

Chi-Square Test

DATA621 Blog 05

Zhi Ying Chen

27 November 2020

Chi-Square test in R is a statistical method which used to determine if two categorical variables have a significant correlation between them. The two variables are selected from the same population. Furthermore, these variables are then categorised as Male/Female, Red/Green, Yes/No etc.

Particularly in this test, I have to check the p-values. Moreover, like all statistical tests, I assume this test as a null hypothesis and an alternate hypothesis.

The main thing is, I will reject the null hypothesis if the p-value that comes out in the result is less than a predetermined significance level, which is 0.05 usually, then I reject the null hypothesis.

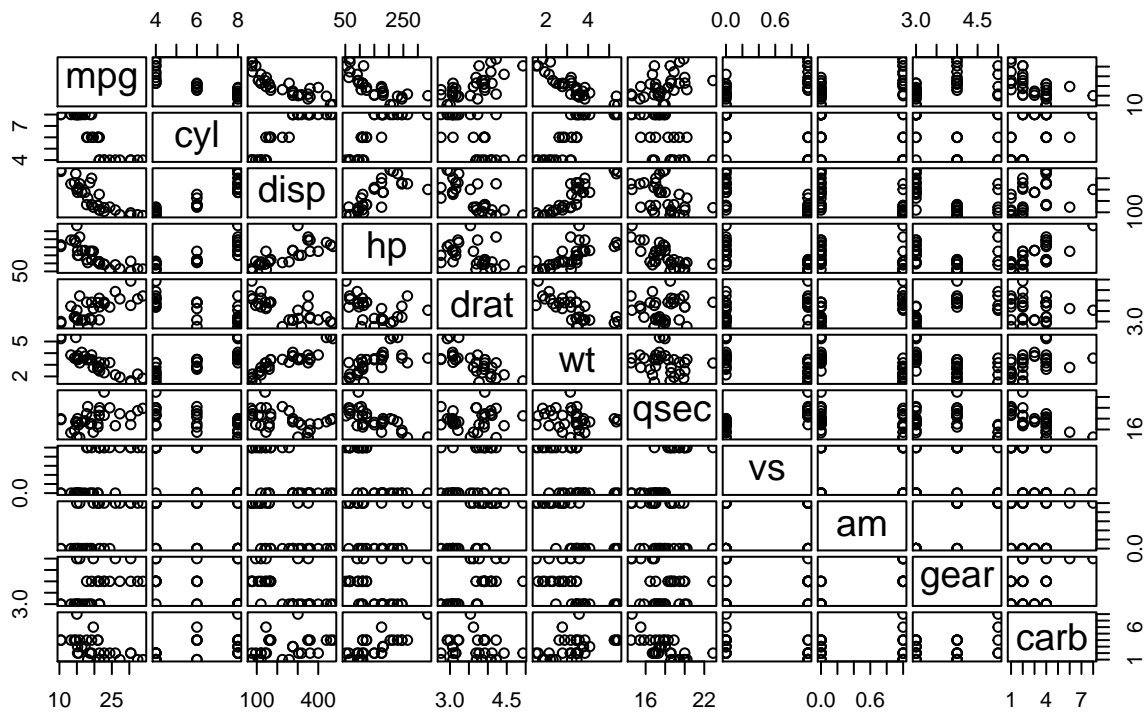
H0: The two variables are independent. H1: The two variables relate to each other.

In the case of a null hypothesis, a chi-square test is to test the two variables that are independent.

```
data("mtcars")
```

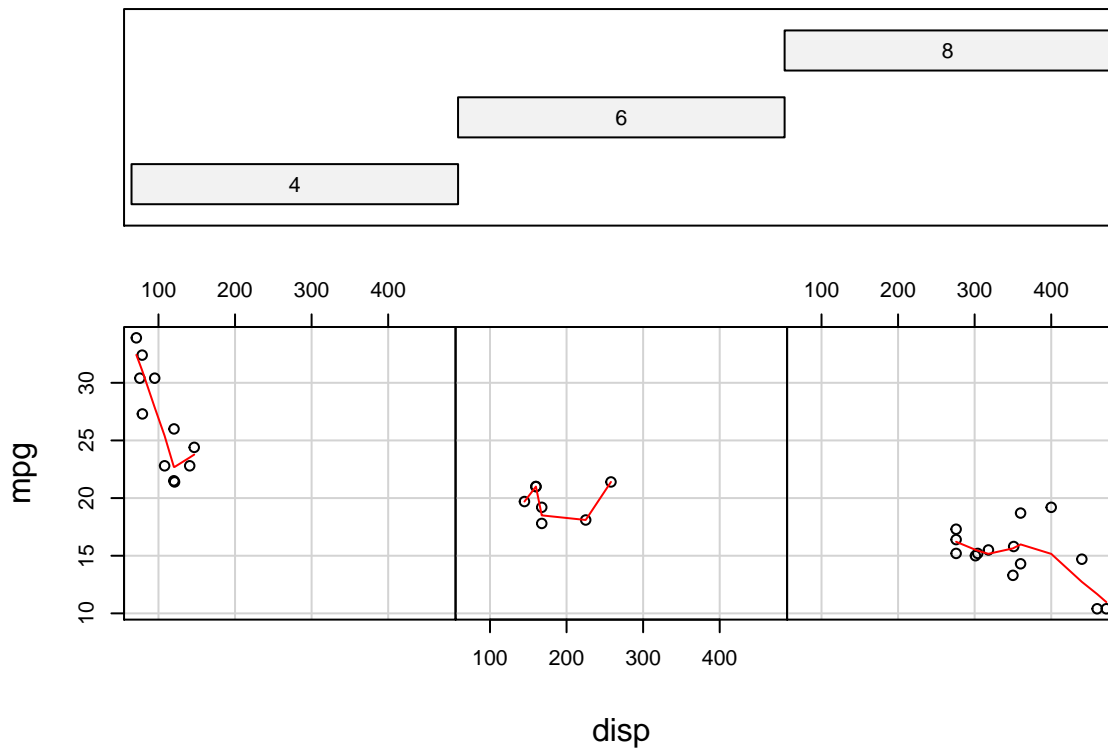
```
pairs(mtcars, main = "mtcars data", gap = 1/4)
```

mtcars data



```
coplot(mpg ~ disp | as.factor(cyl), data = mtcars,
       panel = panel.smooth, rows = 1)
```

Given : as.factor(cyl)



```
table(mtcars$carb, mtcars$cyl)
```

```
##
##      4 6 8
##  1 5 2 0
##  2 6 0 4
##  3 0 0 3
##  4 0 4 6
##  6 0 1 0
##  8 0 0 1
```

```
chisq.test(mtcars$carb, mtcars$cyl)
```

```
## Warning in chisq.test(mtcars$carb, mtcars$cyl): Chi-squared approximation may be
## incorrect
```

```
##
## Pearson's Chi-squared test
##
## data:  mtcars$carb and mtcars$cyl
## X-squared = 24.389, df = 10, p-value = 0.006632
```

We have a high chi-squared value and a p-value as 0.00632 which is less than 0.05 significance level. So I reject the null hypothesis and conclude that carb and cyl have a significant relationship.