Chi-square distribution

The sum of k independent squared standard normal variables is a Chi-square random variable with k degrees of freedom. That is, $\chi^2 = Z_1^2 + Z_2^2 + \ldots + Z_k^2$

Chi-square distribution

The sum of k independent squared standard normal variables is a Chi-square random variable with k degrees of freedom. That is, $\chi^2 = Z_1^2 + Z_2^2 + \ldots + Z_k^2$

- The curve is non symmetrical and skewed to the right
- The curve differs for each degrees of freedom

Chi-square distribution

The sum of k independent squared standard normal variables is a Chi-square random variable with k degrees of freedom. That is, $\chi^2 = Z_1^2 + Z_2^2 + \ldots + Z_k^2$

- The curve is non symmetrical and skewed to the right
- The curve differs for each degrees of freedom

Applications

- Goodness of fit
- Test for independence

Test for independence

Procedure

1. Formulate the null and alternate hypothesis:
   H_0: Two variables are independent
   H_1: Two variables are not independent

Test for independence

Procedure

1. Formulate the null and alternate hypothesis:
   H_0: Two variables are independent
   H_1: Two variables are not independent
2. Calculate the Expected frequencies
   $E =$ (row total)(column total) / sample size
   Note that each expected value must be greater than or equal to 5 for the chi square test to be valid

Test for independence

Procedure

1. Formulate the null and alternate hypothesis:
   H_0: Two variables are independent
   H_1: Two variables are not independent
2. Calculate the Expected frequencies
   $E =$ (row total)(column total) / sample size
   Note that each expected value must be greater than or equal to 5 for the chi square test to be valid
3. Calculate the test statistic:

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

Test for independence

4. Find the degrees of freedom

$$df = (r - 1)(c - 1)$$

where r is the number of rows and c is the number of columns

Test for independence

4 Find the degrees of freedom

$$df = (r - 1)(c - 1)$$

where r is the number of rows and c is the number of columns

5 Calculate the critical value (cv) at the given LOS

Test for independence

4 Find the degrees of freedom

$$df = (r-1)(c-1)$$

where r is the number of rows and c is the number of columns

5 Calculate the critical value (cv) at the given LOS

6 Conclusion: If $\chi^2 < cv$, then accept H_0 (Variables are independent)
else reject H_0 (Variables are not independent)

Problems

1. The side effects of a new drug are being tested against a placebo. A simple random sample of 565 patients yields the results below. At a significance level $\alpha = 0.10$, is there enough evidence to conclude that the treatment is independent of the side effect of nausea?

Result	Drug	Placebo	Total
Nausea	36	13	49
No Nausea	254	262	516
Total	290	275	565

Problems

1. The side effects of a new drug are being tested against a placebo. A simple random sample of 565 patients yields the results below. At a significance level $\alpha = 0.10$, is there enough evidence to conclude that the treatment is independent of the side effect of nausea?

Result	Drug	Placebo	Total
Nausea	36	13	49
No Nausea	254	262	516
Total	290	275	565

Solution

H_0: The treatment and the response are independent.

H_1: The treatment and the response are dependent.

Problems

1. The side effects of a new drug are being tested against a placebo. A simple random sample of 565 patients yields the results below. At a significance level $\alpha = 0.10$, is there enough evidence to conclude that the treatment is independent of the side effect of nausea?

Result	Drug	Placebo	Total
Nausea	36	13	49
No Nausea	254	262	516
Total	290	275	565

Solution

H_0: The treatment and the response are independent.
H_1: The treatment and the response are dependent.
$\alpha = 0.10$

Problems

solution contd.

Expected frequencies

Result	Drug	Placebo	Total
Nausea	25.15	23.85	49
No Nausea	264.85	251.15	516
Total	290	275	565

Problems

solution contd.

Expected frequencies

Result	Drug	Placebo	Total
Nausea	25.15	23.85	49
No Nausea	264.85	251.15	516
Total	290	275	565

degrees of freedom $df = (2-1)(2-1) = 1$

Problems

solution contd.

Expected frequencies

Result	Drug	Placebo	Total
Nausea	25.15	23.85	49
No Nausea	264.85	251.15	516
Total	290	275	565

degrees of freedom $df = (2-1)(2-1) = 1$
Test statistic: $\chi^2 = 10.53$

Problems

solution contd.

Expected frequencies

Result	Drug	Placebo	Total
Nausea	25.15	23.85	49
No Nausea	264.85	251.15	516
Total	290	275	565

degrees of freedom $df = (2-1)(2-1) = 1$
Test statistic: $\chi^2 = 10.53$
Critical value is 2.71 ($df = 1$, $\alpha = 0.10$)

Problems

solution contd.

Expected frequencies

Result	Drug	Placebo	Total
Nausea	25.15	23.85	49
No Nausea	264.85	251.15	516
Total	290	275	565

degrees of freedom $df = (2-1)(2-1) = 1$
Test statistic: $\chi^2 = 10.53$
Critical value is 2.71 ($df = 1$, $\alpha = 0.10$)
Since $\chi^2 > 2.71$, there is enough evidence to reject H_0. Hence, there is a relation between the treatment and response.

Problems

2. Suppose the undergraduate degrees are BA, BE, BBA, and several others. There are three possible majors for the MBA students which are accounting, finance, and marketing. Can the statistician conclude that the undergraduate degree affects the choice of major from the given table?

UG/ MBA	Accounting	Finance	Marketing	Total
BA	31	13	16	60
BE	8	16	7	31
BBA	12	10	17	39
Other	10	5	7	22
Total	61	44	47	152

Solution

1. H_0: The undergraduate degree and MBA major are independent
   H_1: The undergraduate degree and MBA major are dependent
2. Expected frequencies:

UG/MBA	Accounting	Finance	Marketing	Total
BA	24.08	17.37	18.55	60
BE	12.44	8.97	9.59	31
BBA	15.65	11.29	12.06	39
Other	8.83	6.37	6.8	22
Total	61	44	47	152

Solution contd.

3 Test statistic:

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

$$\chi^2 = \frac{(31 - 24.08)^2}{24.08} + \frac{(13 - 17.37)^2}{17.37} + \ldots + \frac{(7 - 6.8)^2}{6.8} = 14.7$$

Solution contd.

3 Test statistic:

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

$$\chi^2 = \frac{(31-24.08)^2}{24.08} + \frac{(13-17.37)^2}{17.37} + \ldots + \frac{(7-6.8)^2}{6.8} = 14.7$$

4 degrees of freedom $df = (r-1)(c-1) = (4-1)(3-1) = 6$

Solution contd.

3 Test statistic:

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

$$\chi^2 = \frac{(31 - 24.08)^2}{24.08} + \frac{(13 - 17.37)^2}{17.37} + \ldots + \frac{(7 - 6.8)^2}{6.8} = 14.7$$

4 degrees of freedom $df = (r - 1)(c - 1) = (4 - 1)(3 - 1) = 6$

5 $\alpha = 0.05$, Critical value $\chi^2_{(0.05,6)} = 12.59$

Solution contd.

3. Test statistic:

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

$$\chi^2 = \frac{(31-24.08)^2}{24.08} + \frac{(13-17.37)^2}{17.37} + \ldots + \frac{(7-6.8)^2}{6.8} = 14.7$$

4. degrees of freedom $df = (r-1)(c-1) = (4-1)(3-1) = 6$
5. $\alpha = 0.05$, Critical value $\chi^2_{(0.05,6)} = 12.59$
   The critical region is $\chi^2 > 12.59$.
   Since calculated $\chi^2 > 12.59$, we reject H_0.
6. Conclusion: There is sufficient evidence to the claim that the undergraduate degree and the MBA major are related.

Problems

3. The operations manager of a company that manufactures tyres wants to determine whether there are any differences in the quality of workmanship among the three daily shifts. She randomly selects 496 tyres and carefully inspects them. Each tyre is either classified as perfect, satisfactory, or defective, and the shift that produced it is also recorded. The two categorical variables of interest are : shift and condition of the tyre produced. Do these data provide sufficient evidence at 5% significance level to infer that there are differences in quality among the three shifts?

	Perfect	Satisfactory	Defective	Total
Shift 1	106	114	11	231
Shift 2	67	70	16	153
Shift 3	37	65	10	112
Total	210	249	37	496

Problems

4. Various countries are compared using two variables- composition of economy and growth band as shown in the table.

	High growth	Medium growth	Low growth
Predominant agriculture	20	25	5
Predominant manufacturing	40	5	6
Predominant services	5	55	20

Test whether the predominant function in an economy has an impact on the growth of the economy.