

Project Report for Basic Inferential Data Analysis

Yukai Zou

Overview

This report is the second part of the Coursera 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 will be first given. After that, comparisons of tooth growth by supp and dose will be performed respectively, with the relevant confidence intervals calculated; different hypotheses and assumptions will be formulated and tested using T statistics. Conclusions will be made based on the statistical results.

Exploratory Analyses on ToothGrowth Data

For exploratory plots, see Appendix for reference. Based on the plots, in terms of supplements (Figure 1), OJ seems to contribute to a higher tooth growth compared with VC; in terms of dose (Figure 2), a higher dose seems to increase the length of tooth growth. Both observations require hypothesis testing.

Basic Summary of the Data

##	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

Basic Summary of Tooth Growth by Supp

Basic Summary of Tooth Growth by Dose

Comparisons of Tooth Growth by Supp

Assuming that the data of tooth growth by OJ and VC are both normally distributed, and the population variances are unequal.

H0: there is no difference between tooth growth by VC and tooth growth by OJ ($VC = OJ$).

Ha: the tooth growth by VC is larger than the tooth growth by OJ ($VC > OJ$).

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, and the p-value is not less than 0.05.

Conclusion: The null hypothesis is not rejected. There is no difference between tooth growth by VC and tooth growth by OJ.

Comparisons of Tooth Growth by Dose

Dose 0.5 v.s. Dose 1.0

Assuming that the data of tooth growth by dose 0.5 and dose 1.0 are normally distributed, and the population variances are equal.

H₀: there is no difference between tooth growth by dose 0.5 and tooth growth by dose 1.0 (dose 0.5 = dose 1.0).

H_a: the tooth growth by dose 0.5 is smaller than the tooth growth by dose 1.0 (dose 0.5 < 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, and the p-value is much smaller than 0.05.

Conclusion: The null hypothesis is rejected. The tooth growth by dose 0.5 is smaller than the tooth growth by dose 1.0.

Dose 0.5 v.s. Dose 2.0

Assuming that the data of tooth growth by dose 0.5 and dose 2.0 are normally distributed, and the population variances are not equal.

H0: there is no difference between tooth growth by dose 0.5 and tooth growth by dose 2.0 (dose 0.5 = dose 2.0).

Ha: the tooth growth by dose 0.5 is smaller than the tooth growth by dose 2.0 (dose 0.5 < 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, and the p-value is much smaller than 0.05.

Conclusion: The null hypothesis is rejected. The tooth growth by dose 0.5 is smaller than the tooth growth by dose 2.0.

Dose 0.5 v.s. Dose 2.0

Assuming that the data of tooth growth by dose 1.0 and dose 2.0 are normally distributed, and the population variances are not equal.

H0: there is no difference between tooth growth by dose 1.0 and tooth growth by dose 2.0 (dose 1.0 = dose 2.0).

Ha: the tooth growth by dose 1.0 is smaller than the tooth growth by dose 2.0 (dose 1.0 < 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, and the p-value is much smaller than 0.05.

Conclusion: The null hypothesis is rejected. The tooth growth by dose 1.0 is smaller than the tooth growth by dose 2.0.

Appendix

R Codes

To be added

Exploratory Plots

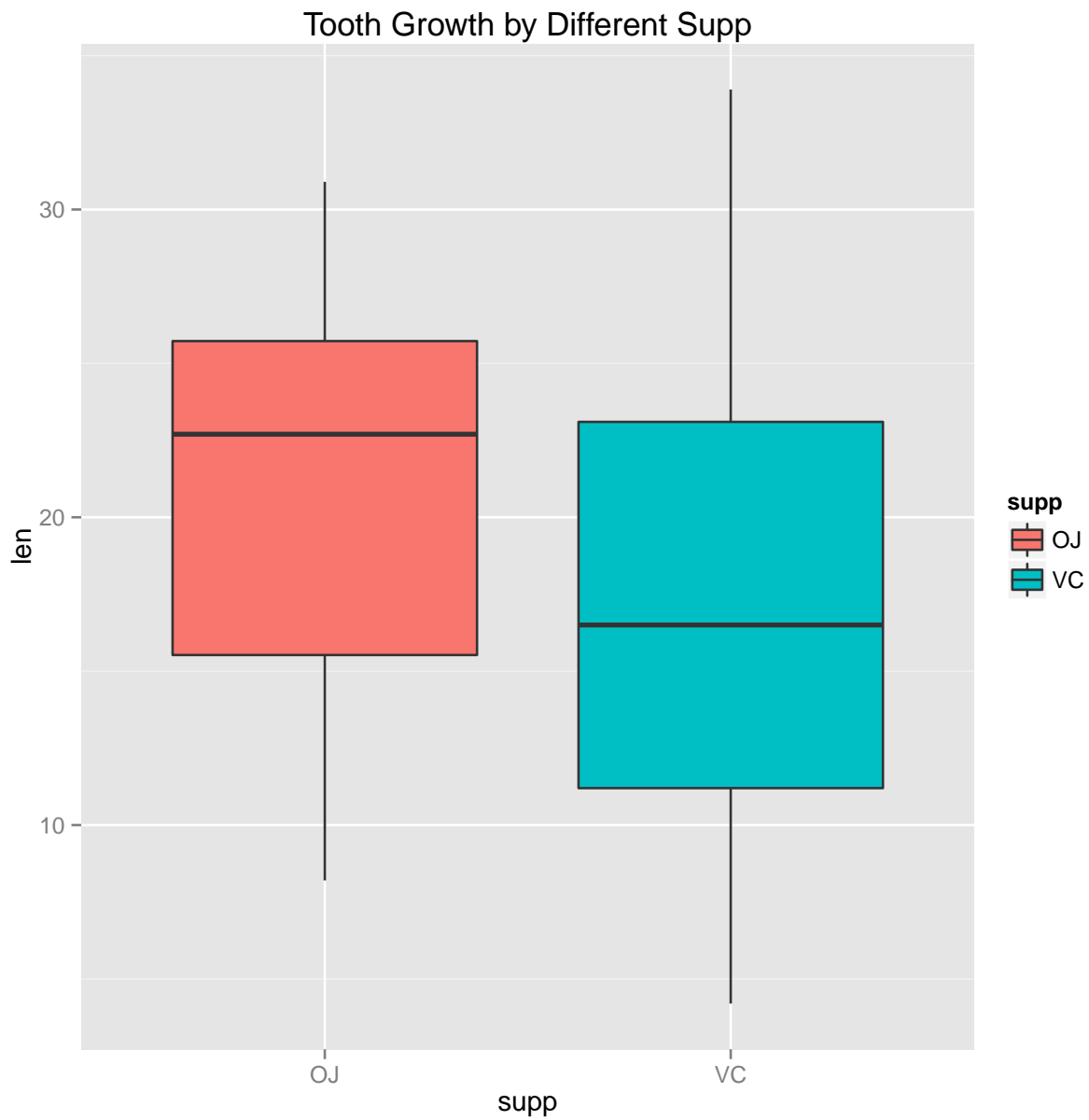


Figure 1: Tooth Growth by Different Supp

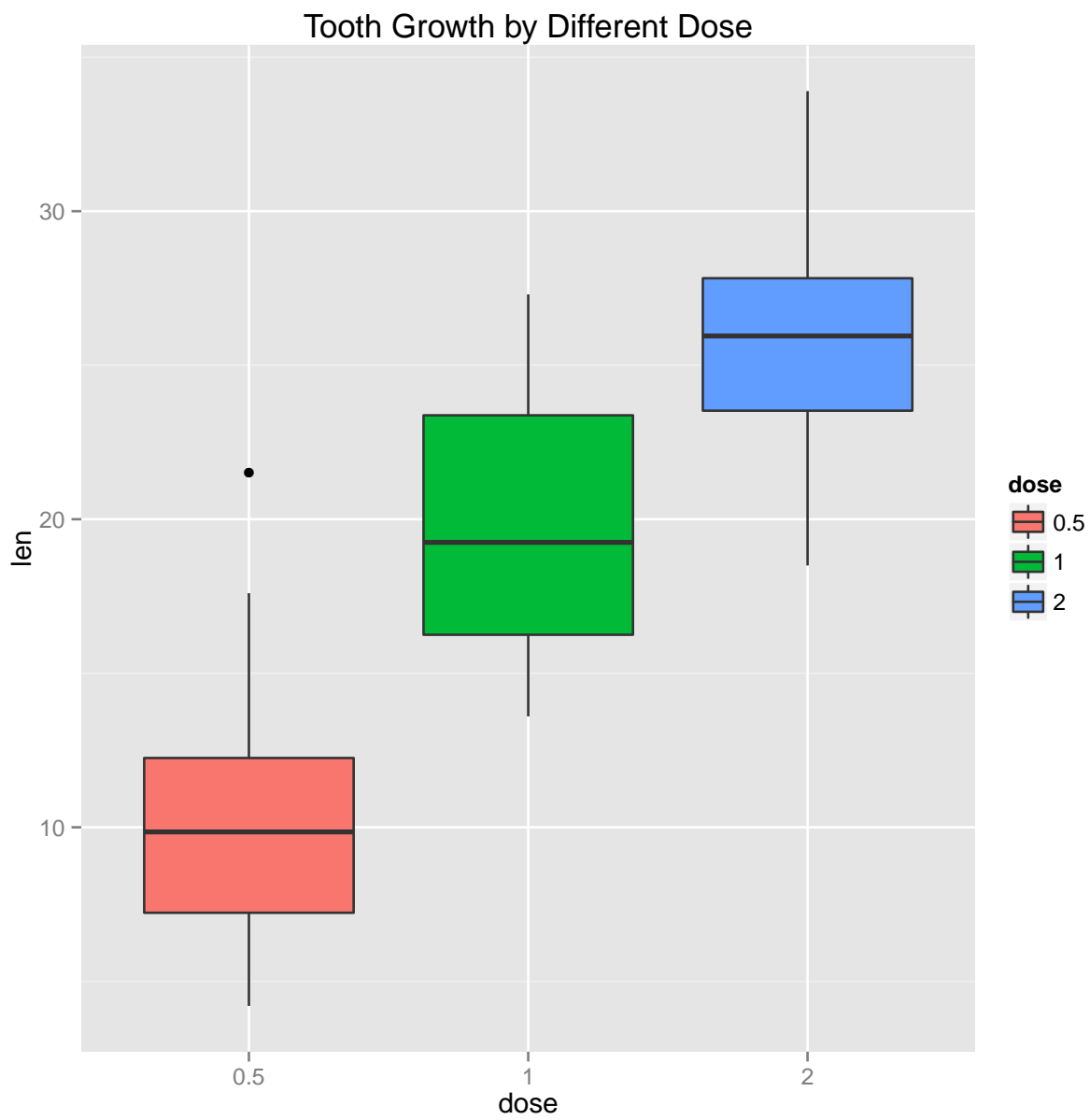


Figure 2: Tooth Growth by Different Dose