Statistical Inference Project

José Gildardo Rojas-Nandayapa

Saturday, September 20, 2014

Part 2: Basic inferential data analysis. Analyze the ToothGrowth data in the R datasets package.

Summary

The data set contains records of the Effect of Vitamin C on Tooth Growth in Guinea Pigs.

The response is the length of odontoblasts (teeth) 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 (orange juice or ascorbic acid).

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

```
library(datasets)
data(ToothGrowth)
attach(ToothGrowth)
# Print the 4 first and last rows
head(ToothGrowth,n=4)
```

```
##    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
```

```
tail(ToothGrowth,n=4)
```

```
##     len supp dose
## 57 26.4   OJ    2
## 58 27.3   OJ    2
## 59 29.4   OJ    2
## 60 23.0   OJ    2
```

2. Provide a basic summary of the data.

```
dim(ToothGrowth)
```

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

```
names(ToothGrowth)
```

```
## [1] "len"  "supp" "dose"
```

Data is stored in a data frame with 60 observations and 3 variables.

- [,1] len numeric Tooth length
- [,2] supp factor Supplement type (VC or OJ).
- [,3] dose numeric Dose in milligrams.

```r
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 ...
```

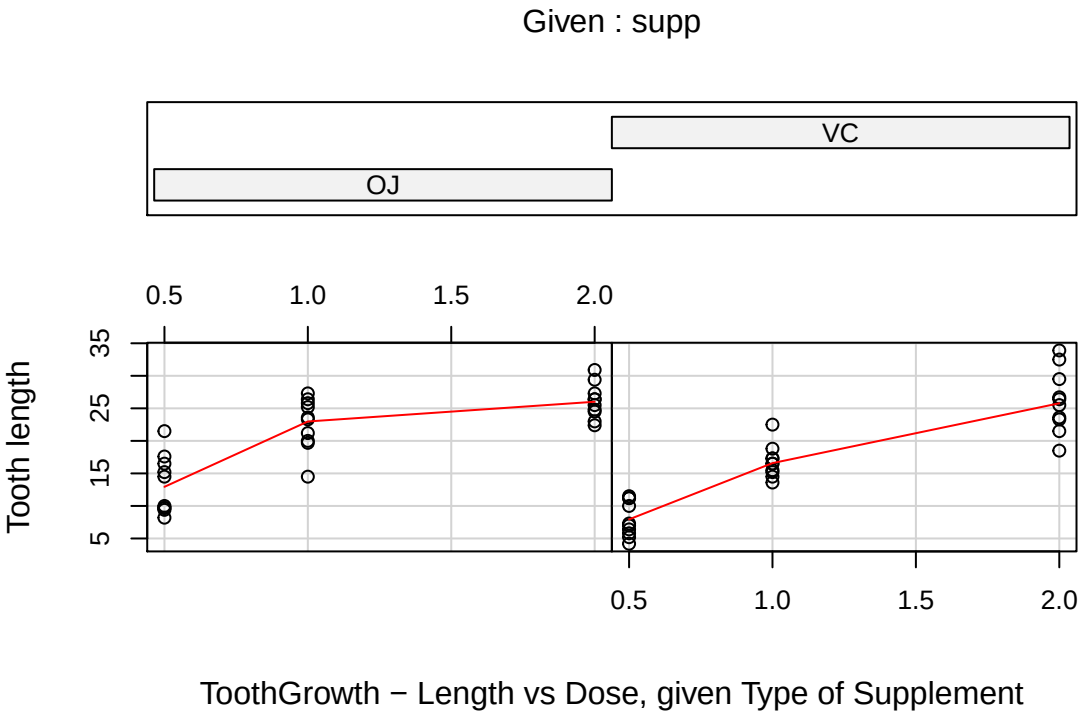
```r
levels(ToothGrowth$supp)
```

```
## [1] "OJ" "VC"
```

```r
summary(ToothGrowth)
```

```
##       len        supp         dose
##  Min.   : 4.2   OJ:30   Min.   :0.50
##  1st Qu.:13.1   VC:30   1st Qu.:0.50
##  Median :19.2           Median :1.00
##  Mean   :18.8           Mean   :1.17
##  3rd Qu.:25.3           3rd Qu.:2.00
##  Max.   :33.9           Max.   :2.00
```

```r
coplot(len ~ dose | supp,data=ToothGrowth,panel=panel.smooth,ylab="Tooth length",
       xlab="ToothGrowth - Length vs Dose, given Type of Supplement",cex.axis=.6)
```



Supplements: OJ:Orange juice, VC:Vitamin C | Doses:0.5, 1.0 and 2.0

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)

IMPORTANT: ***Please see Appendix for hypothesis tests code.***

Mean tooth growth by supplement is bigger for Orange Juice (20.6633) vs. Vitamin C (16.9633), the p-value (0.6063) is slightly bigger than 0.05, we can say that in general Orange Juice over Vitamin C is a better supplement, but the difference is not huge, so more detailed analysis is required.

In general, averages add valuable information about the data groups being anlyzed. while p-values below 0.05 mean a strong presumption against null hypothesis, meaning that there is a difference between both methods, while p-values bigger than 0.05 mean a low presumption of null hypothesis.

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

By suplement: Orange juice vs. Vitamin C

- Orange Juice supplement is better than Vitamin C.

By dosis compared

- 1.0 mg is much better than 0.5 mg.
- 2.0 mg are better than 1.0, but the difference is not exactly proportional, in other words, 2.0 mg doesn't provide twice as much benefit as 1.0mg, specially with Orange Juice results at 1.0 mg.

By dosis, comparing supplements.

- At 0.5 and 1.0 mg, the supplement effect is bigger for Orange Juice than Vitamin C.
- At 2.0 mg, Vitamin C is just slightly better than Orange Juice, but the difference not relevant.

Conclusion

Orange juice is better in general than Vitamin C for guinea pigs odontoblasts (tooth), at 1mg, Orange Juice provides a significant effect, near to 2.0mg doses of any of the two supplements tested, while Vitamin C only provides as much as Orange Juice at 2.0 milligram doses. These conclusions may also be confirmed by inspecting the chart above.

Appendix

```
# Performing T Test by supplemant type
t.test(len ~ supp,data=ToothGrowth)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 1.915, df = 55.31, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.171  7.571
## sample estimates:
## mean in group OJ mean in group VC
##            20.66            16.96
```

```r
# Subsetting ToothGrowth by dose comparing doses
TG_0.5to1.0<-subset(ToothGrowth,dose %in% c(0.5,1.0))
TG_1.0to2.0<-subset(ToothGrowth,dose %in% c(1.0,2.0))
TG_0.5to2.0<-subset(ToothGrowth,dose %in% c(0.5,2.0))
# Performing T test by dose comparing doses
t.test(len ~ dose, data=TG_0.5to1.0)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -6.477, df = 37.99, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.984  -6.276
## sample estimates:
## mean in group 0.5   mean in group 1
##             10.61             19.73
```

```r
t.test(len ~ dose, data=TG_1.0to2.0)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -4.901, df = 37.1, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.996 -3.734
## sample estimates:
## mean in group 1 mean in group 2
##           19.73           26.10
```

```r
t.test(len ~ dose, data=TG_0.5to2.0)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -11.8, df = 36.88, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.16 -12.83
## sample estimates:
## mean in group 0.5   mean in group 2
##             10.61             26.10
```

```r
# Subsetting ToothGrowth by dose, comparing supplements
TG0.5 <- subset(ToothGrowth,dose==0.5)
TG1.0 <- subset(ToothGrowth,dose==1.0)
TG2.0 <- subset(ToothGrowth,dose==2.0)
# Performing T test by doses, comparing supplements
t.test(len ~ supp, data = TG0.5)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 3.17, df = 14.97, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.719 8.781
## sample estimates:
## mean in group OJ mean in group VC
##            13.23             7.98
```

```
t.test(len ~ supp, data = TG1.0)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 4.033, df = 15.36, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802 9.058
## sample estimates:
## mean in group OJ mean in group VC
##            22.70            16.77
```

```
t.test(len ~ supp, data = TG2.0)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = -0.0461, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.798  3.638
## sample estimates:
## mean in group OJ mean in group VC
##            26.06            26.14
```