Tooth Growth Analysis in Pigs using R

Shalini Ruppa Subramanian

2015-09-24

Overview

This report presents the basic inferrential data analysis on tooth growth dataset in pigs.

```r
library(datasets)
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 ...
```

```r
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  
```

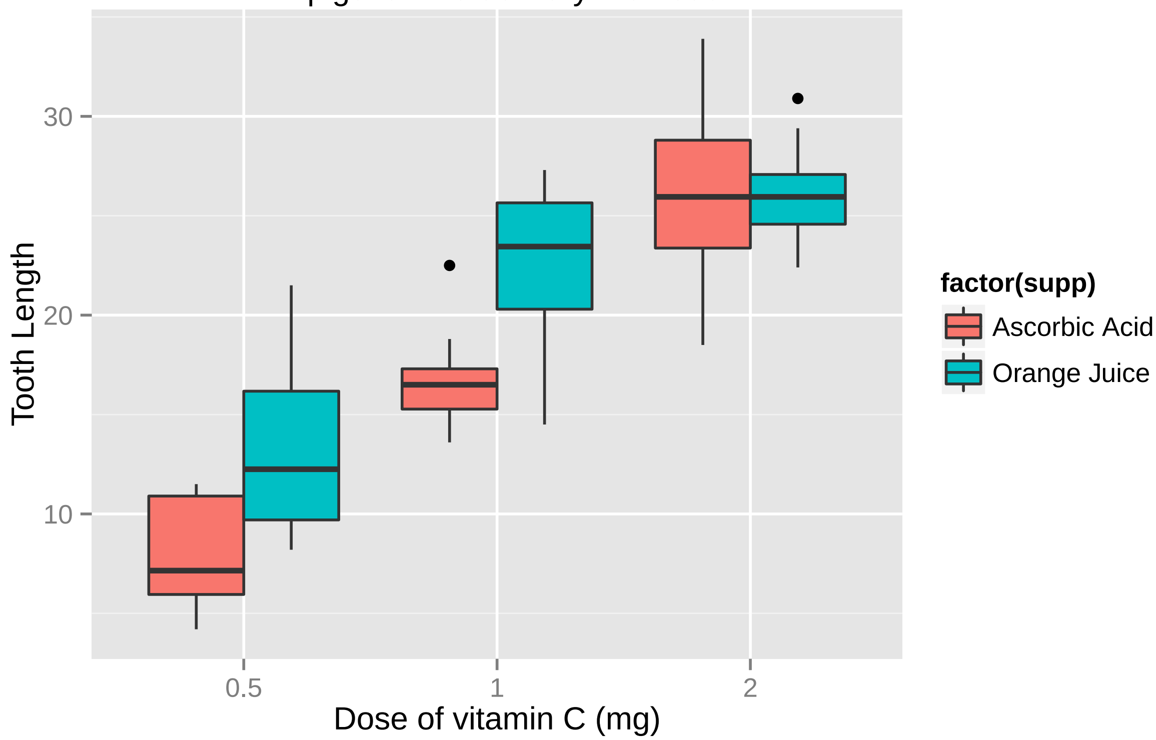
Data Summary

This data shows the results of the effect of vitamin C on tooth growth in each of 10 guinea pigs at each of three dose levels of vitamin C with two types of delivery methods.

The data is in the form of data frame with 60 observations in total on 3 variables. The variabes are tooth length (len), supplement type ascorbic acid or orange juice (VC or OJ) and three types of dose (0.5,1,2) in milligrams (dose).

The figure show how tooth length varies with dose and supplement type. The code used to generate the plots are detailed in the Appendix and additional figures are shown how tooth length varies with dose and how tooth length varies with supplement type. graph fig1-1.pdf

Figure 1: Effect of vitamin C (dose) on tooth length in pigs for two delivery methods



Hypothesis Testing

Referring to the boxplots above, hypothesis testing will be conducted to make sure if there is a significant effect on tooth length based on the factor studied. It involves a null and alternative hypothesis. A 2-sample t-test will be used to test our hypothesis. We assume that the guinea pigs samples are independent and don't have equal variances. We will calculate the 95% confidence interval in which the difference lies. The t-test results are detailed in the appendix. **Null Hypothesis**: The difference in tooth length meansured from two delivery methods is zero. **Alternative Hypothesis**: The difference in tooth length meansured from two delivery methods is not zero.

```
data(ToothGrowth)
supp1 <- (subset(ToothGrowth, supp=="VC"))[,-2]
supp2 <- (subset(ToothGrowth, supp=="OJ"))[,-2]
CI1 <- round((t.test(supp1$len,supp2$len,paired=FALSE,var.equal=FALSE))$conf.int[1:2],3)
```

The 95% confidence interval [-7.571, 0.171] contains zero and we cannot rule out zero as the possibility of the population difference between the two groups. Hence, we do not reject the null hypothesis. **Null Hypothesis**: The difference in tooth length meansured from two different doses is zero. **Alternative Hypothesis**: The difference in tooth length meansured from two different doses is not zero.

```
dose.5 <- (subset(ToothGrowth, dose==0.5))[,-2]
dose1 <- (subset(ToothGrowth, dose==1.0))[,-2]
dose2 <- (subset(ToothGrowth, dose==2.0))[,-2]
CI2A <- round((t.test(dose1$len,dose.5$len,paired=FALSE,var.equal=FALSE))$conf.int[1:2],3)
CI2B <- round((t.test(dose2$len,dose.5$len,paired=FALSE,var.equal=FALSE))$conf.int[1:2],3)
CI2C <- round((t.test(dose2$len,dose1$len,paired=FALSE,var.equal=FALSE))$conf.int[1:2],3)
```

The 95% confidence interval for two independent dose levels are [6.276, 11.984] (0.5 mg and 1 mg), [12.834, 18.156] (0.5 mg and 2 mg), and [3.734, 8.996] (1 mg and 2 mg) respectively. Since these confidence intervals are more than zero, we reject the null hypothesis. Hence, there is correlation between the tooth length and step in dose levels. **Null Hypothesis**: For a given dose, the difference in tooth length meansured from two delivery methods is zero. **Alternative Hypothesis**: For a given dose, the difference in tooth length meansured from two delivery methods is not zero.

```
CI3A <- round((t.test((subset(supp1,dose==0.5))[,1],(subset(supp2,dose==0.5))[,1],
                      paired=FALSE,var.equal=FALSE))$conf.int[1:2],3)
CI3B <- round((t.test((subset(supp1,dose==1.0))[,1],(subset(supp2,dose==1.0))[,1],
                      paired=FALSE,var.equal=FALSE))$conf.int[1:2],3)
CI3C <- round((t.test((subset(supp1,dose==2.0))[,1],(subset(supp2,dose==2.0))[,1],
                      paired=FALSE,var.equal=FALSE))$conf.int[1:2],3)
```

The confidence interval will now be tested for the two supplement methods for same dose level. The 95% confidence interval for two supplement methods are [-8.781, -1.719] (0.5 mg dose), [-9.058, -2.802] (1 mg dose), and [-3.638, 3.798] (2 mg dose) respectively. The 95% confidence intervals for 0.5 mg and 1 mg doses does not contain zero, indicating that there are significant differences between the delivery methods orange juice and ascorbic acid. Therefore, we can reject the null hypothesis. The 95% confidence interval for 2 mg dose contains zero, indicating that there is no significant difference between the two delivery methods. Hence, we accept the null hypothesis.

Assumptions

1. The populations are independent i.e. the variations between the populations are dfferent. A random population is selected for the experiment. A large sample size is used such that different doses (three) and delivery method (two) can be tested independently.
2. The populations comprises similar guinea pigs, measurement error was accounted for with significant digits, and double blind research method is used. Double blind research is where both the tester and the subject are blinded.

Conclusion

1. There is no correlation between tooth length and delivery method.
2. There is coorelation between dosage levels and tooth length.
3. For doses 0.5 mg and 1 mg, there is correlation between the delivery method and tooth length. However, for 2 mg dose, delivery method does not affect tooth length.

Appendix

The code for generating the exploratory boxplots on the tooth growth data is shown below using ggplot2.

```
library(ggplot2)
data(ToothGrowth)
ToothGrowth$supp <- factor(ToothGrowth$supp,levels=c("VC","OJ"),labels=c("Ascorbic Acid","Orange Juice"
Figure1 <- ggplot(ToothGrowth,aes(factor(dose),len))+
  geom_boxplot(aes(fill=factor(supp)))+
  ggtitle("Figure 1: Effect of vitamin C (dose) on tooth length \nin pigs for two delivery methods")+
```

```
  theme(plot.title = element_text(size = 12)) +
  xlab("Dose of vitamin C (mg)")+ylab("Tooth Length")
Figure2 <- ggplot(ToothGrowth,aes(supp,len))+
  geom_boxplot(aes(fill=supp))+
  ggtitle("Figure 2: Effect of delivery method on tooth length in pigs")+
  theme(plot.title = element_text(size = 12))+
  xlab("Delivery Method")+ylab("Tooth Length")
Figure3 <- ggplot(ToothGrowth,aes(factor(dose),len))+
  geom_boxplot(aes(fill=factor(dose)))+
  ggtitle("Figure 3: Effect of Vitamin C dose on tooth length in pigs")+
  theme(plot.title = element_text(size = 12))+
  xlab("Dose of vitamin C (mg)")+ylab("Tooth Length")
```

The t-test results comparing the two delivery methods are below.

```
##
##  Welch Two Sample t-test
##
## data:  supp1$len and supp2$len
## 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:
##  -7.5710156  0.1710156
## sample estimates:
## mean of x mean of y
##  16.96333  20.66333
```

The t-test results comparing any two doses are below. There were three different doses in total.

```
##
##  Welch Two Sample t-test
##
## data:  dose1$len and dose.5$len
## 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:
##   6.276219 11.983781
## sample estimates:
## mean of x mean of y
##    19.735    10.605
```

```
##
##  Welch Two Sample t-test
##
## data:  dose2$len and dose.5$len
## 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:
##  12.83383 18.15617
## sample estimates:
## mean of x mean of y
##    26.100    10.605
```

```
##
##  Welch Two Sample t-test
##
## data:  dose2$len and dose1$len
## 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:
##  3.733519 8.996481
## sample estimates:
## mean of x mean of y
##    26.100    19.735
```

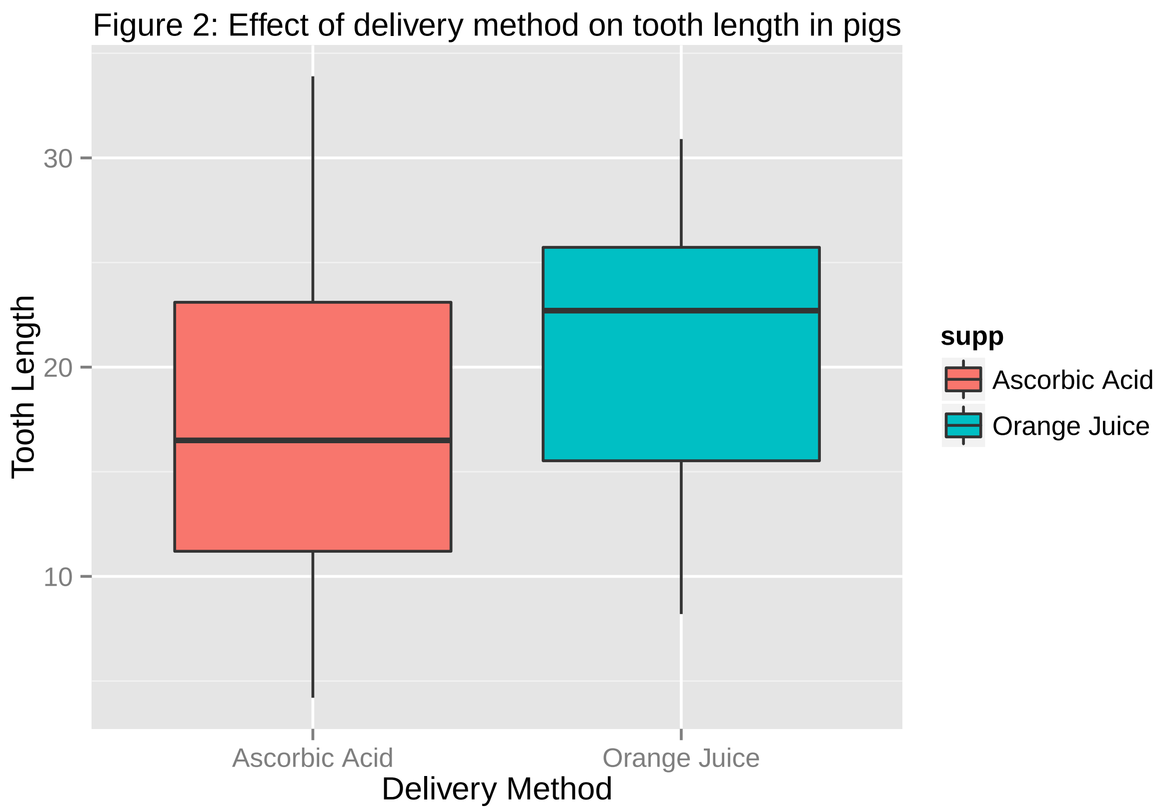
The t-test results comparing the two delivery methods at a particular dose are below. There were three different doses in total.

```
##
##  Welch Two Sample t-test
##
## data:  (subset(supp1, dose == 0.5))[, 1] and (subset(supp2, dose == 0.5))[, 1]
## 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:
##  -8.780943 -1.719057
## sample estimates:
## mean of x mean of y
##      7.98     13.23
```

```
##
##  Welch Two Sample t-test
##
## data:  (subset(supp1, dose == 1))[, 1] and (subset(supp2, dose == 1))[, 1]
## 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:
##  -9.057852 -2.802148
## sample estimates:
## mean of x mean of y
##     16.77     22.70
```

```
##
##  Welch Two Sample t-test
##
## data:  (subset(supp1, dose == 2))[, 1] and (subset(supp2, dose == 2))[, 1]
## 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:
##  -3.63807  3.79807
## sample estimates:
## mean of x mean of y
##     26.14     26.06
```

Figure 2 and 3 show how tooth length varies with supplement type and dose respectively. graph fig2-1.pdf



graph

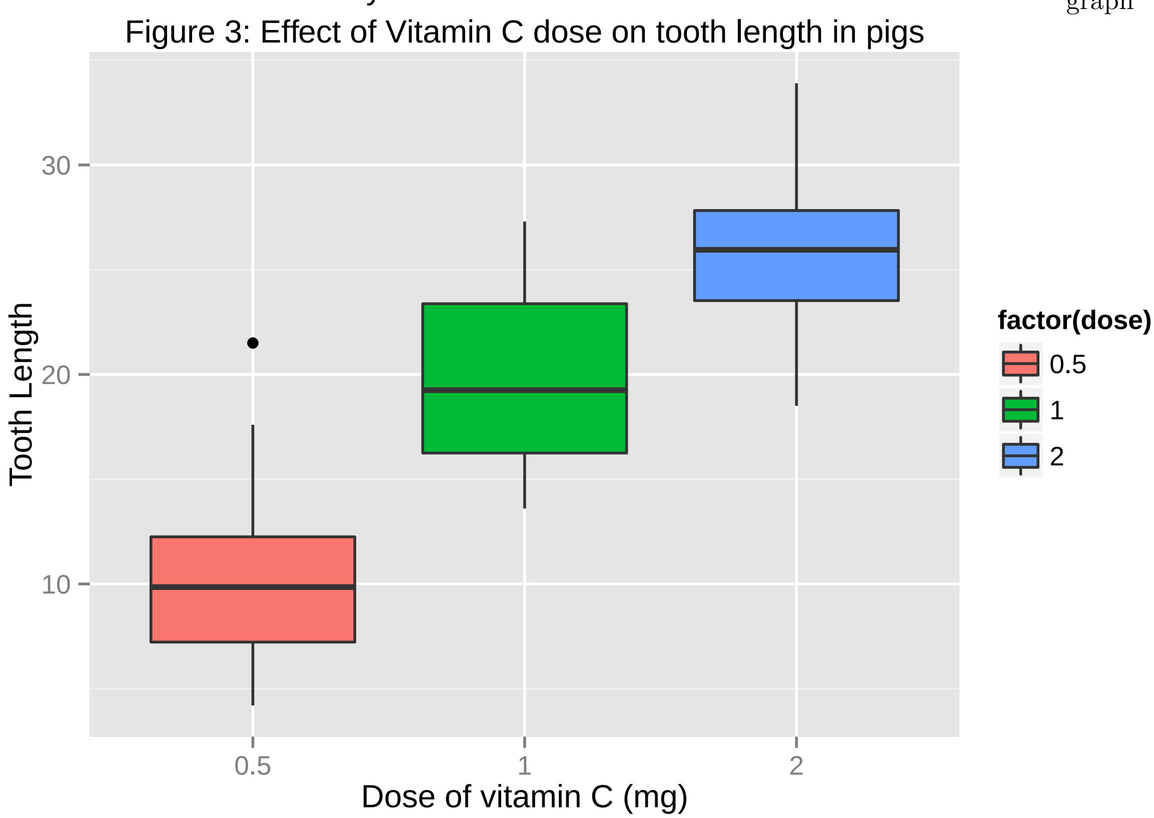


fig3-1.pdf