ToothGrowth An Inferential Study

thenuv

26 July 2018

Overview

This study is using the ToothGrowth data set provided with R. There are about 60 samples collected on guineapigs. There are 2 Delivery methods(supp) **OJ** (Orange Juice) & **VC** (Asorbic acid) which were provided to 30 of the animals. They received 3 different Dose levels **0.5, 1, 2**. With inferential statistical methods this analysis would find significance of tooth growth with Dose / Delivery method.

Data Summary

```
data("ToothGrowth") # Load Data
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 ...
```

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

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

```
summary(ToothGrowth)
```

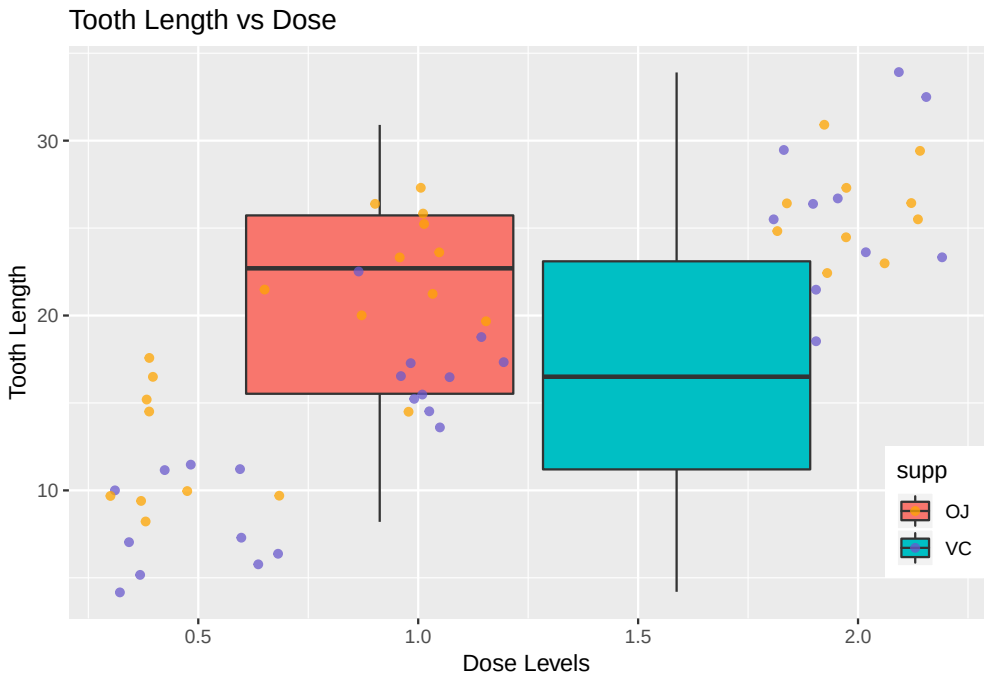
```
##       len        supp         dose
##  Min.   : 4.20   OJ:30   Min.   :0.500
##  1st Qu.:13.07   VC:30   1st Qu.:0.500
##  Median :19.25           Median :1.000
##  Mean   :18.81           Mean   :1.167
##  3rd Qu.:25.27           3rd Qu.:2.000
##  Max.   :33.90           Max.   :2.000
```

Exploratory Analysis

```
suppressMessages(library(ggplot2))
```

```
## Warning: package 'ggplot2' was built under R version 3.5.1
```

```
g <- ggplot(ToothGrowth, aes(dose, len)) +
        geom_boxplot(aes(fill=supp)) +
        geom_jitter(alpha=I(3/4), aes(color=supp)) +
        theme(legend.position=c(1,0.3),legend.justification=c(1,1)) +
        scale_color_manual(values=c("orange","slateblue")) +
        labs(title="Tooth Length vs Dose",x="Dose Levels",y="Tooth Length")
print(g)
```



From the plot it looks the tooth length increases as the dose level goes up. *OJ* method looks to have higher median than the other method. The data has the potential to analyse if there is relationship between dose vs length, method vs length and with the combination of both.

Assumptions

- The sample size represent the population.
- The length of the tooth in the animals are assumed to be fairly same when the test was begun.
- The t-test pairing is set to false as the groups are independent of each other.
- Variance is treated as not equal for the t-test.
- A 95% confidence interval is taken for analysis.

Hypothesis Testing

Tooth Length Vs Delivery

Null Hypothesis: There is no correlation between Tooth length and Delivery method. We will test it with *Student's T* method using Confidence intervals.

```
# Test for Delivery method

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

The ***p-value = 0.06063*** which is above .05 and we *fail to reject this null hypothesis.*

Refer to the Appendix for the t test output

Tooth Length Vs Dose

Null Hypothesis: There is no correlation on toothlength based on dose provided for the animals.

```
# Subset
dose_05_1 <- ToothGrowth[ToothGrowth$dose %in% c(0.5,1), ] # Subset for dose 0.5 & 1
dose_05_2 <- ToothGrowth[ToothGrowth$dose %in% c(0.5,2), ] # Subset for dose 0.5 & 2
dose_1_2 <- ToothGrowth[ToothGrowth$dose %in% c(1,2), ] # Subset for dose 1 & 2

# Test for dose
t.test(len ~ dose, data = dose_05_1, paired = FALSE, var.equal = FALSE )
t.test(len ~ dose, data = dose_05_2, paired = FALSE, var.equal = FALSE )
t.test(len ~ dose, data = dose_1_2, paired = FALSE, var.equal = FALSE )
```

- ***p-value = 0.0000001268*** *: Reject Null Hypothesis*
- ***p-value = 0.00000000000004398*** *: Reject Null Hypothesis*
- ***p-value = 0.00001906*** *: Reject Null Hypothesis*

There is significance on tooth length and dose. The p value for all levels are bellow .05, so we should look for the alternate hypothesis.

Tooth Length Vs Dlivery method for specific Dose

Null Hypothesis: For a specfic dose level, there is no correlation between the Delivery method and tooth length.

```
# Subset
dose_05 <- ToothGrowth[ToothGrowth$dose %in% c(0.5), ] # Subset for dose 0.5
dose_1 <- ToothGrowth[ToothGrowth$dose %in% c(1), ] # Subset for dose 1
dose_2 <- ToothGrowth[ToothGrowth$dose %in% c(2), ] # Subset for dose 2

# Test for Delivery method
t.test(len ~ supp, data = dose_05, paired = FALSE, var.equal = FALSE )
t.test(len ~ supp, data = dose_1, paired = FALSE, var.equal = FALSE )
t.test(len ~ supp, data = dose_2, paired = FALSE, var.equal = FALSE )
```

- ***p-value = 0.006359*** *: Reject Null Hypothesis*
- ***p-value = 0.001038*** *: Reject Null Hypothesis*
- ***p-value = 0.9639*** *: Accecpt Null Hypothesis (p >.05)*

At dose level .05 and 1 "OJ" method shows a higher length (refer to the mean in the output for t test in the appendix). There is not much impact for dose level 2 on either methods.

Conclusion

- Mean of Tooth length for dose **0.5, 1 & 2 are 10.605, 19.735 & 26.100** respectively. This shows there is significance on the length of the tooth with the dose given.
- From the hypothesis test performed, there is *not much impact* on Delivery method on the length of the tooth.
- There is *significance* on length of the tooth with different dose levels.
- For a dose level **0.5 & 1** the Delivery method **Orange Juice** has higher tooth growth. For dose level at **2** there is *not much signicance* with either Delivery method.

Appendix

Test Output for Length Vs Delivery method

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

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## 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 in group OJ mean in group VC
##         20.66333         16.96333
```

Test Output for Length Vs Dose

```
t.test(len ~ dose, data = dose_05_1, paired = FALSE, var.equal = FALSE )
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.983781  -6.276219
## sample estimates:
## mean in group 0.5   mean in group 1
##            10.605            19.735
```

```
t.test(len ~ dose, data = dose_05_2, paired = FALSE, var.equal = FALSE )
```

```
##
##  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
```

```
t.test(len ~ dose, data = dose_1_2, paired = FALSE, var.equal = FALSE )
```

```
##
##  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
```

Test Output for Length Vs Dlivery method for specific Dose

```
t.test(len ~ supp, data = dose_05, paired = FALSE, var.equal = FALSE )
```

```
##
##  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 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, data = dose_1, paired = FALSE, var.equal = FALSE )
```

```
##
##  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 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, data = dose_2, paired = FALSE, var.equal = FALSE )
```

```
##
##  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 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
```
