Peer-graded Assignment: Statistical Inference Course Project

Part 2: Basic Inferential Data Analysis Instructions

Gilbert Toroitich Tarus

27/01/2021

Overview

In this part, we analyze the `ToothGrowth` data in the R `datasets` package. The analysis is aiming investigating Effect of Vitamin C on Tooth Growth in Guinea Pigs.

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

```
library(datasets)
data("ToothGrowth")
```

2. Provide a basic summary of the data.

The top 10 observations

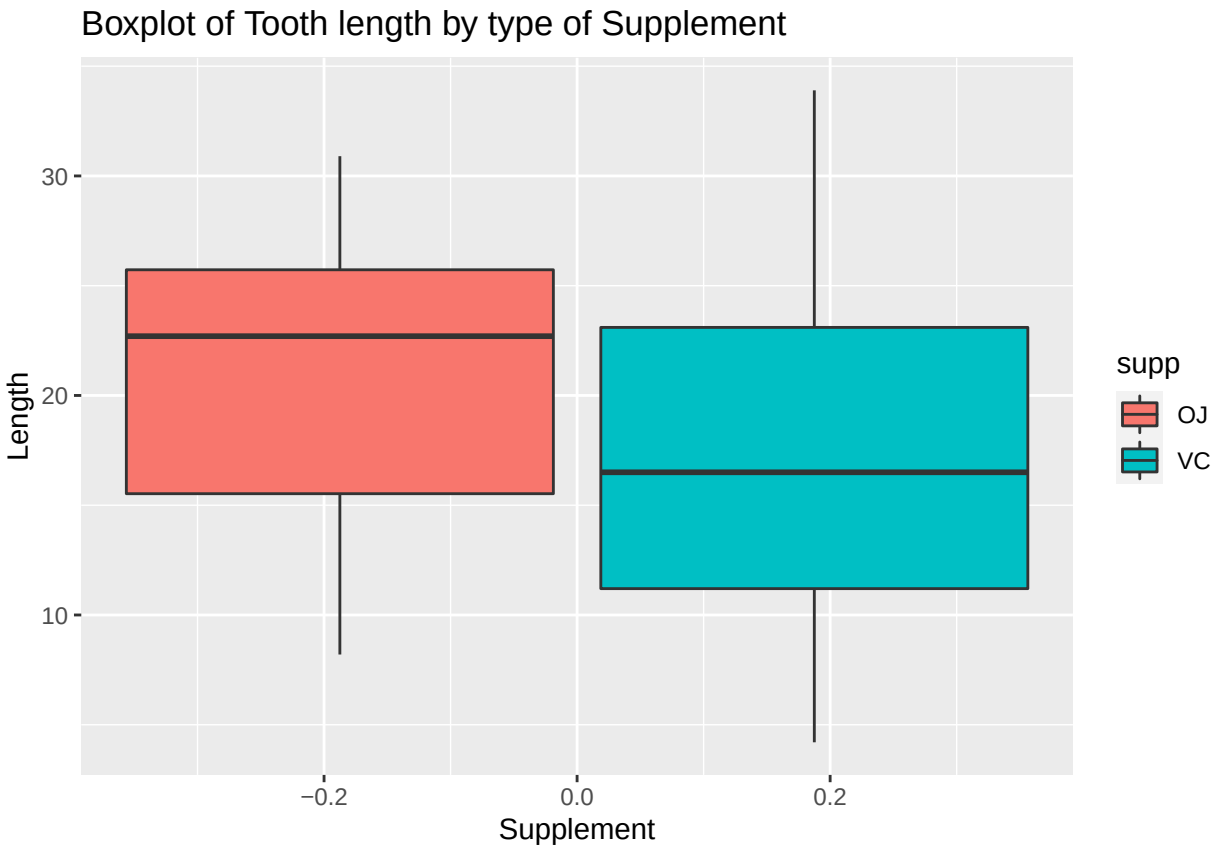
```
head(ToothGrowth,10)
```

```
table(ToothGrowth$supp,ToothGrowth$dose)
```

```
##
##      0.5  1  2
##   OJ  10 10 10
##   VC  10 10 10
```

Box Plot

```
library(ggplot2)
ggplot(ToothGrowth,aes(y = len))+#, y = len,col = supp))+
  geom_boxplot(aes(fill = supp))+
  labs(title = "Boxplot of Tooth length by type of Supplement",x = "Supplement",y = "Length")
```

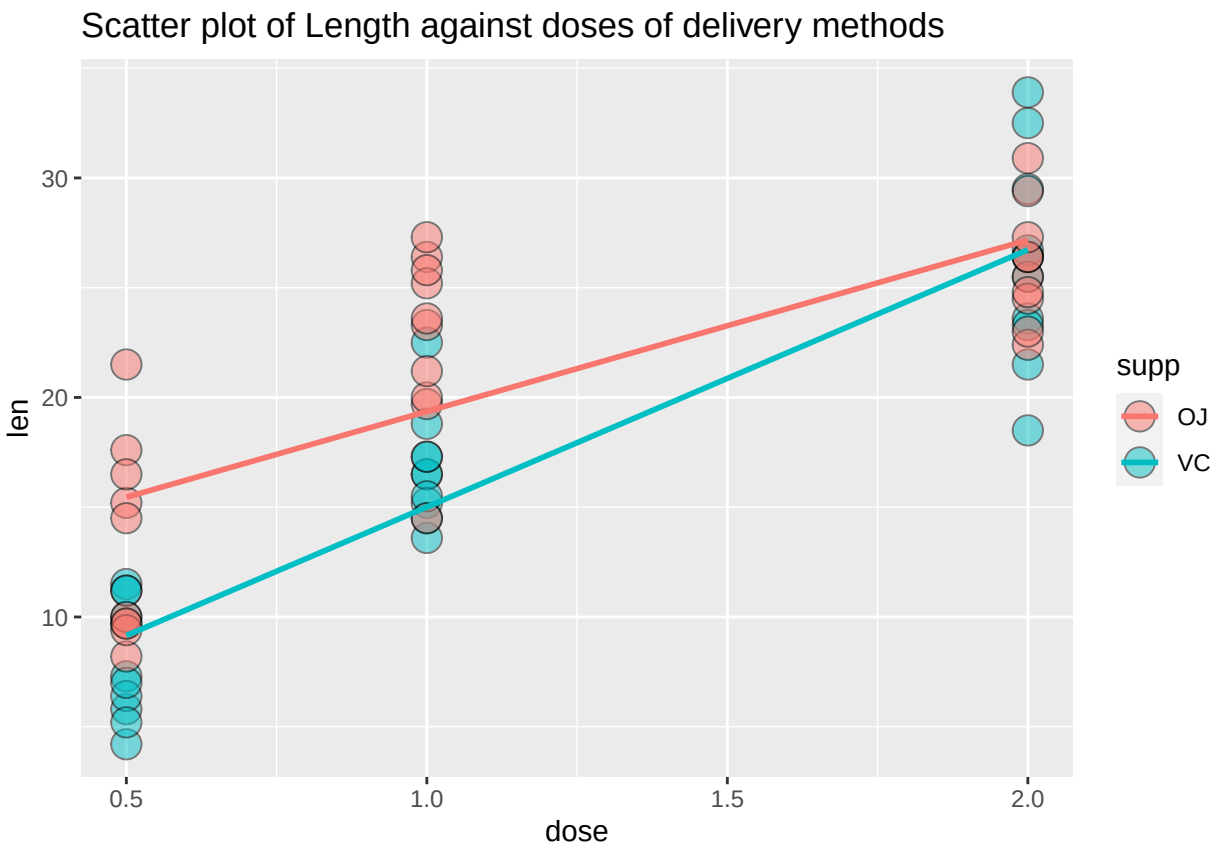


From the above plot, it is clear that the length of odontoblasts in the 60 guinea pigs is higher for those the animals that received Orange Juice (OJ) compared to those who received Vitamin C (VC). This shows that there is a significant difference in the effects of the delivery methods.

Scatter Plot

```
ggplot(ToothGrowth, aes(x = dose, y = len, group = factor(supp),fill = supp))+
  geom_point(size =5, pch = 21,alpha = .5)+
  labs(title = "Scatter plot of Length against doses of delivery methods")+
  geom_smooth(method = "lm",se = FALSE,aes(color = supp))
```

```
## `geom_smooth()` using formula 'y ~ x'
```



From the above plot, there is a significant increase in the length as the dose increase for both orange juice and vitamin C. The Orange juice exhibits higher increase compared to Vitamin C.

Overview

```
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
```

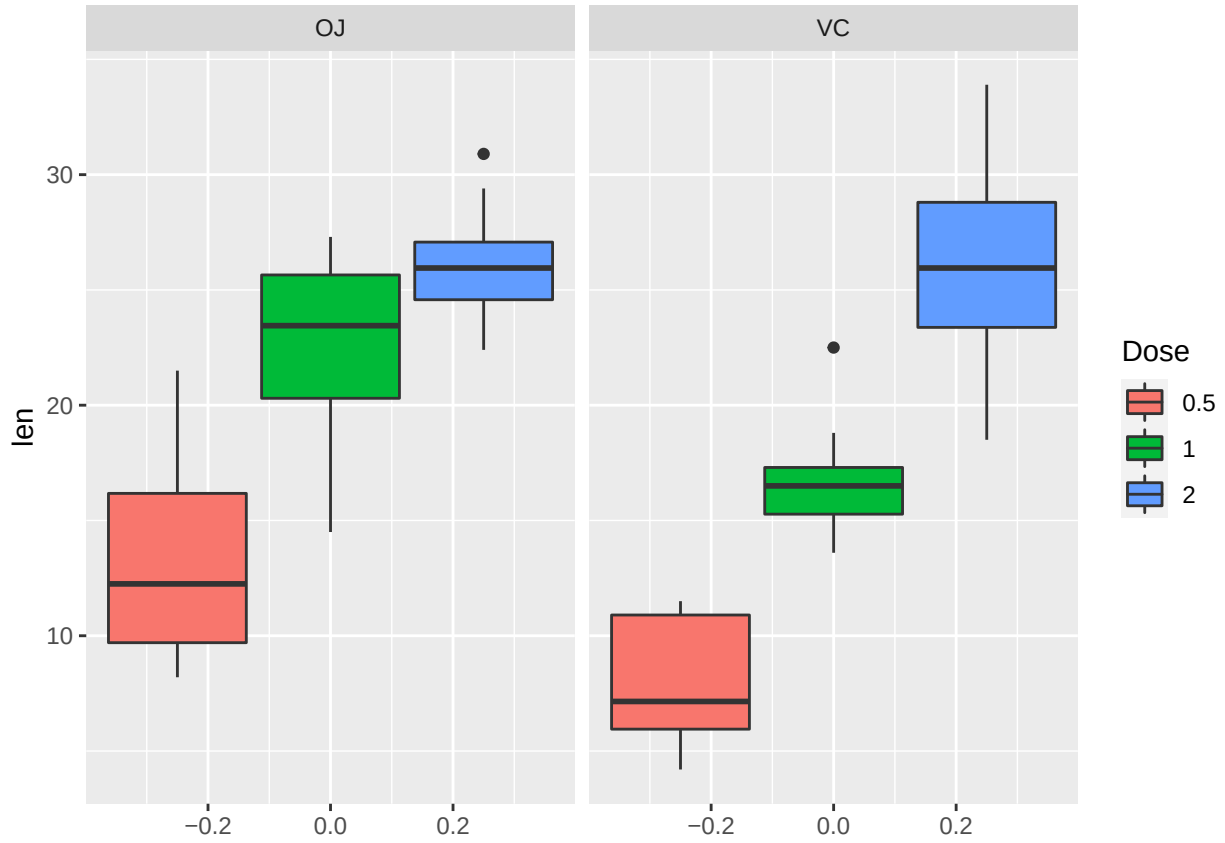
```
cor(ToothGrowth$len,ToothGrowth$dose)
```

```
## [1] 0.8026913
```

There is a strong positive relationship between tooth length and the amount of dose.

Distribution of tooth length by type of supplement and amount of doses

```
ggplot(ToothGrowth,aes(y = len,fill = as.factor(dose)))+
  geom_boxplot()+
  facet_grid(. ~ supp)+
  guides(fill = guide_legend(title = "Dose"))
```



From the plot above, those who received 2 doses have higher length followed by those who received 1 and those with least tooth length are those who received 0.5 doses for both supplements.

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)

Comparing tooth growth by supplemnt

The null hypothesis assumes that there is no significant difference in the mean growth among type supplements.

```
t.test(ToothGrowth$len~ToothGrowth$supp)$conf.int
```

```
## [1] -0.1710156  7.5710156
## attr(,"conf.level")
## [1] 0.95
```

The confidence interval above contains 0. This shows that there is no significant difference in tooth growth by supplement. Therefore we fail to reject the null hypothesis.

Dose 0.5 and 1.0

The null hypothesis assumes that there is no significant difference in the mean growth between dose 0.5 and 1.0.

```
# Subset the data that contains 0.5 and 1 doses only
comb1 <- ToothGrowth[ToothGrowth$dose%in%c(.5,1),]
#Compare their means
t.test(comb1$len~comb1$dose,var.equal = TRUE)$conf.int
```

```
## [1] -11.983748  -6.276252
## attr(,"conf.level")
## [1] 0.95
```

The confidence interval above is below 0 (does not contain 0). This shows that there is a significant difference in tooth growth between dose 0.5 and 1.0. Therefore we reject the null hypothesis.

Dose 0.5 and 2.0

The null hypothesis assumes that there is no significant difference in the mean growth between dose 0.5 and 2.0.

```
# Subset the data that contains 0.5 and 2 doses only
comb2 <- ToothGrowth[ToothGrowth$dose%in%c(.5,2),]
#Compare the means
t.test(comb2$len~comb2$dose,var.equal = TRUE)$conf.int
```

```
## [1] -18.15352 -12.83648
## attr(,"conf.level")
## [1] 0.95
```

The confidence interval above is below 0(does not contain 0). This shows that there is a significant difference in tooth growth between dose 0.5 and 2.0. Therefore we reject the null hypothesis.

Dose 1.0 and 2.0

The null hypothesis assumes that there is no significant difference in the mean growth between dose 1.0 and 2.0.

```
comb3 <- ToothGrowth[ToothGrowth$dose%in%c(1,2),]
t.test(comb3$len~comb3$dose,var.equal = TRUE)$conf.int
```

```
## [1] -8.994387 -3.735613
## attr(,"conf.level")
## [1] 0.95
```

The confidence interval above is below 0 (does not contain 0). This shows that there is a significant difference in tooth growth between dose 1.0 and 2.0. Therefore we reject the null hypothesis.

4. State your conclusions and the assumptions needed for your conclusions.

Assumptions

- When comparing tooth growth by supplement type, the equality of variance is assumed to be false.
- When comparing tooth growth by dose, the equality of variance is assumed to be true.
- The tooth length is assumed to be normally distributed.

Conclusions

- There is no significant differnce in tooth length between Orange Juice (OJ) and Vitamin C (VC). At 95% confidence, with equal doses, there is no significant difference in the tooth length.
- Given different doses of supplement at 95% confidence interval causes a significantly difference in the tooth length.