Statistical Inference Assignment 1 Part 2

MHiero

Wednesday, September 17, 2014

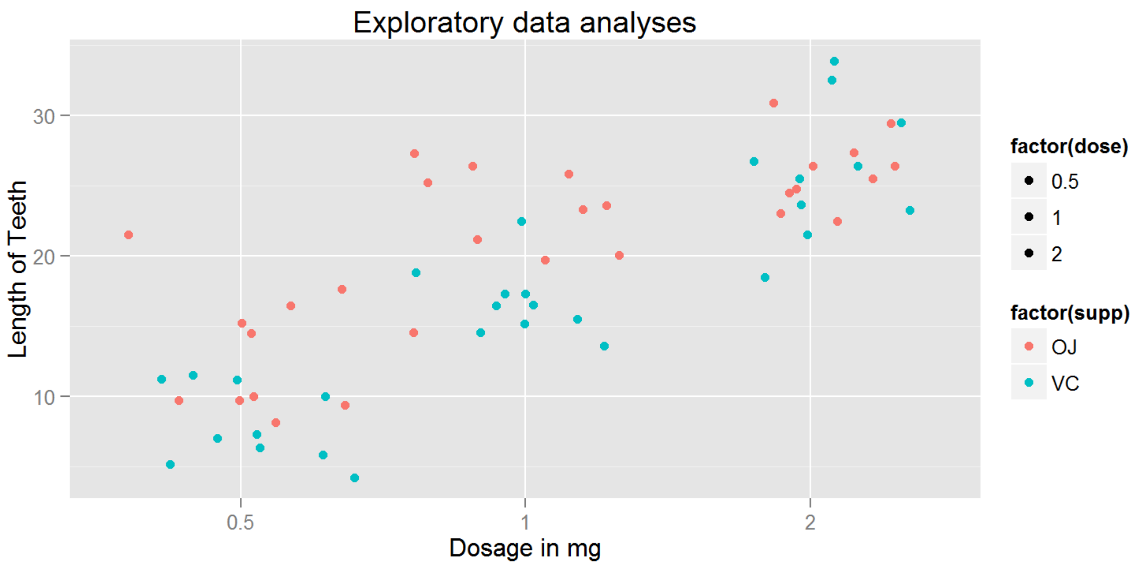
The Effect of Vitamin C on Tooth Growth in Guinea Pigs

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). [...R Documentation on ToothGrowth{datasets}]

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

```
# Prerequisites and loading data "ToothGrowth"
# help(ToothGrowth)

library(ggplot2)
data(ToothGrowth)
ggplot(ToothGrowth, aes(x=factor(dose),y=len,fill=factor(dose)))+
geom_jitter(aes(colour = factor(supp)))+
scale_x_discrete("Dosage in mg") +
scale_y_continuous("Length of Teeth") +
ggtitle("Exploratory data analyses")
```



The help(ToothGrowth) function gives enough information about the variables to make a scatterplot. The scatterplot reveals that longer teeth tend to receive a higher dosage of both supps.

2. Provide a basic summary of the data.

```
head(ToothGrowth)

##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5

summary(ToothGrowth)

##       len        supp         dose      
##  Min.   : 4.20   OJ:30   Min.   :0.50  
##  1st Qu.:13.07   VC:30   1st Qu.:0.50  
##  Median :19.25           Median :1.00  
##  Mean   :18.81           Mean   :1.17  
##  3rd Qu.:25.27           3rd Qu.:2.00  
##  Max.   :33.90           Max.   :2.00  

summary(ToothGrowth$len)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##     4.2    13.1    19.2    18.8    25.3    33.9

summary(ToothGrowth$dose)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##    0.50    0.50    1.00    1.17    2.00    2.00

summary(ToothGrowth$supp)

## OJ VC 
## 30 30
```

In addition to the scatterplot and help function the summary function is used to provide more information on 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)

```
supp.t1 <- t.test(len~supp, paired=F, var.equal=T, data=ToothGrowth)
supp.t2 <- t.test(len~supp, paired=F, var.equal=F, data=ToothGrowth)
supp.result <- data.frame("p-value"=c(supp.t1$p.value, supp.t2$p.value),"Conf-Low"=c(supp.t1$conf[1],supp.t2$conf[1]),
                          "Conf-High"=c(supp.t1$conf[2],supp.t2$conf[2]), row.names=c("Equal Var","Unequal Var"))
supp.result
```

```
##             p.value Conf.Low Conf.High
## Equal Var   0.06039   -0.167     7.567
## Unequal Var 0.06063   -0.171     7.571
```

The p values of both equal variance and unequal variance t tests, in terms of Orange Juice & Vitamin C, are larger than 5 percent and the confidence intervals of both t tests contain zero. That is why null hypothesis cannot be denied. It is not possible to conclude that there are differences between OJ and VC groups.

```
dose.05 <- ToothGrowth[which(ToothGrowth$dose==.5),1]
dose.10 <- ToothGrowth[which(ToothGrowth$dose==1),1]
dose.20 <- ToothGrowth[which(ToothGrowth$dose==2),1]
dose0510.t1 <- t.test(dose.05, dose.10, paired=F, var.equal=T)
dose0510.t2 <- t.test(dose.05, dose.10, paired=F, var.equal=F)
dose0510.result <- data.frame("p-value"=c(dose0510.t1$p.value, dose0510.t2$p.value),"Conf-Low"=c(dose0510.t1$conf[1],dose0510.t2$conf[1]),"Conf-High"=c(dose0510.t1$conf[2],dose0510.t2$conf[2]), row.names=c("Equal Var","Unequal Var"), "Dose"="0.5 to 1")
dose0520.t1 <- t.test(dose.05, dose.20, paired=F, var.equal=T)
dose0520.t2 <- t.test(dose.05, dose.20, paired=F, var.equal=F)
dose0520.result <- data.frame("p-value"=c(dose0520.t1$p.value, dose0520.t2$p.value),"Conf-Low"=c(dose0520.t1$conf[1],dose0520.t2$conf[1]),"Conf-High"=c(dose0520.t1$conf[2],dose0520.t2$conf[2]), row.names=c("Equal Var","Unequal Var"), "Dose"="0.5 to 2")
dose1020.t1 <- t.test(dose.10, dose.20, paired=F, var.equal=T)
dose1020.t2 <- t.test(dose.10, dose.20, paired=F, var.equal=F)
dose1020.result <- data.frame("p-value"=c(dose1020.t1$p.value, dose1020.t2$p.value),"Conf-
```

```
Low"=c(dose1020.t1$conf[1],dose1020.t2$conf[1]),"Conf-
High"=c(dose1020.t1$conf[2],dose1020.t2$conf[2]), row.names=c("Equal Var","Unequal Var"),
"Dose"="1 to 2")
dose.result <- rbind(dose0510.result,dose0520.result,dose1020.result)
dose.result
```

```
##                 p.value Conf.Low Conf.High     Dose
## Equal Var     1.266e-07  -11.984    -6.276 0.5 to 1
## Unequal Var   1.268e-07  -11.984    -6.276 0.5 to 1
## Equal Var1    2.838e-14  -18.154   -12.836 0.5 to 2
## Unequal Var1  4.398e-14  -18.156   -12.834 0.5 to 2
## Equal Var2    1.811e-05   -8.994    -3.736   1 to 2
## Unequal Var2  1.906e-05   -8.996    -3.734   1 to 2
```

The result shows that the dosages have an impact on the tooth growth. All p values are very tiny and the confidence intervals are not zero. That is why the null hypthesis may be denied. It may be concluded that true difference in means is not equal to 0 among the groups with different dosages. It is clear that higher dosage tends to result in higher tooth length.

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

The P-value for drawing an association between the tooth length and the supplement type is larger than between the tooth length and dose.

It is concluded that the dose affects the tooth length a lot more than the supplement type, 2mg dose has larger impact on tooth growth than 1mg and 0.5mg, and 1mg dose has more impact than 0.5mg dose. (...But for accurate conclusion additional tests are required.)

Assumptions for the Analysis are

- The supplements have a treatment effect and there are no other confounding factors.
- Samples are unpaired, with unequal variances.
- The Guinea pigs are essentially identical - spezies, diet, environmental exposure etc.