

Statistical_Infer_Part2.Rmd

Joe Okelly

18/09/2024

Instructions

1. Load the ToothGrowth data and perform some basic exploratory data analyses
2. Provide a basic summary of the data.
3. 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)
4. State your conclusions and the assumptions needed for your conclusions.

Exploratory data Analysis

First me load the packages, and dataset

```
library(ggplot2)
library(knitr)
library(datasets)
```

Load the ToothGrowth data and perform basic Exploratory Data Analysis

```
data(ToothGrowth)
str(ToothGrowth)
```

```
## 'data.frame':    60 obs. of  3 variables:
## $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
## $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
## $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

```
head(ToothGrowth, 4)
```

```
##      len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
```

```
tail(ToothGrowth, 4)
```

```
##      len supp dose
## 57 26.4   OJ    2
## 58 27.3   OJ    2
## 59 29.4   OJ    2
## 60 23.0   OJ    2
```

Calculate the summary of the data

```
summary(ToothGrowth)
```

```
##      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
```

Calculate the mean of the length

```
suppl_mean = split(ToothGrowth$len, ToothGrowth$supp)
sapply(suppl_mean, mean)
```

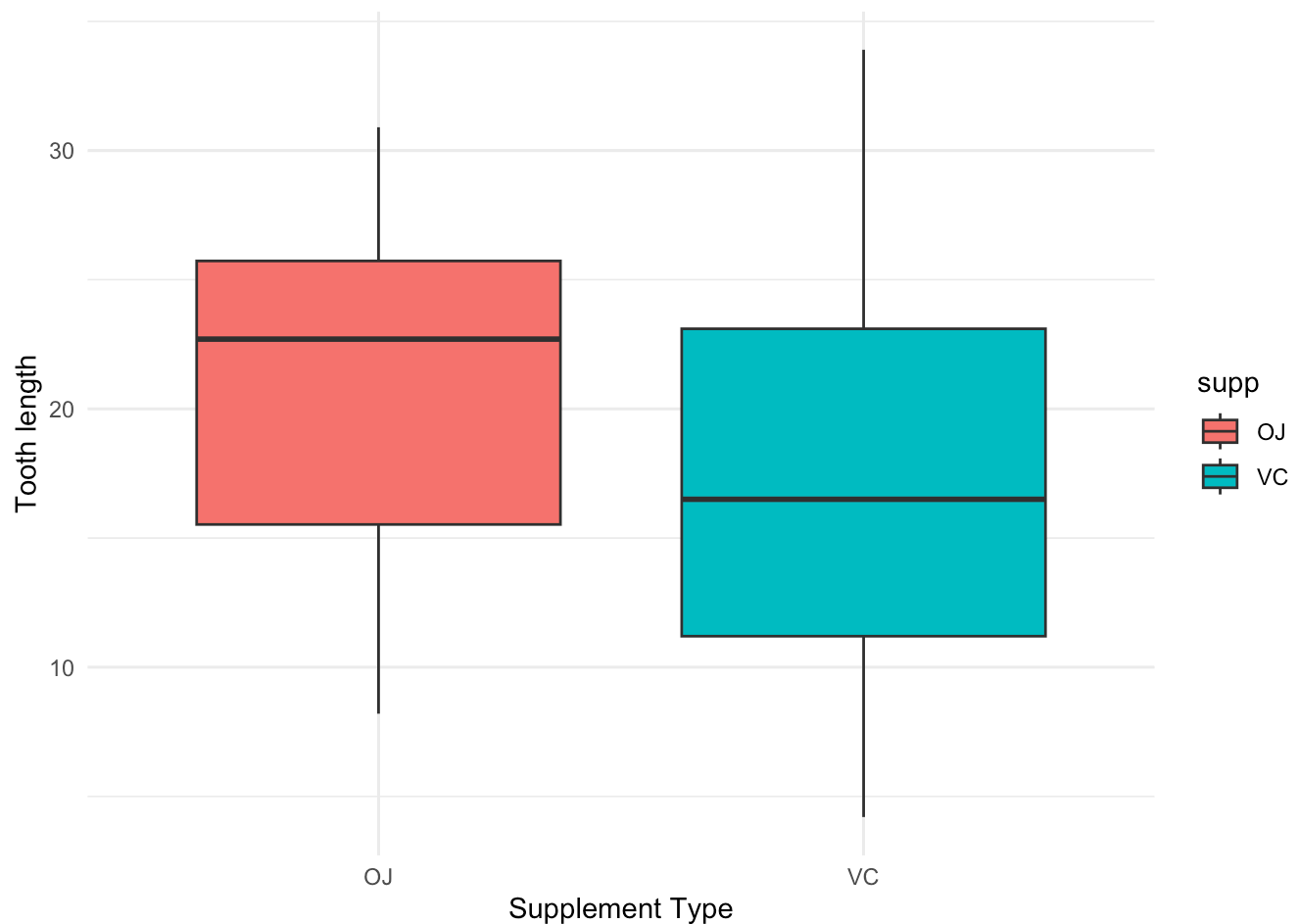
```
##      OJ      VC
## 20.66333 16.96333
```

```
suppl_mean
```

```
## $OJ
## [1] 15.2 21.5 17.6  9.7 14.5 10.0  8.2  9.4 16.5  9.7 19.7 23.3 23.6 26.4 20.0
## [16] 25.2 25.8 21.2 14.5 27.3 25.5 26.4 22.4 24.5 24.8 30.9 26.4 27.3 29.4 23.0
##
## $VC
## [1]  4.2 11.5  7.3  5.8  6.4 10.0 11.2 11.2  5.2  7.0 16.5 16.5 15.2 17.3 22.5
## [16] 17.3 13.6 14.5 18.8 15.5 23.6 18.5 33.9 25.5 26.4 32.5 26.7 21.5 23.3 29.5
```

Basic Exploratory Analysis, Graph below

```
ggplot(aes(x=supp, y=len), data=ToothGrowth) + geom_boxplot(aes(fill=supp))+
  xlab("Supplement Type") + ylab("Tooth length") +
  theme_minimal()
```



Above Plot gives you a basic exploratory visualization showing how tooth length varies with supplement type (supp).

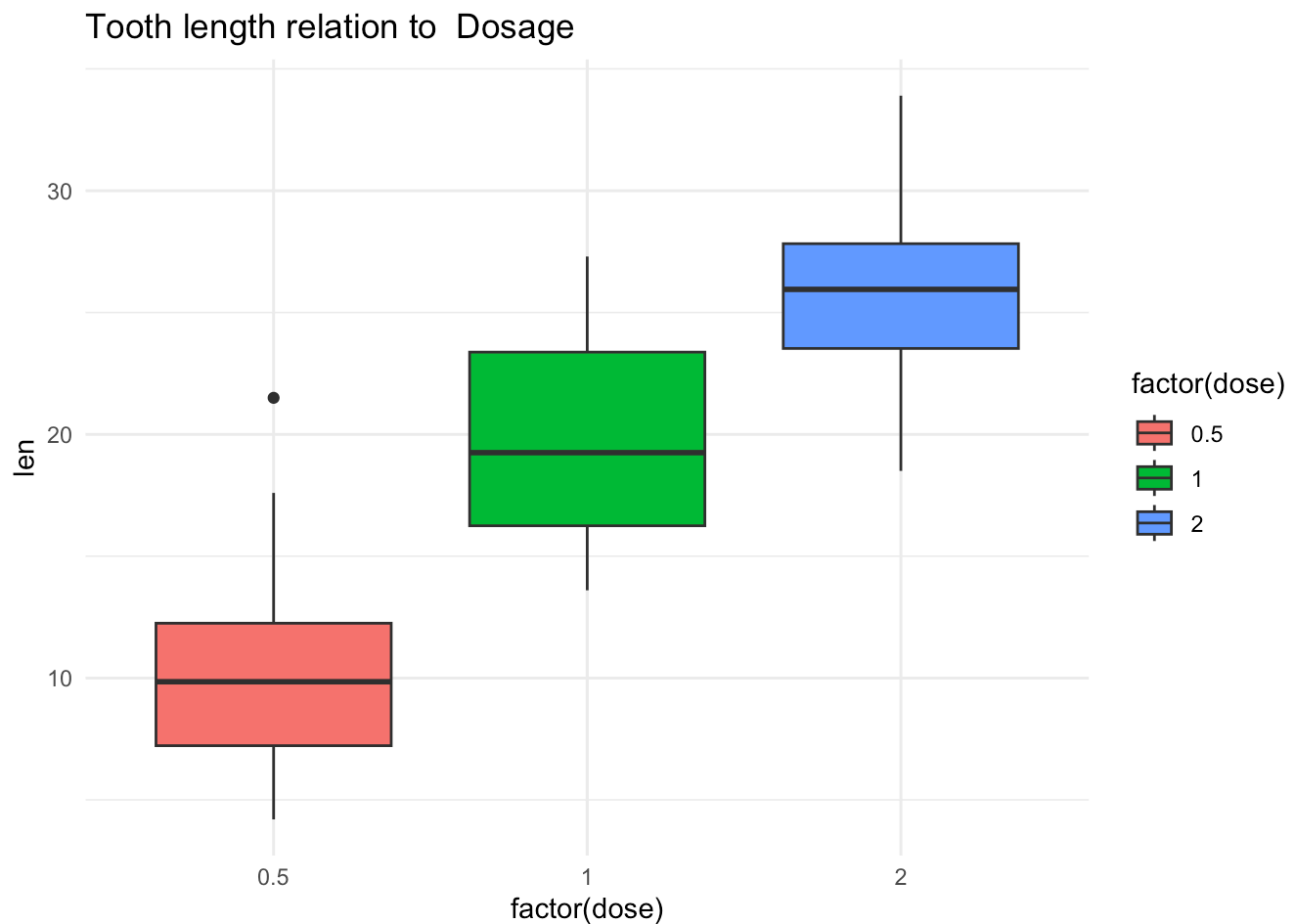
```
unique(ToothGrowth$dose)
```

```
## [1] 0.5 1.0 2.0
```

Unique dose groups are 0.5, 1.0, 2.0

Graph below, shows the relationship between Tooth Length and Dosages

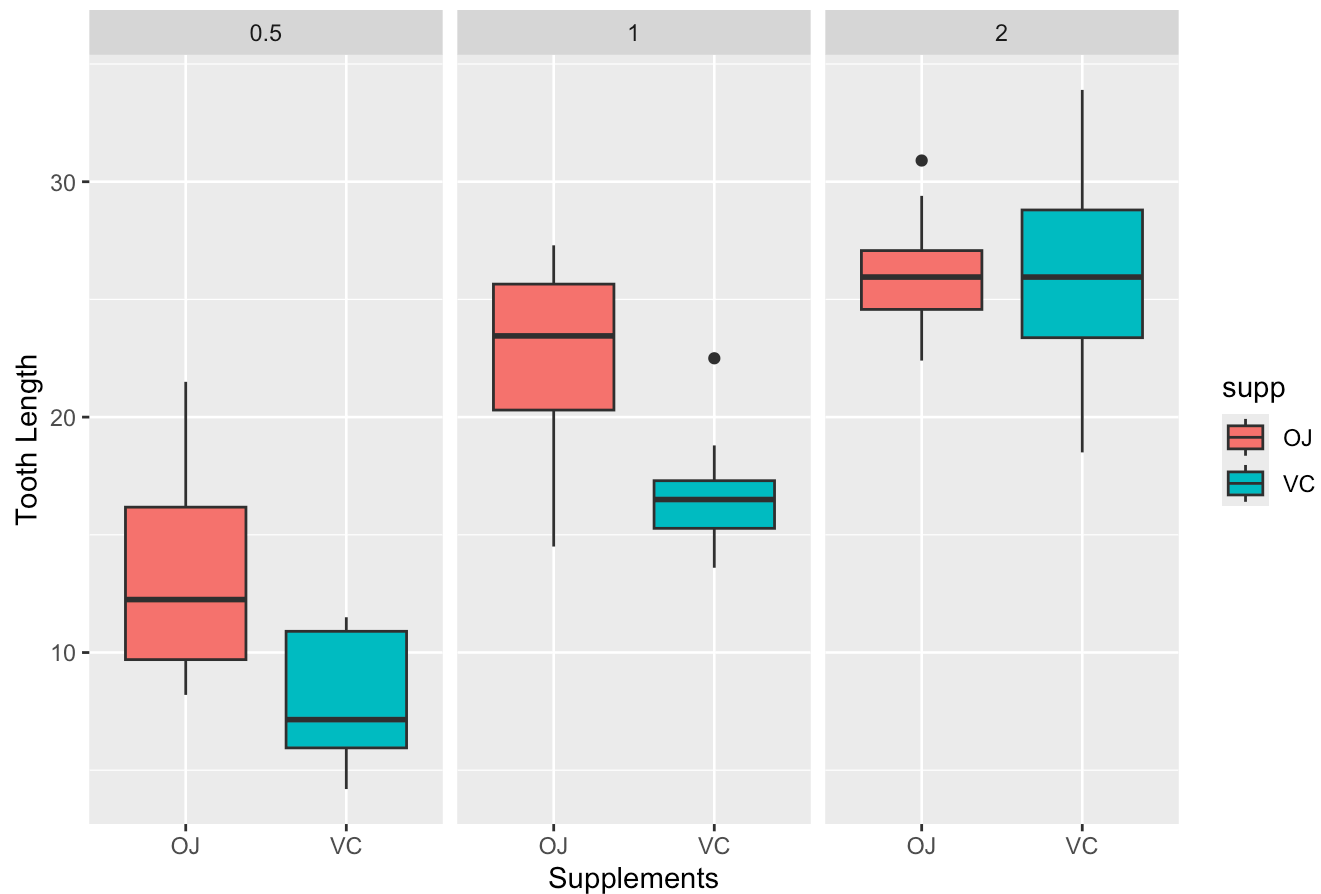
```
ggplot(aes(x = factor(dose), y = len), data = ToothGrowth) +
  geom_boxplot(aes(fill = factor(dose))) +
  ggtitle("Tooth length relation to Dosage") +
  theme_minimal()
```



The above graph show the relationship between tooth length and dosage in the ToothGrowth dataset

```
ggplot(aes(x=supp, y=len), data=ToothGrowth) +
  geom_boxplot(aes(fill=supp)) + xlab("Supplements") +
  ylab("Tooth Length") + facet_grid(~ dose) +
  ggtitle("Tooth length relation dosage of each Supplement")
```

Tooth length relation dosage of each Supplement



Graph above show the tooth Length relation to dosage of each supplement

Hypothesis test defined below :

H_0 : tooth length does not depend of different supplements H_a : tooth length are effected by different supplement

```
#t.test(len ~ supp, ToothGrowth[ToothGrowth$dose == .5, ])
# For dose = 0.5
test_dose_0.5 <- t.test(len ~ supp, data = ToothGrowth[ToothGrowth$dose == 0.5, ])
print("t-test for dose 0.5:")
```

```
## [1] "t-test for dose 0.5:"
```

```
test_dose_0.5
```

```
##
## Welch Two Sample t-test
##
## data: len by supp
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means between group OJ and group VC 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
```

```
#t.test(len ~ supp, ToothGrowth[ToothGrowth$dose == 1, ])
# For dose = 1
test_dose_1 <- t.test(len ~ supp, data = ToothGrowth[ToothGrowth$dose == 1, ])
print("t-test for dose 1:")
```

```
## [1] "t-test for dose 1:"
```

```
test_dose_1
```

```
##
## Welch Two Sample t-test
##
## data: len by supp
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means between group OJ and group VC 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
```

```
#t.test(len ~ supp, ToothGrowth[ToothGrowth$dose == 2, ])
# For dose = 2
test_dose_2 <- t.test(len ~ supp, data = ToothGrowth[ToothGrowth$dose == 2, ])
print("t-test for dose 2:")
```

```
## [1] "t-test for dose 2:"
```

```
test_dose_2
```

```
##  
## Welch Two Sample t-test  
##  
## data: len by supp  
## t = -0.046136, df = 14.04, p-value = 0.9639  
## alternative hypothesis: true difference in means between group OJ and group VC 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
```

Interpretation of Results:

For each t-test, you'll get:

t-value: The test statistic. p-value: Determines whether to reject the null hypothesis. If the p-value is less than your significance level (e.g., 0.05), you reject H_0 and conclude that tooth length is significantly different between supplements for that dose.

Confidence Interval: The range of values that likely contains the true difference in means.

What to Look For: Low p-values ($p < 0.05$): Suggest that the tooth length does depend on the supplement type. High p-values ($p > 0.05$): Suggest that there is no significant difference in tooth length between the supplements.

Conclusion

Since the p-value < 0.05 Reject H_0 This means that there is significant evidence to conclude that the tooth length differs based on the type of supplement (OJ vs VC) for that particular dose.