Statistical Inference Course Project - Part 2

Darrell Nabors

March 21, 2015

The Effect of Vitamin C on Tooth Growth in Guinea Pigs

Now in the second portion of the class, we're going to analyze the ToothGrowth data in the R datasets package.

1. Load the ToothGrowth data and perform some basic exploratory data analyses.
2. Provide a basic summary of the data.
3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there's other approaches worth considering)
4. State your conclusions and the assumptions needed for your conclusions.

A brief description of the Tooth Growth Data is as follows:

The response is the length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1.0, and 2.0 mg) with each of two delivery methods (orange juice or ascorbic acid).

As stated in the aforementioned description, this data is the result of essay testing. These essays tested the effect of Vitamin C on the growth of teeth in guinea pigs. The data is composed of the measured tooth length in 6 groups of 10 guinea pigs, while these are divided into 2 groups depending on the delivery technique of Vitamin C. The 2 delivery techniques are orange juice or ascorbic acid. The 2 groups are also split into 3 groups of doses of Vitamin C (0.5 mg, 1.0 mg or 2.0 mg).

First, let's load the Tooth Growth data from the R package and display the structure of the data.

```
data(ToothGrowth)
str(ToothGrowth)
```

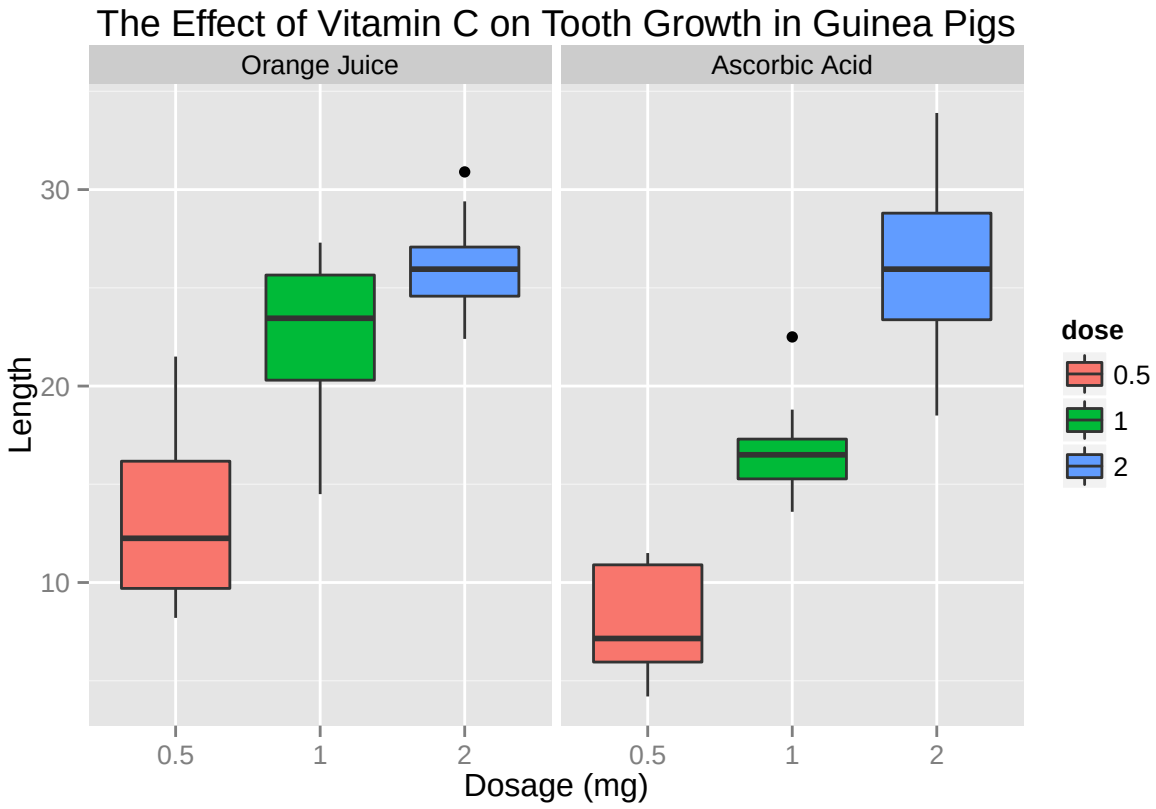
```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

Now, we will have at look at the summary of the data.

```
##       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
```

We can transform the dosages into a `factor` data type and show the relationship of the data.

```r
library(ggplot2)
TG <- ToothGrowth
TG$dose <- factor(TG$dose)
levels(TG$supp) <- c("Orange Juice","Ascorbic Acid")
ggplot(TG,aes(x=dose,y=len)) + geom_boxplot(aes(fill=dose)) + facet_wrap(~supp) +
      ggtitle("The Effect of Vitamin C on Tooth Growth in Guinea Pigs") +
      xlab("Dosage (mg)") + ylab("Length")
```



Observations

Based on the graph, we can make some initial observations:

1. Vitamin C dosage appears to affect the length of teeth in guinea pigs.
2. The delivery technique also impacts tooth length. The technique of orange juice appears to be associated with longer teeth. However, in the highest dosage, the length in teeth for both orange juice and ascorbic acid are very similar.
3. Finally, there appears to be a maximum threshold level between 1.0 and 2.0 mg, in which elevated doses of Vitamin C produce no additional impact on tooth growth.

Hypothesis Testing

From the aforementioned observations, we can perform some hypothesis testing and examine them utilizing 2-group t-tests and p-values. There will be no assumptions made regarding the population variances. We will

implement the safest approach taking into account different variances in the populations. The data will not be considered as paired since guinea pigs in each essay are dissimiliar.

Based on our first observation regarding dosage, the hypothesis will be that the mean tooth length in relation to 2 different doses is identical. To accomplish this, we will examine the hypotheses pairwise. The t-test will be constructed utilizing the population from both delivery techniques.

Let's compare the does of 0.5 mg and 1.0 mg.

```
TG_05_1 <- subset(TG,dose %in% c(0.5,1))
t.test(len~dose,paired=F,var.equal=F,data=TG_05_1)
```

```
##
##  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
```

From the results of the t-test, we obtained a confidence interval of `([-11.984,-6.276])` and a p-value of `1.268e-07`. Furthermore, our interval does not include `0` and the p-value is below 5%. Also, this indicates that the means of the tooth growth cannot be equal. Therefore these figures indicate that we must reject our null hypothesis. However, our first observation is correct in that the 1.0 mg dosed guinea pigs have longer teeth that the 0.5 mg dosed population.

Now, let's construct a test for the 1.0 mg doses and the 2.0 mg doses.

```
TG_1_2 <- subset(TG, dose %in% c(1,2))
t.test(len~dose,paired=F,var.equal=F,data=TG_1_2)
```

```
##
##  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
```

Repeating the test for the 1.0 mg and 2.0 mg doses, we get a similar result, with a slightly higher p-value`(1.906e-05)`. However, this result still leads us to reject the hypothesis on the grounds of identical means and therefore validating the hypothesis of the 2.0 mg dosage yielding even longer teeth than the 1.0 mg doses.

We could construct a third test with the 0.5 mg and 2.0 mg doses. However, the outcome will be the same as our 2 previous tests, since it is a logical consequence of them. Therefore, we will not construct a third test of the mean tooth length in regard to dosage.

For our second observation regarding the delivery technique, we will construct 3 separate tests for each dosage level. This is because we have observed contrasting responses between the lower dosage levels vs. the highest level.

For each test, we can create the null hypothesis stating that the mean for teeth growth for both delivery techniques are identical. We will construct the following tests for each dosage level.

```
# Delivery method testing for 0.5 mg dose
t.test(len~supp,paired=F,var.equal=F,data=TG[TG$dose==0.5,])
```

```
##
##  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 Orange Juice mean in group Ascorbic Acid
##                      13.23                       7.98
```

```
# Delivery method testing for 1.0 mg dose
t.test(len~supp,paired=F,var.equal=F,data=TG[TG$dose==1,])
```

```
##
##  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 Orange Juice mean in group Ascorbic Acid
##                      22.70                       16.77
```

```
# Delivery method testing for 2.0 mg dose
t.test(len~supp,paired=F,var.equal=F,data=TG[TG$dose==2,])
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## 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:
##  -3.79807  3.63807
## sample estimates:
##  mean in group Orange Juice mean in group Ascorbic Acid
##                      26.06                       26.14
```

According to the obtained p-values for the 3 tests, 0.006, 0.001 and 0.964 for doses 0.5 mg, 1.0 mg and 2.0 mg, respectively, we should reject the null hypothesis for doses 0.5 mg and 1.0 mg. However, we should accept the null hypothesis for the 2.0 mg dosage. This means that the first observation is correct in the first 2 cases, indicating that the delivery method does have a significant impact on tooth growth.

However, for the 2.0 mg dose, we obtain a very high p-value. This indicates that we should not reject the null hypothesis. This specific test also indicates that the means of the tooth length are the same. This phenomena demonstrates that our third observation is correct and that there is a correlation between tooth length and dosage in the 0.5 mg and 1.0 mg doses. However, the correlation is not present in the test of the 2.0 mg dosage.

Conclusions and Additional Assumptions

In conclusion, it can be stated that:

1. Dosage does have an impact on tooth growth. Higher doses of Vitamin C correlates with tooth length (This is according to our data in observations up to 2.0 mg. There is no data beyond 2.0 mg).
2. The delivery technique is significant when the dosage applied is within the range of 0.5 mg and 1.0 mg. However, there is no impact on tooth growth when the dosage is 2.0 mg, with orange juice being slightly more effective than ascorbic acid.

Additional Assumptions:

1. The experiment was done with random assignment of guinea pigs to different dose levels and supplement types to contol the confounders that may affect the outcome.
2. Members of the sample population, i.e. the 60 guniea pigs, are representative of the entire population of guinea pigs. This assumption allows for generalization of the results.
3. For the t-tests, the variances are assumed to be different for the 2 groups that are being compared. This assumption is less stronger than the case in which the variances are assumed to be equal.