**STAT 40001/STAT 50001 Statistical Computing Fall 2024**

**Lab-12**

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

> #Q1

> data <- c(62,84,17,16,21)

> data

[1] 62 84 17 16 21

> chisq.test(data)

Chi-squared test for given probabilities

data: data

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

> cat("Reject the null hypothesis i.e. proportion is not the same")

Reject the null hypothesis i.e. proportion is not the same

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.

> #Q2

> us = 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(us) = c("Gold", "Silver", "Bronze")

> Q2 = rbind(us, china, russia, britain, germany)

> Q2

Gold Silver Bronze

us 46 29 29

china 38 27 22

russia 24 25 33

britain 29 17 19

germany 11 19 14

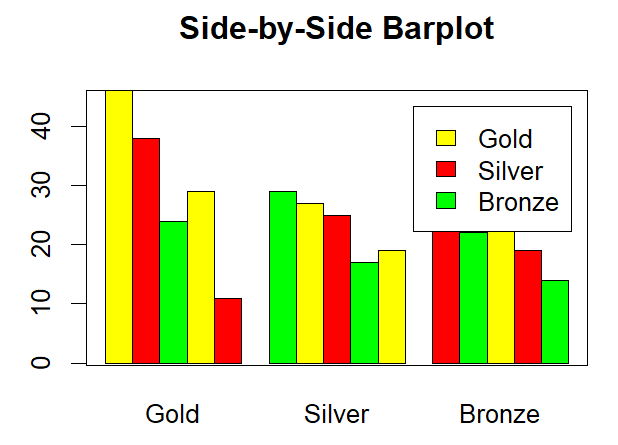
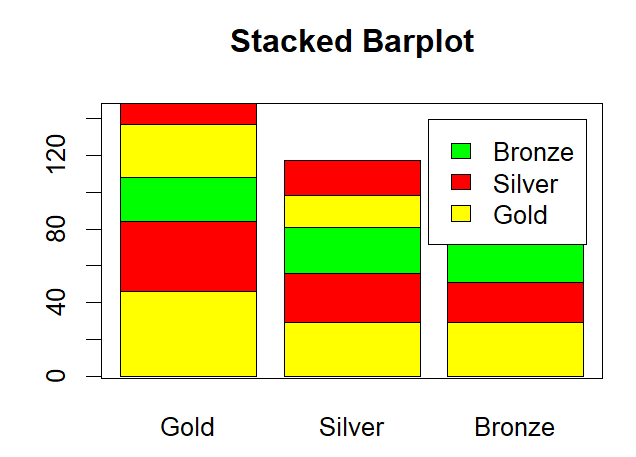
> colnames(Q2) = c("Gold", "Silver", "Bronze")

> barplot(Q2, col = c("yellow", "red", "green"), legend = colnames(Q2), main = "Stacked Barplot", y.lim = c(0, 150), beside = FALSE)

> box()

> barplot(Q2, col = c("yellow", "red", "green"),beside = TRUE, legend = colnames(Q2), main = "Side-by-Side Barplot", y.lim = c(0, 50))

> box()



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.

> #Q3

> very\_happy = c(271,261,82,20)

> pretty\_happy = c(247,567,231,53)

> no\_too\_happy = c(33,103,92,36)

> names(very\_happy) = c("Excellent","Good","Fair","Poor")

> Q3 <- rbind(very\_happy,pretty\_happy,no\_too\_happy)

> Q3

Excellent Good Fair Poor

very\_happy 271 261 82 20

pretty\_happy 247 567 231 53

no\_too\_happy 33 103 92 36

> chisq.test(Q3)

Pearson's Chi-squared test

data: Q3

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

> cat("Reject the null hypothesis i.e we conclude that health and happiness are dependant")

Reject the null hypothesis i.e we conclude that health and happiness are dependant

|  |  |  |
| --- | --- | --- |
|  | Child | |
| Parent | Buckled | Unbuckled |
| Buckled | 56 | 8 |
| Unbuckled | 2 | 16 |

1. An informal survey of seat-belt usage in California examined the relationship between a parent’s uses of a seat-belt with a child’s. The data are provided below

Does the fact that a parent has seatbelt buckled affect the chance that the child’s seat belt will be buckled?

> p\_buckled = c(56,8)

> p\_unbuckled = c(2,16)

> names(p\_buckled) = c("c\_buckled","c\_unbuckled")

> Q4 <- rbind(p\_buckled,p\_unbuckled)

> Q4

c\_buckled c\_unbuckled

p\_buckled 56 8

p\_unbuckled 2 16

> chisq.test(Q4)

Pearson's Chi-squared test with Yates' continuity correction

data: Q4

X-squared = 35.995, df = 1, p-value = 1.978e-09

> cat("Reject the null Hypothesis i.e we conclude that parent buckled dependant on the child buckled")

Reject the null Hypothesis i.e we conclude that parent buckled dependant on the child buckled

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 brown, 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).

> install.packages("UsingR")

> library(UsingR)

> dataset("mandms")

> mandms

blue brown green orange red yellow

milk chocolate 10.0000 30.0000 10.0000 10.0000 20.0000 20.0000

Peanut 20.0000 20.0000 10.0000 10.0000 20.0000 20.0000

Peanut Butter 20.0000 20.0000 20.0000 0.0000 20.0000 20.0000

Almond 16.6667 16.6667 16.6667 16.6667 16.6667 16.6667

kid minis 16.6667 16.6667 16.6667 16.6667 16.6667 16.6667

> data <- c(15,34,7,19,29,24)

> p\_milkchoco = c(1/10,3/10,1/10,1/10,2/10,2/10)

> chisq.test(data,p=p\_milkchoco)

Chi-squared test for given probabilities

data: data

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

> cat("Fail to reject the null hypothesis i.e. we cannot tell it is not from milk chocolate")

Fail to reject the null hypothesis i.e. we cannot tell it is not from milk chocolate

1. Repeat (5) assuming the Peanut Package. Based on the p-value which would you suspect is the true source of candies?

> install.packages("UsingR")

> library(UsingR)

> dataset("mandms")

> mandms

blue brown green orange red yellow

milk chocolate 10.0000 30.0000 10.0000 10.0000 20.0000 20.0000

Peanut 20.0000 20.0000 10.0000 10.0000 20.0000 20.0000

Peanut Butter 20.0000 20.0000 20.0000 0.0000 20.0000 20.0000

Almond 16.6667 16.6667 16.6667 16.6667 16.6667 16.6667

kid minis 16.6667 16.6667 16.6667 16.6667 16.6667 16.6667

> data <- c(15,34,7,19,29,24)

|  |
| --- |
| > p\_peanut = c(2/10,2/10,1/10,1/10,2/10,2/10)  > chisq.test(data,p=p\_peanut)  Chi-squared test for given probabilities  data: data  X-squared = 13.328, df = 5, p-value = 0.02049  > cat("Reject the null hypothesis i.e. we can tell it is not from peanut")  Reject the null hypothesis i.e. we can tell it is not from peanut  > p\_peanutButter = c(2/10,2/10,2/10,0/10,2/10,2/10)  > chisq.test(data,p=p\_peanutButter)  Chi-squared test for given probabilities  data: data  X-squared = Inf, df = 5, p-value < 2.2e-16  > cat("Reject the null hypothesis i.e. we can tell it is not from peanut Butter")  Reject the null hypothesis i.e. we can tell it is not from peanut Butter |
|  |
| |  | | --- | |  | |