Effects of Vitamin C on Tooth Length in Guinea Pigs

JBrand

Synopsis

The dataset `ToothGrowth` provides 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). This analysis will attempt to determine how dosage and/or delivery method affects tooth growth. All code required will be provided in the appendix at the end of this report

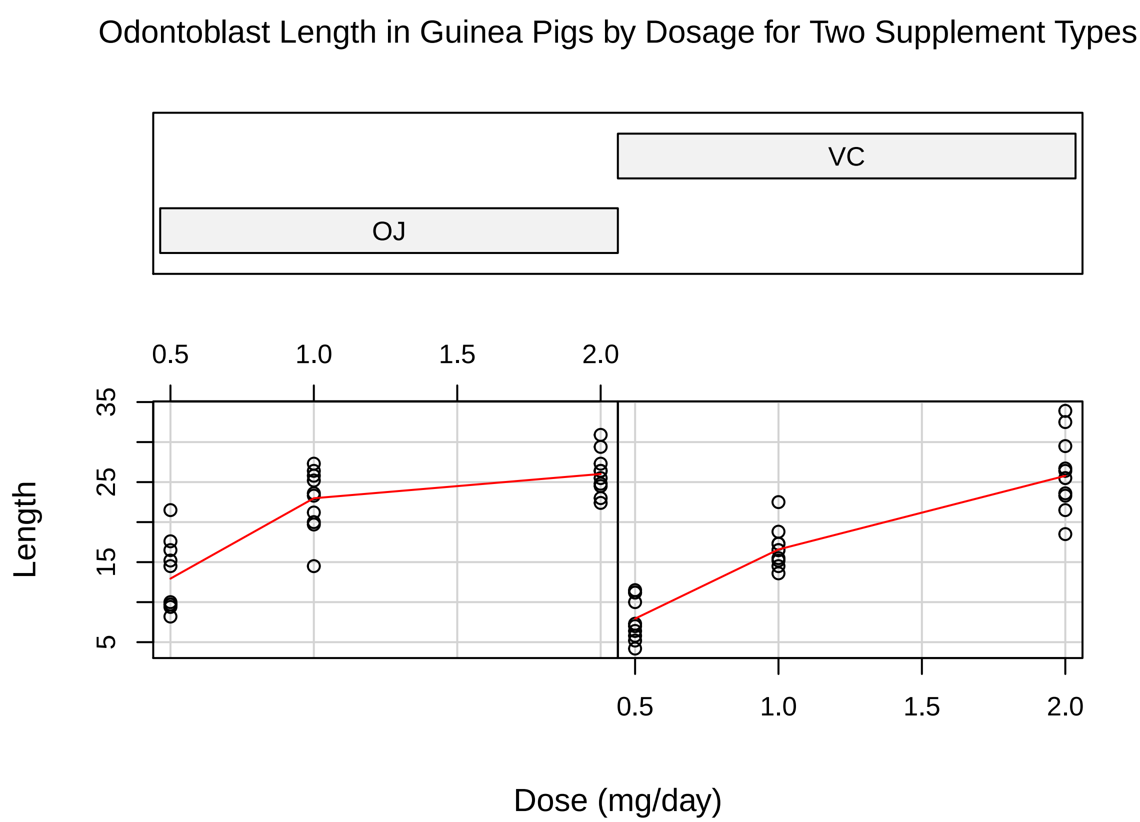
Exploring the Data

The first step will be to look at the dataset.

```
## '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 ...
```

We can see three variables:
1. **len**: Length of the odontoblasts (unknown units)
2. **supp**: Vitamin C supplement source (OJ=Orange Juice, VC=Vitamin C via Ascorbic Acid)
3. **dose**: Dosage level of Vitamin C in mg/day

Based on that, let's take a first look at the relations of these values to see if we can generate hypotheses to test. We'll start by running the sample plot provided with the dataset.



So at first glance, it looks like we're going to be able to verify that tooth length increases with dosage, and that orange juice is a more effective source of Vitamin C than the ascorbic acid at lower dosages.

For one final glance, let's look at some numbers. We will aggregate the mean of length against dosage, agnostic of the supplement, and also the mean of length against supplement, agnostic of dosage.

```
##   dose    len
## 1  0.5 10.605
## 2  1.0 19.735
## 3  2.0 26.100
```

```
##   supp      len
## 1   OJ 20.66333
## 2   VC 16.96333
```

Looks like the numbers support what we saw in the chart: average lengths increase with dosage, and it's possible that the orange juice has a slight edge on the ascorbic acid in general.

Analysis

Now that we've seen a top level summary of the data and have determined hypotheses, we will look at some confidence intervals and hypothesis testing to finalize our conclusions.

We need to do hypothesis testing on two separate categories: supplement choice and dosage size. We will start with supplement choice.

For each of three dosages, we will run a t-test to explore the relationship between the means (DM - Difference in the Means) of tooth growth for orange juice and ascorbic acid. So for each test run, H0: DM = 0, Ha: DM != 0.

Dosage 1 - 0.5 mg/day

```
## 
##  Welch Two Sample t-test
## 
## data:  len by supp
## 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:
##  1.719057 8.780943
## sample estimates:
## mean in group OJ mean in group VC 
##            13.23             7.98
```

So for 0.5 mg/day dosage, it looks as though we have a very low p-value, a nice sized t, and the 95% confidence interval does not contain zero. With that being said, we should reject the null hypothesis in this case. It should also be noted that at the 0.5 mg/day dosage, the mean for OJ is much higher than that of VC.

Dosage 2 - 1.0 mg/day

```
## 
##  Welch Two Sample t-test
## 
## data:  len by supp
## 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:
##  2.802148 9.057852
## sample estimates:
## mean in group OJ mean in group VC 
##            22.70            16.77
```

So for 1.0 mg/day dosage, it looks as though we have a very low p-value, a nice sized t, and the 95% confidence interval does not contain zero. With that being said, we should again reject the null hypothesis in this case. It should also be noted again that at the 1.0 mg/day dosage, the mean for OJ is much higher than that of VC.

Dosage 3 - 2.0 mg/day

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = -0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean in group OJ mean in group VC
##            26.06            26.14
```

So for 2.0 mg/day dosage, it looks as though we have a higher p-value, a small sized t, and the 95% confidence interval now contains zero. In this case, we cannot reject the null hypothesis in this case. It should also be noted again that at the 2.0 mg/day dosage, the means for OJ and VC are almost identical.

Now for the dosage sizes. Since we have three items to compare, we will have to do this in three parts. We will compare Dose 1 to 2, 2 to 3, and 1 to 3. Same hypotheses as before: for each test run, H0: DM = 0, Ha: DM != 0, but this time we are comparing the differences in the means of the dosage sizes.

Dosage 1 (0.5 mg/day) vs. Dosage 2 (1.0 mg/day)

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## 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 in group 0.5   mean in group 1
##            10.605            19.735
```

In this first pairing, we have a very low p-value, a large t-value in absolute terms, and the 95% confidence interval does not contain zero. We can reject the null hypothesis.

Dosage 2 (1.0 mg/day) vs. Dosage 3 (2.0 mg/day)

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## 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:
##  -8.996481 -3.733519
## sample estimates:
## mean in group 1 mean in group 2
##          19.735          26.100
```

In this pairing, we have a very low p-value, a large t-value in absolute terms, and the 95% confidence interval does not contain zero. We can again reject the null hypothesis.

Dosage 3 (0.5 mg/day) vs. Dosage 3 (2.0 mg/day)

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## 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 in group 0.5   mean in group 2
##            10.605            26.100
```

In this pairing, we have a very low p-value, a large t-value in absolute terms, and the 95% confidence interval does not contain zero. We can again reject the null hypothesis.

Assumptions

Before I give my conclusions, here are the items we must assume about the data:

1. The sample population was a random selection from the general guinea pig population and that internally to this experiment the populations were independent, i.e. no genetic relatives so we can eliminate possible pairings.
2. The sample population of 60 guinea pigs are representative of the general population.
3. Proper care was taken in terms of guinea pig breed so that all were equally represented in their test groups.
4. All standard double-blind test/measurement procedures were incorporated to reduce bias.

Conclusions

Based on the initial plot and verified through the hypothesis testing:
- Length of the odontoblasts is definitely affected by dosage level. The higher the dosage, the longer the odontoblasts.
- For lower dosage levels, orange juice certainly is more effective than ascorbic acid. But for the 2.0 mg/day dosage level, there is no statistical proof that either is more effective.

Appendix

```r
## First codeblock showing structure of the dataset
data("ToothGrowth")
str(ToothGrowth)

## Plot
coplot(len ~ dose | supp, data = ToothGrowth, panel = panel.smooth,
       xlab = c("Dose (mg/day)", "Odontoblast Length in Guinea Pigs by Dosage for Two Supplement Types"),
       ylab = "Length")

## Next codeblocks showing the means of the dosages and means of the supplements
doseEffect <- aggregate(len ~ dose, data = ToothGrowth, FUN = mean)
suppEffect <- aggregate(len ~ supp, data = ToothGrowth, FUN = mean)
doseEffect
suppEffect

## t.tests for each dosage to compare the supplements
d1 <- ToothGrowth[ToothGrowth$dose == 0.5,]
t.test(len ~ supp, paired = FALSE, var.equal = FALSE, data = d1)

d2 <- ToothGrowth[ToothGrowth$dose == 1,]
t.test(len ~ supp, paired = FALSE, var.equal = FALSE, data = d2)

d3 <- ToothGrowth[ToothGrowth$dose == 2,]
t.test(len ~ supp, paired = FALSE, var.equal = FALSE, data = d3)

## t.tests for each dosage pairing to compare the dosage levels
d12 <- ToothGrowth[ToothGrowth$dose == 0.5 | ToothGrowth$dose == 1,]
t.test(len ~ dose, paired = FALSE, var.equal = FALSE, data = d12)

d23 <- ToothGrowth[ToothGrowth$dose == 1 | ToothGrowth$dose == 2,]
t.test(len ~ dose, paired = FALSE, var.equal = FALSE, data = d23)

d13 <- ToothGrowth[ToothGrowth$dose == .5 | ToothGrowth$dose == 2,]
t.test(len ~ dose, paired = FALSE, var.equal = FALSE, data = d13)
```