Statistical Inference project

Jochen Schmiedbauer

Sunday, January 25, 2015

Background

we're going to analyze the ToothGrowth data in the R datasets package.

```
Load the ToothGrowth data and perform some basic exploratory data analyses
Provide a basic summary of the data.
Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use th
State your conclusions and the assumptions needed for your conclusions.
```

we assume that the data is evenly distributed and (therefore) approximately symmetric around the mean.

load and explore data

```
data(ToothGrowth)
summary(ToothGrowth)
```

```
##       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
```

```
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
```

as per the documentation, the Toothgrow data set, provides the length of odontoblasts for 10 test persons for 3 doses of vitamin C (0.5, 1 & 2 mg) for the two delivery methods (orange juice or ascorbic acid)

show data by delivery method and dosage

```
library(ggplot2)
plot<-ggplot(ToothGrowth,aes(x=supp,y=len))
plot+geom_point(size=3,aes(color=supp))+facet_grid(~dose,labeller=label_both)+labs(title="Tooth Length
```

calculate the mean and SD for these groups

```
library(plyr)
Toothgrowstats <- ddply(ToothGrowth,.(supp,dose),summarise, mean = mean(len), sd = sd(len))
Toothgrowstats
```

```
##   supp dose  mean    sd
## 1   OJ  0.5 13.23 4.460
## 2   OJ    1 22.70 3.911
## 3   OJ    2 26.06 2.655
## 4   VC  0.5  7.98 2.747
## 5   VC    1 16.77 2.515
## 6   VC    2 26.14 4.798
```

Conclusion

per dose level we will verify the most effective delivery method by looking at the 95% confidence intervals of the difference in means.

For dose level 0.5

```
OrangeJuicemean <- subset(Toothgrowstats,dose==0.5 & supp=='OJ')$mean
VC_ascorbic_acidmean <- subset(Toothgrowstats,dose==0.5 & supp=='VC')$mean
diffmean <- OrangeJuicemean - VC_ascorbic_acidmean
sd1 <- subset(Toothgrowstats,dose==0.5 & supp=='OJ')$sd
sd2 <- subset(Toothgrowstats,dose==0.5 & supp=='VC')$sd
n1 <- n2 <- 10
df <- (sd1^2/n1 + sd2^2/n2)^2 / (((sd1^2/n1)^2/(n1-1)) + ((sd2^2/n2)/(n2-1)))
tdf <- qt(0.975,df)
tConf05 <- diffmean + c(-1,1)*tdf*(sd1^2/n1 + sd2^2/n2)^0.5
```

95 % confidence interval is

```
round(tConf05,2)
```

```
## [1] 1.71 8.79
```

As it is above 0 we conclude that Orange Juice is the more effective delivery method for 0.5 mg doses.

For dose level 1.0

```
OrangeJuicemean <- subset(Toothgrowstats,dose==1.0 & supp=='OJ')$mean
VC_ascorbic_acidmean <- subset(Toothgrowstats,dose==1.0& supp=='VC')$mean
diffmean <- OrangeJuicemean - VC_ascorbic_acidmean
sd1 <- subset(Toothgrowstats,dose==1.0 & supp=='OJ')$sd
sd2 <- subset(Toothgrowstats,dose==1.0 & supp=='VC')$sd
n1 <- n2 <- 10
df <- (sd1^2/n1 + sd2^2/n2)^2 / (((sd1^2/n1)^2/(n1-1)) + ((sd2^2/n2)/(n2-1)))
tdf <- qt(0.975,df)
tConf1 <- diffmean + c(-1,1)*tdf*(sd1^2/n1 + sd2^2/n2)^0.5
```

95 % confidence interval is

```
round(tConf1,2)
```

```
## [1] 2.78 9.08
```

As it is above 0 we conclude that Orange Juice is the more effective delivery method for 1 mg doses.

For dose level 2.0

```
OrangeJuicemean <- subset(Toothgrowstats,dose==2.0 & supp=='OJ')$mean
VC_ascorbic_acidmean <- subset(Toothgrowstats,dose==2.0 & supp=='VC')$mean
diffmean <- OrangeJuicemean - VC_ascorbic_acidmean
sd1 <- subset(Toothgrowstats,dose==2.0 & supp=='OJ')$sd
sd2 <- subset(Toothgrowstats,dose==2.0 & supp=='VC')$sd
```

```
n1 <- n2 <- 10
df <- (sd1^2/n1 + sd2^2/n2)^2 / (((sd1^2/n1)^2/(n1-1)) + ((sd2^2/n2)/(n2-1)))
tdf <- qt(0.975,df)
tConf2 <- diffmean + c(-1,1)*tdf*(sd1^2/n1 + sd2^2/n2)^0.5
```

95 % confidence interval is

```
round(tConf2,2)
```

```
## [1] -3.63  3.47
```

As the range includes 0 we can't conclude that Orange Juice is the more effective delivery method for 2 mg doses. But neither can we say that Vitamin C ascorbic acid is more efficient