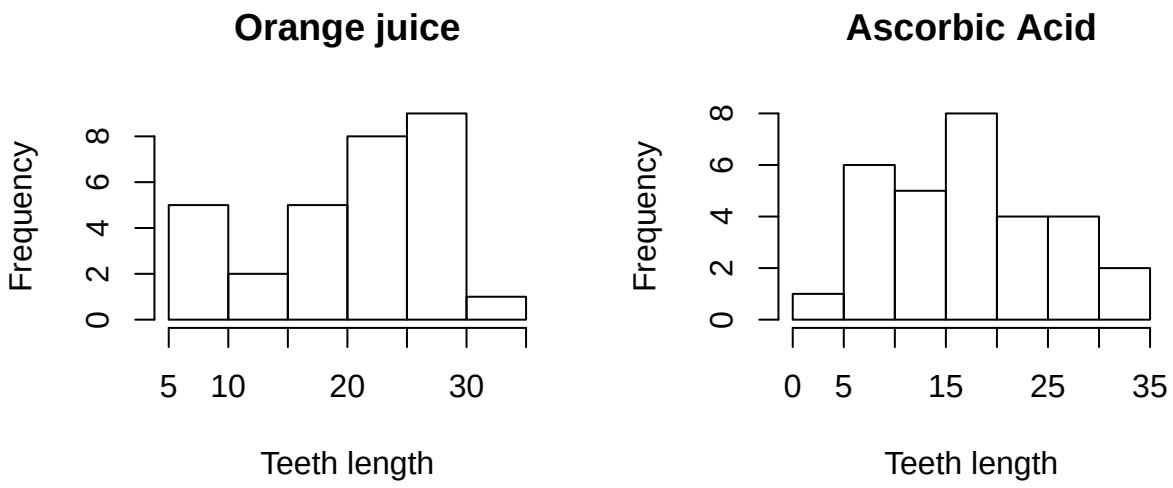

Figure 1:

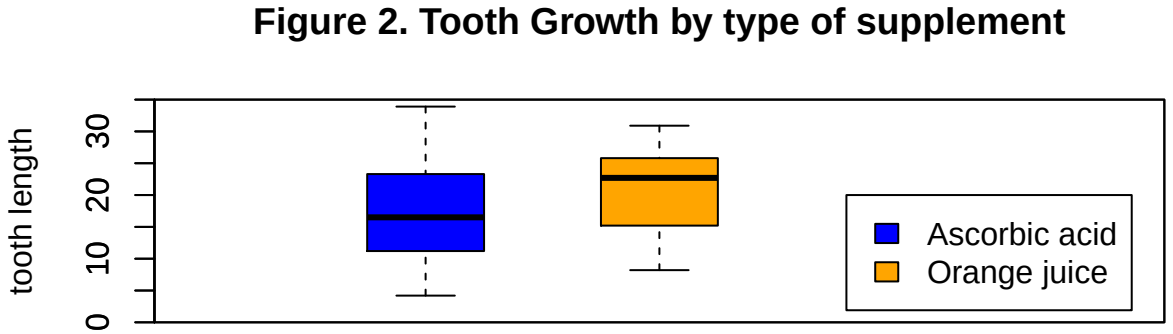

Figure 2:

Figure 4

```
##
##  Welch Two Sample t-test
##
## data:  vc_subset$len and oj_subset$len
## t = -1.9153, df = 55.309, p-value = 0.06063
```

```
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -7.5710156  0.1710156
## sample estimates:
## mean of x mean of y
##  16.96333  20.66333
```

Figure 5 - Dose: 0.5 x 1

```
##
##  Welch Two Sample t-test
##
## data:  dose_05$len and dose_1$len
## t = -6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.983781  -6.276219
## sample estimates:
## mean of x mean of y
##    10.605    19.735
```

Figure 6 - Dose: 0.5 x 2

```
##
##  Welch Two Sample t-test
##
## data:  dose_05$len and dose_2$len
## t = -11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.15617 -12.83383
## sample estimates:
## mean of x mean of y
##    10.605    26.100
```

Figure 7 - Dose: 1 x 2

```
##
##  Welch Two Sample t-test
##
## data:  dose_2$len and dose_1$len
## t = 4.9005, df = 37.101, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  3.733519 8.996481
## sample estimates:
## mean of x mean of y
##    26.100    19.735
```

Figure 8 - Dose: 0.5 Orange Juice x Ascorbic Acid

```
##
##  Welch Two Sample t-test
##
## data:  vc_05$len and oj_05$len
## t = -3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.780943 -1.719057
## sample estimates:
## mean of x mean of y
##      7.98     13.23
```

Figure 9 - Dose: 1 Orange Juice x Ascorbic Acid

```
##
##  Welch Two Sample t-test
##
## data:  vc_1$len and oj_1$len
## t = -4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -9.057852 -2.802148
## sample estimates:
## mean of x mean of y
##     16.77     22.70
```

Figure 10 - Dose: 2 Orange Juice x Ascorbic Acid

```
##
##  Welch Two Sample t-test
##
## data:  vc_2$len and oj_2$len
## 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:
##  -3.63807  3.79807
## sample estimates:
## mean of x mean of y
##     26.14     26.06
```