Statistical Inference Project Part 2

Rich Robinson

Friday, May 22, 2015

The original author's goal was to determine whether odontoblast cell length was a reliable test for Vitamin C intake. Can we answer this question?

The data consists of measurements of the mean length of the odontoblast cells harvested from the incisor teeth of a population of 60 guinea pigs. These animals were divided into 6 groups of 10 and consistently fed a diet with one of 6 Vitamin C supplement regimes for a period of 42 days. The Vitamin C was administered either in the form of Orange Juice (OJ) or chemically pure Vitamin C in aqueous solution (VC). Each animal received the same daily dosage of Vitamin C (either 0.5, 1.0 or 2.0 milligrams) consistently. Since each combination of supplement type and dosage was given to 10 randomly selected animals this required a total of 60 animals for the study. After 42 days, the animals were euthanized, their incisor teeth were harvested and subject to analysis via optical microscopy to determine the length (in microns) of the odontoblast cells (the layer between the pulp and the dentine). The ToothGrowth data set therefore consists of 60 observations of the 3 variables - mean length of odontoblasts (microns), supplement type (OJ or VC) and Vitamin C dosage (milligrams/day).

T-tests seem an obvious choice for the numerical analysis because the sample size of each group is small (10 subjects) with 6 groups, so only 60 independent subjects in total. We should avoid using the normal distribution and z-tests. If the sample size was much larger we could use a z-test to make things simpler and approximate the solution because we know by the Central Limit Theorem that as the sample size of each sub-population grows, the mean of each sub-population behaves more like a normal distribution.

All numerical/data manipulation was performed using the `dplyr` package in `R` .

Assumptions

When performing this data analysis of tooth growth in Guinea Pigs we have assumed that all the Guniea Pigs are individual and were sufficiently randomly chossen to represent a good sample of the population of all Guinea Pigs.
Also during the t-tests we have assumed that all the values are not paired (from individual subjects) and variances between the sample groups are not equal.

Understanding our dataset

```
library(datasets)

data(ToothGrowth)

dim(ToothGrowth)
```

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

```
head(ToothGrowth)
```

```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

```
unique(ToothGrowth$supp)
```

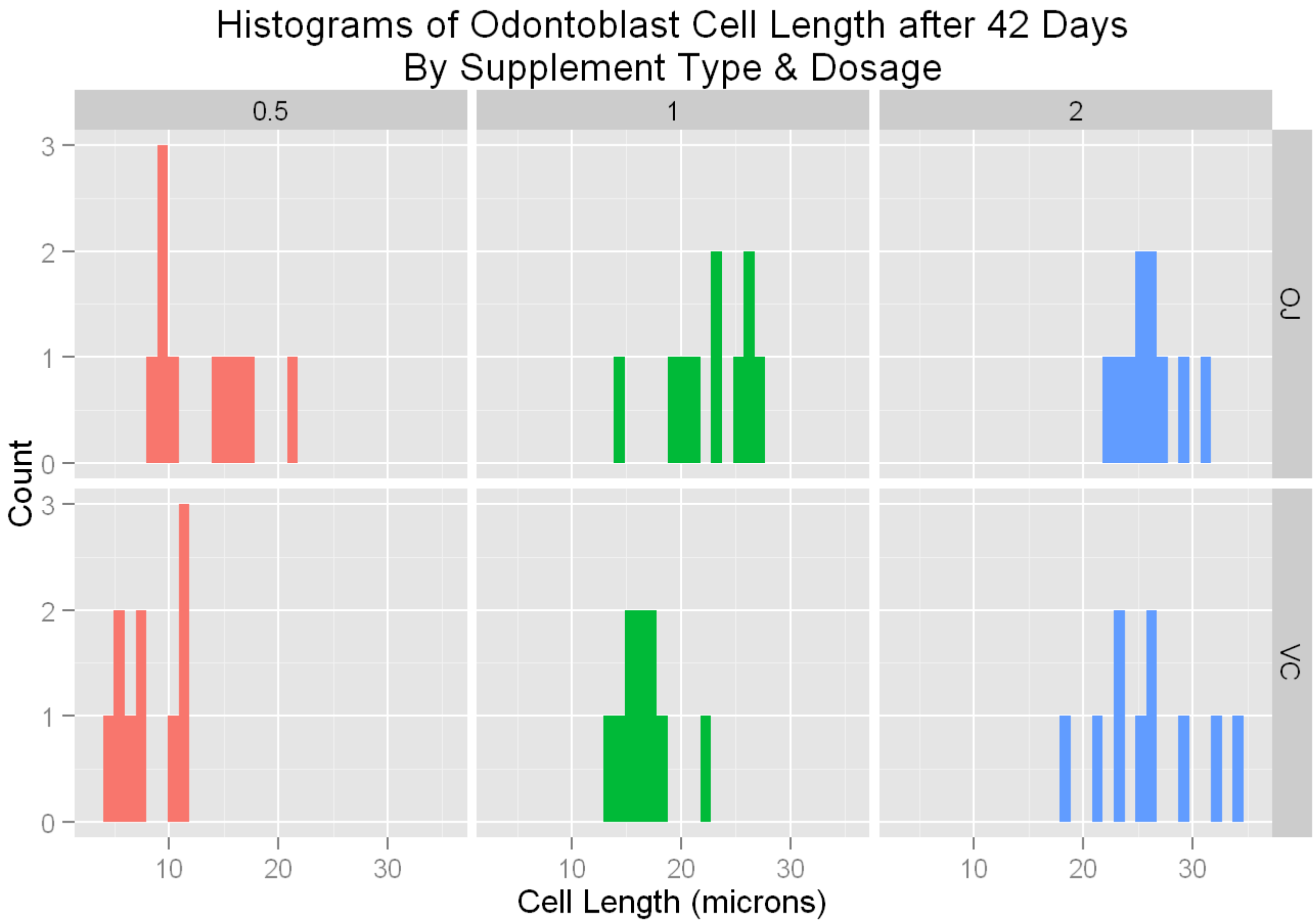
```
## [1] VC OJ
## Levels: OJ VC
```

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

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

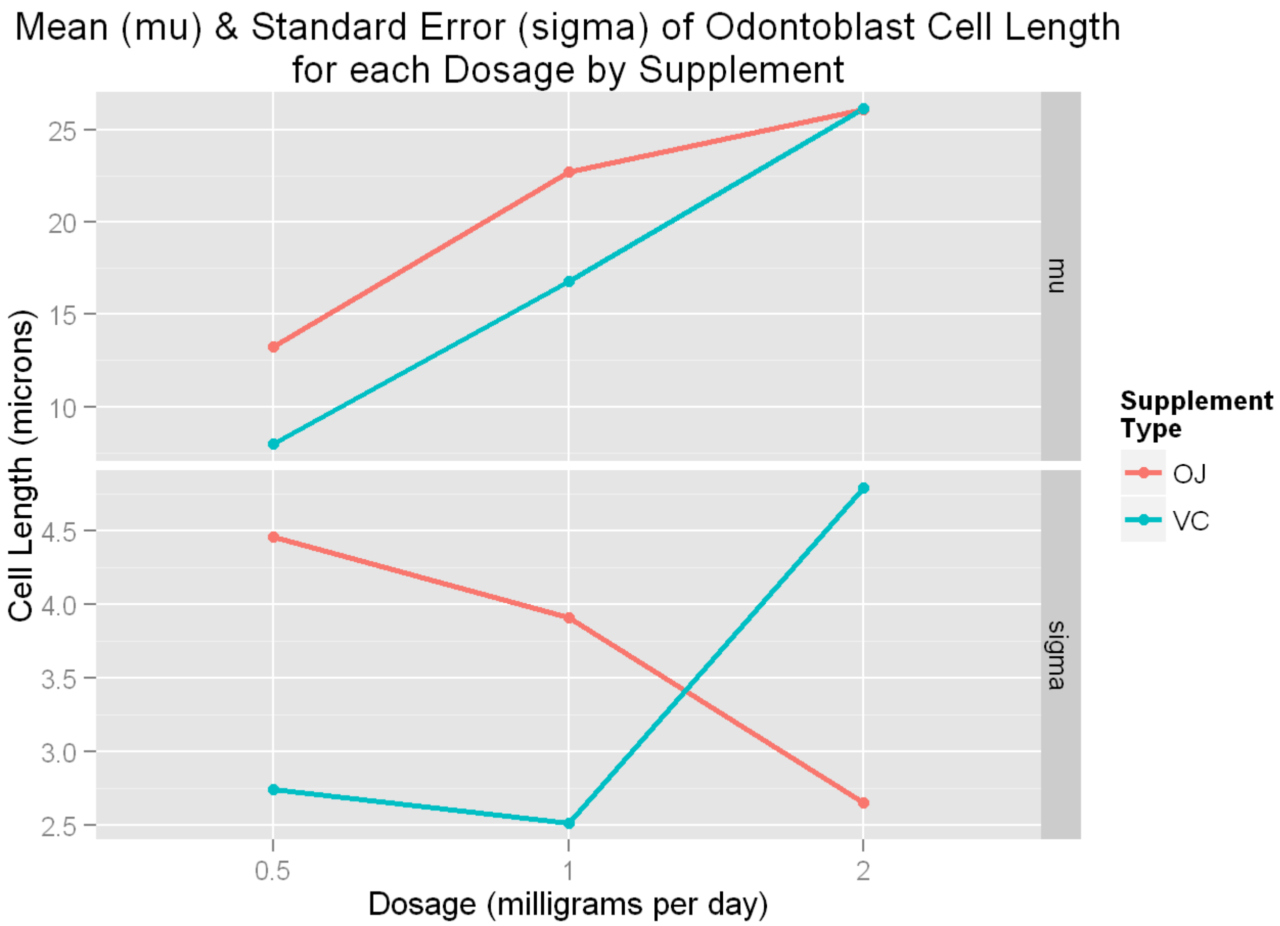
Comparing Tooth Growth by Supplement and Dosage

```
library(dplyr)
library(ggplot2)

mu_by_grp <- ToothGrowth %>% group_by(supp, dose) %>% summarise_each(funs(mea
n))

var_by_grp <- ToothGrowth %>% group_by(supp, dose) %>% summarise_each(funs(s
d))

g1 <- ggplot(ToothGrowth, aes(x=len, fill = factor(dose))) + facet_grid(supp
~ dose)
g1 <- g1 + geom_histogram()
g1 <- g1 + labs(title = "Histograms of Odontoblast Cell Length after 42 Day
s\nBy Supplement Type & Dosage") + labs(x = "Cell Length (microns)", y = "Cou
nt") + theme(legend.position="none")
g1
```



This conditional plot of histograms shows a quick overview of our dataset. By visual comparision it can be seen that there is a general trend that as the dosage of vitamin C increases then so too does the length of the Odontoblast cells. This is true for both vitamin C sources, however the measurements for pure vitamin C (VC) seem a little more variable.

Next we should perform a numerical statistical analysis to support the pattern shown in the histograms.



T-Test of Mean Difference by Supplement Type

From visual analysis it can be seen that there is some relation between dosage size and the amount of tooth growth observed. What is less clear is if there is a relationship between the supplement type and observed tooth growth. To evauate this we will ignore the dosage values in our data set. Stating this more formally we have the following hypothesis:

$H_0 =$ Supplement type does not affect tooth growth

```
t1 <- t.test(len~supp, paired = FALSE, var.equal = FALSE, data = ToothGrowth)
t1
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## 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:
##  -0.1710156  7.5710156
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

```
t1_df <- t1$parameter

t1_95 <- qt(0.95, t1_df)
```

From the t-test comparing the mean values of both supplement types (irrespective of dosage) we discover the following characteristics:

```
##      t_stat      p_val   low_conf high_conf
## t 1.915268 0.06063451 -0.1710156  7.571016
```

The test statistic t = 1.915 which is greater than 1.673, the 95th percentile of the T-distribution with the same degrees of freedom, df = 55.309. This means that we would reject the null hypothesis. However the confidence interval is between -0.171 and 7.571 which contains zero so we cannot rule out the possibility of no difference in means for the two groups and would result that we fail to reject the null hypothesis. Thus the above test is inconclusive.

T-Test of Mean Difference by Supplement Type Individually

To analyise this further let perform the same t-test as above but on each dosage level separately.

```
##       t_stat p_val low_conf high_conf   t95
## 0.5mg  3.170 0.006    1.719     8.781 1.753
## 1.0mg  4.033 0.001    2.802     9.058 1.750
## 2.0mg -0.046 0.964   -3.798     3.638 1.761
```

From the results shown above we can conclude that in the case of dosages of 0.5mg and 1.0mg there is significant difference between the supplement types. as shown by the positive values for their t statistics which are greater than the corresponding value of the 95th percentile. This is supported by the t confidence intervals which do not contain zero, meaning that there is significant difference between the supplement means. Thus we **reject** the null hypothesis. Also because the test statistics are positive this concludes that the mean tooth growth for delivery method `OJ` is greater than that of `VC`, which is not surprising when compared to the earlier histograms.

However in the case of 2.0mg dosages of vitamin C there is no significant differences in the mean values because the test statistic is less than the 95th percentile. Also the t confidence interval contains zero meaning we cannot rule out the possibility of no difference in mean values between delivery methods. Thus in this case we **fail to reject** the null hypothesis.

T-Test for Mean Difference by Dosage

With this t-test we can assess if there is significant difference in Odontoblast cell length depending on the dosage level of vitamin C irrespective of delivery method. This suggests the following null hypothesis:

$H_0 =$ Dosage level does not affect tooth growth

In this analysis, the different dosage levels will be cross-evaluated against each other (0.5mg against 1.0mg & 2.0mg, 1.0mg against 2.0mg).

```
##                  t_stat        p_val   low_conf  high_conf      t95
## 0.5mg v 1mg  -6.476648 1.268301e-07 -11.983781  -6.276219 1.685970
## 0.5mg v 2mg -11.799046 4.397525e-14 -18.156167 -12.833833 1.687232
## 1mg v 2mg    -4.900484 1.906430e-05  -8.996481  -3.733519 1.686976
```

In all three scenarious above we **reject the null hypothesis**, because with significant difference between the mean values for each groups it supports that the dosage amount does affect the mean level of tooth growth. The criteria the must be satisfied to reject the null hypothesis are:

- The absolute value of the test statistis (`t_stat`) is greater than the 95th percentile of the T-distribution for the sample.
- The T confidence interval (`[low_conf, high_conf]`) does not contain zero, ruling out the possibility of zero difference in means.
- The p-value (`p_val`) is small and is small than the α value where α = 1 - significance level = 1 - 0.95 = 0.05 in our case.

Conclusions

From our analysis of the different delivery methods (supplement types) it has been shown that when the dosage is small (0.5mg/day or 1.0mg/day) then there is a difference in observed tooth growth, namely that the vitamin C source from orange juice yields better results than vitamin C from an chemical source disolved in water. This was particularly so with a dosage level of 0.5mg/day. However for the larger dosage level (2.0mg/day) we were unable to significantly show a difference in tooth growth, so we conclude that the supplement source does not affect tooth growth.

In the analysis of the dosage levels it was clearly shown, as expected, that they do yield siginficant differences in affects, namely that the second group in the tests (which was consistantly that of the higher dosage level) yields a higher mean tooth growth than the first group.

End of Analysis