Hypothesis testing

library(BSDA)

## Loading required package: lattice

##   
## Attaching package: 'BSDA'

## The following object is masked from 'package:datasets':  
##   
## Orange

library(nortest)  
library(readr)  
  
###### test for significance difference in the diameter of the cutlet between two units  
cutlet<-read.csv("E:\\assignments data\\hypothesis testing\\Cutlets.csv")  
View(cutlet)  
attach(cutlet)  
head(cutlet)

## Unit.A Unit.B  
## 1 6.8090 6.7703  
## 2 6.4376 7.5093  
## 3 6.9157 6.7300  
## 4 7.3012 6.7878  
## 5 7.4488 7.1522  
## 6 7.3871 6.8110

str(cutlet)

## 'data.frame': 35 obs. of 2 variables:  
## $ Unit.A: num 6.81 6.44 6.92 7.3 7.45 ...  
## $ Unit.B: num 6.77 7.51 6.73 6.79 7.15 ...

## normality test  
shapiro.test(Unit.A)

##   
## Shapiro-Wilk normality test  
##   
## data: Unit.A  
## W = 0.96495, p-value = 0.32

shapiro.test(Unit.B)

##   
## Shapiro-Wilk normality test  
##   
## data: Unit.B  
## W = 0.97273, p-value = 0.5225

## p values of unitA and unit B are >0.05, so data is in normal format  
## variance test  
var.test(Unit.A,Unit.B)

##   
## F test to compare two variances  
##   
## data: Unit.A and Unit.B  
## F = 0.70536, num df = 34, denom df = 34, p-value = 0.3136  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
## 0.3560436 1.3974120  
## sample estimates:  
## ratio of variances   
## 0.7053649

## p>0.05 , unit A and Unit B have equal variances  
## two sample t test for equal variances  
t.test(Unit.A, Unit.B,mu=0, alternative="two.sided",conf.level=0.95, var.equal=TRUE)

##   
## Two Sample t-test  
##   
## data: Unit.A and Unit.B  
## t = 0.72287, df = 68, p-value = 0.4722  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -0.09646454 0.20605311  
## sample estimates:  
## mean of x mean of y   
## 7.019091 6.964297

## p>0.05, accept Ho  
##there is no significant difference in the diameter of the cutlet between two units

###test for finding difference in average TAT among the different laboratories   
labTAT<-read.csv("file:///E:/assignments data/hypothesis testing/LabTAT.csv")  
View(labTAT)  
attach(labTAT)  
str(labTAT)

## 'data.frame': 120 obs. of 4 variables:  
## $ Laboratory.1: num 185 170 193 177 193 ...  
## $ Laboratory.2: num 166 186 195 183 170 ...  
## $ Laboratory.3: num 177 198 201 200 205 ...  
## $ Laboratory.4: num 166 161 185 176 153 ...

head(labTAT)

## Laboratory.1 Laboratory.2 Laboratory.3 Laboratory.4  
## 1 185.35 165.53 176.70 166.13  
## 2 170.49 185.91 198.45 160.79  
## 3 192.77 194.92 201.23 185.18  
## 4 177.33 183.00 199.61 176.42  
## 5 193.41 169.57 204.63 152.60  
## 6 179.45 197.00 181.51 161.12

## variance test  
data<-stack(labTAT)  
View(data)  
shapiro.test(data$values)

##   
## Shapiro-Wilk normality test  
##   
## data: data$values  
## W = 0.99494, p-value = 0.1175

## all the variables have p>0.05, so data is normally distributed  
## variance test  
library(car)

## Loading required package: carData

##   
## Attaching package: 'carData'

## The following objects are masked from 'package:BSDA':  
##   
## Vocab, Wool

leveneTest(data$values~data$ind, data=labTAT)

## Levene's Test for Homogeneity of Variance (center = median)  
## Df F value Pr(>F)   
## group 3 2.5996 0.05161 .  
## 476   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

### p>0.05, variances are equal  
## Anova test  
anova\_results<-aov(data$values~ind, data=data)  
summary(anova\_results)

## Df Sum Sq Mean Sq F value Pr(>F)   
## ind 3 79979 26660 118.7 <2e-16 \*\*\*  
## Residuals 476 106905 225   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

### p<0.05, accept Ha  
## there is difference in average TAT among the different laboratories

###test for finding if male-female buyer rations are similar across regions  
BuyerRatio<-read.csv("E:\\assignments data\\hypothesis testing\\BuyerRatio.csv")  
View(BuyerRatio)  
str(BuyerRatio)

## 'data.frame': 2 obs. of 5 variables:  
## $ Observed.Values: Factor w/ 2 levels "Females","Males": 2 1  
## $ East : int 50 435  
## $ West : int 142 1523  
## $ North : int 131 1356  
## $ South : int 70 750

stacked\_data<-stack(BuyerRatio)

## Warning in stack.data.frame(BuyerRatio): non-vector columns will be ignored

View(stacked\_data)  
table(stacked\_data$ind, stacked\_data$values)

##   
## 50 70 131 142 435 750 1356 1523  
## East 1 0 0 0 1 0 0 0  
## West 0 0 0 1 0 0 0 1  
## North 0 0 1 0 0 0 1 0  
## South 0 1 0 0 0 1 0 0

###chisquare test  
chisq.test(table(stacked\_data$ind, stacked\_data$values))

## Warning in chisq.test(table(stacked\_data$ind, stacked\_data$values)): Chi-squared  
## approximation may be incorrect

##   
## Pearson's Chi-squared test  
##   
## data: table(stacked\_data$ind, stacked\_data$values)  
## X-squared = 24, df = 21, p-value = 0.2931

### p>0.05, accept Ho  
##male-female buyer rations are similar across regions  
  
  
  
cof<-read.csv("E:\\assignments data\\hypothesis testing\\Costomer+OrderForm.csv")  
View(cof)  
attach(cof)  
stacked\_data<-stack(lapply(cof, as.character))  
View(stacked\_data)  
table(stacked\_data$ind, stacked\_data$values)

##   
## Defective Error Free  
## Phillippines 29 271  
## Indonesia 33 267  
## Malta 31 269  
## India 20 280

## chisquare test  
chisq.test(table(stacked\_data$ind, stacked\_data$values))

##   
## Pearson's Chi-squared test  
##   
## data: table(stacked\_data$ind, stacked\_data$values)  
## X-squared = 3.859, df = 3, p-value = 0.2771

### p>0.05, acceptHo  
##4 centers haved equal efective %   
  
  
  
###test on finding male versus females walking in to the store differ based on day of the week  
data<-read.csv("E:\\assignments data\\hypothesis testing\\Faltoons.csv")  
View(data)  
table(data)

## Weekend  
## Weekdays Female Male  
## Female 167 120  
## Male 66 47

prop.test(table(data), alternative = "two.sided", conf.level = 0.95, correct = TRUE)

##   
## 2-sample test for equality of proportions with continuity correction  
##   
## data: table(data)  
## X-squared = 2.605e-30, df = 1, p-value = 1  
## alternative hypothesis: two.sided  
## 95 percent confidence interval:  
## -0.1116865 0.1073080  
## sample estimates:  
## prop 1 prop 2   
## 0.5818815 0.5840708

### two sample proportion test  
prop.test(table(data), alternative = "two.sided", conf.level = 0.95, correct = TRUE)

##   
## 2-sample test for equality of proportions with continuity correction  
##   
## data: table(data)  
## X-squared = 2.605e-30, df = 1, p-value = 1  
## alternative hypothesis: two.sided  
## 95 percent confidence interval:  
## -0.1116865 0.1073080  
## sample estimates:  
## prop 1 prop 2   
## 0.5818815 0.5840708

### p>0.05, acceptHo  
##there is no difference male versus females walking in to the store based on day of the week