Hypothesis Testing on ToothGrowth Data

Arkosnato Neogy

Sunday 26 April 2015

In this report we shall investigate the ToothGrowth data (from datasets library in R). The dataset relates to increase in tooth length of guinea-pigs who were given 3 different doses of 2 possible supplements (orange juice and ascorbic acid). There were 10 candidate guinea pigs for each test, and 6 possible combinations of tests.

```
library(data.table)
library(datasets)
mydt= data.table(ToothGrowth)
dim(mydt)
```

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

```
head(mydt)
```

```
##     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
```

We are interested to compare the effects for dose and supplement on tooth growth. In order to do this we shall run a t CI-based hypothesis test on the data by choosing one pair of contrasting treatments (like dose of 1mg vs 2mg) and holding other effects constant (like setting the supplement as 'OJ' while contrasting dosage).

We are making a few assumptions here: 1. that the data values are IID samples from a normally distributed population for the t-test to hold 2. that the variance across the groups is the same

Further while running the tests we shall imagine the values to be unpaired. This makes sense because to be paired we would need to run 2 treatments on the same guinea-pig. That sounds absurd because once its teeth have grown, it would need to be brought back to the initial size, or it would just keep growing... neither of which sound practical.

Lets start by looking at the nature of the data in general. Here we plot the growth by dose for a given mode of supplement.
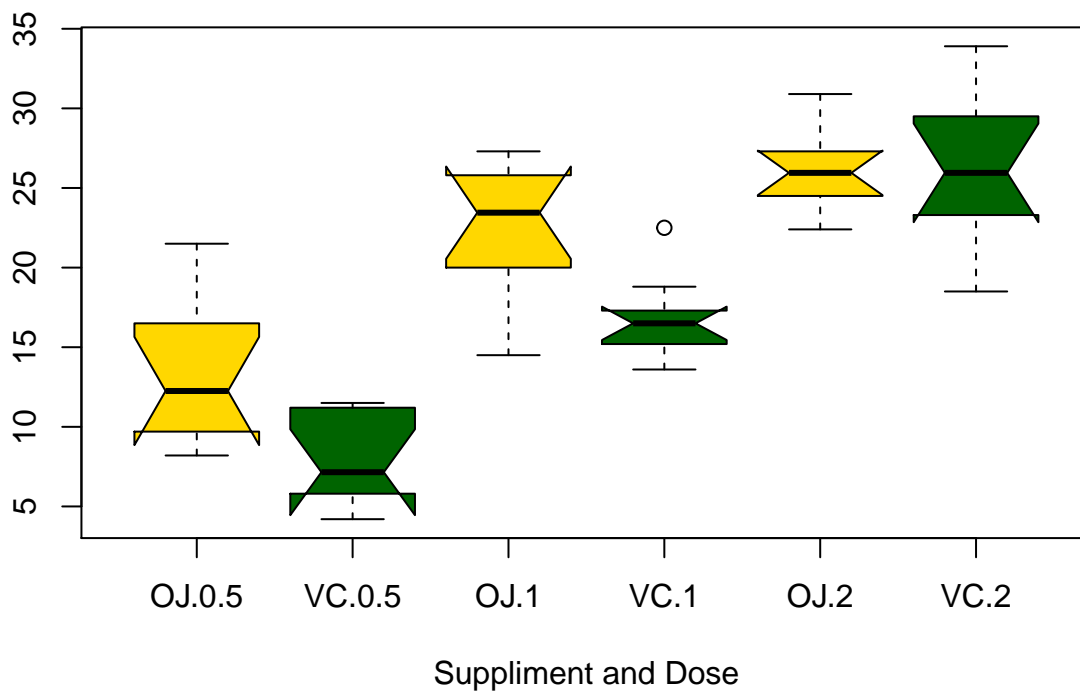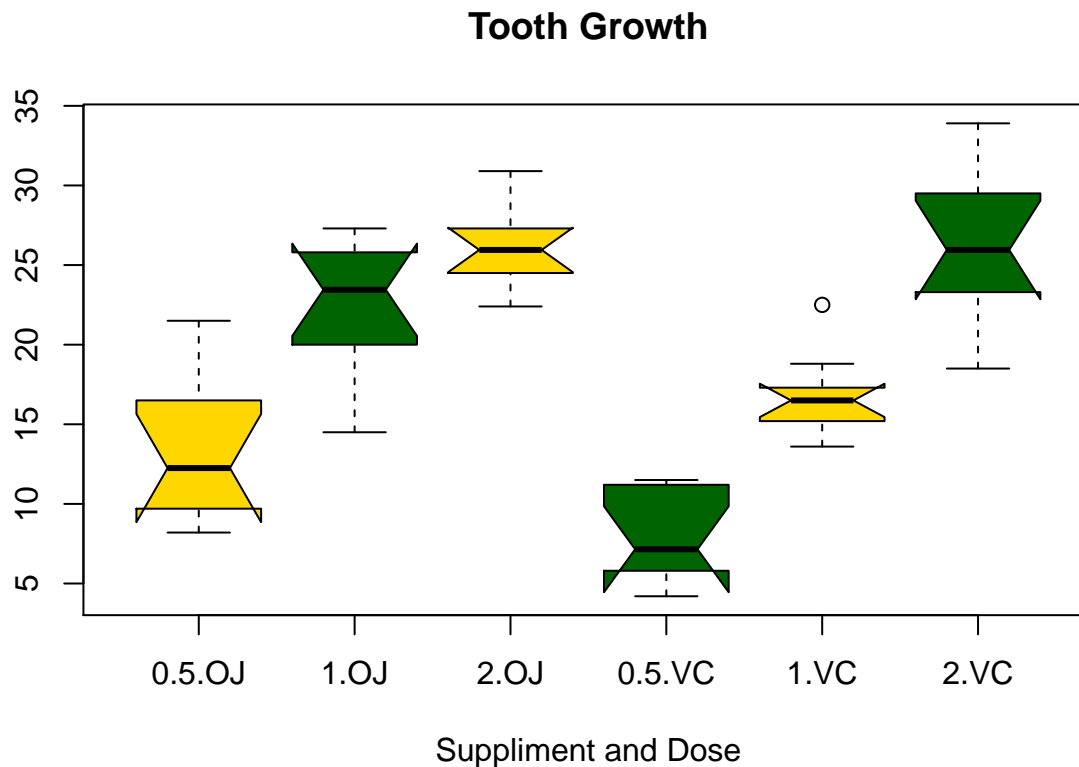
Given : supp

VC

OJ

len
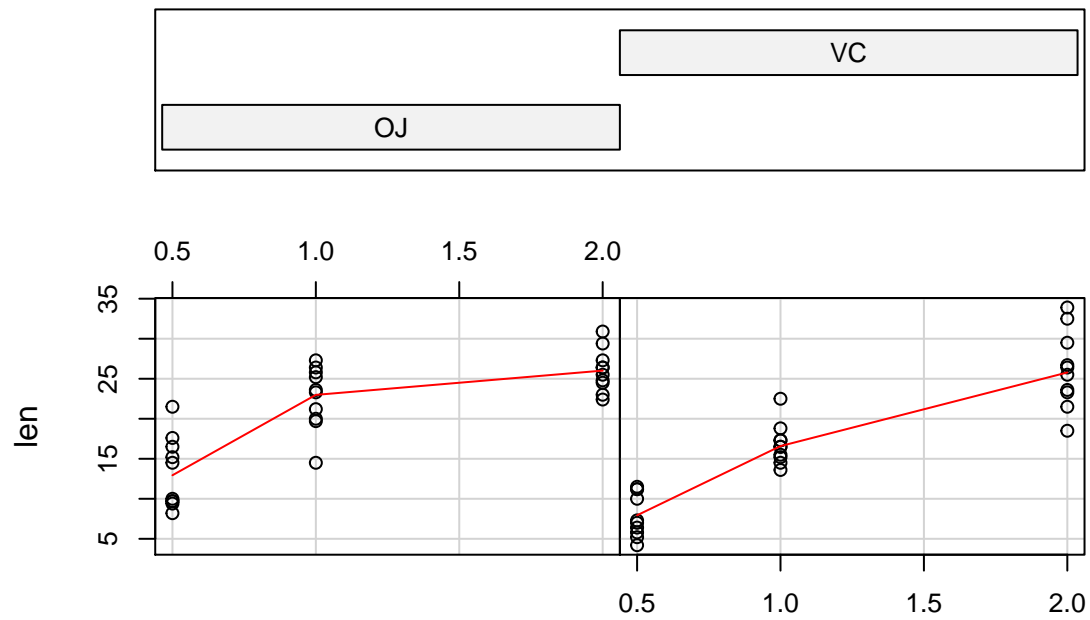
ToothGrowth data: length vs dose, given type of supplement

And here we have notched box plots of growth vs supplement for each dose and vice-versa.

Tooth Growth

OJ.0.5   VC.0.5   OJ.1   VC.1   OJ.2   VC.2

Suppliment and Dose

Tooth Growth

Suppliment and Dose

From the plots its quite clear that increasing doses produce greater growth. Also, in general VC (ascorbic acid) seems a poorer supplement than OJ (orange juice), but for a dosage of 2mg they seem to have similar effects. Let us try to verify these insights through sounds statistics.

Let us run the test for dose 2 mg vs 1 mg when the supplement is OJ. This can be done by the following commands.

```
test.dt= mydt[supp == 'OJ' & dose != 0.5]
test.dt= test.dt[, dose:= as.factor(dose)]
result= t.test(len~I(relevel(dose,2)), paired=F, data= test.dt)
result
```

```
##
##  Welch Two Sample t-test
##
## data:  len by I(relevel(dose, 2))
## t = 2.2478, df = 15.842, p-value = 0.0392
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.1885575 6.5314425
## sample estimates:
## mean in group 2 mean in group 1
##           26.06           22.70
```

The test result shows that the 95% confidence interval for the difference of means (dose 1 mg being reference) lies completely above 0. Alternatively, the t-statistic for the test is positive, and the p-value is less than the significance level of the test. Thus, dose of 2 mg produces statistically significant greater tooth growth than dose of 1 mg.

We would not like to do this repeatedly for all the different combinations of dose and supplement, so we need to automate the testing. We define a couple of testing functions to do all our tests. The conclusions will be printed based on the p-values and t-statistics.

```r
performAllTests = function(mydt)
{
  sups= unique(mydt$supp)
  dss= unique(mydt$dose)

  # all dose tests
  for (sup in sups) {
    for (ds in dss) {
      # setup the test data
      dss.rel= dss[dss != ds]
      dss.test= max(dss.rel)
      dss.ref= min(dss.rel)
      print(paste('Testing dose', dss.test, 'vs', dss.ref,
                  'keeping supplement fixed as', sup))

      test.dt= mydt[supp == sup & dose %in% dss.rel]
      test.dt[, dose:= as.factor(dose)]

      # run the test
      test.result = t.test(len ~ I(relevel(dose, 2)), paired= F, data= test.dt)
      printConclusions(test.result, dss.ref, dss.test, type= 'dose')
    }
  }

  # all supplement tests
  for (ds in dss) {
    print(paste('Testing supplment OJ vs VC keeping dose fixed at', ds))
    test.result = t.test(len ~ supp, paired= F, data= mydt[dose == ds])
    printConclusions(test.result, 'VC', 'OJ', 'supplement')
  }
}

printConclusions = function (result, refLevel, testLevel, type, sigLevel= 0.05)
{
  if (result$p.value < sigLevel){
    print (paste('Evidence to reject H0 at', sigLevel, 'level of significance!'))
    if (result$statistic > 0 ){
      print(paste('Treatment of', type, testLevel,
                  'produces BETTER results than', type, refLevel))
    } else {
      print(paste('Treatment of', type, testLevel,
                  'produces WORSE results than', type, refLevel))
    }
  } else {
    print (paste('No clear evidence favoring H1 at', sigLevel, 'level of significance'))
  }
}
```

We have automated things a bit here to help in repeated testing. Note that `dss.rel` picks out the 2 doses to test with, and assigns the higher of them to `dss.test` (the one being tested) while the lower goes to `dss.ref`

4

which serves as reference. For the supplement part of the test we looked at the data and figured out that OJ was being labeled as the highest factor, so no 'relevel' was required.

So now we run the battery of tests. . .

```
performAllTests(mydt)
```

```
## [1] "Testing dose 2 vs 1 keeping supplement fixed as VC"
## [1] "Evidence to reject H0 at 0.05 level of significance!"
## [1] "Treatment of dose 2 produces BETTER results than dose 1"
## [1] "Testing dose 2 vs 0.5 keeping supplement fixed as VC"
## [1] "Evidence to reject H0 at 0.05 level of significance!"
## [1] "Treatment of dose 2 produces BETTER results than dose 0.5"
## [1] "Testing dose 1 vs 0.5 keeping supplement fixed as VC"
## [1] "Evidence to reject H0 at 0.05 level of significance!"
## [1] "Treatment of dose 1 produces BETTER results than dose 0.5"
## [1] "Testing dose 2 vs 1 keeping supplement fixed as OJ"
## [1] "Evidence to reject H0 at 0.05 level of significance!"
## [1] "Treatment of dose 2 produces BETTER results than dose 1"
## [1] "Testing dose 2 vs 0.5 keeping supplement fixed as OJ"
## [1] "Evidence to reject H0 at 0.05 level of significance!"
## [1] "Treatment of dose 2 produces BETTER results than dose 0.5"
## [1] "Testing dose 1 vs 0.5 keeping supplement fixed as OJ"
## [1] "Evidence to reject H0 at 0.05 level of significance!"
## [1] "Treatment of dose 1 produces BETTER results than dose 0.5"
## [1] "Testing supplment OJ vs VC keeping dose fixed at 0.5"
## [1] "Evidence to reject H0 at 0.05 level of significance!"
## [1] "Treatment of supplement OJ produces BETTER results than supplement VC"
## [1] "Testing supplment OJ vs VC keeping dose fixed at 1"
## [1] "Evidence to reject H0 at 0.05 level of significance!"
## [1] "Treatment of supplement OJ produces BETTER results than supplement VC"
## [1] "Testing supplment OJ vs VC keeping dose fixed at 2"
## [1] "No clear evidence favoring H1 at 0.05 level of significance"
```

As expected, the results show higher doses produce statistically significant better results that lower, and OJ shows better results than VC, except for the dosage of 2 mg where the test fails to reject the null hypothesis.