Stats 332 Assignment 3

Junqiao Lin

Mar 06, 2019

```r
library(survey)
```

Question 1.

```r
#importing the dataset
Q1.dat <- read.csv("tomato.csv")
Q1.dat <- Q1.dat[!is.na(Q1.dat$mass),]
```

```r
Q1.trtmean <- tapply(Q1.dat$mass,Q1.dat$treatment, mean)
Q1.trtsize <- tapply(Q1.dat$mass,Q1.dat$treatment, length)
Q1.mean <- mean(Q1.trtmean)
```

a)

```r
#Calculating SStrt
Q1.SSTRT = sum(Q1.trtsize*(Q1.trtmean - Q1.mean)^2)
Q1.SSTRT
```

```
## [1] 0.04753454
```

```r
Q1.MSSTrt<-Q1.SSTRT/5
Q1.MSSTrt
```

```
## [1] 0.009506908
```

```r
#Cakculating SSres
Q1.SSTOT = var(Q1.dat$mass)* (sum(Q1.trtsize) - 1)
Q1.SSRES = Q1.SSTOT - Q1.SSTRT
Q1.SSTOT
```

```
## [1] 0.1039377
```

```r
Q1.SSRES
```

```
## [1] 0.05640318
```

```r
Q1.MSSRES<-Q1.SSRES/26
Q1.MSSRES
```

```
## [1] 0.002169353
```

```r
Q1.Fobs <- Q1.MSSTrt/Q1.MSSRES
Q1.Fobs
```

```
## [1] 4.38237
```

Hence we can fill out the following ANOVA table:

Source	SS	df	MSS	F-stats
Treatment	0.04753454	5	0.009506908	4.38237
Residual	0.05640318	26	0.002169353	
Total	0.1039377	31		

b)

Under H_0 we have $\tau_1 = \tau_2 = ...\tau_6 = 0$ (there is no difference among the six treatment).

```r
qf(0.95, 5, 26)
```

```
## [1] 2.58679
```

Since $F_{obs} > 2.58679$, we reject H_0 at $\alpha = 0.05$.

c)

```r
Q1.lst1 <- Q1.dat[which(Q1.dat$treatment=="A1"),]$mass
Q1.lst2 <- Q1.dat[which(Q1.dat$treatment=="B3"),]$mass
Q1c.meandiff = mean(Q1.lst1) - mean(Q1.lst2)
df = 8
v.hat <- 0.002169353

test <- Q1c.meandiff/sqrt(v.hat*(1/6 + 1/4))
```

Note under the t distribution:

```r
Q1c.meandiff
```

```
## [1] 0.09083333
```

```r
qt(c(0.025,0.975),8)
```

```
## [1] -2.306004  2.306004
```

```r
abs(Q1c.meandiff) - 2.306004 *sqrt(v.hat*(1/6 + 1/4))
```

```
## [1] 0.0215036
```

```r
abs(Q1c.meandiff) + 2.306004 *sqrt(v.hat*(1/6 + 1/4))
```

```
## [1] 0.1601631
```

Hence the 95% confidence interval for the difference bewteen A_1 and B_3 is (0.0215036,0.1601631). From the confidence interval, there is significant evidence that treatment A_1 and B_3.

```r
#importing the dataset
Q2.dat <- read.csv("cloth.csv")
```

a) We can define the following statistical model for the model as the following:

$$Y_{ij} = \mu + \tau_i + \beta_j + R_{ij}, \quad i = 1, ..., 4 \quad j = 1, ..., 5 \quad R_{ij} \overset{i.i.d.}{\sim} N(0, \sigma^2)$$

Where τ_i denote the treatment of each chemical and β_j denote the block effect from the different bolts of cloth.

Under this model, we are assuming the effect of the cloth being in different bolt would be independent from the effect of the different chemical. b)

```r
#Overall variance
Q2.SSTOT = var(Q2.dat$strength)* 19
#Sample varaince with treatment
Q2.SSTrt = var(tapply(Q2.dat$strength,Q2.dat$chemical,mean))* 3
#Sample varaince with block
Q2.SSblock = var(tapply(Q2.dat$strength,Q2.dat$bolt,mean))* 4

#Sample varaince with block
Q2.SSRes = Q2.SSTOT - Q2.SSTrt - Q2.SSblock
Q2.SSTOT
```

```
## [1] 191.75
```

```r
Q2.SSTrt
```

```
## [1] 2.59
```

```r
Q2.SSblock
```

```
## [1] 39.25
```

```r
Q2.SSRes
```

```
## [1] 149.91
```

Hence we have the following ANOVA table

Source	SS	df	MSS	F-stats
Treatment	2.59	3	0.86333	0.0691078
Block	39.25	4	9.8125	0.78547
Residual	149.91	12	12.4925	
Total	191.75	19		

Let $H_0 : \tau_1 = ... = \tau_4 = 0$, note

```r
qf(0.95, 3, 19)
```

```
## [1] 3.12735
```

Since $F_{obs} < 3.12735$, hence we fail to reject H_0 under the significant level of $\alpha = 0.05$

Question 3.

```r
#importing the dataset
Q3.dat <- read.csv("tv.csv")
```

a) We can define the following statistical model for the model as the following:

$$Y_{ij} = \mu + \tau_i + \alpha_j + \gamma_{ij}, \quad i = 1,2 \quad j = 1,2 \quad R_{ij} \overset{i.i.d.}{\sim} N(0, \sigma^2)$$

Where τ_i is the treatment effect from factor A and α_j is the treatment effect from factor B, (1 being low and 2 being high). γ_{ij} denote the interaction effect bewteen Factor A and Factor B.

b) Alright please don't kill me for using xkcd style graph.

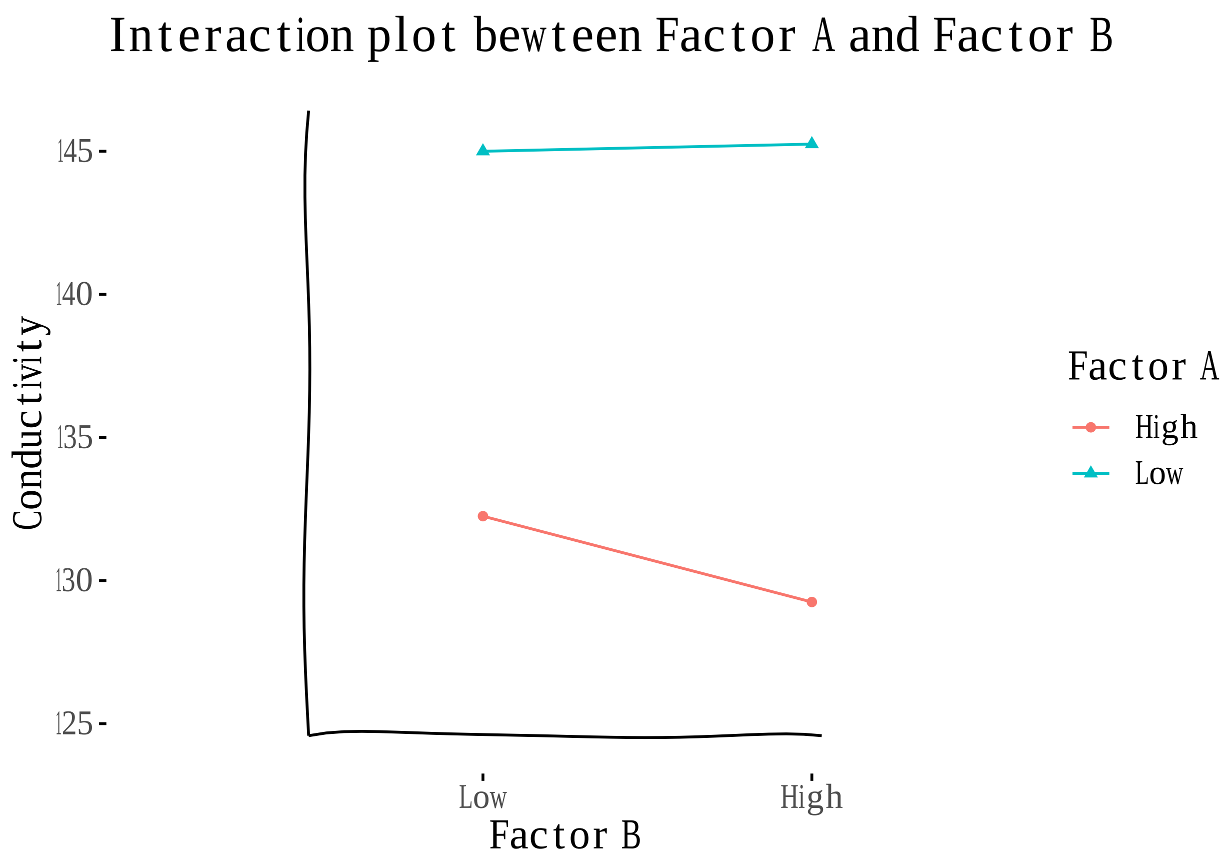
```r
#A bit Lazy :p
Q3.ll <- mean(Q3.dat[c(1:4),]$conduct)
Q3.lh <- mean(Q3.dat[c(5:8),]$conduct)
Q3.hl <- mean(Q3.dat[c(9:12),]$conduct)
Q3.hh <- mean(Q3.dat[c(13:16),]$conduct)

Q3.dataf=data.frame(facA=factor(c("Low","Low","High","High")),  facB=factor(c("Low","High","Low","High")
#datalines.pltl <- data.frame(xbegin=c(1),ybegin=c(Q3.ll), xend=c(2), yend=c(Q3.lh))
#datalines.plth <- data.frame(xbegin=c(1),ybegin=c(Q3.hl), xend=c(2), yend=c(Q3.hh))
```

```r
library(xkcd)
library(extrafont)
```

```r
windowsFonts("xkcd" = windowsFont("xkcd Script"))
Q3.plot <- ggplot(data=Q3.dataf, aes(x = facB, y = cond))  + geom_point(aes(shape=facA,colour=facA))  +
                            breaks=c("High", "Low"),
                            labels=c("High", "Low")) +
      scale_shape_discrete(name  ="Factor A",
                           breaks=c("High", "Low"),
                           labels=c("High", "Low"))+ labs(title="Interaction plot bewteen Factor A and F
#+ xkcdline(aes(xbegin,ybegin, xend,yend),datalines.pltl, xjitteramount = 0.12)

Q3.plot
```



c)

```
summary(aov(Q3.dat$conduct ~ Q3.dat$A*Q3.dat$B))
```

```
##                    Df Sum Sq Mean Sq F value   Pr(>F)
## Q3.dat$A            1  826.6   826.6  41.984 3.03e-05 ***
## Q3.dat$B            1    7.6     7.6   0.384    0.547
## Q3.dat$A:Q3.dat$B   1   10.6    10.6   0.537    0.478
## Residuals          12  236.3    19.7
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

i) Base on the ANOVA table, we have that factor A is significant while factor B seems to be insignifican with the effect of the other factor present.

```
(826.6+7.6+ 10.6)/(3*19.7)
```

```
## [1] 14.29442
```

```
qf(0.95, 3, 12)
```

```
## [1] 3.490295
```

Note since $F_{obs} > 3.490295$,there is significant evidence bewteen the variable.

ii) Note since $0.478 > 0.05$ we fail to reject the null hypothesis such that $\gamma_{ij} = 0$

iii) For factor B begin low, under the null hyphosis assume factor A is the same. Hence we have the following for H_0 (let ct_i be coat type i):

$$(\hat{ct}_1 - \hat{ct}_3) - (\hat{ct}_2 - \hat{ct}_4) = 0$$

```r
theta = (Q3.ll - Q3.hl) - (Q3.lh - Q3.hh)
theta
```

```
## [1] -3.25
```

from the ANOVA table we have $\sigma^2 = 19.7$

```r
SE <- sqrt(19.7) * sqrt(4/16)
theta/SE
```

```
## [1] -1.464469
```

```r
qt(c(0.025,0.975),12)
```

```
## [1] -2.178813  2.178813
```

Since $|-1.464469| < 2.178813$ We fail to reject the null hyphosis such that the effect of factor A is independent from factor B.