Statistical Inference Course Project - Part 2

Norman Norris

May 24, 2015

Confidence Intervals and Hypothesis Testing

Overview

This project analyzes the ToothGrowth data in the R datasets package. The data consists of measurements of the mean tooth length from a population of 60 guinea pigs. The animals were divided into 6 groups of 10 and fed a diet with one of 6 Vitamin C supplements for a period of 42 days. The Vitamin C was administered in the form of Orange Juice (OJ) or Vitamin C (VC). Each animal received the same daily dose of Vitamin C (either 0.5, 1.0 or 2.0 milligrams/day) consistently. The ToothGrowth dataset consists of 60 observations of 3 variables - mean tooth length (microns), supplement type (OJ or VC) and Vitamin C dose (milligrams/day).

Objectives

Analyze the ToothGrowth data:

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.

Results

Objectives 1 and 2:

```
## Load packages and dataset
library(datasets)
data(ToothGrowth)

## Perform some basic exploratory data analyses
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 ...
```

We see the “dose” variable is a numeric. Since the dose can only takes on 3 discrete values, we convert it to a Factor and summarize the data.

```
## Convert the dose to a Factor, as it only takes on 3 discrete values
ToothGrowth$dose <- as.factor(ToothGrowth$dose)

## Summarize the data
summary(ToothGrowth)
```

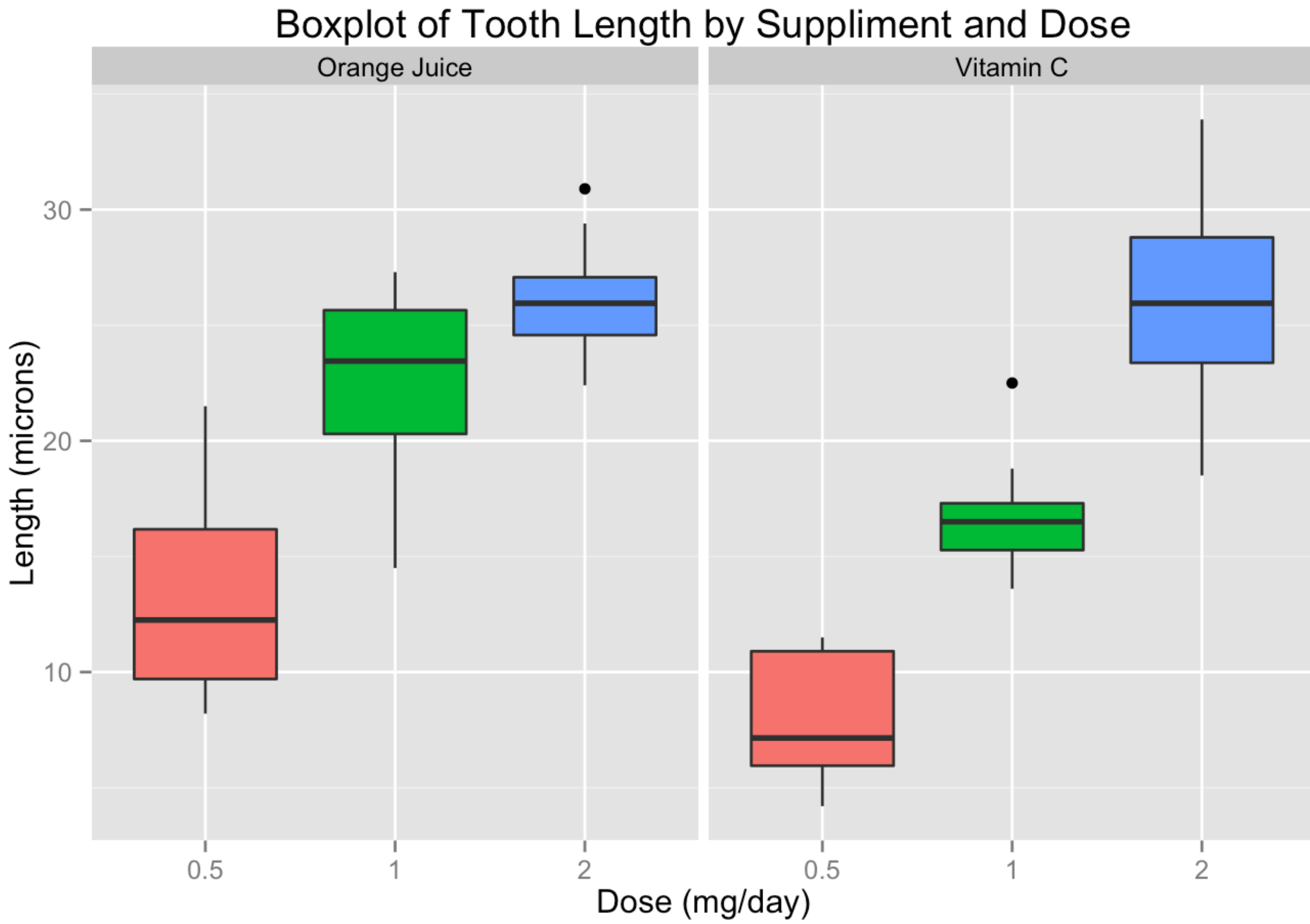
```
##       len        supp     dose   
##  Min.   : 4.20   OJ:30   0.5:20  
##  1st Qu.:13.07   VC:30   1  :20  
##  Median :19.25           2  :20  
##  Mean   :18.81                   
##  3rd Qu.:25.27                   
##  Max.   :33.90                   
```

Then we perform a quick boxplot of tooth length by dose, faceting on suppliment type.

```
## Create a boxplot of tooth length vs dose, faceting by suppliment type
g <- ggplot(ToothGrowth, aes(x = dose, y = len))

p1 <- g + geom_boxplot(aes(fill = dose)) + facet_grid(. ~ supp, labeller = fips_label
ler) +
    labs(title = "Boxplot of Tooth Length by Suppliment and Dose") +
    labs(x = "Dose (mg/day)", y = "Length (microns)") +
    guides(fill = F)

print(p1)
```



Objective 3:

Based on these observations, we propose the following hypotheses:

1. Suppliment dose improves tooth length, since a higher dose is associated with longer teeth for both suppliment types.
2. Suppliment type also has an affect tooth length, since tooth lengths associated with Orange Juice supplimentation seem to be longer than those with Vitamin C supplimentation (questionable for the highest dosage of OJ/VC, where tooth lengths are very close).
3. The fact the maximum dose produced similar tooth lengths might indicate a threshold level between 1 and 2 mg, which marks a “maximum effect point”, over which higher doses of suppliment produce no additional effect.

Hypothesis #1 - Higher dose improves tooth length

We calculate the T-test for the different doses, regardless of suppliment type. (2 tests performed, 0.5 mg vs 1.0 mg, 1.0 mg vs 2.0 mg)

Dose 0.5 mg/day vs 1.0 mg/day

```
## Subset data and then compare dose of 0.5 mg/day to 1.0 mg/day
dose_05v10 <- subset(ToothGrowth, dose %in% c(0.5, 1))
t.test(len ~ dose, paired = F, var.equal = F, data = dose_05v10)
```

```
##
##  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
```

Dose 1.0 mg/day vs 2.0 mg/day

```
## Subset data and then compare dose of 1.0 mg/day to 2.0 mg/day\
dose_10v20 <- subset(ToothGrowth, dose %in% c(1, 2))
t.test(len ~ dose, paired = F, var.equal = F, data = dose_10v20)
```

```
##
##  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
```

Hypothesis #2 - Suppliment type matters (OJ is better than VC)

We calculate the T-test for the two suppliments, while holding the dose constant. (3 tests performed, one for each dose)

Dose = 0.5 mg/day

```
## Compare suppliment type for dose = 0.5 mg/day
t.test(len ~ supp, paired = F, var.equal = F, data = ToothGrowth[ToothGrowth$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 OJ mean in group VC
##            13.23             7.98
```

Dose = 1.0 mg/day

```
## Compare suppliment type for dose = 1.0 mg/day
t.test(len ~ supp, paired = F, var.equal = F, data = ToothGrowth[ToothGrowth$dose ==
1.0, ])
```

```
##
##  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
```

Dose = 2.0 mg/day

```
## Compare suppliment type for dose = 2.0 mg/day
t.test(len ~ supp, paired = F, var.equal = F, data = ToothGrowth[ToothGrowth$dose ==
2.0, ])
```

```
##
##  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 OJ mean in group VC
##            26.06            26.14
```

Conclusions

Objective 4:

1. Dose influences tooth length. Higher doses produce more tooth growth.

- Looking at the T-tests for 0.5 mg/day vs 1.0 mg/day and 1.0 mg/day vs 2.0 mg/day, both have completely negative confidence intervals (at 95%) and very small p-values (both less than 0.1%). This tells us a value of 0 is not in the confidence interval, therefore we should reject the null hypothesis and conclude higher doses are responsible for greater tooth length.

2. Suppliment type influences tooth length. Orange Juice is a better suppliment for tooth growth than Vitamin C, below 2.0 mg/day.

- At doses of 0.5 mg/day and 1.0 mg/day, both T-tests show completely positive confidence intervals (at 95%) and small p-values (~ 0.6% and 0.1% respectively), therefore we are forced to reject the null hypothesis in these cases and conclude that suppliment (i.e. Orange Juice), at these doses, are responsible for greater tooth length. However, at a dose of 2.0 mg/day, zero is contained in the confidence interval and the p-value is very high, therefore we cannot reject the null hypothesis for a dose of 2.0 mg/day and conclude that at this dose, the type of suppliment does not affect tooth growth.

3. More data would be required to determine if there is a “threshold” dose, above which further supplimentation does not improve tooth length.

Assumptions

1. The experiment states each Guinea pig was randomly assigned to a combination of dose and supplement type, so the tests we performed used the independent samples methodology, meaning the tests assume different population variances and “not paired” subjects.
2. The sample of 60 Guinea pigs is assumed to be representative of the population, therefore allowing us to to generalize our conclusions.