Exercise 6 - Chi Square Test

S.DEIVANAYKI (21BCS003)

26.03.2024

#Include Necessary Data “mtcars”

```
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

head(mtcars)

##                    mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

#2. Create Contingency Table

```
cyl_carb <- data.frame(mtcars$cyl,mtcars$carb)
cars_data <- table(cyl_carb)
cars_data

##             mtcars.carb
## mtcars.cyl 1 2 3 4 6 8
##          4 5 6 0 0 0 0
##          6 2 0 0 4 1 0
##          8 0 4 3 6 0 1
```

#Apply Chi-Square Test without API

```
e<-matrix(nrow=3,ncol=6)
for (i in 1:3)
{
  for(j in 1:6)
  {
```

```
    e[i,j]<-(sum(cars_data[i,])*sum(cars_data[,j]))/sum(cars_data)
  }
}
e

##          [,1]   [,2]    [,3]   [,4]    [,5]    [,6]
## [1,] 2.40625 3.4375 1.03125 3.4375 0.34375 0.34375
## [2,] 1.53125 2.1875 0.65625 2.1875 0.21875 0.21875
## [3,] 3.06250 4.3750 1.31250 4.3750 0.43750 0.43750

chi_squ=0.0
for (i in 1:3)
{
  for (j in 1:6)
  {
    chi_squ<- chi_squ + ((e[i,j]-cars_data[i,j])^2 / e[i,j])
  }
}
chi_squ

## [1] 24.38887

if(chi_squ>10.843)
{
  print("cyl and carb are dependent")
}

## [1] "cyl and carb are dependent"

if(chi_squ<=10.843)
{
  print("cyl and carb are dependent")
}
```

#Apply Chi Square Test with API

```
chisq.test(cars_data)

## Warning in chisq.test(cars_data): Chi-squared approximation may be
incorrect

##
##  Pearson's Chi-squared test
##
## data:  cars_data
## X-squared = 24.389, df = 10, p-value = 0.006632
```

#Find Covariance between cyl and carb without API

```
A=mean(mtcars$cyl)
B=mean(mtcars$carb)
res=0
```

```
res <- res + (mtcars$cyl-A)*(mtcars$carb-B)
res <- res/nrow(cyl_carb)
covariance <- sum(res)
covariance

## [1] 1.472656
```

#Find Covariance between cyl and carb with API

```
cov(mtcars$cyl,mtcars$carb)

## [1] 1.520161
```

#Find Correlation between cyl and carb without API

```
vara <- sqrt(sum((mtcars$cyl-A)^2)/nrow(cyl_carb))
varb <- sqrt(sum((mtcars$carb-B)^2)/nrow(cyl_carb))
correlaton <- covariance/(vara*varb)
correlaton

## [1] 0.5269883
```

#Find Correlation between cyl and carb with API

```
cor(mtcars$cyl,mtcars$carb)

## [1] 0.5269883
```

#Apply Two Sample t-test.

```
t.test(mtcars$cyl,mtcars$carb)

## 
##  Welch Two Sample t-test
## 
## data:  mtcars$cyl and mtcars$carb
## t = 7.9286, df = 61.384, p-value = 5.609e-11
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.523917 4.226083
## sample estimates:
## mean of x mean of y 
##    6.1875    2.8125
```

#Apply Annova Test

```
mtcars_aov <- aov(mtcars$cyl~mtcars$carb)
mtcars_aov

## Call:
##    aov(formula = mtcars$cyl ~ mtcars$carb)
## 
## Terms:
```

```
##                  mtcars$carb Residuals
## Sum of Squares     27.45923  71.41577
## Deg. of Freedom           1        30
##
## Residual standard error: 1.542895
## Estimated effects may be unbalanced
```

#Include Necessary Data “survey”

```
library(MASS)

head(survey)

##      Sex Wr.Hnd NW.Hnd W.Hnd    Fold Pulse    Clap Exer Smoke Height
M.I
## 1 Female   18.5   18.0 Right  R on L    92    Left Some Never 173.00
Metric
## 2   Male   19.5   20.5  Left  R on L   104    Left None Regul 177.80
Imperial
## 3   Male   18.0   13.3 Right  L on R    87 Neither None Occas     NA
<NA>
## 4   Male   18.8   18.9 Right  R on L    NA Neither None Never 160.00
Imperial
##      Age
## 1 18.250
## 2 17.583
## 3 16.917
## 4 20.333
```

#2. Create Contingency Table

```
smoke_exer <- data.frame(survey$Smoke,survey$Exer)
survey_data <- table(smoke_exer)
survey_data

##             survey.Exer
## survey.Smoke Freq None Some
##        Heavy    7    1    3
##        Never   87   18   84
##        Occas   12    3    4
##        Regul    9    1    7
```

#Apply Chi-Square Test without API

```
e<-matrix(nrow=4,ncol=3)
for (i in 1:4)
{
  for(j in 1:3)
  {
    e[i,j]<-(sum(survey_data[i,])*sum(survey_data[,j]))/sum(survey_data)
  }
```

```
}
e

##           [,1]      [,2]      [,3]
## [1,]  5.360169  1.072034  4.567797
## [2,] 92.097458 18.419492 78.483051
## [3,]  9.258475  1.851695  7.889831
## [4,]  8.283898  1.656780  7.059322

chi_squ=0.0
for (i in 1:4)
{
  for (j in 1:3)
  {
    chi_squ<- chi_squ + ((e[i,j]-survey_data[i,j])^2 / e[i,j])
  }
}
chi_squ

## [1] 5.488546
```

#Apply Chi Square Test with API

```
chisq.test(survey_data)

## Warning in chisq.test(survey_data): Chi-squared approximation may be
incorrect

##
##  Pearson's Chi-squared test
##
## data:  survey_data
## X-squared = 5.4885, df = 6, p-value = 0.4828
```

#Find Covariance between smoke and exer without API

```
smoke_exer <- data.frame(Smoke=survey$Smoke,Exer=survey$Exer)
smoke_exer$Smoke<-as.numeric(smoke_exer$Smoke)
smoke_exer$Exer<-as.numeric(smoke_exer$Exer)
smoke_exer <- smoke_exer[!is.na(smoke_exer$Smoke),]
A=mean(smoke_exer$Smoke)
B=mean(smoke_exer$Exer)
res=0.0
res <- res + (smoke_exer$Smoke-A)*(smoke_exer$Exer-B)
res <- res/nrow(smoke_exer)
covariance <- sum(res)
covariance

## [1] -0.02107871
```

#Find Covariance between cyl and carb with API

```r
cov(smoke_exer$Smoke,smoke_exer$Exer)

## [1] -0.02116841
```

#Find Correlation between cyl and carb without API

```r
vara <- sqrt(sum((smoke_exer$Smoke-A)^2)/nrow(cyl_carb))
varb <- sqrt(sum((smoke_exer$Exer-B)^2)/nrow(cyl_carb))
correlaton <- covariance/(vara*varb)
correlaton

## [1] -0.004871593
```

#Find Covariance between cyl and carb with API

```r
cor(smoke_exer$Smoke,smoke_exer$Exer)

## [1] -0.035928
```

#Apply Two Sample t-test.

```r
t.test(smoke_exer$Smoke,smoke_exer$Exer)

##
##  Welch Two Sample t-test
##
## data:  smoke_exer$Smoke and smoke_exer$Exer
## t = 3.3862, df = 404.87, p-value = 0.000778
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.104863 0.395137
## sample estimates:
## mean of x mean of y
##  2.177966  1.927966
```

#Apply Annova Test.

```r
mtcars_aov <- aov(mtcars$disp~factor(mtcars$gear))
summary(mtcars_aov)

##                     Df Sum Sq Mean Sq F value   Pr(>F)
## factor(mtcars$gear)  2 280221  140110   20.73 2.56e-06 ***
## Residuals           29 195964    6757
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```