An Analysis of ToothGrowth Data

Luke Coughlin

5/29/2020

Synopsis

This report gives a brief analysis of the tooth growth in guinea pigs based off the data from *Crapton, E. W.'s 1947 publication in The Journal of Nutrition.* We find that orange juice correlates to higher tooth growth in the sample of guinea pigs, and that higher doses also correlate to greater tooth growth.

Load Data

The `ToothGrowth` data set looks at the effect of vitamin C on tooth growth in guinea pigs. Each guinea pig received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, orange juice or ascorbic acid (a form of vitamin C). The length of their odontoblasts (the cells responsible for tooth growth) were then recorded. We will now load this data set.

Summary and Distribution

Below is a box-plot of the tooth growth data subsetted by dose level in each row and delivery method in each column(OJ for orange juice and VC for ascorbic acid). Following the plot is a table of some statistics from the data, which will be explained after.
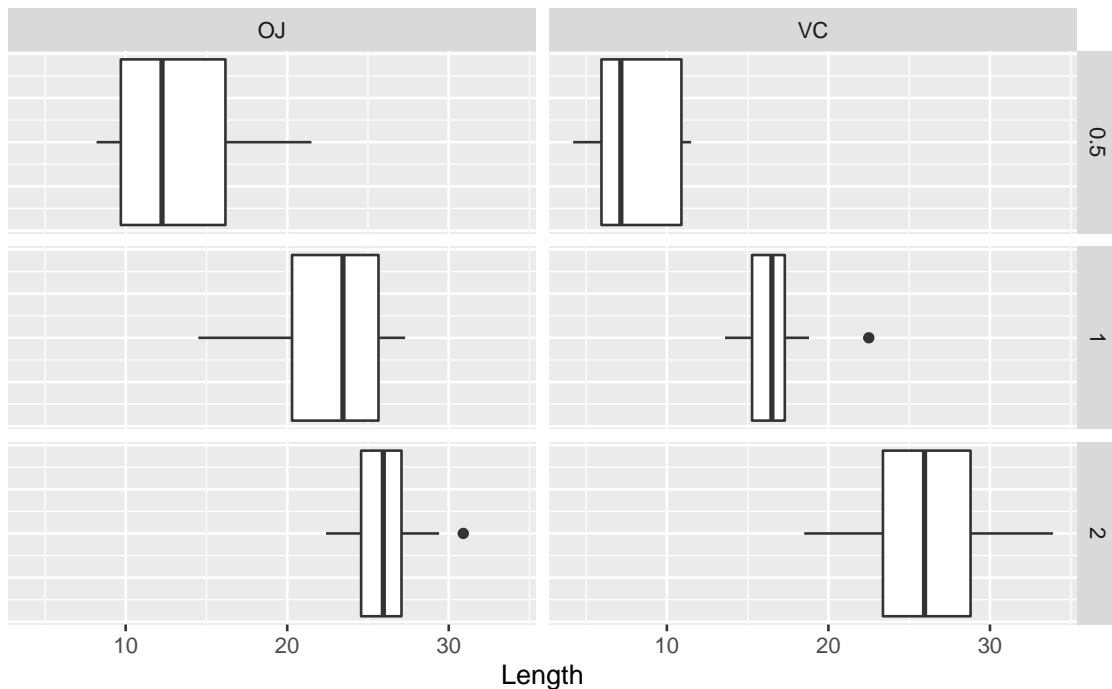
dose	supp	Count	Mean	Median	Range	Shaprio__Wilk__Test
0.5	OJ	10	13.23	12.25	[8.2, 21.5]	0.182
0.5	VC	10	7.98	7.15	[4.2, 11.5]	0.170
1.0	OJ	10	22.70	23.45	[14.5, 27.3]	0.415
1.0	VC	10	16.77	16.50	[13.6, 22.5]	0.270
2.0	OJ	10	26.06	25.95	[22.4, 30.9]	0.815
2.0	VC	10	26.14	25.95	[18.5, 33.9]	0.919

Explanation of Statistics:

`dose` & `supp` - These are both categorical variables describing the amount and source of vitamin C, respectively, given to each animal.

`Count` - This is how many observations are present in each group, all of them have 10 subjects.

`Mean` - This is the mean length of odontoblasts for each group.

`Median` - This is the median length of odontoblasts for each group.

`Range` - This is the range of lengths for each group.

`Shaprio_Wilk_Test` - This is the p-value of a **Shapiro-Wilk test of normality**. The null hypothesis in this test is that the sample comes from a normally distributed population. Since none of our subsets have a p-value less than 0.05 we can assume that the data is approximately normal.

Hypothesis Tests

We will compare tooth growth by overlaying the distribution of the three dose amounts by their delivery mechanism. We will also plot the respective confidence intervals for the population mean at $\alpha = 0.05$
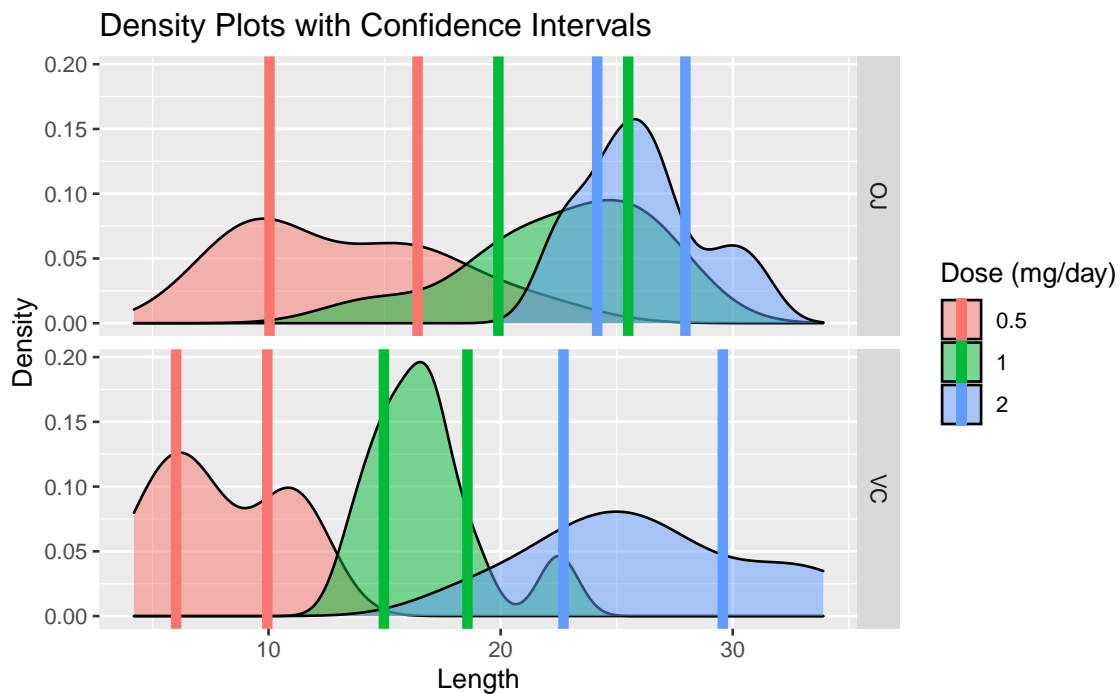
The above plots show the distribution of length for each delivery method by dose. We can say with 95% confidence that the true population mean lies in between each of the respective lines. With that in mind we can see that as more vitamin C is administered to a guinea pig the more the teeth grow. However, we cannot confidently say that administering 1 versus 2 mg/day of orange juice made a significant difference in the population's tooth growth.

As for the difference in affect between orange juice (OJ) and ascorbic acid (VC), for doses less than

2 mg/day, the orange juice correlated to observing a greater tooth growth. For a dose of 2 mg/day we cannot confidently state that either orange juice or ascorbic acid has a higher correlation to tooth growth in our population. We will now include all the confidence intervals for clarity.

Lower_Bound	Upper_Bound	supp	dose
6.02	9.94	VC	0.5
14.98	18.56	VC	1.0
22.71	29.57	VC	2.0
10.04	16.42	OJ	0.5
19.91	25.49	OJ	1.0
24.17	27.95	OJ	2.0

Conclusions

Under the assumption that the data is representative of the population and normally distributed, it appears that dosages of orange juice correlates to a greater tooth growth.

More observations could be done to see at what point higher dosages begin having diminishing returns, however these data didn't contain enough dose levels to investigate this. These data did not contain a control so investigating how these doses alter a baseline could be further researched with observations of a control group.

Appendix

Luke Coughlin

5/29/2020

```
## ----setup, include=FALSE--------------------------------------------------
knitr::opts_chunk$set(echo = FALSE)


## ----Load_Data, message = FALSE--------------------------------------------
library(datasets); library(tidyverse)
library(gridExtra); library(knitr)
phi <- (1+sqrt(5))/2 #For fig dimensions
data("ToothGrowth")
dat <- ToothGrowth


## ----Summary, fig.height = 4, fig.width= 4*phi-----------------------------
grouped <- group_by(dat, dose, supp)
ggplot(grouped, aes(len)) +
        geom_boxplot() +
        theme(axis.ticks.y = element_blank(),
              axis.text.y = element_blank()) +
        facet_grid(dose~supp) +
        xlab("Length")
```
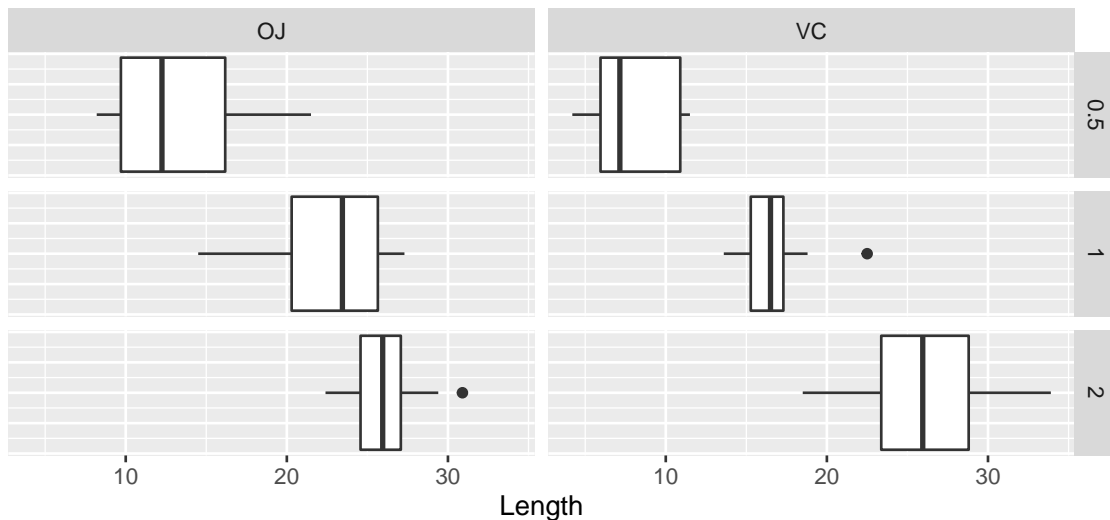


```
summary <- summarize(grouped,
        Count = sum(!is.na(len)), Mean = mean(len), Median = median(len),
        Range = paste0("[",range(len)[1], ", ", range(len)[2], "]"),
        Shaprio_Wilk_Test = round(shapiro.test(len)$p.value, 3))
kable(summary)
```

dose	supp	Count	Mean	Median	Range	Shaprio_Wilk_Test
0.5	OJ	10	13.23	12.25	[8.2, 21.5]	0.182

dose	supp	Count	Mean	Median	Range	Shaprio_Wilk_Test
0.5	VC	10	7.98	7.15	[4.2, 11.5]	0.170
1.0	OJ	10	22.70	23.45	[14.5, 27.3]	0.415
1.0	VC	10	16.77	16.50	[13.6, 22.5]	0.270
2.0	OJ	10	26.06	25.95	[22.4, 30.9]	0.815
2.0	VC	10	26.14	25.95	[18.5, 33.9]	0.919

```
## ----t_tests----------------------------------------------------------------
subsets <- matrix(rep(rep(NA, 10), 6), nrow = 6, ncol = 10)
CIs <- matrix(rep(rep(NA, 2), 6), nrow = 6, ncol = 2)
a <- 0.05
for (i in 0:5) {
        subsets[i + 1,] <- seq(1 + 10*i, 10*(i + 1), by = 1)
        con <- t.test(dat[subsets[i + 1,],]$len,
                        alternative = "two.sided", conf.level = 1 - a)$conf.int
        CIs[i + 1, 1] <- con[1]
        CIs[i + 1, 2] <- con[2]
}
CIs <- data.frame(LB = CIs[,1], UB = CIs[,2],
                  supp = rep(c("VC", "OJ"), each = 3),
                  dose = rep(c(0.5,1.0,2.0), times = 2))

## ----Denisty_Plot, fig.height = 4, fig.width = 4*phi------------------------
ggplot(grouped, aes(len)) +
        facet_grid(supp~.) +
        geom_density(aes(group = dose, fill = as.factor(dose)), alpha = 0.5) +
        geom_vline(data = CIs[CIs$supp == "VC",], lwd = 2,
                  aes(xintercept = LB,
                        group = dose, color = as.factor(dose))) +
        geom_vline(data = CIs[CIs$supp == "VC",], lwd = 2,
                  aes(xintercept = UB,
                        group = dose, color = as.factor(dose))) +
        geom_vline(data = CIs[CIs$supp == "OJ",], lwd = 2,
                  aes(xintercept = LB,
                        group = dose, color = as.factor(dose))) +
        geom_vline(data = CIs[CIs$supp == "OJ",], lwd = 2,
                  aes(xintercept = UB,
                        group = dose, color = as.factor(dose))) +
        labs(x = "Length", y = "Density",
            fill = "Dose (mg/day)", color = "Dose (mg/day)") +
        ggtitle("Density Plots with Confidence Intervals")
```

Density Plots with Confidence Intervals
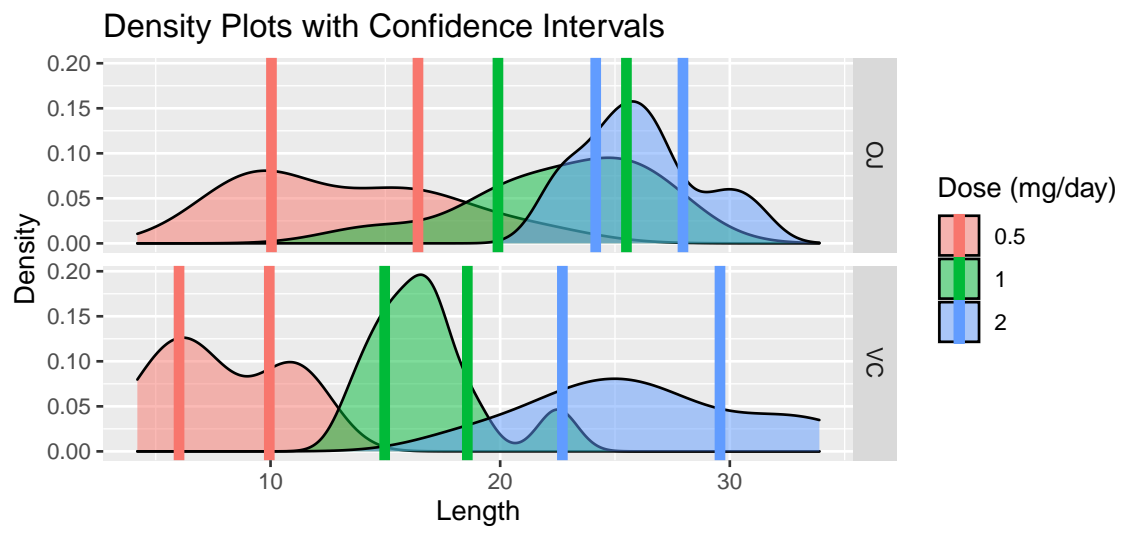
```
## -----------------------------------------------------------------------------
#Ceiling and floor-ing values to maintain confidence level
CIs$LB <- ceiling(CIs$LB*100)/100
CIs$UB <- floor(CIs$UB*100)/100
names(CIs) <- c("Lower_Bound", "Upper_Bound",
                "supp", "dose")
kable(CIs)
```

Lower_Bound	Upper_Bound	supp	dose
6.02	9.94	VC	0.5
14.98	18.56	VC	1.0
22.71	29.57	VC	2.0
10.04	16.42	OJ	0.5
19.91	25.49	OJ	1.0
24.17	27.95	OJ	2.0