BAB 5
UJI CHI-SQUARE

Introduction

The chi-square distribution is the result of the square of the normal distribution. The chi-square distribution has the characteristic that the value is always positive. This distribution can be used to test statistics, which is called the chi-square test. The chi-square test is divided into 3, namely the goodness of fit test, independence test and homogeneity test. In the discussion this time will only discuss the goodness of fit and independence test.

Goodness of Fit Test

The goodness of fit test is often called the suitability test, which is a statistical test to determine the suitability between the expected frequency and the observed frequency. In other words, in this test, we want to know whether there is a significant difference between the observed frequency and the expected frequency. The statistical test formula is:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

Information:

χ^2 : chi-square value with degrees of freedom $v = k - 1$

k : number of classes / cells in testing, with $i = 1,2,\dots,k$

O_i : Observed frequency

E_i : Expected Frequency

In general, the stages in the chi-square test are:

1. Creating a contingency table
2. Calculating the expected frequency
3. Determining the H_0 and H_1
4. Determining the significant level
5. Determining the critical region

6. Calculating the statistical test (chi-square)
7. Deducing a hypothesis

Example:

A coin is tossed 100 times. The result of the throwing was that 58 appeared on the front side and 42 appeared on the back side. Test whether the coin is symmetrical or not! (use α = 5%)

$n = 100$

What is expected is a symmetrical front and back, so that:

$$P(front\ side) = \frac{1}{2}$$

$$P(back\ side) = \frac{1}{2}$$

From that probability, so the expected frequency are:

$$e_1 = 100 \times \frac{1}{2} = 50$$

$$e_2 = 100 \times \frac{1}{2} = 50$$

1. Creating a contingency table:

appears the side of the coin	Front side	Back side
Observed frequency (o_i)	58	42
Expected frequency (e_i)	50	50

2. Determining a hypothesis:

   $$H_0: P(front\ side) = P(back\ side) = \frac{1}{2}$$

   $$H_1: P(front\ side) \neq P(back\ side)$$

3. Significant level:

   $\alpha = 5\%$

4. Critical region:

   Rejecting H_0, if $\chi^2 > \chi^2{}_{\alpha/2}$ or $\chi^2 < \chi^2{}_{1-\alpha/2}$

5. Calculating statistical test:

   $$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} = \frac{(58-50)^2}{50} + \frac{(42-50)^2}{50} = 1.28 + 1.28 = 2.56$$

$v = k - 1 = 2 - 1 = 1$

${\chi^2}_{0.025,1} = 5.024$

${\chi^2}_{0.975,1} = 0.001$

6. Conclusion

$\chi^2 > {\chi^2}_{0.975,1}$ and $\chi^2 < {\chi^2}_{0.025,1}$

So that the chi-square value is between${\chi^2}_{0.975,1}$ and ${\chi^2}_{0.025,1}$

The decision is failing to reject H_0 so that a coin is symmetrical.

From these examples it can be seen that basically the goodness of fit test is a test that is expected to show the conformity between the observations made and the desired expectations, therefore the goodness of fit is said to be a test of suitability.

Independence Test

The independence test is used to determine whether there is a relationship (association) or not between two variables / two factors. This test is often said to be a test of freedom between two factors. If the test results show that H_0 fails, it can be said that the two variables / two factors tested are independent of each other. This independence test is used for categorical variables and is arranged in the form of a contingency table. The general form of the contingency table in this test is:

		Variable_2				n_{io}
		1	2	$\cdots$	k	
Variable_1	1	O_{11}	O_{12}		O_{1k}	n_{1o}
	2	O_{21}	O_{22}		O_{2k}	n_{2o}
	$\vdots$					
	r	O_{r1}	O_{r2}		O_{rk}	n_{ro}
n_{ok}		n_{o1}	n_{o2}		n_{ok}	n

The table above contains the observed frequency (O_{ij}) and the expected frequency (E_{ij}). The expected frequency is found by the formula:

$$E_{ij} = \frac{(n_{io})(n_{oj})}{n}$$

And the statistical test is:

$$\chi^2 = \sum_{i,j=1}^{k,r} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

χ^2 : chi-square value with degrees of freedom $v = (k-1)(r-1)$

k : the number of columns

r : the number of rows

O_{ij} : the observed frequency in row i and column j

E_{ij} : the expected frequency in row i and column j

example:

The research was conducted to obtain information about the relationship between learning achievement and the timeliness of completing the UAS for FRI students. Test whether the UAS completion time has anything to do with student achievement results! (use α = 5%)

Achievement	Time		Total
	On Time	Time is over	
Satisfactory	55	20	75
Not satisfactory	10	15	25
Total	65	35	100

Determining the expected frequency

$$E_{11} = \frac{(75)(65)}{100} = 48.75$$

$$E_{12} = \frac{(75)(35)}{100} = 26.25$$

$$E_{21} = \frac{(25)(65)}{100} = 16.25$$

$$E_{22} = \frac{(25)(35)}{100} = 8.75$$

Achievement	Time		Total
	On time	Time is over	
Satisfactory	$O_{11} = 55$ $E_{11} = 48.75$ $(O_{11} - E_{11}) = 6.25$	$O_{12} = 20$ $E_{12} = 26.25$ $(O_{12} - E_{12}) = -6.25$	75
Not satisfactory	$O_{21} = 10$ $E_{21} = 16.25$ $(O_{21} - E_{21}) = -6.25$	$O_{22} = 15$ $E_{22} = 8.75$ $(O_{22} - E_{22}) = 6.25$	25
Total	65	35	100

The hypothesis:

$H_0: there\ is\ no\ relationship\ between\ the\ timeliness\ of\ UAS\ and\ learning\ achievement$

$H_1: there\ is\ a\ relationship\ between\ the\ punctuality\ of\ UAS\ with\ learning\ achievement$

$\alpha = 5\%$

Critical region:

Rejecting H_0 if $\chi^2 > \chi^2{}_{\alpha/2}$ or $\chi^2 < \chi^2{}_{1-\alpha/2}$

$$\chi^2 = \sum_{i,j=1}^{k,r} \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}} = \frac{(6.25)^2}{48.75} + \frac{(-6.25)^2}{26.25} + \frac{(-6.25)^2}{16.25} + \frac{(6.25)^2}{8.75}$$

$\chi^2 = 0.801 + 1.488 + 2.404 + 4.464 = 9.157$

$v = (k-1)(r-1) = (2-1)(2-1) = 1$

$\chi^2{}_{0.025,1} = 5.024$

$\chi^2{}_{0.975,1} = 0.001$

$\chi^2 > \chi^2{}_{0.025,1}$

conclusion:

$\chi^2 > \chi^2{}_{0.025,1}$, the decision is rejecting H_0 So it can be said that there is a relationship between the timeliness of completing the UAS with student achievement.