

Project Report for Basic Inferential Data Analysis

Yukai Zou

1. Overview

This report is the second part of the Statistical Inference course project. In this report, the ToothGrowth data in the R datasets will be analyzed using some basic inferential statistics. Exploratory data analyses and summary of the data are first given. After that, comparisons of tooth growth by supp and dose are performed respectively, in which different assumptions, null hypotheses (H_0), and alternative hypotheses (H_a) are set up and tested using T test. Finally, conclusions are made based on the T statistics.

2. Exploratory Analyses on ToothGrowth Data

For exploratory plots, see Appendix for reference. Based on the plots, in terms of different supp (Figure 1), VC and OJ seem to have different effects on the length of tooth growth, though it is not very easy to distinguish; in terms of dose (Figure 2), a higher dose seems to increase the length of tooth growth. Both observations require two-sided hypothesis testing, which will be explained in details later.

3. Basic Summary of the Data

3.1 General information:

##	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

3.2 Basic summary of tooth growth by supp:

##	\$OJ					
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	8.20	15.52	22.70	20.66	25.72	30.90
##						
##	\$VC					
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	4.20	11.20	16.50	16.96	23.10	33.90

3.3 Basic summary of tooth growth by dose:

```
## $`0.5`
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##    4.200   7.225   9.850  10.600  12.250  21.500
##
## $`1`
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##    13.60   16.25   19.25   19.74   23.38   27.30
##
## $`2`
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##    18.50   23.52   25.95   26.10   27.83   33.90
```

4. Comparisons of Tooth Growth by Supp

Assuming that “len” is an iid variable, the data of tooth growth by VC and OJ are both normally distributed, and the population variances are unequal.

H_0 : The length of tooth growth by VC is equivalent to the length of tooth growth by OJ ($VC = OJ$).

H_a : The length of tooth growth by VC is not equivalent to the length of tooth growth by OJ ($VC \neq OJ$).

Hypothesis testing:

```
##
## Welch Two Sample t-test
##
## data:  OJ and VC
## 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 of x mean of y
## 20.66333 16.96333
```

According to the statistical test, the confidence interval did not exclude zero, the absolute value of the test statistic is not greater than $qt(.975, 55.309) = 2.0038$, and the p-value is not less than the 0.05 threshold.

Conclusion: The null hypothesis is not rejected. The length of tooth growth by VC is equivalent to the length of tooth growth by OJ ($VC = OJ$).

5. Comparisons of Tooth Growth by Dose (Need replacement for one-sided test)

5.1 $Dose_{0.5}$ v.s. $Dose_{1.0}$

Assuming that “len” is an iid variable, the data of tooth growth by dose 0.5 and dose 1.0 are normally distributed, and the population variances are equal.

H_0 : The length of tooth growth by a dose of 0.5 is equivalent to the length of tooth growth by a dose of 1.0 ($Dose_{0.5} = Dose_{1.0}$).

H_a : The length of tooth growth by a dose of 0.5 is not equivalent to the length of tooth growth by a dose of 1.0 ($Dose_{0.5} \neq Dose_{1.0}$).

Hypothesis testing:

```
##
## Two Sample t-test
##
## data: len by dose
## t = -6.4766, df = 38, p-value = 1.266e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -11.983748 -6.276252
## sample estimates:
## mean in group 0.5 mean in group 1
## 10.605 19.735
```

According to the statistical test, both the lower and upper limit of the confidence interval is smaller than zero, the absolute value of the test statistic is much smaller than $qt(.025, 38) = -2.0244$, and the p-value is much smaller than the 0.05 threshold.

Conclusion: The null hypothesis is rejected. The length of tooth growth by a dose of 0.5 is not equivalent to the length of tooth growth by a dose of 1.0 ($Dose_{0.5} \neq Dose_{1.0}$).

5.2 $Dose_{0.5}$ v.s. $Dose_{2.0}$

Assuming that “len” is an iid variable, the data of tooth growth by dose 0.5 and dose 2.0 are normally distributed, and the population variances are not equal.

H_0 : The length of tooth growth by a dose of 0.5 is equivalent to the length of tooth growth by a dose of 2.0 ($Dose_{0.5} = Dose_{2.0}$).

H_a : The length of tooth growth by a dose of 0.5 is not equivalent to the length of tooth growth by a dose of 2.0 ($Dose_{0.5} \neq Dose_{2.0}$).

Hypothesis testing:

```
##
## 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
```

According to the statistical test, both the lower and upper limit of the confidence interval is smaller than zero, the absolute value of the test statistic is much smaller than $qt(.025, 36.883) = -2.0264$, and the p-value is much smaller than the 0.05 threshold.

Conclusion: The length of tooth growth by a dose of 0.5 is not equivalent to the length of tooth growth by a dose of 2.0 ($Dose_{0.5} \neq Dose_{2.0}$).

5.3 $Dose_{1.0}$ v.s. $Dose_{2.0}$

Assuming that “len” is an iid variable, the data of tooth growth by dose 1.0 and dose 2.0 are normally distributed, and the population variances are not equal.

H_0 : The length of tooth growth by a dose of 1.0 is equivalent to the length of tooth growth by a dose of 2.0 ($Dose_{1.0} = Dose_{2.0}$).

H_a : The length of tooth growth by a dose of 1.0 is not equivalent to the length of tooth growth by a dose of 2.0 ($Dose_{1.0} \neq Dose_{2.0}$).

Hypothesis testing:

```
##
## 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
```

According to the statistical test, both the lower and upper limit of the confidence interval is smaller than zero, the absolute value of the test statistic is smaller than $qt(.025, 37.101) = -2.0260$, and the p-value is much smaller than the 0.05 threshold.

Conclusion: The null hypothesis is rejected. The length of tooth growth by a dose of 1.0 is not equivalent to the length of tooth growth by a dose of 2.0 ($Dose_{1.0} \neq Dose_{2.0}$).

Appendix

Exploratory Plots

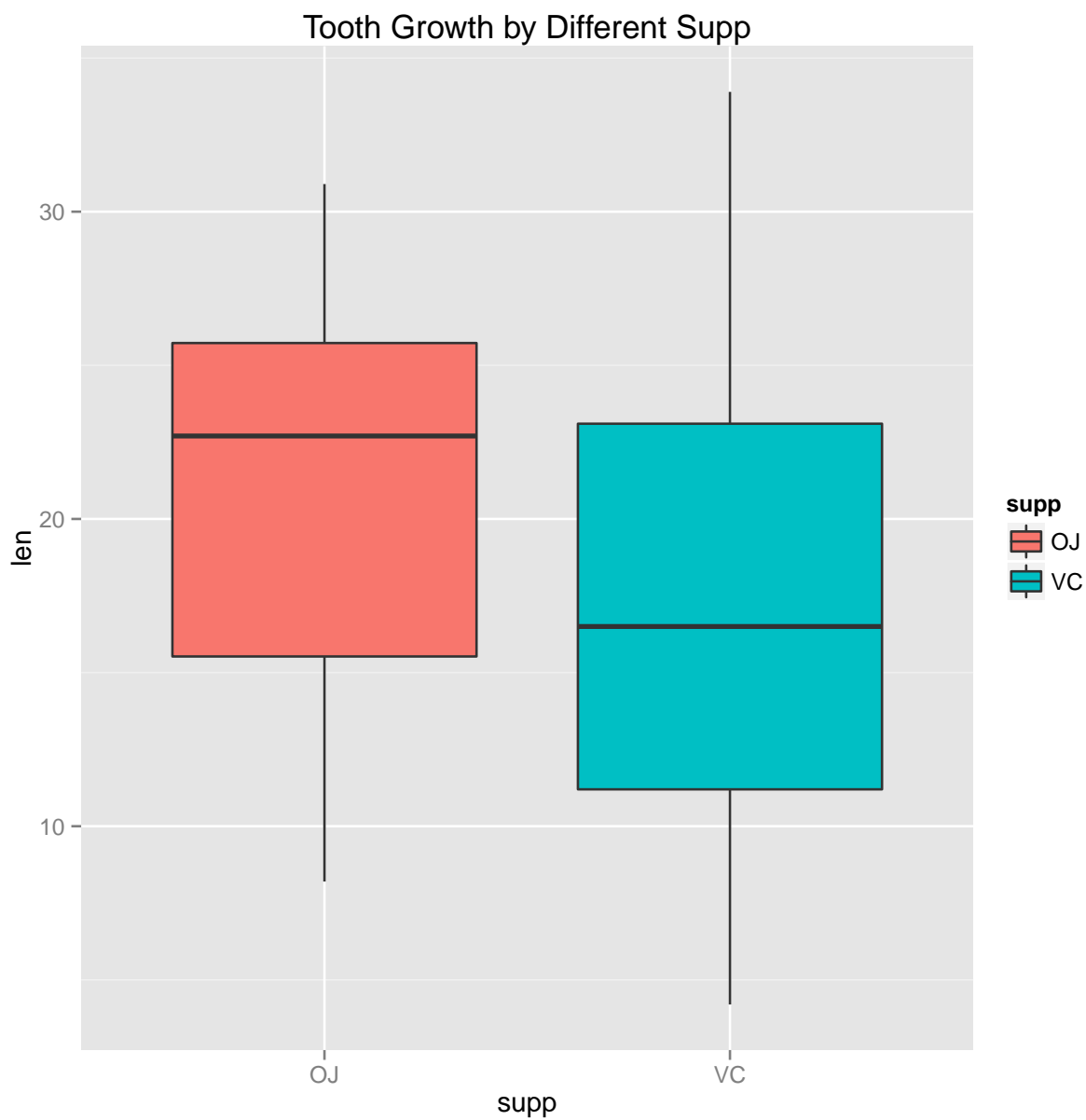


Figure 1: Tooth Growth by Different Supp

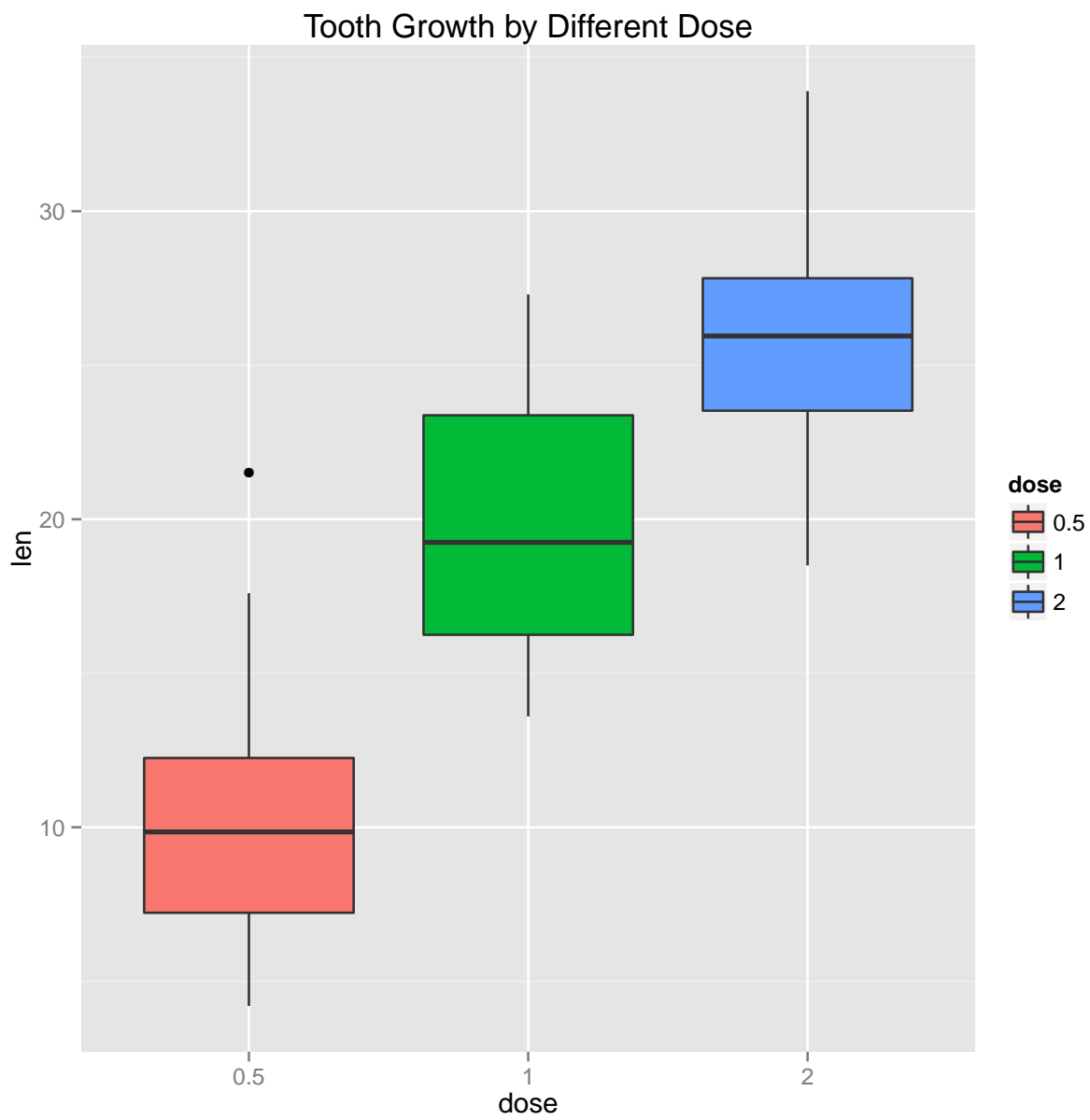


Figure 2: Tooth Growth by Different Dose