

Analysis of “ToothGrowth” Dataset from R Datasets

Brett A. Ochs

February 28, 2016

```
library(ggplot2); library(gridExtra); library(dplyr); library(datasets); library(DT)
```

Dataset description from datasets() package in R

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, and 2 mg) with each of two delivery methods (orange juice or ascorbic acid).

First we will load the dataset into a dataframe, rename columns for better understanding, and evaluate summary statistics.

```
## Load "ToothGrowth" dataset into dataframe
df <- ToothGrowth;
## Change variable names within dataframe for better understanding and summarize
names(df) <- c("Tooth.Length", "Delivery.Method", "Dosage"); summary(df)
```

```
##   Tooth.Length   Delivery.Method      Dosage
##   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
```

```
## Set up variables for dosage and delivery method.
dosages <- c(0.5, 1, 2); deliverys <- c("OJ", "VC")
```

Exploratory Data Analysis

Summary Table

Sometimes we may want to see a simple table with means and standard deviations for each treatment to better understand if differences may be likely. This table can be sorted and filtered to make individual comparisons to the tooth length means of that treatment, suggested to filter first by Delivery.Mechanism then by Dosage to see rows close together for Dosage.

```
df.summary <- tbl_df(df)%>%
  group_by(Delivery.Method, Dosage) %>%
  summarise(Count = n(),
            Mean.Tooth.Length = round(mean(Tooth.Length), digits=2),
            Stdev.Tooth.Length = round(sd(Tooth.Length), digits = 2))
datatable(df.summary, rownames = FALSE, filter="top",
          caption="Table 1. Interactive summary table of different dosage and deli
very mechanism statistics for Tooth Length")
```

Show

10

entries

Search:

Table 1. Interactive summary table of different dosage and delivery mechanism statistics for Tooth Length

Delivery.Method	Dosage	Count	Mean.Tooth.Length	Stdev.Tooth.Length
<div>All</div>	<div></div>	<div>.</div>	<div>All</div>	<div>All</div>
OJ	0.5	10	13.23	4.46
OJ	1	10	22.7	3.91
OJ	2	10	26.06	2.66
VC	0.5	10	7.98	2.75
VC	1	10	16.77	2.52
VC	2	10	26.14	4.8

Showing 1 to 6 of 6 entries

Previous

1

Next

We see very little difference between delivery mechanisms at the highest dosage, but do see some differences of interest between delivery mechanism for both dosage or 0.5 and 1.0.

Visuals

One of the best visual EDA for this type of dataset is to evaluate boxplots of the different dosages and delivery methods. To understand the visualization I like to facet by different variables for different look. I see trends within delivery mechanisms that the higher dosage administered, the longer the Tooth Length means. I do not see strong trends for differences in delivery methods either within or between dosages.

```

## Simple boxplot to see if tooth length trends are similar across delivery methods
within dosage
box1 <- ggplot(data=df) + geom_boxplot(aes(x=Delivery.Method, y=Tooth.Length, fill
=Delivery.Method)) +
  scale_fill_manual(name="Delivery\nMethod", breaks=c("OJ", "VC"), labels=c("O
J", "VC2"),
                    values=c("orangered3", "green4")) +
  labs(title="Fig.1- Boxplot of Tooth Length by\nDelivery Method Within Dosage",
x="Delivery Method", y="Tooth Length") +
  theme(axis.text.x=element_text(angle=30, vjust=0.5, size=8, face="bold"),
        axis.text.y=element_text(size=10, face="bold"), axis.title=element_text
(size=10, face="bold"),
        plot.title=element_text(size=12, face="bold"), legend.text=element_text
(size=6)) +
  facet_grid(. ~ Dosage)

## Simple boxplot to see if tooth length trends are similar across dosage within d
elivery methods
box2 <-ggplot(data=df) + geom_boxplot(aes(x=as.factor(Dosage), y=Tooth.Length, fil
l=Delivery.Method)) +
  scale_fill_manual(name="Delivery\nMethod",
                    breaks=c("OJ", "VC"),
                    labels=c("OJ", "VC2"),
                    values=c("orangered3", "green4")) +labs(title="Fig.1- Boxplo
t of Tooth Length by\nDosage Within Delivery Method", x="Dosage", y="Tooth Lengt
h") +
  theme(axis.text.x=element_text(angle=30, vjust=0.5, size=8, face="bold"),
        axis.text.y=element_text(size=10, face="bold"), axis.title=element_text
(size=10, face="bold"),
        plot.title=element_text(size=12, face="bold"), legend.text=element_text
(size=6)) +
  facet_grid(. ~ Delivery.Method)
grid.arrange(box1, box2, ncol=2)

```

Fig.1- Boxplot of Tooth Length by Delivery Method Within Dosage

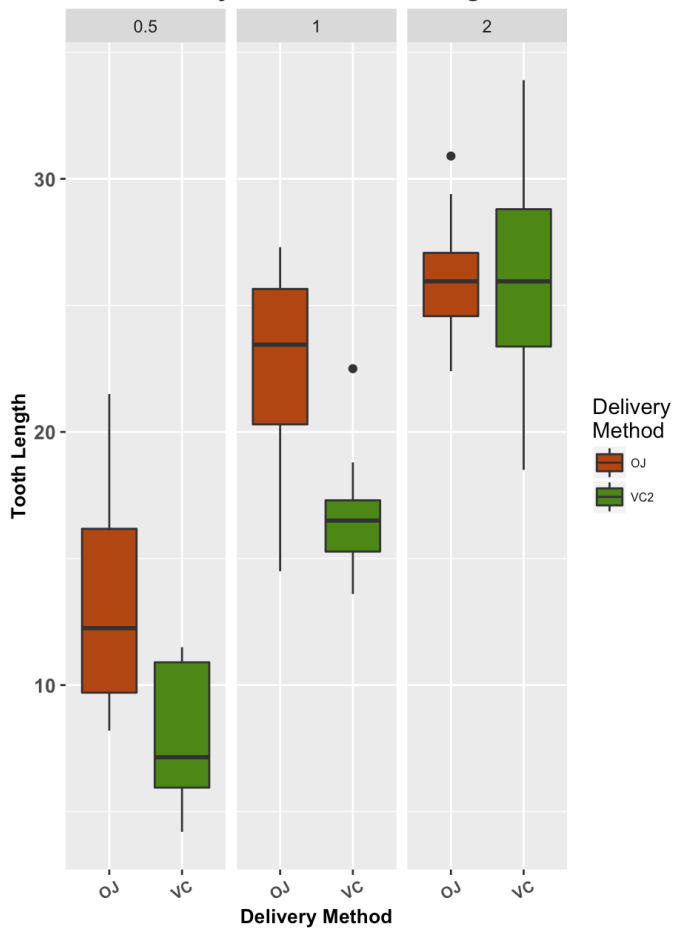
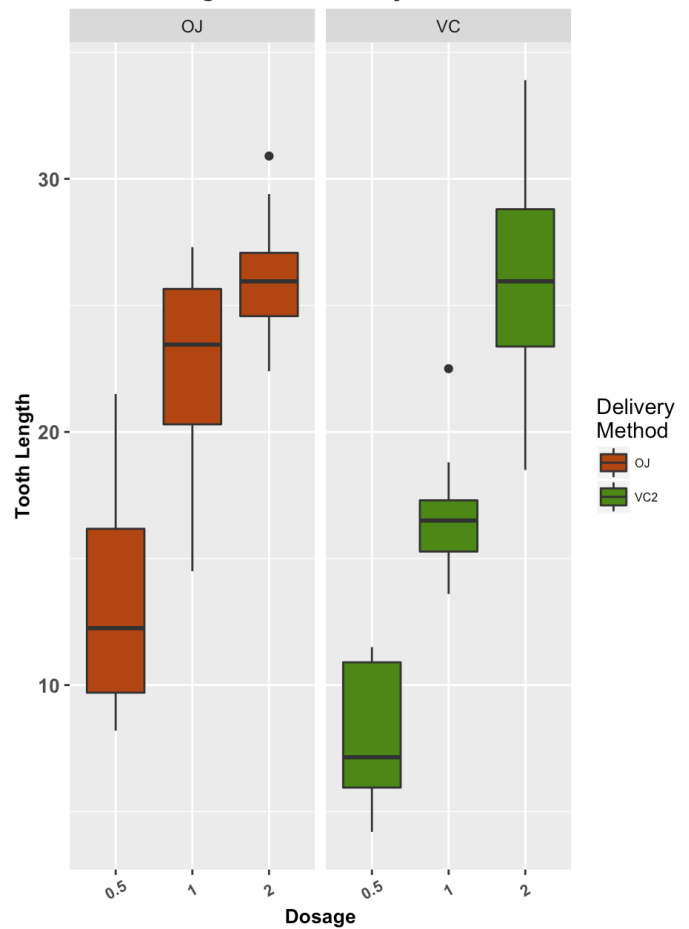


Fig.1- Boxplot of Tooth Length by Dosage Within Delivery Method



Hypothesis Testing

To get started lets do a simple t-test to determine if there are differences in delivery mechanism when we evaluate all dosage levels. Above we saw that visually it did not look like a trend that delivery mechanisms were different when looking at all dosage levels.

```
## Evalaute differences in delivery method for teeth length across dosage
delivery.t <- t.test(Tooth.Length ~ Delivery.Method, paired=FALSE, alternative="two.sided", data=df)
delivery.t
```

```
##
## Welch Two Sample t-test
##
## data: Tooth.Length by Delivery.Method
## 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
```

We get a p-value of **0.0606** so we fail to reject the Null Hypothesis that the means of delivery mechanism are equal across all dosage levels. We also see the 95% confidence interval is **-0.171** to **7.571** which encompasses zero (but barely so) which also would lead us to fail to reject Null Hypothesis.

Since the p-value is close to 0.05 though we will dig in further to dosage levels and see if there are differences between delivery mechanism within individual dosage levels. Below is code and resulting t-test information for comparison of delivery mechanism within each dosage level of 0.5 mg., 1 mg., and 2 mg. Lets set up 3 Null Hypothesis to test: 1. Means of delivery mechanisms are equal for dosage **0.5 mg**. 2. Means of delivery mechanisms are equal for dosage **1 mg**. 3. Means of delivery mechanisms are equal for dosage **2 mg**.

```
## Loop to evaluate effect of delivery method within same dosage
result <- vector("list", 6)
for(i in 1:length(dosages)){
  df.temp <- df[df$Dosage == dosages[i], ]
  t.test.temp <- t.test(Tooth.Length ~ Delivery.Method, data=df.temp)
  print(paste0("Comparison of delivery mechanism t-test of ", dosages[i], "mg dosage"))
  result[[i]] <- t.test.temp
  print(t.test.temp)
}
```

```
## [1] "Comparison of delivery mechanism t-test of 0.5mg dosage"
##
## Welch Two Sample t-test
##
## data: Tooth.Length by Delivery.Method
## 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
##
## [1] "Comparison of delivery mechanism t-test of 1mg dosage"
##
## Welch Two Sample t-test
##
## data: Tooth.Length by Delivery.Method
## 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
##
## [1] "Comparison of delivery mechanism t-test of 2mg dosage"
##
## Welch Two Sample t-test
##
## data: Tooth.Length by Delivery.Method
## 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
```

1. For dosage of **0.5 mg.** we reject the Null Hypothesis of equal delivery mechanism means since p-value of **0.00636** \leq **0.05**. In this case we do see differences in Delivery Mechanism within that low dose, similar to trends within boxplots above.
2. For dosage of **1 mg.** we reject the Null Hypothesis of equal delivery mechanism means since p-value of **0.00104** \leq **0.05**. In this case we do see differences in Delivery Mechanism within that middle dose, similar to trends within boxplots above.
3. For dosage of **2 mg.** we fail to reject the Null Hypothesis of equal delivery mechanism means since p-value of **0.96385** \geq **0.05**. In this case we do not see differences in Delivery Mechanism within that high dose, similar to trends within boxplots above.