Inferential data analysis

Sunday, October 26, 2014

Task

Analyze the ToothGrowth data in the R datasets package.

1. Load the ToothGrowth data and perform some basic exploratory data analyses
2. Provide a basic summary of the data.
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)
4. State your conclusions and the assumptions needed for your conclusions.

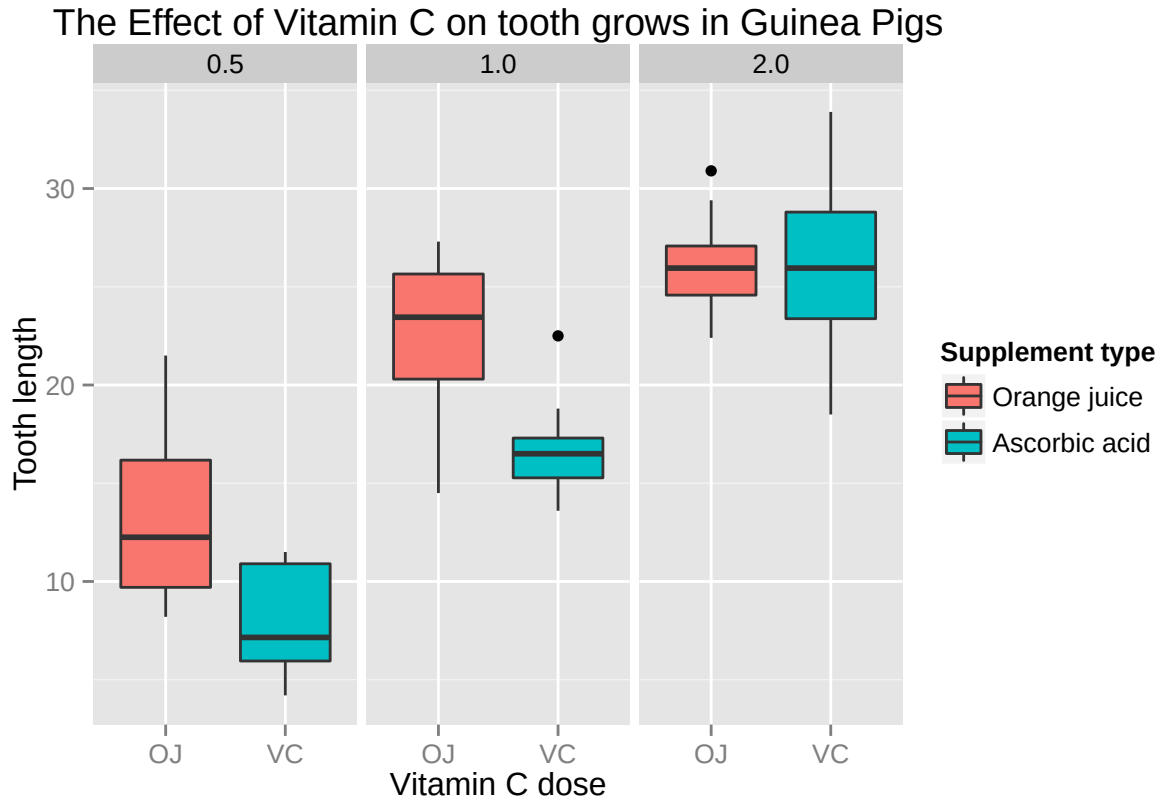
Lets try to answer the questions:

1. Does vitamin C affects on tooth growth in guinea pigs?
2. Does supplement type have affect on tooth growth in guinea pigs?

Loading and exploring the data

```
library(datasets)
library(ggplot2)
tg <- ToothGrowth

ggplot(data = ToothGrowth, aes(x = supp, y = len)) +
        geom_boxplot(aes(fill = supp)) +
        facet_wrap(~ dose) +
        ggtitle("The Effect of Vitamin C on tooth grows in Guinea Pigs") +
        ylab("Tooth length") +
        xlab("Vitamin C dose") +
        scale_fill_hue(name="Supplement type",
                   breaks=c("OJ", "VC"),
                   labels=c("Orange juice", "Ascorbic acid"))
```



Basic summary of the data

```
str(tg)
```

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

```
summary(tg)
```

```
##       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
```

Hypothesis testing

Qestion 1 At first, lets try to find out does supplement type affects on tooth growth in guinea pigs. Assume the null hypothesis $H_0 : \mu_{OJ} = \mu_{VC}$ and alternative hypothesis $H_a : \mu_{OJ} \neq \mu_{VC}$.

```r
t.test(len ~ supp, paired = FALSE, var.equal = FALSE, data = tg)$conf
```

```
## [1] -0.171  7.571
## attr(,"conf.level")
## [1] 0.95
```

As we can see 95% confidence interval contains zero, so we cannot reject the null hypothesis. Lets compare supplement types for each dose amount.

```r
tg05 <- tg[tg$dose == 0.5,]
tg10 <- tg[tg$dose == 1.0,]
tg20 <- tg[tg$dose == 2.0,]
rbind(t.test(len ~ supp, paired = FALSE, var.equal = FALSE, data = tg05)$conf
     ,t.test(len ~ supp, paired = FALSE, var.equal = FALSE, data = tg10)$conf
     ,t.test(len ~ supp, paired = FALSE, var.equal = FALSE, data = tg20)$conf)
```

```
##        [,1]  [,2]
## [1,]  1.719 8.781
## [2,]  2.802 9.058
## [3,] -3.798 3.638
```

We can reject the null hypothesis in case of small doses - 0.5, 1.0 mg.

Question 2 Try to answer does vitamin C influences on tooth growth. Assume the null hypothesis $H_0 : \mu_i = \mu_j$, where $i, j \in \{0.5, 1.0, 2.0\}$. And alternative hypothesis $H_a : \mu_i > \mu_j, i > j$ (larger dose - longer teeth).

```r
rbind(t.test(tg20$len - tg10$len, paired = FALSE, var.equal = FALSE)$conf
     ,t.test(tg20$len - tg05$len, paired = FALSE, var.equal = FALSE)$conf
     ,t.test(tg10$len - tg05$len, paired = FALSE, var.equal = FALSE)$conf)
```

```
##        [,1]   [,2]
## [1,]  3.472  9.258
## [2,] 12.623 18.367
## [3,]  6.387 11.873
```

Calculating the 95% confidence intervals for pairwise differences of dose amounts we can see, that null hypothesis can be rejected for all cases.

Conclusions

1. There is no significant difference in supplement type for large doses of vitamin C. Meanwhile, oranje juice has more effect on teeth length at low doses of vitamin C. In total, there is no significant difference.
2. Vitamin C dose amount directly affects of teeth length. More vitamin C - the longer teeth.

Assumtions

1. All guinea pigs are randomly selected.
2. There are no other factors which can deal confounding effect.