**12.11 R 语言应用**

**# 陈文贤 着 《大话统计学》 清华大学出版社 2022年**

if(!require(DescTools)){install.packages("DescTools")} ; library(DescTools)

if(!require(vcd)){install.packages("vcd")} ; library(vcd)

data <- read.csv("https://goo.gl/j6lRXD") ; str(data) # 读入CSV 数据

T <- table(data$treatment, data$improvement) ; T

Xsq <- chisq.test(T, correct=FALSE) ; Xsq

K <- Xsq$statistic ; S <- sum(Xsq$observed)

Phi(T) ; sqrt(K/S)

ContCoef(T) ; sqrt(K/(K+S))

a <- nrow(T) ; b <- ncol(T) ; m <- min(a,b)

CramerV(T) ; sqrt(K/(S\*(m-1)))

assocstats(T)

# Chap12\_2

Y <- c(13, 24, 31, 18, 11, 3)

P <- c(0.1353, 0.2707, 0.2707, 0.1804, 0.0902, 0.0527)

k <- chisq.test(Y, p=P) ; k # 卡方检验

x <- k$statistic

pchisq(x, 4, lower.tail = FALSE) # 自由度减 1 的 p 值

# Chap12\_4

M <- as.table(rbind(c(199, 119, 174, 214), c(130, 166, 536, 685)))

dimnames(M) <- list(Survived = c("Yes", "No"), Class = c("1st", "2nd", "3rd", "Crew"))

(ChiSq <- chisq.test(M)) # 卡方检验

# Chap12\_9

x<-matrix(c(48,5,12,35), nrow = 2,

dimnames = list("A公司" = c("Yes", "No"), "B公司" = c("Yes", "No"))) ; x

mcnemar.test(x) # 修正 卡方检验值 Qc

mcnemar.test(x,correct=F) # 未修正 卡方检验值 Q

# Chap12\_10 家庭分工专案资料

if(!require(gplots)){install.packages("gplots")} ; library(gplots)

if(!require(graphics)){install.packages("graphics")} ; library(graphics)

if(!require(vcd)){install.packages("vcd")} ; library(vcd)

if(!require(corrplot)){install.packages("corrplot")} ; library(corrplot)

if(!require(profvis)){install.packages("profvis")} ; library(profvis)

data = read.delim("C:/大话统计学 网络资源/StatData/Chap12\_10.txt", row.names = 1)

tab <- as.table(as.matrix(data)) ; tab # 转成表格

balloonplot(t(tab), main ="家庭分工项目 列联表 气球图", xlab ="家庭", ylab="项目",

label = FALSE, show.margins = FALSE) # 列联表 气球图

mosaicplot(tab,shade=T, las=2,main="马赛克图", ylab ="家庭", xlab="项目")

assoc(tab, shade = TRUE, las=3, ylab ="家庭", xlab="项目")

chisq <- chisq.test(tab) ; chisq

chisq$observed # 样本观察值

round(chisq$expected,2) # 理论期望值

round(chisq$residuals, 3) # 残差 = (观察值 - 期望值) / sqrt(期望值)

corrplot(chisq$residuals, is.cor = FALSE) # 卡方残差相关图

### 卡方检验 比例相等

success <- c(9,4)

total <- c(12,13)

prop.test(success, total) # 两比例相等检验 , 有连续性修正

binom.test(success, total) # 两比例相等检验 , 二项分布检验

mat <- matrix(c(9,4,3,9),2) ; mat

fisher.test(mat) #两比例相等Fisher检验

chisq.test(mat) #两比例相等 卡方检验 , Yates 连续性修正

### 卡方检验 独立性/同构性

if(!require(ISwR)){install.packages("ISwR")} ; library(ISwR)

caesar.shoe ; shoe <- caesar.shoe # 孕产妇鞋号(鞋子尺码)和剖腹生产

yes <- shoe["Yes",]

total <- margin.table(shoe,2)

prop.test(yes, total)

prop.trend.test(yes, total)

### 卡方检验 独立性/同构性

Mat <- matrix(c(652,1537,598,242,36,46,38,21,218 ,327,106,67), nrow=3,byrow=T)

colnames(Mat) <- c("A","B","C","D")

rownames(Mat) <- c("X","Y","Z")

Mat

chisq.test(Mat)

chisq.test(Mat)$expected

chisq.test(Mat)$observed

E <- chisq.test(Mat)$expected

O <- chisq.test(Mat)$observed

(O-E)^2/E

X <- matrix(c(83, 80, 90, 50, 86, 93, 136, 82) , nrow=2, byrow=T)

colnames(X) <- c("A","B","C","D")

rownames(X) <- c("smokers","patients")

smokers <- X[1, ] ; patients <- X[2, ]

prop.test(smokers, patients) # 检验 H0 : p1=p2=p3=p4 ; H1 : pi <> pj 有不相等

pairwise.prop.test(smokers, patients) # 成对比较

# 检验 H0 : p1=p2=p3=p4 ; H1 : p1>p2>p3>p4 or p1<p2<p3<p4

prop.trend.test(smokers, patients)

# 检验 H0 : p1=p2=p3=p4 ; H1 : p2>p3>p4>p1 or p2<p3<p4<p1

prop.trend.test(smokers, patients, c(2,3,4,1))

# 检验 H0 : p1=p2=p3=p4 ; H1 : p4>p1>p3>p2 or p4<p1<p3<p2

prop.trend.test(smokers, patients, c(4,1,3,2))

if(!require(lawstat)){install.packages("lawstat")} ; library(lawstat)

data(Titanic, package = "datasets") # 泰坦尼克号存活数据

str(Titanic)

TitanicCSY <- margin.table(Titanic, c(2,4,1))

TitanicCSY # 2×2×4 表格 (性别, 存活, 身分)

TitanicCY <- margin.table(Titanic, c(1,4)) #

TitanicCY # 4×2 表格 (身分, 存活)

TitanicSY <- margin.table(Titanic, c(2,4))

TitanicSY # 2×2 表格 (性别, 存活)

TitanicAY <- margin.table(Titanic, c(3,4))

TitanicAY # 2×2 表格 (年龄, 存活)

(MS1 <- margin.table(TitanicSY, 1)) # 边际列表 行(性别)总和

TitanicSY[, 2] # 性别存活数

(MA1 <- margin.table(TitanicAY, 1)) # 边际列表 行(年纪别)总和

TitanicAY[, 2] # 年纪别存活率

(MC1 <- margin.table(TitanicCY, 1)) # 边际列表 行(身分别)总和

TitanicCY[, 2] # 身分别存活率

# 比例检验

prop.test(TitanicCY) # 四总体比例相等检验

prop.test(TitanicCY[, 2], MC1)

prop.test(TitanicSY) # 两总体比例相等检验

prop.test(TitanicSY[, 2], MS1)

prop.test(TitanicAY) # 两总体比例相等检验

prop.test(TitanicAY[, 2], MA1)

binom.test(TitanicSY[, 2], MS1)

binom.test(TitanicCY[, 2], MC1) # 不适用四总体比例相等检验

binom.test(TitanicAY[, 2], MA1)

# 卡方独立性检验 m×n 表格

chisq.test(TitanicCY)

chisq.test(TitanicSY)

chisq.test(TitanicAY)

# Fisher精确检验 2×2 表格, 检验 优势比(odds ratio) = 1 , 独立性检验

fisher.test(TitanicSY)

fisher.test(TitanicAY)

# Cochran-Mantel-Haenszel检验 2×2×k 表格, 检验 优势比(odds ratio) = 1 , 独立性检验

mantelhaen.test(TitanicCSY)

cmh.test(TitanicCSY)

# 优势比，[比值比](https://baike.baidu.com/item/%E6%AF%94%E5%80%BC%E6%AF%94) (odds ratio)

if(!require(samplesizeCMH)){install.packages("samplesizeCMH")} ; library(samplesizeCMH)

odds.ratio(TitanicSY)

odds.ratio(TitanicAY)

apply(TitanicCSY, 3, odds.ratio)