Stastical Inference - Assignment 1 - Part 2

DomR

Saturday, September 13, 2014

Part 2

Now in the second portion of the class, we're going to analyze the ToothGrowth data in the R datasets package.

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

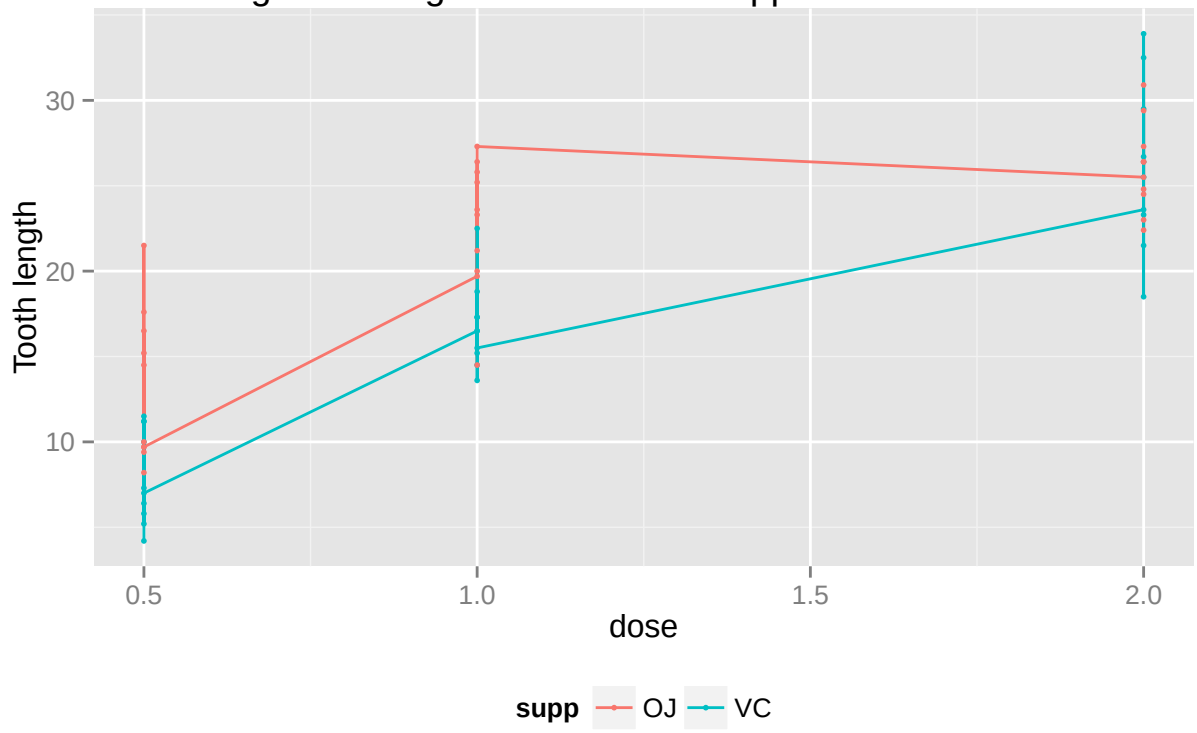
```
data(ToothGrowth)
```

2. Provide a basic summary of the data.

ToolGrowth dataset has data from a study of growth of tooth as in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods/supplements[OJ: Orange juice VC: ascorbic acid]. Column names are len, supp and dose. [Ref: ?ToothGrowth]

```
ggplot(ToothGrowth, aes(x=dose, y=len, colour=supp)) +
    geom_line(aes(group=supp)) +
    geom_point(size=1)+ theme(legend.position="bottom") +ylab("Tooth length") +
    ggtitle("Fig 1. Tooth growth based on supplement and dose")
```

Fig 1. Tooth growth based on supplement and dose



3. Use confidence intervals and hypothesis tests to compare tooth growth by supp and dose. (Use the techniques from class even if there's other approaches worth considering)

- Perform a two sample test using dose as grouping factor. Ho: There exists no difference between the supplements for tooth growth (i.e. mean of both samples is 0)
  - Confidence Interval and pValue

```
tTestSupp <- t.test(len~supp, var.equal=T, data=ToothGrowth)
tTestSupp
```

```
##
##  Two Sample t-test
##
## data:  len by supp
## t = 1.915, df = 58, p-value = 0.06039
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.167  7.567
## sample estimates:
## mean in group OJ mean in group VC
##            20.66            16.96
```

+ At 95% condfidence level, since p value ` r attr(tTestSupp$conf.int, "conf.level")` is greater than

- Perform a two sample test for various doses for each supplement. Ho: There exists no difference between the dosage levels

```
tTestOJDosePt5And1Pt0 <- t.test(len~dose, var.equal=T,
    data=subset(ToothGrowth, ToothGrowth$supp=='OJ' & ToothGrowth$dose %in% c(0.5,1)))
tTestVCDosePt5And1Pt0 <- t.test(ToothGrowth$len, ToothGrowth$dose, var.equal=T,
    data=subset(ToothGrowth, ToothGrowth$supp=='VC' & ToothGrowth$dose %in% c(0.5,1)))


tTestOJDosePt5And2Pt0 <- t.test(len~dose, var.equal=T,
  data=subset(ToothGrowth, ToothGrowth$supp=='OJ' & ToothGrowth$dose %in% c(0.5,2.0)))
tTestVCDosePt5And2Pt0 <- t.test(ToothGrowth$len, ToothGrowth$dose, var.equal=T,
  data=subset(ToothGrowth, ToothGrowth$supp=='VC' & ToothGrowth$dose %in% c(0.5,2.0)))

tTestOJDose1Pt0And2Pt0 <- t.test(len~dose, var.equal=T,
    data=subset(ToothGrowth, ToothGrowth$supp=='OJ' & ToothGrowth$dose %in% c(1.0,2.0)))
tTestVCDose1Pt0And2Pt0 <- t.test(ToothGrowth$len, ToothGrowth$dose, var.equal=T,
    data=subset(ToothGrowth, ToothGrowth$supp=='VC' & ToothGrowth$dose %in% c(1.0,2.0)))
```

- tValues and pValues for various doses

```
options("scipen"=100, "digits"=4)
tResultsVariousDoses <-  rbind(
c("OJ - 0.5 and 1.0",signif(tTestOJDosePt5And1Pt0$statistic,digits=3),
    signif(tTestOJDosePt5And1Pt0$p.value,digits=3),tTestOJDosePt5And1Pt0$p.value<0.05),
c("VC - 0.5 and 1.0",signif(tTestVCDosePt5And1Pt0$statistic,digits=3),signif(tTestOJDosePt5And1Pt0$p.va
c("OJ - 0.5 and 2.0",signif(tTestOJDosePt5And2Pt0$statistic,digits=3),
  signif(tTestOJDosePt5And2Pt0$p.value,digits=3),tTestOJDosePt5And2Pt0$p.value<0.05),
c("VC - 0.5 and 2.0",signif(tTestVCDosePt5And2Pt0$statistic,digits=3),
  signif(tTestOJDosePt5And2Pt0$p.value,digits=3),tTestOJDosePt5And2Pt0$p.value<0.05),
c("OJ - 1.0 and 2.0",signif(tTestOJDose1Pt0And2Pt0$statistic,digits=3),
  signif(tTestOJDose1Pt0And2Pt0$p.value,digits=3),tTestOJDose1Pt0And2Pt0$p.value<0.05),
c("VC - 1.0 and 2.0",signif(tTestVCDose1Pt0And2Pt0$statistic,digits=3),
  signif(tTestOJDose1Pt0And2Pt0$p.value,digits=3),tTestOJDose1Pt0And2Pt0$p.value<0.05))
colnames(tResultsVariousDoses ) <-c("Test", "tValue", "pValue","Reject Null?")
tResultsVariousDoses
```

```
##      Test               tValue  pValue       Reject Null?
## [1,] "OJ - 0.5 and 1.0" "-5.05" "0.0000836"  "TRUE"
## [2,] "VC - 0.5 and 1.0" "17.8"  "0.0000836"  "TRUE"
## [3,] "OJ - 0.5 and 2.0" "-7.82" "0.00000034" "TRUE"
## [4,] "VC - 0.5 and 2.0" "17.8"  "0.00000034" "TRUE"
## [5,] "OJ - 1.0 and 2.0" "-2.25" "0.0374"     "TRUE"
## [6,] "VC - 1.0 and 2.0" "17.8"  "0.0374"     "TRUE"
```

As seen from the above table, Null hypothesis of no difference between dosage levels cannot be rejected at 95 percent confidence level i.e. there exists difference in tooth growth for various dosage levels for each supplement

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

As shown in Figure 1 as well as proven using Null Hypothesis, there exists a difference in tooth growth for various dosage levels for each supplement. However, use of supplement type(OC or VC) has no impact on tooth growth. All tests were conducted at 95 percent confidence level.