Statistical Inference - Course Project - Question 2

2014-09-21

Synopsis

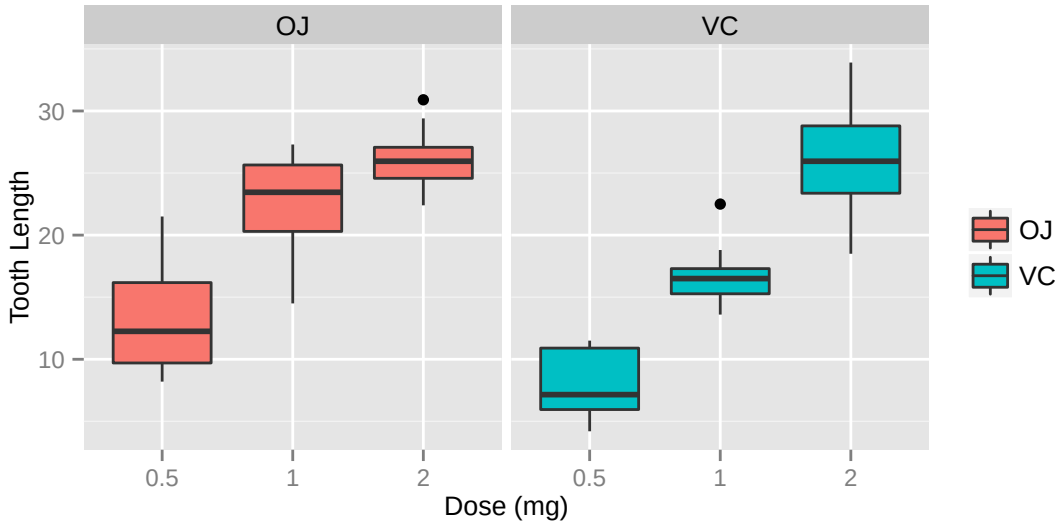
In this report, I will review and analyze the ToothGrowth data in the R datasets package. This dataset details The Effect of Vitamin C on Tooth Growth in Guinea Pigs. The response is the length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods (orange juice or ascorbic acid). The Data Dictionary is as follow:

Field	Data Type	Definition
`len`	numeric	Tooth length
`supp`	factor	Supplement type (OJ or VC).
`dose`	numeric	Dose in milligrams.

1 & 2. Load the ToothGrowth data and perform some basic exploratory data analyses. Provide a basic summary of the data.

We call `data(ToothGrowth)` to load the data set followed by `str(ToothGrowth)` to get our first impression of the data structure. Next we use comparative box plots to get a feel for the relationship between `len`, `supp` and `dose` and close our exploration off with a `summary(ToothGrowth)` function call.

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




```
##       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
```

3. Use Confidence Intervals (CI) and hypothesis tests to compare tooth growth by supp and dose.

A review of the plot above seems to show that supplement type OJ (orange juice) achieves better growth at lower doses (0.5mg and 1.0mg) than VC (ascorbic acid) but at higher doses they realize comparable growth. It also seems clear dose and growth are positively correlated. Using a Welch Two Sample t-test we can review tooth growth attributed to supplement and dose by calling `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
```

Next we execute T Tests on the dose specific subsets of the `ToothGrowth` dataset. Our findings are tabulated below:

T Test	Lower 95% CI	Upper 95% CI	t	p-value
Length ~ Supplement [OJ vs. VC]	-0.171	7.571	1.92	0.0606
Length ~ Dose [0.5mg vs. 1.0mg]	-11.984	-6.276	-6.48	1.26830072017385e-07
Length ~ Dose [1.0mg vs. 2.0mg]	-8.996	-3.734	-4.9	1.9064295136718e-05

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

Assumptions: I'm assuming that all significant factors (i.e. habitat and feeding) were controlled for the duration of the experiment. The analysis was also done under the assumption that the populations exhibit differing variances.

Conclusions:

- Supplement [OJ vs. VC]: The null hypothesis is accepted. There is insufficient proof for rejecting the null hypothesis that OJ and VC yield similar growth. The p-value is in excess of 0.05 at 0.0606.
- Dose [0.5mg vs. 1.0mg]: The null hypothesis is rejected. There is overwhelming proof that a dose increase from 0.5mg to 1.0mg will yield more growth. The p-value is far below 0.05 at 1.26830072017385e-07.
- Dose [1.0mg vs. 2.0mg]: The null hypothesis is rejected. There is overwhelming proof that a dose increase from 1.0mg to 2.0mg will yield more growth. The p-value is far below 0.05 at 1.9064295136718e-05.