Statistical Inference Project: part 2

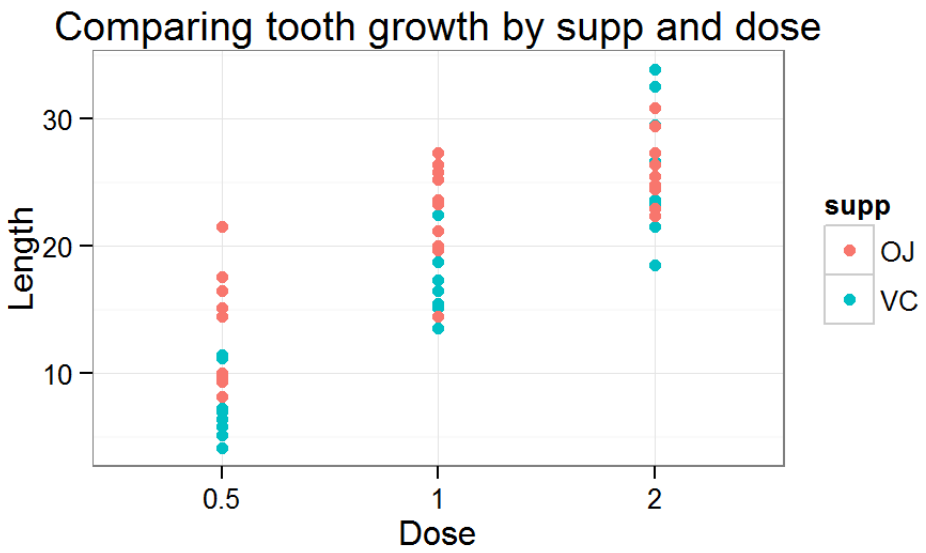
Instruction: Now in the second portion of the class, we're going to analyze the ToothGrowth data in the R datasets package.

1. Load the ToothGrowth data and perform some basic exploratory data analyses

```
data(ToothGrowth); str(ToothGrowth); library(ggplot2)
```

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

```
ggplot(data=ToothGrowth, aes(x=factor(dose), y=len, color=supp))  + geom_point()+ xlab("Dose")+
ylab("Length")+theme_bw()+ggtitle("Comparing tooth growth by supp and dose")
```



2. Provide a basic summary of the data. *Summary:*

- The data has 60 records and three variables(len, supp, dose).
- There are three doses (0.5, 1, 2) under the dose varialble and 2 factors ("VC", "OJ") under the supp variable.
- The len variable has the tooth growth data, which can be divided into to six group according to the variables dose and supp.
- Below is a table reshaped to show the mean and standard deviation of each group.

	dose	supp	mean	sd	Group
1	0.50	OJ	13.23	4.46	1
2	0.50	VC	7.98	2.75	2
3	1.00	OJ	22.70	3.91	3
4	1.00	VC	16.77	2.52	4

5	2.00	OJ	26.06	2.66	5
6	2.00	VC	26.14	4.80	6

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)

- calculate the 95% confidence interval between the two groups as stated in the table below. we use unpaired t-test assuming the groups are independent.
- We first tried unequal variance t-test. we also tried t.test assuming equal variance. Both equal or unequal variances reach a same conclusion.

```
conFun <- function(m=1, n=2, df=dat) {temp<-t.test(df[m,3:12],df[n,3:12],paired=F, var.equal=F)
$conf[1:2];
    return(data.frame(Data_1=paste("group_",m,": dose=",df[m,1],", supp=",df[m,2],sep=""),
    Data_2=paste("group_",n,": dose=",df[n,1],", supp=",df[n,2],sep=""),Lowlimit=temp[1],Uplimi
t=temp[2]))}
print(xtable(rbind(conFun(1,2),conFun(3,4),conFun(5,6))),type='html')
```

	Data_1	Data_2	Lowlimit	Uplimit
1	group_1: dose=0.5, supp=OJ	group_2: dose=0.5, supp=VC	1.72	8.78
2	group_3: dose=1, supp=OJ	group_4: dose=1, supp=VC	2.80	9.06
3	group_5: dose=2, supp=OJ	group_6: dose=2, supp=VC	-3.80	3.64

```
print(xtable(rbind(conFun(1,3),conFun(3,5),conFun(2,4),conFun(4,6))),type='html')
```

	Data_1	Data_2	Lowlimit	Uplimit
1	group_1: dose=0.5, supp=OJ	group_3: dose=1, supp=OJ	-13.42	-5.52
2	group_3: dose=1, supp=OJ	group_5: dose=2, supp=OJ	-6.53	-0.19
3	group_2: dose=0.5, supp=VC	group_4: dose=1, supp=VC	-11.27	-6.31
4	group_4: dose=1, supp=VC	group_6: dose=2, supp=VC	-13.05	-5.69

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

- The 95% confidence intervals (or CI) between "OJ" and "VC" for dose =0.5 and 1 is above zero, however the 95% CI between "OJ" and "VC" for the dose of 2 includes zero.
- We have 95% confidence that "OJ"" has more effect on the tooth growth that "VC" at lower dose (0.5 and 1).
- "OJ" and "VC" have similar effect on tooth growth at a higher dose of 2.