Statistical Inference Course Project

Part 2: Basic Inferential Data Analysis

Lief Pagalan

January 3, 2023

Overview

Now in the second portion of the project, I'm going to analyze the ToothGrowth data in the R datasets package.

Load Data

Load the ToothGrowth data and perform some basic exploratory data analyses.

```
library(tidyverse)
library(skimr)
data("ToothGrowth")
```

The response variable is the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. Each animal 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 and coded as VC). The data frame contains 60 observations on 3 variables.

Descriptive Analysis

```
skim_without_charts(ToothGrowth)
```

Table 1: Data summary

Name	ToothGrowth
Number of rows	60
Number of columns	3
Column type frequency:	
factor	1
numeric	2
Group variables	None

Variable type: factor

skim_variable	n_missing	complete_rate	ordered	n_unique	top_counts
supp	0	1	FALSE	2	OJ: 30, VC: 30

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
len	0	1	18.81	7.65	4.2	13.07	19.25	25.27	33.9
dose	0	1	1.17	0.63	0.5	0.50	1.00	2.00	2.0

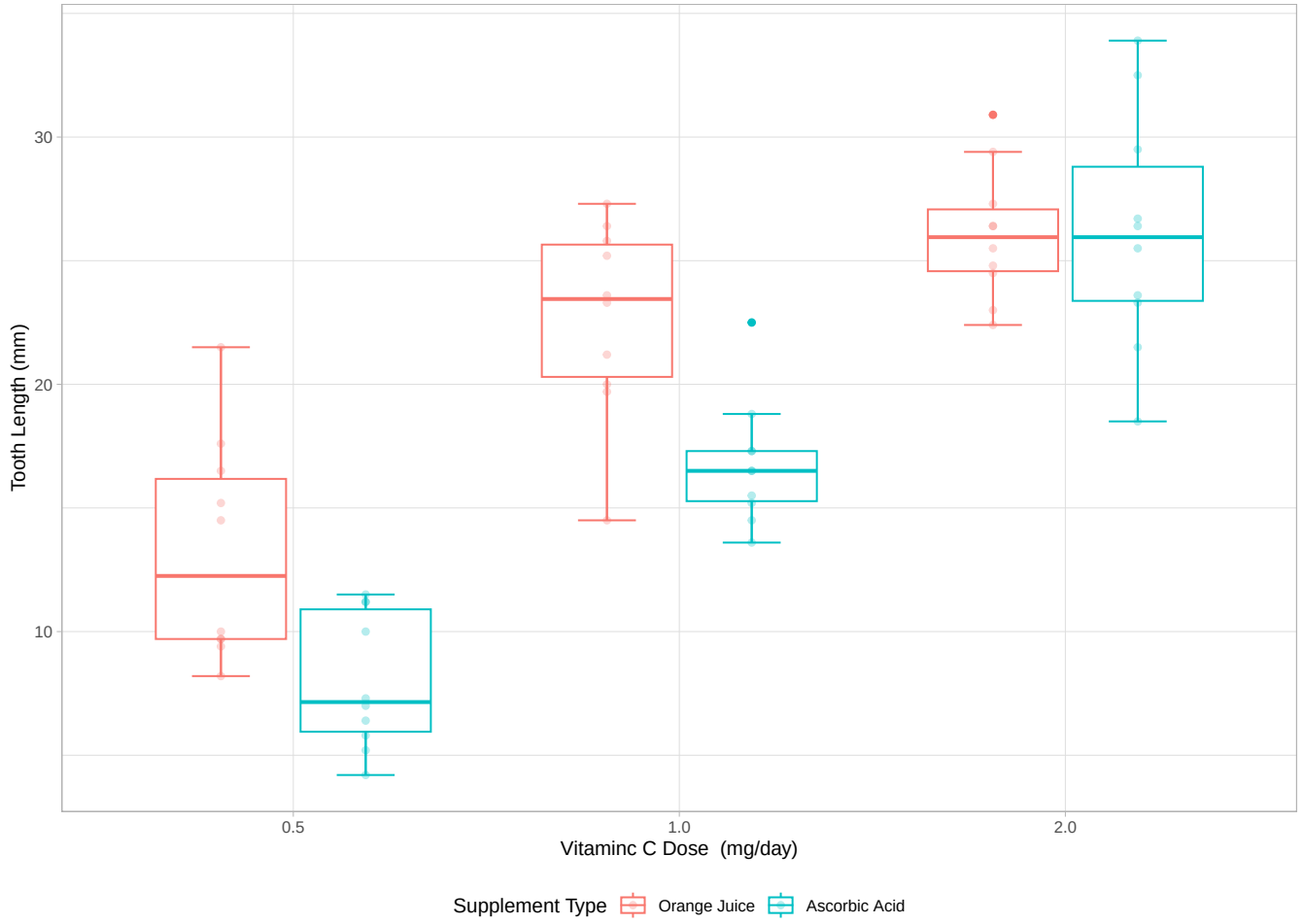

Figure 1: Odontoblasts Length by Vitaminc C Dose and Supplement Type

Hypothesis Testing

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).

The t-tests is used to determine whether the means of two groups are equal. The null hypothesis is that the two means are equal, and the alternative is that they are not equal.

Orange Juice vs. Ascorbic Acid: 0.5 mg/day Dose Level

```
(t_05 <- t.test(len ~ supp, data = ToothGrowth, subset = dose == "0.5", var.equal = TRUE))
```

```
    Two Sample t-test

data:  len by supp
t = 3.1697, df = 18, p-value = 0.005304
alternative hypothesis: true difference in means between group OJ and group VC is not equal to 0
95 percent confidence interval:
 1.770262 8.729738
sample estimates:
mean in group OJ mean in group VC
           13.23             7.98
```

Orange Juice vs. Ascorbic Acid: 1.0 mg/day Dose Level

```
(t_10 <- t.test(len ~ supp, data = ToothGrowth, subset = dose == "1.0", var.equal = TRUE))
```

```
    Two Sample t-test

data:  len by supp
t = 4.0328, df = 18, p-value = 0.0007807
alternative hypothesis: true difference in means between group OJ and group VC is not equal to 0
95 percent confidence interval:
 2.840692 9.019308
sample estimates:
mean in group OJ mean in group VC
           22.70             16.77
```

Orange Juice vs. Ascorbic Acid: 2.0 mg/day Dose Level

```
(t_20 <- t.test(len ~ supp, data = ToothGrowth, subset = dose == "2.0", var.equal = TRUE))
```

```
    Two Sample t-test

data:  len by supp
t = -0.046136, df = 18, p-value = 0.9637
alternative hypothesis: true difference in means between group OJ and group VC is not equal to 0
95 percent confidence interval:
 -3.722999  3.562999
sample estimates:
mean in group OJ mean in group VC
           26.06             26.14
```

Conclusions

At lower dose levels of vitamin C, 0.5 and 1.0 mg/day, orange juice was on average more effective than ascorbic acid at increasing tooth length (p-values: < 0.01). However, at the 2.0 mg/day dose level there was no difference in average tooth length between orange juice or ascorbic acid (p-value: 0.96). (See Table 4.)

Table 4: T-Test for Difference in Group Means

	Tooth Length (mm)				
Dose (mg/day)	Orange Juice	Ascorbic Acid	T-Test Statistic (95% CI)	P-Value	Intepretation
0.5	13.23	7.98	3.17 (1.77, 8.73)	0.0053	Difference in Means
1.0	22.70	16.77	4.03 (2.84, 9.02)	0.0008	Difference in Means
2.0	26.06	26.14	-0.05 (-3.72, 3.56)	0.9637	No Difference in Means

Assumptions

The assumption for the t-test is that both groups are sampled from normal distributions with equal variances. There were no significant differences in tooth length variance between orange juice and ascorbic acid for each dose level (see Table 5), meeting the equal variance requirement of the t-test. Equal variance was tested using the F-test, where $H_0 : \sigma_A^2 = \sigma_B^2$ (i.e., variances are equal) and $H_a : \sigma_A^2 \neq \sigma_B^2$ (i.e., variances are not equal).

Table 5: F-Test for Equal Variance

Vitamin C Dose (mg/day)	Ratio of Variance (95% CI)	P-Value	Interpretation
0.5	2.64 (0.65, 10.61)	0.16	Equal variance
1.0	2.42 (0.60, 9.73)	0.20	Equal variance
2.0	0.31 (0.07, 1.23)	0.09	Equal variance