Statistical Inference on Tooth Growth

Charles Westby

11/13/2017

Synopsis

In this paper we will look at the results of a dataset called ToothGrowth in R. This dataset contains the data from a test that was done on guinea pigs. The guinea pigs were given Vitamin C from two different sources. One source was absorbic acid and the other was orange juice. There were also three different dosage levels. Either they were give 0.5, 1.0 or 2.0 mg/day of the Vitamin C. From this dataset we are going to see if there is any statistical difference in the length of odontoblasts (cells responsible for tooth growth), as a result of the different delivery methods for Vitamin C or the different dosages.

Data Processing

Loading Data

Here I loaded the packages that I would use to manipulate the data. Then I also loaded the dataset ToothGrowth

```
library(dplyr)
library(ggplot2)
library(gridExtra)
data("ToothGrowth")
```

Understanding The Data

```
summary(ToothGrowth)
```

```
##       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 see that the mean growth in odontoblasts is 18.81 and the median growth in odontoblasts is 19.25. We also see that 30 guinea pigs were given orange juice and 30 guinea pigs were given absorbic acid. However, we also see a maximum and minumum value for dose when it should be a factor variable. So we convert that column to a factor variable.

```
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
```

We then check out the data again

```
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
```

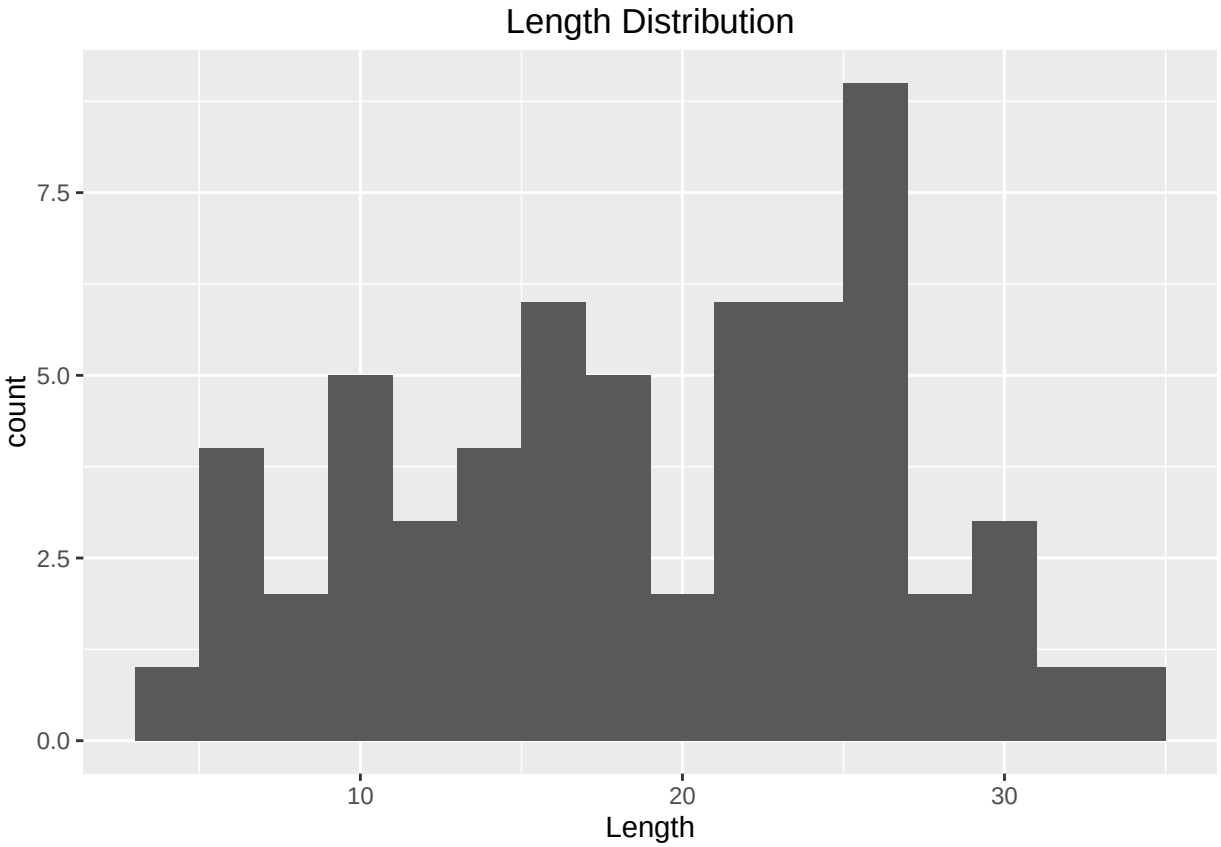
We now have one continuous variable, len, to represent the length of odontoblasts and two factor variables, supp and dose, to represent supplement and dose respectively.

Results

Visualizing The Data

First we looked at a histogram of the length variable to get an idea for the distribution of the length variable

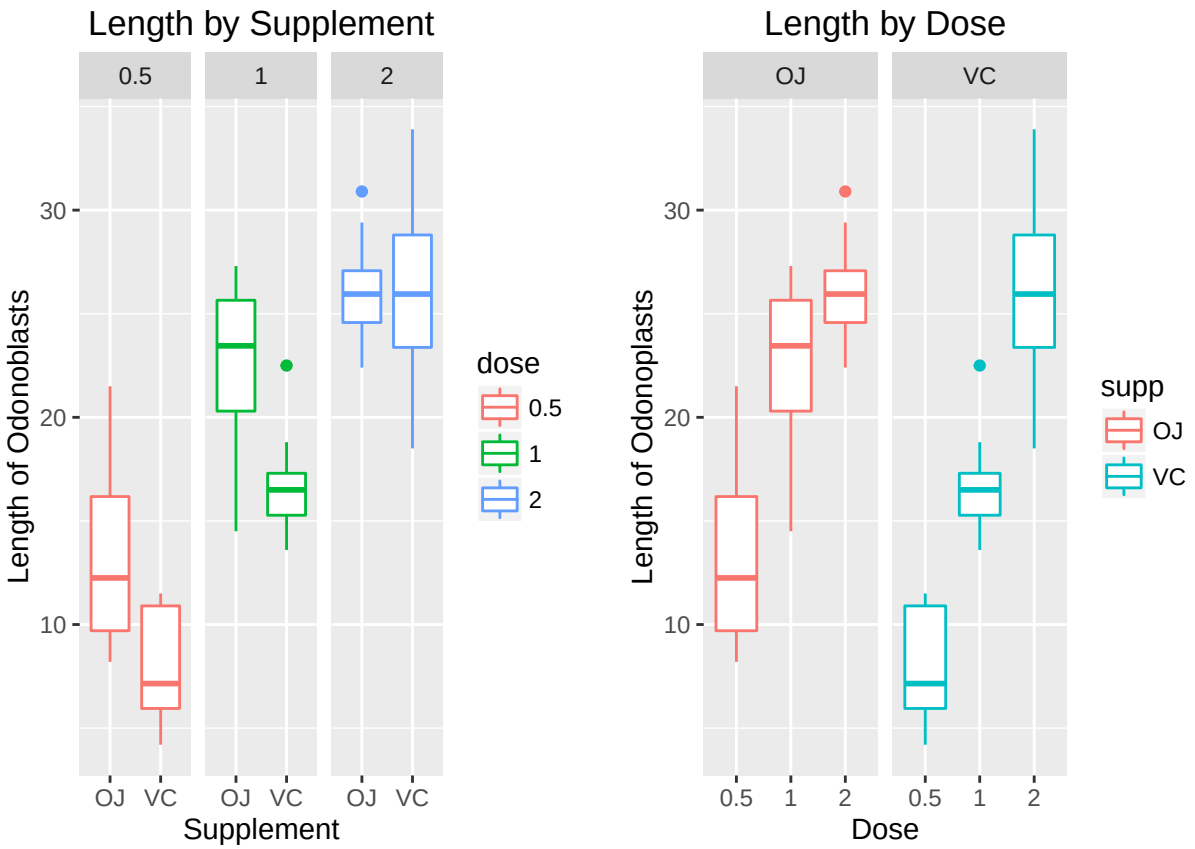
```
ggplot(ToothGrowth, aes(x = len)) +
  geom_histogram(binwidth = 2) + labs(title = "Length Distribution", x = "Length")
```



Then we looked at the change in the length of odonoblast. First we looked by supplment and then we looked by dose

```
supplement_gr <- ggplot(ToothGrowth, aes(x = supp, y = len, col = dose)) +
  geom_boxplot() + facet_grid(~ dose) + labs(x = "Supplement", y = "Length of Odonoblasts", title = "Le
dose_gr <-ggplot(ToothGrowth, aes(x = dose, y = len, col = supp)) +
```

```
  geom_boxplot() + facet_grid(~ supp) + labs(x = "Dose", y = "Length of Odonoplasts", title = "Length by
grid.arrange(supplement_gr, dose_gr, ncol = 2)
```



Here we see that supplement does not seem like much of a factor in the growth of the odonoblasts. But we do see dosage having an effect. In order to quantify these differences we see the mean of growth when grouped by supplement and when grouped by dose.

```
by_supp <- ToothGrowth %>%
  group_by(supp) %>%
  summarize(supp_mean = mean(len))
by_supp
```

```
## Source: local data frame [2 x 2]
##
##     supp supp_mean
##   (fctr)     (dbl)
## 1     OJ  20.66333
## 2     VC  16.96333
```

```
by_dose <- ToothGrowth %>%
  group_by(dose) %>%
  summarize(dose_mean = mean(len))
by_dose
```

```
## Source: local data frame [3 x 2]
##
##     dose dose_mean
##   (fctr)     (dbl)
```

```
## 1    0.5    10.605
## 2      1    19.735
## 3      2    26.100
```

We see a small difference when looking at the means compared by supplement. Where we see the most change in the mean is when we see the change in means by dosage. However, we will do t-tests to see if we can make any inferences from the data.

Statistical Inference

T-Test Length by Supplement

```
t.test(len ~ supp, ToothGrowth)
```

```
##
##  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
```

T-Test Length by Dosages

Here the dosages have to be broken up into different groups because the t-test are for one on one comparisons.

```
ToothGrowth0.5 <- ToothGrowth %>%
  filter(dose %in% c(0.5, 1.0))
t.test(len ~ dose, ToothGrowth0.5)
```

```
##
##  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
```

```
ToothGrowth1.0 <- ToothGrowth %>%
  filter(dose %in% c(0.5, 2.0))
t.test(len ~ dose, ToothGrowth1.0)
```

```
##
##  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
```

```
ToothGrowth2.0 <- ToothGrowth %>%
  filter(dose %in% c(1.0, 2.0))
t.test(len ~ dose, ToothGrowth2.0)
```

```
##
##  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
```

Conclusion

From these t-tests we can draw a few inferences. From the first t-test of length of odonoblasts by supplement, we can say with 95% probability that the difference in the effects of the supplements is between -0.17 and 7.57 units of length for odonoblasts. However the confidence interval includes 0, therefore we cannot reject the null hypothesis that the supplement causes a difference in length of odonoblasts. The p-value, 0.603, is greater than 0.05, which also shows that we cannot reject the null hypothesis. Therefore we do no conclude that the difference in length is a result of the different supplements. However the t-tests that were done on the different dosages show that we can reject the null hypothesis, that the different dosages cause no difference in the length of odonoblasts. All of the t-statistics fall within the rejection region. Also all the p-values are well below the 0.05 threshold. Therefore when it comes to dosages we can reject the null hypothesis that there is no difference between doses.