project1-part2

Miguel Baquero

Sunday, September 21, 2014

Part 2

Load the ToothGrowth data and perform some exploratory data analysis:

```
data(ToothGrowth)
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
```

```
head(ToothGrowth)
```

```
##    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
```

First question comes to my mind is which method is more effective for tooth growth, lets do a first approach:

```
#Split by separate methods (factor "supp") and divide lenght by doses
(to "normalize" for doses):
sapply(split(x=ToothGrowth,f=ToothGrowth$supp),function(x)
mean(x$len/x$dose))
```

```
##    OJ    VC
## 20.73 15.27
```

So at first sight, looks like the orange juice is more effective as a delivery method, but lets do some serious statistical inference to confirm this.

If we use t.test to compare the variable len / dose by the factor supp:

```
t.test(len/dose~supp,data=ToothGrowth)
```

```
##
##  Welch Two Sample t-test
##
## data:  len/dose by supp
## t = 3.371, df = 42.61, p-value = 0.001603
```

```
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.194 8.733
## sample estimates:
## mean in group OJ mean in group VC
##            20.73            15.27
```

The result is that with 95% confidence the Orange Juice is better delivery and makes growht to be at least:

```
t.test(len/dose~supp,data=ToothGrowth)$conf.int[1]
```

```
## [1] 2.194
```

more than with VC.

But is it important to note that this is under the assumption that the pigs where selected as iid...basically meaning that their characteristics (weight, breed, environemnt or food for example) should be randomly distributed.