Project Statistical Inference part 2

Question B: Explore Tooth Growth Data

```
library(datasets);data(ToothGrowth);library(ggplot2)
data <- ToothGrowth

#load function to show 2 ggplots sode by side, from http://gettinggeneticsdone.blogspot.com/2010/03/arr
require(grid)
vp.layout <- function(x, y) viewport(layout.pos.row=x, layout.pos.col=y)
arrange_ggplot2 <- function(..., nrow=NULL, ncol=NULL, as.table=FALSE) {
  dots <- list(...)
  n <- length(dots)
  if(is.null(nrow) & is.null(ncol)) { nrow = floor(n/2) ; ncol = ceiling(n/nrow)}
    if(is.null(nrow)) { nrow = ceiling(n/ncol)}
    if(is.null(ncol)) { ncol = ceiling(n/nrow)}
        ## NOTE see n2mfrow in grDevices for possible alternative
grid.newpage()
pushViewport(viewport(layout=grid.layout(nrow+1,ncol,heights = unit(c(0.5,5),"null")  ) ))
    ii.p <- 1
    for(ii.row in seq(1, nrow)){
    ii.table.row <- ii.row
    if(as.table) {ii.table.row <- nrow - ii.table.row + 1}
        for(ii.col in seq(1, ncol)){
            ii.table <- ii.p
            if(ii.p > n) break
            print(dots[[ii.table]], vp=vp.layout(ii.table.row+1, ii.col))
            ii.p <- ii.p + 1
        }
    }
}
```

```
ch1<-ggplot(data,aes(x=factor(dose),y=len,fill=factor(dose)))+geom_boxplot(notch=T,notchwidth = 0.2)+the
ch1 <- ch1+facet_wrap(~supp)+ theme(legend.position="none")+scale_x_discrete("Dosage in mg")+scale_y_co
ch2<-ggplot(data,aes(x=dose,y=len,color=supp,group=supp))+geom_point()+stat_smooth(method="lm",alpha=0.
ch2 <-ch2+theme_bw()+theme(legend.justification=c(1,0),legend.position=c(1, 0))+scale_color_discrete("S
ch2<-ch2+scale_x_continuous("Dosage in mg")+scale_y_continuous("Length of Teeth")
```

Summary of Tooth Growth Data and Exploratory Data Analysis

```
library(dplyr);library(reshape2);library(knitr)
```

```
##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:stats':
##
##     filter, lag
##
## The following objects are masked from 'package:base':
```

```
##
##     intersect, setdiff, setequal, union
```

```
aux <- data %.% group_by(supp,dose) %.% summarise(Mean=mean(len),Min=min(len),Max=max(len),"Std. Dev"=sd
aux2<-dcast(melt(aux,id.vars=c("supp","dose")),dose~supp+variable,fun.aggregate=sum)
aux2<- round(aux2,2)
names(aux2)[1] <- "Dosage"
kable(aux2, format = "markdown")
```

Dosage	OJ_Mean	OJ_Min	OJ_Max	OJ_Std. Dev	VC_Mean	VC_Min	VC_Max	VC_Std. Dev
0.5	13.23	8.2	21.5	4.46	7.98	4.2	11.5	2.75
1.0	22.70	14.5	27.3	3.91	16.77	13.6	22.5	2.52
2.0	26.06	22.4	30.9	2.66	26.14	18.5	33.9	4.80

The table above shows some sumamry statistics for the Tooth Growth data. It suggest that orange juice is more effective at lower dosages but that bth supplements are similalry effective at the 2mg dose. It also suugests that Vitamon C has less variability in results for lower doses and more for the higher dose

```
arrange_ggplot2(ch1,ch2)
grid.text("Exploratory Analysis for Tooth Growth Data", vp = viewport(layout.pos.row = 1, layout.pos.co
```

The figure above shows the impact of the supplement and the dosage on the length of teeth of Guinea Pigs. The left panel shows the distributions (in boxplots) for the different combinations of supplement and dosage, the fact that the boxplot notched do not overlap is strong evidence that their medians are different and if we assume normality which is logical for bilogical patterns like growth that the means are different (as the median is the same as the mean for a noral distribution) The right panel shows a scatter plot of the data with linear regression fits of teeth length as a function of dosage for each supplement. Both Panels show that higher dosages appears to be related to longer teeth (however there appears there may be some diminishig returns for larger doeses of orange juice). It is also possible to see that fr dosage of 0.5 and 1 mg, orange Juice appears to be far superior, whereas Vitamin C could potentially lead to longer teeth at the 2mg dose.

Compare Peformance and Dosage

- Is Orange Juice more effective than Vitamin C?

```
t.test(len~supp,paired=F,var.equal=T,data=data,alternative="g")$p.value
t.test(len~supp,paired=F,var.equal=F,data=data,alternative="g")$p.value

#adjusted values
t.test(len~supp,paired=F,var.equal=T,data=data,alternative="g")$p.value*50
t.test(len~supp,paired=F,var.equal=F,data=data,alternative="g")$p.value*50
```

```
## [1] 0.0302
```

```
## [1] 0.03032
```

```
## [1] 1.51
```

```
## [1] 1.516
```

It appears from performing t-test (either with equal or unequal variance) that overall orange juice had a greater impact on tooth growth than Vitamin C as the confidence intervals do not contain zero. However if we use the Bonferroni adjusted values we can't conclude that there is any difference in the 2 supplements

- Does Dosage Matter?

```
results = data.frame(supplement=character(),dose_1=numeric(),dose_2=numeric(),ll=numeric(),
                     ul=numeric(),pval=numeric(),variance_equal=logical(),test=character(),stringsAsFac
k=1
for (i in levels(data$supp)) {
  for (j in unique(data$dose)) {
    for (test in c("greater","less")) {
      for (variance in c(T,F)) {
        data1 <- subset(data,supp==i & dose!=j)
        temp <- t.test(len~dose,paired=F,var.equal=variance,data=data1,alternative=test)
        results[k,"supplement"] <- i; results[k,"dose_1"] <- unique(data1$dose)[1];
        results[k,"dose_2"] <- unique(data1$dose)[2]; results[k,"ll"] <- temp$conf.int[1]
        results[k,"ul"] <- temp$conf.int[2]; results[k,"pval"] <- temp$p.value
        results[k,"variance_equal"] <- variance; results[k,"test"] <- test
        k <- k+1
      }
    }
  }
}

results$adj_pval <- results$pval*50
results$outcome <- ifelse(results$pval<=0.05,"Difference","No Difference")
results[c(4:6,9)] <- round(results[c(4:6,9)],3)
results2 <- subset(results,outcome=="Difference")
results2 <- results2[,c(1:6,9,7,8,10)]
kable(results2, format = "markdown")
```

	supplement	dose_1	dose_2	ll	ul	pval	adj_pval	variance_equal	test	outcome
3	OJ	1.0	2	-Inf	-0.768	0.019	0.934	TRUE	less	Difference
4	OJ	1.0	2	-Inf	-0.749	0.020	0.980	FALSE	less	Difference
7	OJ	0.5	2	-Inf	-9.984	0.000	0.000	TRUE	less	Difference
8	OJ	0.5	2	-Inf	-9.948	0.000	0.000	FALSE	less	Difference
11	OJ	0.5	1	-Inf	-6.217	0.000	0.002	TRUE	less	Difference
12	OJ	0.5	1	-Inf	-6.214	0.000	0.002	FALSE	less	Difference
15	VC	1.0	2	-Inf	-6.399	0.000	0.001	TRUE	less	Difference
16	VC	1.0	2	-Inf	-6.347	0.000	0.002	FALSE	less	Difference
19	VC	0.5	2	-Inf	-15.129	0.000	0.000	TRUE	less	Difference
20	VC	0.5	2	-Inf	-15.086	0.000	0.000	FALSE	less	Difference
23	VC	0.5	1	-Inf	-6.748	0.000	0.000	TRUE	less	Difference
24	VC	0.5	1	-Inf	-6.747	0.000	0.000	FALSE	less	Difference

The table above shows the dosage comparisons for each supplement for which a statistical difference at 95% was observed. It shows that higher dosages are more effective that lower dosages for either supplement. However, when we use the Bonferroni adjustment for p-values (uisng n=50 as in general we coducted 50 comparisons), we can't conclude anymore that the 2mg dosage for oramge juice is more effective than the 1mg dosage

- Which Supplement is Better at a Given Dosage?

```
results_a = data.frame(dose=numeric(),supplement1=character(),supplement2=character(),ll=numeric(),
                      ul=numeric(),pval=numeric(),variance_equal=logical(),test=character(),stringsAsFact
k=1
for (j in unique(data$dose)) {
    for (test in c("greater","less")) {
      for (variance in c(T,F)) {
        data1 <- subset(data, dose==j)
        temp <- t.test(len~supp,paired=F,var.equal=variance,data=data1,alternative=test)
        results_a[k,"supplement1"] <- levels(data1$supp)[1]; results_a[k,"supplement2"] <- levels(data1$
        results_a[k,"dose"] <- j; results_a[k,"ll"] <- temp$conf.int[1]
        results_a[k,"ul"] <- temp$conf.int[2]; results_a[k,"pval"] <- temp$p.value
        results_a[k,"variance_equal"] <- variance; results_a[k,"test"] <- test
        k <- k+1
      }
    }
}

results_a$adj_pval <- results_a$pval*50
results_a$outcome <- ifelse(results_a$pval<=0.05,"Difference","No Difference")
results_a[c(4:6,9)] <- round(results_a[c(4:6,9)],3)
results_a2 <- subset(results_a,outcome=="Difference")
results_a2 <- results_a2[,c(1:6,9,7:8,10)]
kable(results_a2, format = "markdown")
```

	dose	supplement1	supplement2	ll	ul	pval	adj_pval	variance_equal	test	outcome
1	0.5	OJ	VC	2.378	Inf	0.003	0.133	TRUE	greater	Difference
2	0.5	OJ	VC	2.346	Inf	0.003	0.159	FALSE	greater	Difference
5	1.0	OJ	VC	3.380	Inf	0.000	0.020	TRUE	greater	Difference
6	1.0	OJ	VC	3.356	Inf	0.001	0.026	FALSE	greater	Difference

The table above shows the supplement comparisons for each dosage for which a statistical difference at 95% was observed. It shows that orange juice is more effective at 0.5 and 1 mg but that both supplements are equally effecive at 2mg. When we use the Bonferroni adjustment on the p-values, we can only conclude that orange juice is more effective at 1mg

- Which Dosage is Better?

```
results_b = data.frame(dose1=numeric(),dose2=numeric(),ll=numeric(),ul=numeric(),pval=numeric(),
                        variance_equal=logical(),test=character(),stringsAsFactors=F)
k=1
for (j in unique(data$dose)) {
```

```r
    for (test in c("greater","less")) {
      for (variance in c(T,F)) {
        data1 <- subset(data, dose!=j)
        temp <- t.test(len~dose,paired=F,var.equal=variance,data=data1,alternative=test)
        results_b[k,"dose1"] <- unique(data1$dose)[1]; results_b[k,"dose2"] <- unique(data1$dose)[2]
        results_b[k,"ll"] <- temp$conf.int[1]; results_b[k,"ul"] <- temp$conf.int[2];
        results_b[k,"pval"] <- temp$p.value
        results_b[k,"variance_equal"] <- variance; results_b[k,"test"] <- test
        k <- k+1
      }
    }
}
results_b$adj_pval <- results_b$pval*50
results_b$outcome <- ifelse(results_b$pval<=0.05,"Difference","No Difference")
results_b[c(3:5,8)] <- round(results_b[c(3:5,8)],3)
results_b$difference <- ifelse(results_b$outcome=="Difference",T,F)
results_b2 <- subset(results_b,outcome=="Difference")
results_b2 <- results_b2[c(1:5,8,6,7,9:10)]
kable(results_b2, format = "markdown")
```

	dose1	dose2	ll	ul	pval	adj_pval	variance_equal	test	outcome	difference
3	1.0	2	-Inf	-4.175	0	0	TRUE	less	Difference	TRUE
4	1.0	2	-Inf	-4.174	0	0	FALSE	less	Difference	TRUE
7	0.5	2	-Inf	-13.281	0	0	TRUE	less	Difference	TRUE
8	0.5	2	-Inf	-13.279	0	0	FALSE	less	Difference	TRUE
11	0.5	1	-Inf	-6.753	0	0	TRUE	less	Difference	TRUE
12	0.5	1	-Inf	-6.753	0	0	FALSE	less	Difference	TRUE

The table above shows the comparisons across supplements for each dosage for which a statistical difference at 95% was observed. It shows that 2mg is superior to 1 and 0.5 mgs, and that 1mg is superior to 0.5 mg. The results are valid even if we use a Bonferroni adjustmnt on the p-values