**STAT 40001/MA59800 Statistical Computing Fall 2017**

**Lab-14**

1. The flu season in southern Nevada for 2005–2006 ran from December to April, the coldest months of the year. The Southern Nevada Health District reported the numbers of vaccine-preventable influenza cases shown in Table below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| December 2005 | January 2006 | February 2006 | March 2006 | April 2006 |
| 62 | 84 | 17 | 16 | 21 |

Test whether the numbers of flu cases in the district are equally distributed among the five flu season months. That is, we wish to know if flu cases follow a uniform distribution

Let p1 , p2 , p3, p4 and p5 are probability of different month respectively. We would like to test   
 H0 : p1 = p2 = p3 = p4 = p5

Ha : at least one of the probabilities is different from others

x = c(62, 84, 17,16,21)

chisq.test(x)

Chi-squared test for given probabilities

data: x

X-squared = 97.15, df = 4, p-value < 2.2e-16

p-value < 0.05, reject null hypothesis.

1. Table below provides data on the top 5 Olympic medal winners in 2016 Olympic

|  |  |  |  |
| --- | --- | --- | --- |
| Country | Gold | Silver | Bronze |
| United States | 46 | 29 | 29 |
| China | 38 | 27 | 22 |
| Russia | 24 | 25 | 33 |
| Britain | 29 | 17 | 19 |
| Germany | 11 | 19 | 14 |

Display the information by creating stack barplot and side-by-side barplot.

USA = c(46, 29, 29)

China = c(38, 27, 22)

Russia = c(24,25,33)

Britain = c(29, 17, 19)

Germany = c(11,19,14)

names(USA) = c("Gold","Silver","Bronze")

names(China) = c("Gold","Silver","Bronze")

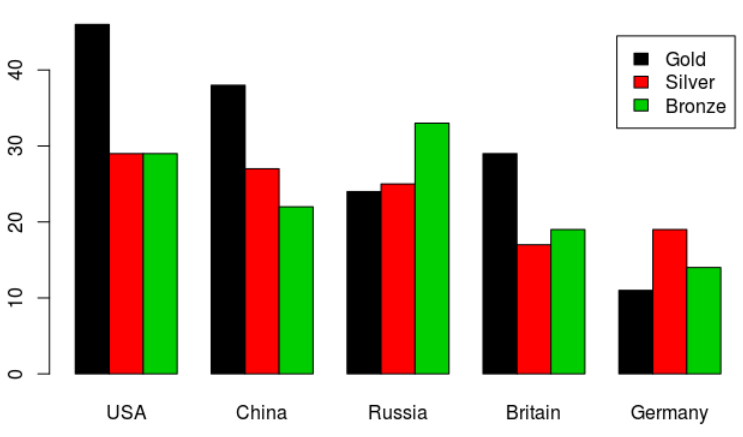
names(Russia) = c("Gold","Silver","Bronze")

names(Britain) = c("Gold","Silver","Bronze")

names(Germany) = c("Gold","Silver","Bronze")

medal = cbind(USA,China,Russia,Britain,Germany)

barplot(medal, legend = rownames(medal), col = c(1,2,3), beside = TRUE)



1. Are health and happiness related? The following data represent the level of happiness and level of health for a random sample of individuals from the General Social Survey.

|  | Health | | | | |
| --- | --- | --- | --- | --- | --- |
|  |  | Excellent | Good | Fair | Poor |
| Happiness | Very Happy | 271 | 261 | 82 | 20 |
| Pretty Happy | 247 | 567 | 231 | 53 |
| Not Too Happy | 33 | 103 | 92 | 36 |

Does the evidence suggest that health and happiness are related? Use the *α*=0.05  level of significance.  
VeryHappy = c(271,261,82,20)

PrettyHappy = c(247,567,231,53)

NotTooHappy = c(33,103,92,36)

table = rbind(VeryHappy,PrettyHappy,NotTooHappy)

data=data.frame(table)

names(data) = c('Excellent', 'Good', 'Fair', 'Poor')

data

Excellent Good Fair Poor

VeryHappy 271 261 82 20

PrettyHappy 247 567 231 53

NotTooHappy 33 103 92 36

chisq.test(data)

Pearson's Chi-squared test

data: data

X-squared = 182.17, df = 6, p-value < 2.2e-16

p-value < 0.05, reject null hypothesis, at the α=0.05 level of significance, happiness and healthy are related.

1. A package of M&M candies is filled from batches that contain a specified percentage of each of six colors. These percentage are given in mandms dataset in UsingR package. Assume a package of candies contains the following color distribution: 15 blue, 34 dbrown, 7 green, 19 orange, 29 red, and 24 yellow. Perform a chi-squared test with the null hypothesis that the candies are from *milkchocolate* group (category).

library(UsingR)

mandms

x = c(15,34,7,19,29,24)

p = c(0.1,0.3,0.1,0.1,0.2,0.2)

chisq.test(x,p=p)

Chi-squared test for given probabilities

data: x

X-squared = 7.0651, df = 5, p-value = 0.2158

p-value > 0.05, do not have enough evidence to reject the null hypothesis. Do not have enough evidence to say that the candies are from milkchocolate group.

1. Repeat (3) assuming the Peanut Package. Based on the p-value which would you suspect is the true source of candies?

x = c(15,34,7,19,29,24)

p1 = c(0.1,0.3,0.1,0.1,0.2,0.2)

p2 = c(0.2,0.2,0.1,0.1,0.2,0.2)

p3 = c(0.2,0.2,0.2,0.0,0.2,0.2)

p4 = c(1/6,1/6,1/6,1/6,1/6,1/6)

p5 = c(1/6,1/6,1/6,1/6,1/6,1/6)

groups = c('milk chocolate','Peanut','Peanut Butter','Almond','kid minis')

p = data.frame(p1,p2,p3,p4,p5)

a = chisq.test(x,p=p1)

print(a$p.value)

table=cbind(NULL,NULL)

for (i in 1:length(p)){

if(i==3){

pvalue = chisq.test(c(15,34,7,29,24),p=c(0.2,0.2,0.2,0.2,0.2))$p.value

table = rbind(table,cbind(groups[i],pvalue))

next

}

pvalue = chisq.test(x,p=p[,i])$p.value

table = rbind(table,cbind(groups[i],pvalue))

}

table

pvalue

[1,] "milk chocolate" "0.215843285741492"

[2,] "Peanut" "0.0204904392927517"

[3,] "Peanut Butter" "0.000241118724005941"

[4,] "Almond" "0.000444257597923391"

[5,] "kid minis" "0.000444257597923391"

the minimal p-value is 0.000241118724005941 which is from Peanut Butter