

# Statistical Inference Project Part 2

## Ron He

This analysis project will use the ToothGrowth data in R to tell whether supps and doses have impacts on the lens. This project will first do some exploratory analysis to the data. Then it will perform t-tests to those intervals to show whether there are actual impacts.

## Load and Explore the data

The ToothGrowth data set is loaded and explored by graphs as following. It turns out that the data has three columns and 60 rows. We set len as the predicting variable and supp and dose as the independent variables to produce some plots.

```
library(ggplot2)
library(lattice)

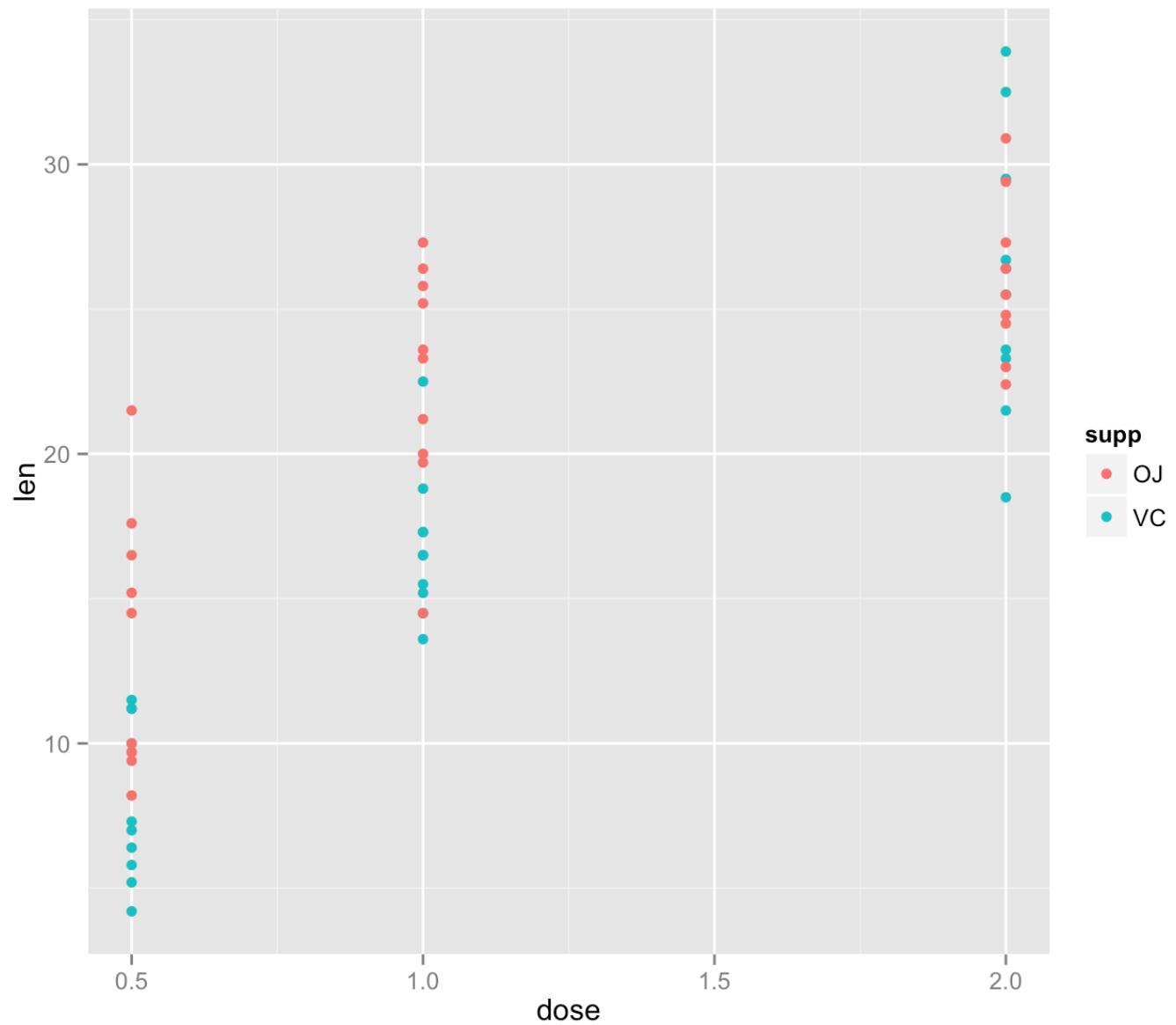
data("ToothGrowth")
head(ToothGrowth)
```

```
##      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
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

```
dim(ToothGrowth)
```

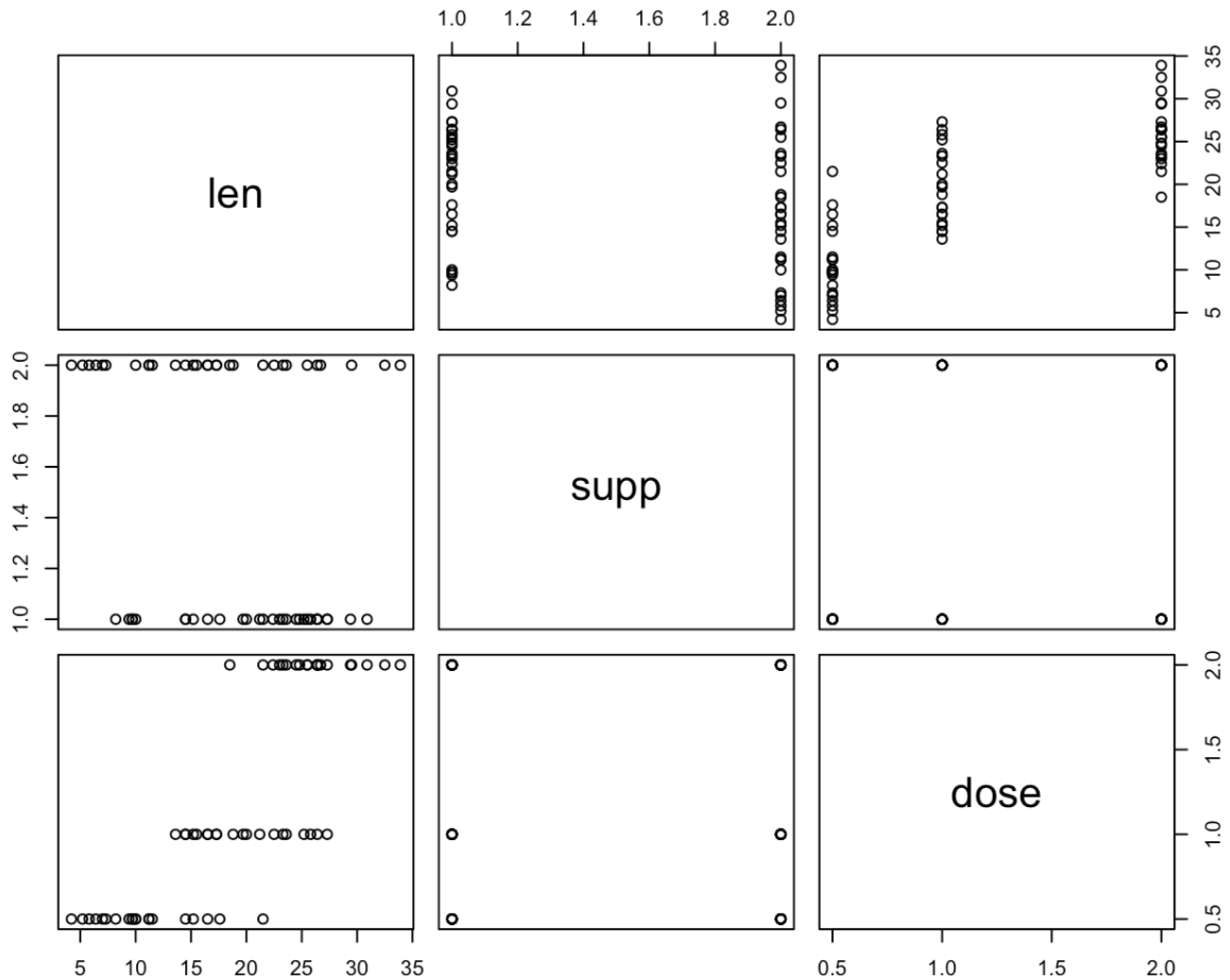
```
## [1] 60  3
```

```
qplot(x=dose, y=len, colour=supp, data=ToothGrowth)
```



```
pairs(~len+supp+dose, data=ToothGrowth, main="Scatterplot of len over supp and dose")
```

## Scatterplot of len over supp and dose



The pairs plot shows that there are two different supps and three different doses. The len vs. supp plot does not seem to have major difference. the len vs. does plot does show a significant increasement of len while increasing dose. T-tests are to performed to confirm these observations:

## Confidence Intervals and Hypothesis Test

Two groups of supp are directly compared to each other.

```
t.test(len ~ supp, paired = FALSE, var.equal = TRUE, data = ToothGrowth)
```

```
##
## Two Sample t-test
##
## data: len by supp
## t = 1.9153, df = 58, p-value = 0.06039
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.1670064 7.5670064
## sample estimates:
## mean in group OJ mean in group VC
## 20.66333 16.96333
```

Since there are three independent groups of doses, t-test analysis will be performed to each of these two groups.

```
tg0.5_1 <-ToothGrowth[which(ToothGrowth$dose==0.5|ToothGrowth$dose==1), ]
t.test(len ~ dose, paired = FALSE, var.equal = TRUE, data = tg0.5_1)
```

```
##
## 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
```

```
tg1_2 <-ToothGrowth[which(ToothGrowth$dose==1|ToothGrowth$dose==2), ]
t.test(len ~ dose, paired = FALSE, var.equal = TRUE, data = tg1_2)
```

```
##
## Two Sample t-test
##
## data: len by dose
## t = -4.9005, df = 38, p-value = 1.811e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -8.994387 -3.735613
## sample estimates:
## mean in group 1 mean in group 2
## 19.735 26.100
```

```
tg0.5_2 <-ToothGrowth[which(ToothGrowth$dose==0.5|ToothGrowth$dose==2), ]
t.test(len ~ dose, paired = FALSE, var.equal = TRUE, data = tg0.5_2)
```

```
##
## Two Sample t-test
##
## data: len by dose
## t = -11.799, df = 38, p-value = 2.838e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -18.15352 -12.83648
## sample estimates:
## mean in group 0.5 mean in group 2
## 10.605 26.100
```

## Difference in between suppliers:

The 95% confidence interval of these two groups contains 0. Also, the p-value is larger than 0.05. Therefore, it is not sufficient to reject the null hypothesis, and there is no difference between the two groups.

## Difference in between 0.5 and 1 does:

The 95% confidence interval of these two groups does not contain 0. Also, the p-value is less than 0.05. Therefore, it is sufficient to reject the null hypothesis, and there is significant difference between the groups of 6 to 11.

## Difference in between 1 and 1.5 does:

The 95% confidence interval of these two groups does not contain 0. Also, the p-value is less than 0.05. Therefore, it is sufficient to reject the null hypothesis, and there is significant difference between the groups of 4 to 8.

## Difference in between 0.5 and 1 does:

The 95% confidence interval of these two groups does not contain 0. Also, the p-value is less than 0.05. Therefore, it is sufficient to reject the null hypothesis, and there is significant difference between the groups of 13 to 18.

# Conclusion

From the t-test results, it shows that lens do not impacted by the supps. However, it shows that lens are impacted by different doeses. The conclusion is based on the Assumption that the testing groups are independent ones and are identically distributed. Aslo we assume that the groups of samples have the same variance within each group.