Multivariate Statistics

By Saba Akram

1/2/2021

# Two Sample T-Test for difference in means.  
 ## Inputing the summary statistics from Chapter 5 table 5.1.  
xbar1<-1.5714  
xbar2<-1.5361  
s1<-1.3126  
s2<-1.3251  
n1<-2529  
n2<-804  
dfs<-min(n1-1,n2-1)  
tdata<-(xbar1-xbar2)/sqrt((s1^2/n1)+(s2^2/n2))  
tdata

## [1] 0.6594724

## Greater that 0 i.e. positive  
pvalue<-2\*pt(tdata,df=dfs,lower.tail = FALSE)  
pvalue

## [1] 0.5097815

# Two Sample Z Test for difference in proppoortions.  
## Inputing the summary statistics.  
x1<-707  
x2<-215  
n1<-2529  
n2<-804  
p1<-x1/n1  
p2<-x2/n2  
ppooled<-(x1+x2)/(n1+n2)  
zdata<-(p1-p2)/sqrt((ppooled\*(1-ppooled)\*(1/n1)+(1/n2)))  
zdata

## [1] 0.3338909

pvalue<-2\*pnorm(abs(zdata),lower.tail = FALSE)  
pvalue

## [1] 0.7384619

# Chi square test for homogenity of proprtions  
table5.2<-as.table(rbind(c(410,340,250),c(95,85,70)))  
table5.2

## A B C  
## A 410 340 250  
## B 95 85 70

dimnames(table5.2)<-list(Data.Set=c("Training Set","Test Set"),Status=c("Married","Single","Other"))  
table5.2

## Status  
## Data.Set Married Single Other  
## Training Set 410 340 250  
## Test Set 95 85 70

Xsq\_data<-chisq.test(table5.2)  
Xsq\_data

##   
## Pearson's Chi-squared test  
##   
## data: table5.2  
## X-squared = 1.1487, df = 2, p-value = 0.5631

Xsq\_data$statistic

## X-squared   
## 1.14867

Xsq\_data$p.value

## [1] 0.5630793

Xsq\_data$expected

## Status  
## Data.Set Married Single Other  
## Training Set 404 340 256  
## Test Set 101 85 64

# Chi-Square goodness of Fit of Multinomial data.  
## Populations proportins  
p\_status<-c(0.40,0.35,0.25)  
## Observed frequencies.  
o\_status<-c(36,35,29)  
chisq.test(o\_status,p=p\_status)

##   
## Chi-squared test for given probabilities  
##   
## data: o\_status  
## X-squared = 1.04, df = 2, p-value = 0.5945

# ANOVA  
a<-c(30,40,50,60)  
b<-c(25,30,50,55)  
c<-c(25,30,40,45)  
ab<-append(a,b)  
datavalue<-append(ab,c)  
datavalue

## [1] 30 40 50 60 25 30 50 55 25 30 40 45

datalabels<-factor(c(rep("a",length(a)),rep("b",length(b)),rep("c",length(c))))  
datalabels

## [1] a a a a b b b b c c c c  
## Levels: a b c

anova.results<-aov(datavalue~datalabels)  
anova.results

## Call:  
## aov(formula = datavalue ~ datalabels)  
##   
## Terms:  
## datalabels Residuals  
## Sum of Squares 200 1400  
## Deg. of Freedom 2 9  
##   
## Residual standard error: 12.47219  
## Estimated effects may be unbalanced