F1

Auzzer\_Pang

6/11/2020

#读取数据，并将原始数据命名为data1  
setwd("/Users/auzzer\_pang")  
data = read.csv("f1.csv")

#先对部门进行分析,先是政府(1)然后是私人(2)  
data\_1 = data[1:10,]  
data\_2 = data[11:20,]  
shapiro.test(data\_1[,3])

##   
## Shapiro-Wilk normality test  
##   
## data: data\_1[, 3]  
## W = 0.93243, p-value = 0.4722

shapiro.test(data\_2[,3])

##   
## Shapiro-Wilk normality test  
##   
## data: data\_2[, 3]  
## W = 0.92266, p-value = 0.3797

# 再对性别进行检验,先是男性(3)再是女性(4)  
data\_3 = cbind.data.frame(data[1:5,],data[11:15,])  
shapiro.test(cbind(data\_3[,3],data\_3[,6]))

##   
## Shapiro-Wilk normality test  
##   
## data: cbind(data\_3[, 3], data\_3[, 6])  
## W = 0.89838, p-value = 0.2103

data\_4 = cbind.data.frame(data[6:10,],data[16:20,])  
shapiro.test(cbind(data\_4[,3],data\_4[,6]))

##   
## Shapiro-Wilk normality test  
##   
## data: cbind(data\_4[, 3], data\_4[, 6])  
## W = 0.95766, p-value = 0.759

#P值均大于显著性水平a=0.05,因此不能拒绝原假设，说明数据在因子A的三个水平下都是来自正态分布的。

#接下来进行方差齐性检验  
x = factor(rep(1:4,each=5))  
y = data[1:20,3]  
bartlett.test(y~x,data=data)

##   
## Bartlett test of homogeneity of variances  
##   
## data: y by x  
## Bartlett's K-squared = 1.3337, df = 3, p-value = 0.7211

# 由于P值远远大于显著性水平a=0.05，因此不能拒绝原假设，我们认为不同水平下的数据是等方差的。

#各组均值  
aggregate(data\_1[,3], by=list(bumen = data\_1[,1]), FUN=mean)

## bumen x  
## 1 0 998.3

aggregate(data\_2[,3], by=list(bumen = data\_2[,1]), FUN=mean)

## bumen x  
## 1 1 1183.5

aggregate(data\_3[,3], by=list(bumen = data\_3[,1]), FUN=mean)

## bumen x  
## 1 0 1018

aggregate(data\_4[,3], by=list(bumen = data\_4[,1]), FUN=mean)

## bumen x  
## 1 0 978.6

#各组标准差  
aggregate(data\_1[,3], by=list(department = data\_1[,1]), FUN=sd)

## department x  
## 1 0 64.81264

aggregate(data\_2[,3], by=list(department = data\_2[,1]), FUN=sd)

## department x  
## 1 1 97.82552

aggregate(data\_3[,3], by=list(department = data\_3[,1]), FUN=sd)

## department x  
## 1 0 61.42068

aggregate(data\_4[,3], by=list(department = data\_4[,1]), FUN=sd)

## department x  
## 1 0 68.6207

#方差分析,若P值小于0.05，则拒绝原假设，认为有显著影响  
aov\_sex = aov(data[,3]~data[,2] , data = data)  
summary(aov\_sex)

## Df Sum Sq Mean Sq F value Pr(>F)   
## data[, 2] 1 52224 52224 3.865 0.0649 .  
## Residuals 18 243206 13511   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

aov\_department = aov(data[,3]~data[,1] , data = data)  
summary(aov\_department)

## Df Sum Sq Mean Sq F value Pr(>F)   
## data[, 1] 1 171495 171495 24.91 9.47e-05 \*\*\*  
## Residuals 18 123935 6885   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#交互作用  
aov\_intere = aov(data[,3]~data[,1]\*data[,2] , data = data)  
summary(aov\_intere)

## Df Sum Sq Mean Sq F value Pr(>F)   
## data[, 1] 1 171495 171495 52.777 1.89e-06 \*\*\*  
## data[, 2] 1 52224 52224 16.072 0.00101 \*\*   
## data[, 1]:data[, 2] 1 19719 19719 6.068 0.02548 \*   
## Residuals 16 51991 3249   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1