Pearson's Chi-squared Test for Count Data

chisq.test(x, y = NULL, correct = TRUE, p = rep(1/length(x), length(x)), rescale.p = FALSE, simulate.p.value = FALSE, B = 2000)

Arguments:

x: a vector or matrix.

y: a vector; ignored if 'x' is a matrix.

correct: a logical indicating whether to apply continuity correction
when computing the test statistic for 2x2 tables: one half is
subtracted from all |O-E| differences. No correction is done
if 'simulate.p.value = TRUE'.

p: a vector of probabilities of the same length of 'x'. An error is given
if any entry of 'p' is negative.

clarification

- **If 'x' is a matrix with one row or column, or if 'x' is a vector**

and 'y' is not given, then a "goodness-of-fit test" is performed ("'x' is treated as a one-dimensional contingency table"). In this case, the hypothesis tested is whether the population probabilities equal those in 'p', or are all equal if 'p' is not given.

- **If 'x' is a matrix with at least two rows and columns, it is taken**

as a two-dimensional contingency table.

OUTPUT:

statistic: the value the chi-squared test statistic.
parameter:the df of the approximate chi-squared distribution
p.value: the p-value for the test.
method: a character string indicating the type of test performed,
data.name: a character string giving the name(s) of the data.
observed: the observed counts.
expected: the expected counts under the null hypothesis.
residuals: the Pearson residuals, '(observed - expected) /sqrt(expected)'.

Example:

Testing for population probabilities [[**whether the population probabilities equal those in 'p', or are all equal if 'p' is not given.]]**

x <- c(A = 20, B = 15, C = 25) #**H0:P(A)=P(B)=P(C)**
chisq.test(x)   # **same as chisq.test(x,p=c(1/3,1/3,1/3))**

	ذكور	إناث
يشربون الشاي	40	33
لا يشربون الشاي	3	12

H_0: لا توجد علاقة بين شرب الشاي والنـــوع

H_1: توجد علاقة بين شرب الشاي والـنــوع

```
x<-matrix(c(40,3,33,12),2);x
###
chisq.test(x,correct=T)$p.value
> chisq.test(x, correct = T)$p.value
[1] 0.02986929
###
chisq.test(x,correct=F)$p.value
> chisq.test(x, correct = F)$p.value
[1] 0.01407356
###
fisher.test(x)$p.value
> fisher.test(x)$p.value
[1] 0.02161659
```

القرار:

نرفض فرض العدم

الاستنتاج:

توجد علاقة بين شرب الشاي والنوع وذلك عند $\alpha=.05$

H_0: لا توجد علاقة بين شرب الشاي والنـــوع

H_1: توجد علاقة بين شرب الشاي والـنــوع

```
x<-matrix(c(40,3,33,12),2);x
###
chisq.test(x,correct=T)$p.value
> chisq.test(x, correct = T)$p.v
[1] 0.02986929
###
chisq.test(x,correct=F)$p.value
> chisq.test(x, correct = F)$p.v
[1] 0.01407356
###
fisher.test(x)$p.value
> fisher.test(x)$p.value
[1] 0.02161659
```

القرار:

نرفض فرض العدم

الاستنتاج:

توجد علاقة بين شرب الشاي والنوع وذلك عند $\alpha=.05$

٨٥

Using Prop.test

```
x=c(40,33)
n=c(43,45)
prop.test(x,n)

        2-sample test for equality of proportions with continuity correction

data:  x out of n
X-squared = 4.7168, df = 1, p-value = 0.02987
alternative hypothesis: two.sided
95 percent confidence interval:
 0.02418799 0.36961046
sample estimates:
   prop 1    prop 2
0.9302326 0.7333333
```

chisquare goodness of fit test

```
x=c(51,11,5,53)
p=c(.4,.1,.05,.45)
chisq.test(x,p=p)

        Chi-squared test for given probabilities

data:  x
X-squared = 0.456, df = 3, p-value = 0.9284
```

Example 4 page(249) 101 book [weekly accidents for 100 weeks]

H0: weekly accidents have poisson distribution

```
counts=c(50,30,12,5,2,1)
 l=0:5;lambda=l*counts
 lmb=sum(lambda)/sum(counts)
 lmb
[1] 0.82
 step_function <- dpois(l, lambda=lmb)
 step_function
[1] 0.440431655 0.361153957 0.148073122 0.040473320
0.008297031 0.001360713
expected= step_function *sum(counts)
[1] 44.0431655 36.1153957 14.8073122  4.0473320  0.8297031
0.1360713
 chisq.test(counts,step_function , rescale.p = TRUE)
 Chi-squared test for given probabilities

data:  counts
X-squared = 9.7315, df = 5, p-value =
0.08321

Warning message:
Chi-squared approximation may be incorrect
in:chisq.test(counts, p = step_function,
rescale.p = TRUE)
```