# inference.Rmd

*Damian*

*May 30, 2016*

## *Vitamin C and Tooth Growth in Guinea Pigs*

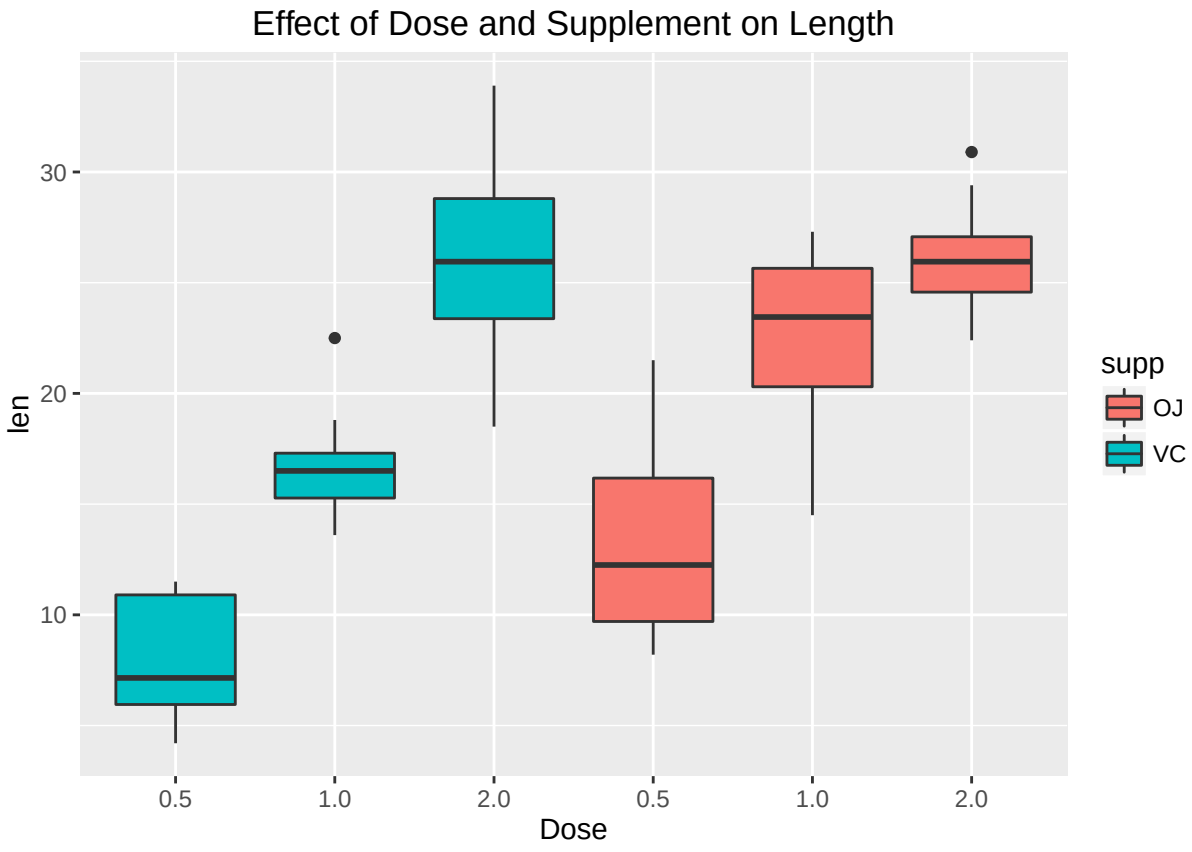
load the data

```
library(ggplot2)
library(datasets)
data("ToothGrowth")
```

### Exploration

I crossed the dose and supplement factors, and plotted tooth length for each level

```
ToothGrowth$DS <- interaction(ToothGrowth$supp, ToothGrowth$dose)
ggplot(data = ToothGrowth, aes(y=len, x=sort(DS), fill = supp)) + geom_boxplot() + scale_x_discrete(labe
```




## Analysis

I wrote a function that would do a two-sample t-test on pairs of tooth length vectors
the null hypothesis for these tests will be that the difference of the means is 0

```
do_t_tests <- function(a_list,alt_hyp){
        # performs two-sample t-tests on all list items
        for(group in seq_along(1:(length(a_list) - 1))){
                for(other_group in (group + 1):length(a_list)){
                    t_test <- t.test(a_list[[other_group]],a_list[[group]],alternative=alt_hyp)
                    t_tests[[paste((names(a_list))[other_group]," vs. ",names(a_list)[group])]] <<- t_t
        }
        }
}
```

I made a list of 6 tooth-length vector, each for one dose-supplement factor level

```
supp_dose_combos <- list()
for(supp_dose_pair in levels(ToothGrowth$DS)){
        supp_dose_combos[[supp_dose_pair]] <- ToothGrowth[(ToothGrowth$DS == supp_dose_pair),]$len
}
```

I split the 6 vectors into 2 groups
* one for guinea pigs that got OJ
* one for gp's that got VC

```
OJ_groups <- supp_dose_combos[grepl(pattern="OJ",x=names(supp_dose_combos))]
VC_groups <- supp_dose_combos[grepl(pattern="VC",x=names(supp_dose_combos))]

supp_group <- list(OJ_groups=OJ_groups, VC_groups=VC_groups)
```

I also split the 6 vectors into 3 groups
one for each dose level: 0.5, 1, and 2

```
half_groups <- supp_dose_combos[grepl(pattern="0.5",x=names(supp_dose_combos))]
one_groups <- supp_dose_combos[grepl(pattern="1",x=names(supp_dose_combos))]
two_groups <- supp_dose_combos[grepl(pattern="2",x=names(supp_dose_combos))]

dose_group <- list(half_groups=half_groups,one_groups=one_groups,two_groups=two_groups)
```

I performed two-sample t-tests on all samples that had the same level for exactly one factor

```
t_tests <- list()

for(each_group in supp_group){
    do_t_tests(each_group,alt_hyp = "greater")
}
for(each_group in dose_group){
    do_t_tests(each_group,alt_hyp = "two.sided")
}
```

I displayed preliminary results
if 0 did not fall in the 95% confidence interval...
then the groups were said to be significantly different

```
for(ind in 1:length(t_tests)) {ifelse(sum(0 > t_tests[[ind]]$conf.int)==0 | sum(0 > t_tests[[ind]]$conf
```

```
## [1] "OJ.1  vs.  OJ.0.5 are significantly different"
## [1] "OJ.2  vs.  OJ.0.5 are significantly different"
## [1] "OJ.2  vs.  OJ.1 are significantly different"
## [1] "VC.1  vs.  VC.0.5 are significantly different"
## [1] "VC.2  vs.  VC.0.5 are significantly different"
## [1] "VC.2  vs.  VC.1 are significantly different"
## [1] "VC.0.5  vs.  OJ.0.5 are significantly different"
## [1] "VC.1  vs.  OJ.1 are significantly different"
## [1] "VC.2  vs.  OJ.2 are not significantly different"
```

I applied the Benjamini-Hochberg correction to fix the FDR at 5%

```
p_scores <- numeric()
for(t_tesst in t_tests){p_scores <<- c(p_scores,t_tesst$p.value)}
adjusted_p_scores <- p.adjust(p_scores,method="BH")
for(ind in 1:length(t_tests)) {ifelse(adjusted_p_scores[ind] < .05 , print(paste(names(t_tests)[[ind]],
```

```
## [1] "OJ.1  vs.  OJ.0.5 are significantly different"
## [1] "OJ.2  vs.  OJ.0.5 are significantly different"
## [1] "OJ.2  vs.  OJ.1 are significantly different"
## [1] "VC.1  vs.  VC.0.5 are significantly different"
## [1] "VC.2  vs.  VC.0.5 are significantly different"
## [1] "VC.2  vs.  VC.1 are significantly different"
## [1] "VC.0.5  vs.  OJ.0.5 are significantly different"
## [1] "VC.1  vs.  OJ.1 are significantly different"
## [1] "VC.2  vs.  OJ.2 are not significantly different"
```

I plotted the p-scores of two sample t-tests on samples with fixed dose

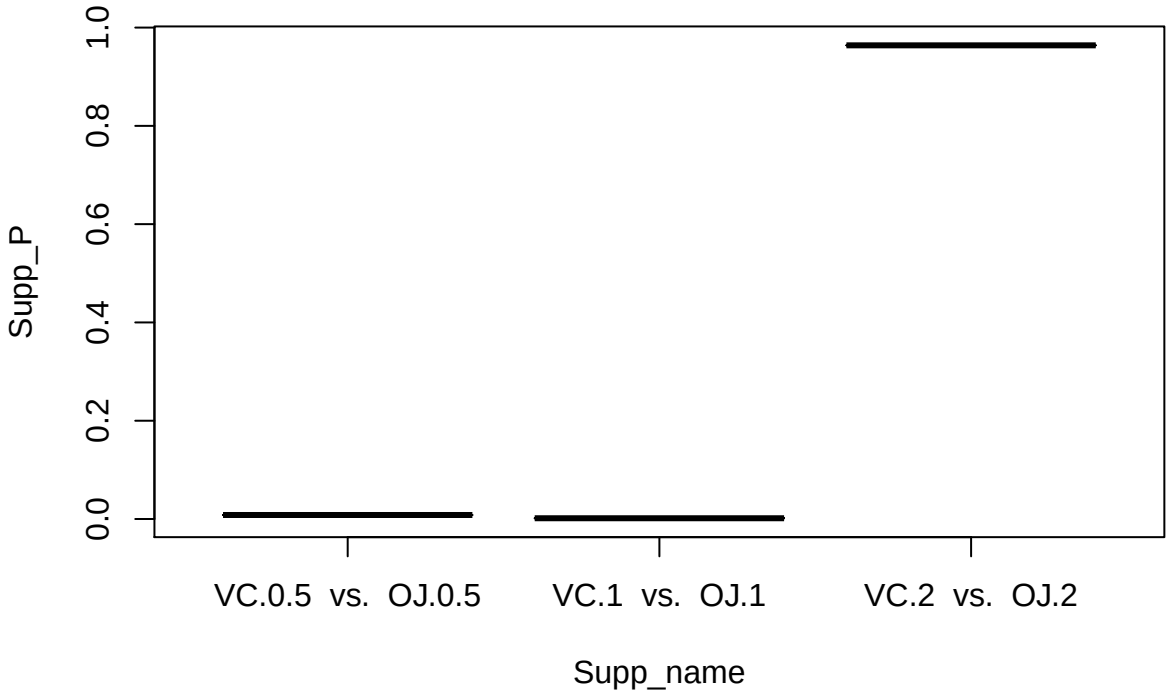
```
the_names <- names(t_tests)

Supp_P <- adjusted_p_scores[7:9]
Supp_name <- the_names[7:9]


sup_df <- data.frame(Supp_P,Supp_name)


par(mfrow=c(1,1))
with(sup_df,plot(Supp_P~Supp_name,main="probability of no significant difference between supplement eff
```

**ity of no significant difference between supplement effectiveness at dif**




## Conclusion

- it appears that larger vitamin C doses significantly increase tooth growth in guinea pigs
- OJ is a more effective supplement at low doses
- at high doses, neither supplement is better Assumptions
- all environmental factors were kept constant across groups
- all guinea pigs were from the same population
- the GP's in the study are representative of the GP population