Analysis of Tooth Growth Data

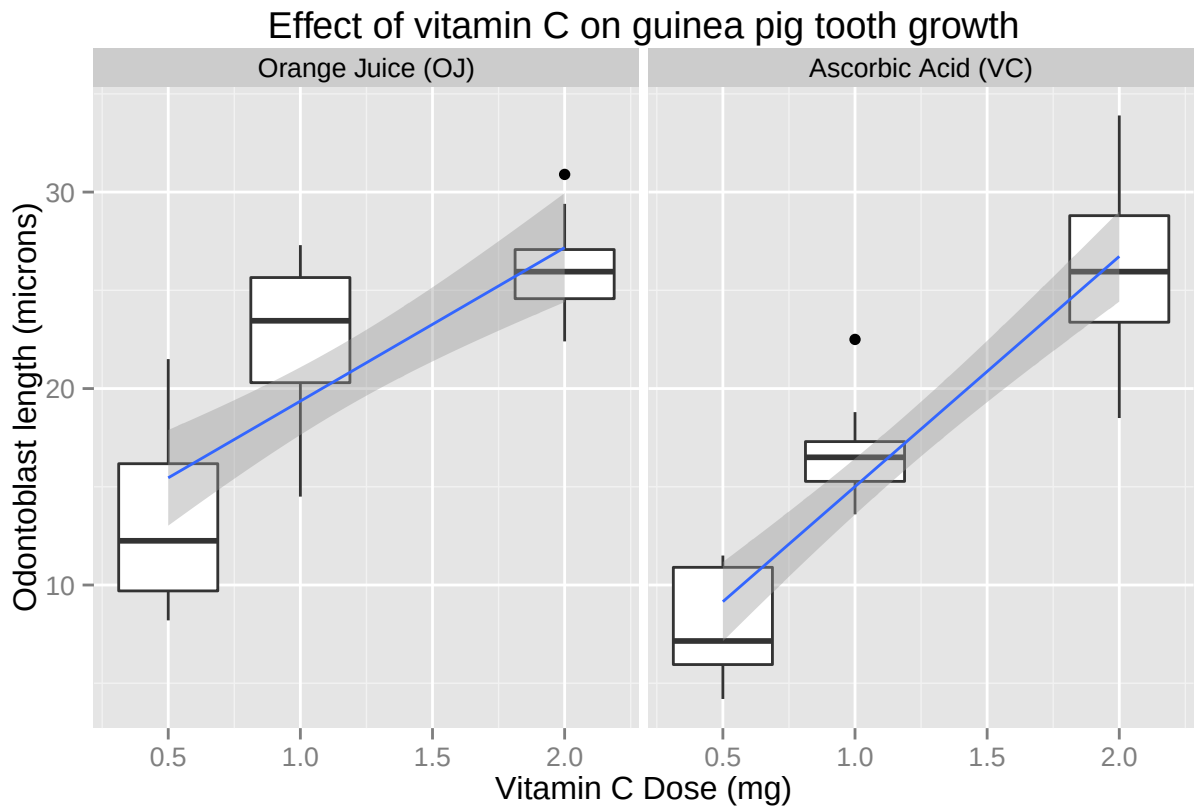
Chris Lill

Saturday, June 20, 2015

Introduction

The Tooth Growth data set assesses the effect of Vitamin C on the growth of the Guinea Pig teeth, specifically the odontoblast cells at the end of the incisor tooth. There are sixty measurements, with groups of ten guinea pigs measured across three doses of Vitamin C (0.5, 1, and 2 mg) and two delivery methods - orange juice (OJ) or ascorbic acid (VC). The study was carried out for the Canadian Government during the Second World War to assess alternative sources of Vitamin C for the armed forces, and is documented in http://jn.nutrition.org/content/33/5/491.full.pdf.

Exploratory analysis



The graphs above show that although each sample size is small, there is a clear relationship between tooth growth and vitamin C doses at these levels. Inspection of the confidence interval for the best fit line (using linear regression) suggests that it is possible that the two delivery methods could be equally effective. This analysis will investigate both hypotheses using statistical methods.

Confidence intervals for tooth growth by dosage

A simple hypothesis is that vitamin C has an effect on tooth growth (L), with the null hypothesis being that it has no effect.

$$H_0 : L = 0$$
$$H_a : L \neq 0$$

This analysis will calculate 95% confidence intervals for the expected odontoblast length at each dose level. This will be calculated regardless of delivery method, and will be calculated using a two-sided T distribution with 19 degrees of freedom.

Assumptions

1. **The T distribution assumes that the underlying data are iid Gaussian.** This is supported by the plots in the appendix which show that the data is roughly symmetric and mound shaped.
2. **Orange juice and ascorbic acid data can be grouped together.** Even though the end of this paper shows that ascorbic acid is not as effective as orange juice at low doses, this is still a reasonable assumption given the similarity of the data. More advanced statistical techniques may be available to address this assumption.
3. **A two-sided T distribution is appropriate.** This conveniently provides us with a confidence interval for the mean for each dose. It is a conservative approach compared with a one-sided test that we could also use to answer whether vitamin C has a positive effect on tooth growth.

Analysis

```
ci.by.dose <- ToothGrowth %>%
    mutate(dose = factor(dose)) %>%
    group_by(dose) %>%
    summarise(lower.ci = round(t.test(len)$conf[1], 1),
              mean     = round(mean(len), 1),
              upper.ci = round(t.test(len)$conf[2], 1))
data.frame(ci.by.dose)
```

```
##   dose lower.ci mean upper.ci
## 1  0.5      8.5 10.6     12.7
## 2    1     17.7 19.7     21.8
## 3    2     24.3 26.1     27.9
```

Conclusion

The 95% confidence intervals for each dose do not contain the value zero, which shows that null hypothesis can be rejected.

For example, the confidence interval for tooth growth with a 2mg dose of vitamin C extends from 24.3 to 27.9 microns.

We can reject the null hypothesis. Therefore vitamin C has an effect on tooth growth.

Confidence intervals for tooth growth by delivery method (supp)

This analysis will investigate whether there is any significant difference between tooth growth (L), depending on whether the vitamin C is produced from Orange Juice (OJ) or Ascorbic Acid (VC).

$$H_0 : L_{OJ} - L_{VC} = 0$$
$$H_a : L_{OJ} - L_{VC} \neq 0$$

To investigate this, confidence intervals for $L_{OJ} - L_{VC}$ will be calculated for each of the dose levels using a two-sided T distribution.

Assumptions

1. **The T distribution assumes that the underlying data are iid Gaussian.** This is supported by the plots in the appendix which show that the data is roughly symmetric and mound shaped.
2. **The data for the different doses have unequal variances.** This is a conservative assumption which uses an approximation for the degrees of freedom in the T distribution.

Analysis

```
tg.oj <- filter(ToothGrowth, supp == "OJ")
tg.vc <- filter(ToothGrowth, supp == "VC")

# Take advantage of the dataset being ordered with a consistent sample size
# to create a single dataframe.
tg.by.supp <- data.frame(dose = tg.oj$dose,
                         len.oj = tg.oj$len,
                         len.vc = tg.vc$len)

# Calculate confidence intervals for each dose by carrying out a two sample
# t-test with paired = FALSE and var.equal = FALSE (both are default values)
ci.by.supp <- tg.by.supp %>%
    group_by(dose) %>%
    summarise(lower.ci = round(t.test(len.oj, len.vc)$conf[1], 2),
              upper.ci = round(t.test(len.oj, len.vc)$conf[2], 2))
data.frame(ci.by.supp)
```

```
##   dose lower.ci upper.ci
## 1  0.5     1.72     8.78
## 2  1.0     2.80     9.06
## 3  2.0    -3.80     3.64
```

Conclusions

The 95% confidence interval for a 2mg dose of vitamin C contains zero, therefore we fail to reject the null hypothesis. At 2mg, Ascorbic Acid is as effective as Orange Juice in supporting tooth growth.

The 95% confidence intervals for 1mg and 0.5mg doses of vitamin C do not contain zero, therefore we reject the null hypothesis. At lower doses, Ascorbic Acid is less effective than Orange Juice in supporting tooth growth.

Appendix: Are the samples approximately normal?

The T distribution is robust to the assumption that samples are iid gaussian. If the sample data is roughly symmetric and mound shaped, then this is a good approximation. These plots show the samples used in the confidence interval calculations. They are roughly symmetric and mound shaped, and support the use of the T distribution.