STA2023 R lab 9 Chi-Square tests.

Goodness of fit: does the data fit a given distribution?

1. Is the die loaded?

Die Value	Assumed Distribution	Observed Frequency
1	1/6	9
2	1/6	15
3	1/6	9
4	1/6	8
5	1/6	6
6	1/6	13

```
> x <- c(9,15,9,8,6,13)
> p <- c(1/6,1/6,1/6,1/6,1/6,1/6)
> chisq.test(x = x, p=p)

        Chi-squared test for given probabilities

data:  x
X-squared = 5.6, df = 5, p-value = 0.3471
```

```
> # toss a coin 100 times; observe 41 heads, 59 tails. is it a fair coin?
> x<-c(41,59)
> p <- c(1/2,1/2
> chisq.test(x=x, p=p)

        Chi-squared test for given probabilities

data:  x
X-squared = 3.24, df = 1, p-value = 0.07186
```

2. During the presidential election of 2008, the Pew Research Center colle cted survey data that suggested that 24% of registered voters were libe ral, 38% were moderate, and 38% were conservative. Is the distribution of political views different this year? This year: 300 voters: 80 libe rals, 105 moderates, 115 conservatives.

```
> y <- c(80,105,115)
> prob <- c(0.24, .38, .38)
> chisq.test(x=y, p=prob)

        Chi-squared test for given probabilities

data:  y
X-squared = 1.6082, df = 2, p-value = 0.4475
```

Test of independence.

```
> #Ho: the two categorical variables are independent; H1: the two categorical
variables are not independent of each other
> #Example: Is the voter preference for a political party in the US independe
nt of gender?
```

Gender \ Party Preference	Republican	Democrat
Male	200	150
Female	250	300

```
> # chisq.test (matrix of values)
> # Create a matrix, using a vector
> x <- c(200,250,150,300)
> data<-matrix(x, 2,2) # vector x, values by default arrange by columns. 2,2
is the num of rows, num of columns.
> data
     [,1] [,2]
[1,]  200  150
[2,]  250  300

> # in order to arrange by rows, it must be stated:
> y <- c(200,150,250,300) # enter the data by rows
> data1 <- matrix(y, byrow = TRUE, 2,2) # specify byrow=T
> data1
     [,1] [,2]
[1,]  200  150
[2,]  250  300

> # label rows and columns as follow we don't need to name rows and columns i
n order to run chisq.test, but you may wish to do it):

> rownames(data1) <- c("Male", "Fem")
> colnames(data1) <- c("Rep", "Dem")

> data1
     Rep Dem
Male 200 150
Fem  250 300

> chisq.test(data1)

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

data:  data1
X-squared = 11.225, df = 1, p-value = 0.0008068

> # p-value << than alpha (any plausible alpha), therefore we rejet the Null
Hyp of independence.
```