Coursera - Statistical Inference Course Project - Part 2

Goh LW

Sunday, October 25, 2015

Overview

For this part of the project we are going to work on dataset ToothGrowth that is available in R. This data is available through the variable ToothGrowth.

according to the help file ?ToothGrowth, the dataset is observing the effect of Vitamin C on tooth growth in Guinea Pigs. The ToothGrowth dataset recorded the relation between the growth of teeth of guinea pigs at each of three dose levels of Vitamin C (0.5, 1 and 2 mg) with each of two delivery methods(orange juice and ascorbic acid).

Load the ToothGrowth data and perform some basic exploratory data analyses

```
##    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
```

```
## '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 ...
```

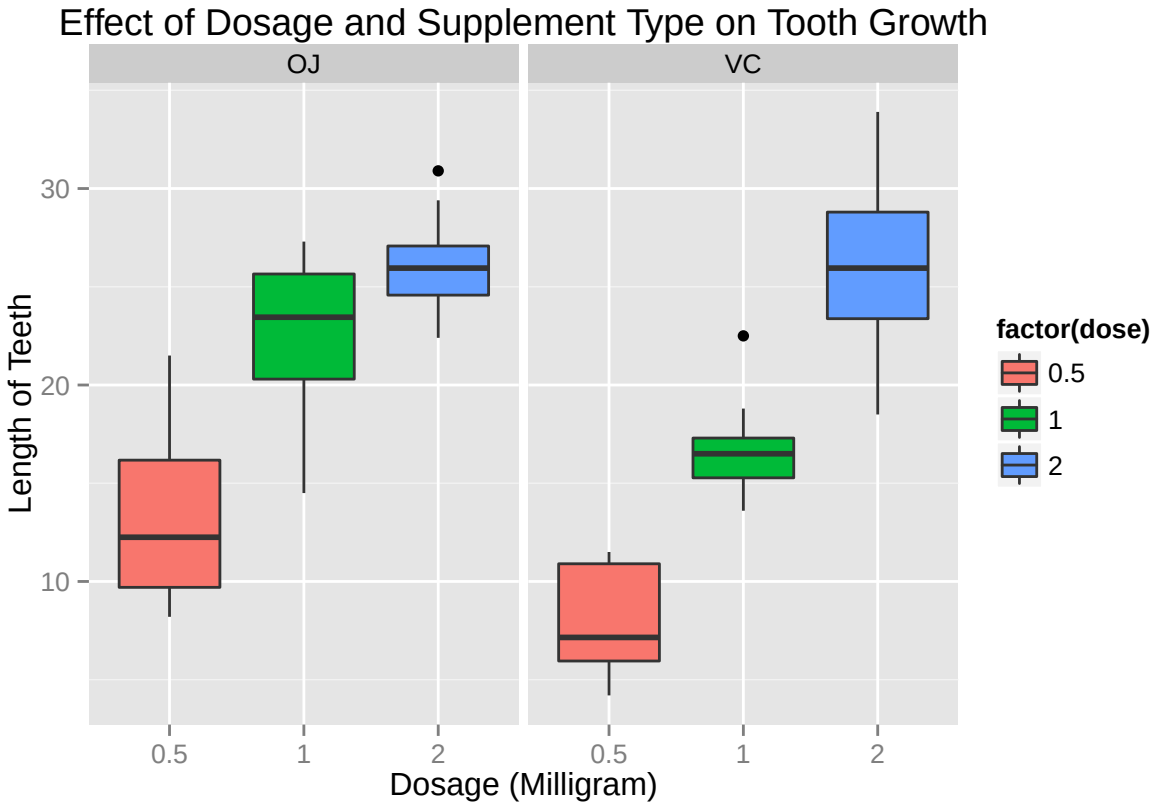
Provide a basic summary of the data.

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

```
##       len        supp     dose
##  Min.   : 4.20   OJ:30   0.5:20
##  1st Qu.:13.07   VC:30   1  :20
##  Median :19.25           2  :20
##  Mean   :18.81
##  3rd Qu.:25.27
##  Max.   :33.90
```

```
## [1] 18.81333
```

```
## [1] 7.649315
```



From the box plot above we see that the higher the dosage the longer the tooth is. The graph also shows the median of tooth grow for 2mg dosage is near for both supplements.

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)

```
##
##  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
```

```
##
##  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

##
##  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

##
##  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

##
##  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

##
##  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
```

```
##
##  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
```

For dose 0.5, the p-value of OJ in comparison to VC is 0.0064. Since it is less than 0.05 (strong presumption against null hypothesis), it indicates that there is a difference between both methods.

For dose 1.0, the p-value of OJ in comparison to VC is 0.001. Since it is less than 0.05 (strong presumption against null hypothesis), it indicates that there is a difference between both methods.

For dose 2.0, the p-value of OJ in comparison to VC is 0.064. Since it is greater than 0.05 (low presumption against null hypothesis), it indicates that there is a not much difference between both methods.

State your conclusions and the assumptions needed for your conclusions.

Assumption

- Sample is used instead of whole population
- For t-test regarding tooth length per supplement type, the variances are assumed to be different for the two groups being compared.
- For t-tests regarding tooth length per dosage level, the variances are assumed to be equal for the three combinations of the two groups being compared.

Conclusion

At 0.5mg and 1mg doses, orange juice causes more growth than the ascorbic acid. At 2mg doses, there is no significant difference.

Also, we see that tooth growth of guinea pigs correlate with the dosage of vitamin C that been given.