

Experiment – 10

Applying Chi-square test for goodness of fit test and Contingency test to real dataset

Aim: To understand Chi-square test for goodness of fit and independent of attributes using R

Introduction

A Pearson's chi-square test is a statistical test for categorical data. It is used to determine whether your data are significantly different from what you expected. There are two types of Pearson's chi-square tests:

- The **chi-square goodness of fit test** is used to test whether the frequency distribution of a categorical variable is different from your expectations.
- The **chi-square test of independence** is used to test whether two categorical variables are related to each other.

Both of Pearson's chi-square tests use the same formula to calculate the test statistic, chi-square (X^2):

$$X^2 = \sum \frac{(O - E)^2}{E}$$

Where:

- X^2 is the chi-square test statistic
- O is the observed frequency
- E is the expected frequency

Procedure:

- Import the data set
- Determine the critical value and sample statistic using R functions
- Conclude the problem using R functions

Problem: 1

Five coins are tossed 256 times. The number of heads observed by binomial distribution is given below. Examine if the coins are unbiased by employing chi-square goodness of fit.

No. of heads	0	1	2	3	4	5
Frequency	5	35	75	84	45	12

Codes and Results:

```
# Problem : 1
# Goodness of fit

# Number of coins
n=5
n

## [1] 5

# level of significance
alpha=0.05
alpha

## [1] 0.05

N=256 # Total number of tosses
N

## [1] 256

P = 0.5 # probability of getting head
P

## [1] 0.5

x = c(0:n);x

## [1] 0 1 2 3 4 5

obf = c(5,35,75,84,45,12)# observed frequencies
obf

## [1] 5 35 75 84 45 12

exf = (dbinom(x,n,P)*256) # expected frequencies
exf

## [1] 8 40 80 80 40 8

# check the condition if the observed and expected frequencies sum are equal
sum(obf)

## [1] 256

sum(exf)

## [1] 256

# output using Chisq-distribution
chisq<-sum((obf-exf)^2/exf)
cv = chisq;cv
```

```
## [1] 4.8875

# critical value using Chisq-distribution
tv = qchisq(1-alpha,n);tv

## [1] 11.0705

# Hypothesis conclusion
if(cv <= tv){print("Accept H0/Fit is good")} else{print("Reject H0/Fit is not
good")}}

## [1] "Accept H0/Fit is good"
```

Problem: 2

From the following information state whether the condition of the child is associated with the condition of the house.

Condition of child	Condition of house Clean	Condition of house dirty
Clean	69	51
Fairly clean	81	20
dirty	35	44

Codes and Results:

```
# Problem : 2

# Independent of attributes
# Input the data
data<-matrix(c(69,51,81,20,35,44),ncol=2,byrow=T)
data

##      [,1] [,2]
## [1,]   69   51
## [2,]   81   20
## [3,]   35   44

# number of data
l=length(data);l

## [1] 6

# output by Chisq-distribution
cv=chisq.test(data)
cv

##
## Pearson's Chi-squared test
```

```
##  
## data:  data  
## X-squared = 25.629, df = 2, p-value = 2.721e-06  
  
# p-value  
cv=cv$p.value  
cv  
  
## [1] 2.72114e-06  
  
# Hypothesis conclusion  
if(cv >alpha){print("Attributes are independent")} else{print("Attributes are  
not independent")}  
  
## [1] "Attributes are not independent"
```

Conclusion: Chi-squared test has been explored through R functions.