CHAPTER 11. GOODNESS OF FIT AND CONTINGENCY TABLES

The chi-square distribution was discussed in Chapter 4. We now turn to some applications of this distribution. As previously discussed, chi-square is a continuous distribution, however, its application is not limited to continuous data. In fact it is the most important distribution used for the evaluation of discrete or categorical data, for example, the classification of experimental units as dead or alive, sick or healthy, white, green or blue, scores of 1 to 5, etc.

11.1 Goodness of fit for discrete distributions

Goodness of fit involves a comparison of the frequency observed in the sample with the expected frequency based on some theoretical model. If the differences between the observed and the expected frequencies are so great that they are unlikely to be due to chance alone, we conclude that the sample is not taken from the population that was used to calculate the expected frequencies. Suppose we have k observed frequencies, $O_1, O_2, ..., O_k$ and their corresponding expected frequencies, $E_1, E_2, ..., E_k$, then the expression

$$\chi^2 = \sum_{i=1}^{k} (0_i - E_i)^2 / E_i$$

is approximately χ^2 distributed with k-1 degrees of freedom. The closer the agreement between the observed and expected values, the smaller will be the value of χ^2, and a value of zero indicates perfect agreement. Calculated values of χ^2 exceeding those in Appendix Table A-5 indicate that, at the corresponding probability level, there is significant disagreement between observed and expected values.

Chi-square, defined in terms of observed and expected frequencies, becomes a discrete variable and can only assume certain, but not all, non-negative values. As the number of classes (k) increases, χ^2 as defined above approaches a continuous variable. When frequencies for only two classes are involved, a correction must be made for the non-continuity, and the observed χ^2 is adjusted as,

$$\text{adj.}\chi^2 = \sum_{i=1}^{2} \frac{(|0_i - E_i| - 1/2)^2}{E_i} \qquad \text{with df} = 1$$

In adjusted χ^2 the absolute value of each of the differences between observed and expected frequencies is reduced by 1/2 before being squared. The correction 1/2 is applicable only in the case of one degree of freedom (k=2). For more degrees of freedom the corrections are more complicated and are not generally used, but for 1 df the adjusted value of χ^2 is always appropriate.

An example in the use of adjusted χ^2

It is hypothesized that blood groups are inherited in a simple Mendelian manner so that a cross of parents both of whom have AB blood should give 3/4 type AB or AA children and 1/4 type BB children (theoretical model). Suppose that among 400 children from such parents, 292 are of type AB or AA. Does the observation conform to the theoretical hypothesis?

Blood Type:	AB or AA	BB
observed frequency	292	108
expected frequency	300	100

$$\text{adj. } \chi^2 = \frac{(|292-300|-.5)^2}{300} + \frac{(|108-100|-.5)^2}{100}$$
$$= 0.1875 + 0.5625 = 0.75$$

Referring to Appendix Table A-5 with 1 df, we find that the observed χ^2 is not significant ($p < 0.10$).

Thus the observed frequencies of the sample support the hypothetical ratio.

χ^2 with 3 or more classes.

In a cross between ivory and red snapdragons the following counts were observed in the F_2 generation.

Phenotype	Number of plants
Red	20
Pink	55
Ivory	25
	100

On the basis of these data, can segregation be assumed to occur in the simple Mendelian ratio of 1:2:1?

Color	Red	Pink	Ivory
Observed frequency	20	55	25
Expected frequency	25	50	25

$$\chi^2 = \frac{(20-25)^2}{25} + \frac{(55-50)^2}{50} + \frac{(25-25)^2}{25}$$
$$= 1.0 + 0.5 + 0.0 = 1.5$$

Note since this χ^2 has 2 df, the adjustment for noncontinuity is not necessary. Again, the calculated χ^2 is much smaller than the tabular χ^2 at the 10% level. Therefore we conclude that the observed color frequencies could conform with Mendelian ratio.

11.2 Goodness of fit for continuous distributions

It is often of interest to know whether a given set of data approximates a continuous distribution such as the normal or chi-square.

To illustrate the procedure, we will use the % sucrose data presented in Table 2-1, summarized in Table 2-2, and graphed in Figure 2-1. We will test to see if these data can be considered to be normally distributed. To compute the expected frequencies for each class interval, we need to determine the probability associated with each interval. This procedure is summarized in Table 11-1 and the computational steps follow the table.

Table 11-1. Observed and expected frequencies to test the goodness of fit of percent sucrose values to a normal distribution.

Class	Mid-point	Y_i End-point	O_i Obs. freq.	Est.Z $\frac{Y-\bar{Y}}{S}$	Cum. prob.	Int. prob.	E_i Exp. freq.	Contri-bution to χ_2
1	4.8	5.6	1	-2.54	0.0055	0.0055	0.6	0.27
2	6 3	7.1	4	-1.96	0.0250	0.0195	2.0	2.00
3	7.8	8.6	4	-1.38	0.0838	0.0588	5.9	0.61
4	9.3	10.1	13	-0.81	0.2090	0.1252	12.5	0.02
5	10.8	11.6	10	-0.23	0.4090	0.2000	20.0	5.00
6	12.3	13.1	24	0.35	0.6368	0.2278	22.8	0.06
7	13.8	14.6	23	0.92	0.8186	0.1818	18.2	1.27
8	15.3	16.1	17	1.50	0.9332	0.1146	11.5	2.63
9	16.8	17.6	4	2.08	0.9812	0.0480	4.8	0.13

$\bar{Y} = 12.2$, $S = 2.6$, $\chi^2 = 11.99$ with 9 - 3 = 6 df

1. Columns 1 through 4 are recorded from the frequency table, Table 2-2.
2. Standardize the class interval end points, $(Y - \bar{Y})/S$, where $\bar{Y} = 12.2$ and $S = 2.6$ as previously calculated.

3. Determine the cumulative probability for each standardized value from Appendix Table A-4. For example,

   $p(Z < -2.54) = p(Z > 2.54) = 0.0055$
   $p(Z < 0.35) = 1$ p $p(Z > 0.35) = 1 - 0.3632 = 0.6368$
   $p(Z < 2.08) = 1$ p $p(Z > 2.08) = 1 - 0.0188 = 0.9812$

4. Calculate the probability for each class interval, e.g., for class 2, the interval probability = 0.0250 - 0.0055 = 0.0195.

5. The expected frequency for each class is calculated by multiplying the interval probability by the sample size, e.g., for class 2, the expected frequency = (0.0195) • (100) = 1.95 ~ 2.00.

6. The contribution of each class to the overall 2 is equal to

$$\frac{(\text{Observed frequency - expected frequency})^2}{\text{expected frequency}} = \frac{(O - E)^2}{E}$$

7. The calculated χ^2 is the sum of each class contribution

$$\chi^2 = 0.27 + 2.00 + ... + 0.13 = 11.99$$

8. The degrees of freedom for χ^2 depends on the number of parameters that must be estimated for computing the expected frequencies. In this case, we have 9 classes, therefore there are 8 df for classes. This is further reduced by a degree of freedom for mean and a degree of freedom for standard deviation. Thus, there are 6 df for the calculated χ^2 .

For Appendix Table A-5, $\chi^2_{0.05,6}$ = 12.592. Although the calculated χ^2 is not significant at the 5% level, it is nearly so. Therefore we cannot be too sure that the data of Figure 2-1 are normally distributed. However, we can conclude that the data are near enough to being normally distributed to have no effect on the AOV procedures we are using in the evaluation of this variable.

11.3 Contingency Tables

A closely related application of a χ^2 distribution is the test of independence, also known as Pearson's test for association. This test is very similar to the test of goodness of fit and some people prefer to treat them as the same test with minor variation. In this section, we are concerned with the hypothesis of statistical independence between two variables, each of which is classified into a number of categories or attributes.

Suppose a group of persons or set of objects is classified according to two criteria of classification, one criterion being entered in rows and the other in column. This two-way table is called a contingency table. If there are j rows and k columns, the table is known as a j x k table. From such a table we are interested in determining whether a relationship exists between the two criteria of classification or if they are independent. For a j x k table there are (j-1)(k-1) degrees of freedom.

A 2x2 contingency table

The following data show the effect of a certain type of fumigation on fruit spoilage.

	Spoiled	Unspoiled	Totals
Unfumigated	8	16	24
Fumigated	2	14	16
Totals	10	30	40

Does the amount of fruit depend upon whether it has been fumigated?

In any contingency table, we set up the hypothesis that the two criteria of classification are independent. The marginal totals are accepted as part of the hypothesis. For a population in which the distribution in the classes is shown by the marginal totals and the classes are independent, we are asking what proportion of a large series of samples similar to the one under consideration will deviate as much or more from the theoretical as the one observed. On the basis of the marginal total the expected entry in the upper left-hand corner would be (24) • (10)/40 = 6. After this entry or any other has been calculated, all the remaining entries can be obtained by subtraction from the marginal totals. Since only one value need be calculated, for a 2 x 2 table, there is only one degree of freedom. This checks with the (j - 1) (k - 1) degrees of freedom for a jxk table since in this case j = k = 2. The expected (theoretical) values are given below:

	Spoiled	Unspoiled	Totals
Unfumigated	6	18	24
Fumigated	4	12	16
Totals	10	30	40

Therefore,

$$\chi^2_{\text{adjusted}} = \frac{(|8-6|-0.5)^2}{6} + \frac{(|16-18|-0.5)^2}{18} + \frac{(|4-2|-0.5)^2}{4} + \frac{(|12-14|-0.5)^2}{12}$$

$$= \frac{(1.5)^2}{6} + \frac{(1.5)^2}{18} + \frac{(1.5)^2}{4} + \frac{(1.5)^2}{12}$$

$$= 0.27 + 0.13 + 0.56 + 0.19$$

$$= 1.25$$

Note the use of the correction for non-continuity for 1 df. From Appendix Table A-5, $\chi^2_{0.05,1} = 3.84$. Since $\chi^2 = 1.25 < 3.84$, there is no reason to reject the null hypothesis of independence. It appears, therefore, that fumigation has had no significant effect in reducing spoilage.

A 2x3 contingency table

The following are tabulated data on 82 strains of oats divided into 2 groups according to the presence or absence of awns, and into 3 groups according to yield. Do these data permit the conclusion that more of the awned strains occur in the highest yielding classes than do awnless strains?

	Yield Class (weight in grams)		
	151-200	201-250	251-325
Awned	6	7	21
Awnless	18	21	9

The χ^2 is calculated as follows:

Yield Class (expected values in parentheses)				
	151-200	201-250	251-325	Total
Awned	6(10)	7(11.6)	21(12.4)	34
Awnless	18(14)	21(16.4)	9(17.6)	48
Totals	24	28	30	82

$$\chi^2 = \frac{(6-10)^2}{10} + \frac{(18-14)^2}{14} + \frac{(7-11.6)^2}{11.6} + \frac{(21-16.4)^2}{16.4}$$

$$+ \frac{(21-12.4)^2}{12.4} + \frac{(9-17.6)^2}{17.6}$$

$$= \frac{16}{10} + \frac{16}{14} + \frac{21.16}{11.6} + \frac{21.16}{16.4} + \frac{70.56}{12.4} + \frac{70.56}{17.6}$$

$$= 1.60 + 1.14 + 1.82 + 1.29 + 5.69 + 4.01$$

$$= 15.55$$

Since $\chi_2 = 15.55 > \chi^2_{0.05,2} = 5.99$, the observed values are not distributed as expected on the basis of the marginal totals, and the awned strains do occur in the highest yielding classes more frequently.

SUMMARY

1. The chi-square distribution can be used to test the goodness of fit of data to a hypothesized model.

$$\chi^2 = \Sigma \frac{(0_i - E_i)^2}{E_i}$$

Where O_i is the observed frequency and E_i is the expected frequency from the hypothesized model. Assume observations are classified into k frequency classes, then

a) χ^2 has k-1 df if no parameter of the hypothesized model needs to be estimated from the data.

b) χ^2 has k-1-n df in n parameters of the model are to be estimated from the data.

2. The chi-square distribution can also be used to test the hypothesis of statistical independence between two variables, each of which is classified into a number of categories or attributes. The two-way table of the classified frequencies is called a contingency table.

$$\chi^2 = \Sigma \frac{(0_{ij} - E_{ij})^2}{E_{ij}}$$

where O_{ij} is the observed frequency in the i^{th} row and the j^{th} column, with expected frequency E_{ij}. If there are j rows and k columns, the χ^2 has (j - 1) (k - 1) df.

3. In case the calculated ^{2} has 1 df, the above formula needs to be modified as

$$\text{adj. } \chi^2 = \Sigma \frac{(|0_i - E_i| - 0.5)^2}{E_i}$$

EXERCISES

1. Suppose 50 years rainfall data are summarized in the table below. Test whether the data approximate a normal distribution.

Class	Midpoint (inches)	Frequency	Endpoint
1	8	3	9.5
2	11	9	12.5
3	14	14	15.5
4	17	10	18.5
5	20	7	21.5
6	23	0	24.5
7	26	4	27.5
8	29	2	30.5
9	32	1	33.5

$$\overline{Y} = 16.5 \text{ and } S = 5.4$$

2. Of 64 offspring of a certain cross of guinea pigs, 34 are red, 10 are black and 20 are white. According to the genetic model, these numbers should be in the ratio of 9:3:4. Are the data consistent with the model?

3. In an experiment involving the crossing of two hybrids of a species of flower the results shown below are observed. Are these results consistent with the expected population 9:3:3:1?

Magenta Flower Green Stigma	Magenta Flower Red Stigma	Red Flower Green Stigma	Red Flower Red Stigma
120	49	36	12

4. It was hypothesized that the F_2 generation of a particular barley cultivar would show a 9 hooded to 3 long-awned to 4 short-awned ratio. The observed data showed 348 hooded to 115 long-awned to 157 short-awned individuals. Test the hypothesis.

5. Suppose in a particular orange growing area of California that on the average 25% of the oranges are graded as best grade, 40% are placed in the above average grade, 25% are graded as average and 10% are graded as poor. A random sample of 500 bushels from one orchard in the region yields 100 bushels of best grade, 300 bushels of above average

grade, 50 bushels of average grade and 50 bushels of poor grade oranges. Is the quality of oranges from this orchard representative of this area?

(χ^2 = 167.5)

6. A study was conducted to establish whether there is a difference in susceptibility to mildew between two types of pasture grass. The following data were obtained:

Type	Mildewed	Not Mildewed	Total
A	107	289	396
B	291	81	372
Total	398	370	768

Test the hypothesis that there is no difference between the two grasses in their susceptibility to mildew. (χ^2 = 199.40)

7. Two barley fields containing two different varieties of barley were examined for rust. The results are presented below.

Variety	Rust	No Rust	Total
1	470	30	500
2	436	64	500
Total	906	94	1000

Is there evidence of heterogeneity in response to rust? (χ^2 = 12.787)

8. During an epidemic of cholera the following data on the effectiveness of inoculation as a means of preventing the disease were obtained.

	Not Attacked	Attacked
Inoculated	192	4
Not inoculated	113	34

Do these data indicate the effectiveness of the inoculation on the basis of the 1% level of significance? (Yes, since adj. χ^2 = 35.79)

9. Twenty-two animals are suffering from a disease, the severity of which is about the same in each case. In order to test the therapeutic value of a serum, it is administered to 10 of the animals; 12 remain uninoculated as a control. The results are shown below.

	Recovered	Died
Inoculated	7	3
Not inoculated	3	9

Has inoculation been effective? (5% level) (No, since adj. χ^2 = 2.82)

10. It is suspected that different combinations of temperature and humidity affect the number of defective articles produced in a certain workroom. Do the following data confirm this suspicion? (5% level)(Yes, since adj. χ^2 = 5.95)

		Humidity	
		Low	High
Temperature	Low	10	5
	High	4	16

11. A plant breeder wants to know if the sterility of rice is a genetic problem. Samples were taken from a large field study of 400 plots and the sterility of each plot was rated as follows:

	Genotypes			
Sterility	A	B	C	D
No problem	20	15	12	10
Moderate	70	60	80	50
Severe	10	25	8	40

Test the hypothesis that the severity of sterility is independent of genetic make-up or genotype. (χ^2 = 43.807)

12. An alfalfa breeder conducted a study on inheritance of resistance to anthracnose. The following data were obtained:

Growth Habit	Resistant	Not Resistant	Total
Standard	55	45	100
Intermediate standard	70	30	100
Intermediate alpha	82	18	100
Alpha	88	12	100
Total	298	102	400

Can he conclude that resistance to crown rot and growth habit are independent?

(χ^2 = 33.636)

13. A pastry chef wishes to determine whether the proportion of unsatisfactory bear claws is affected by oven temperature. He has baked batches of them at 350 , 400 , and 450 , and has obtained the following results:

	350	400	450	Total
Number satisfactory	132	128	111	371
Number unsatisfactory	14	17	35	66
Totals	146	145	146	437

If the chef wishes to test the null hypothesis of identical proportions at an = 0.05 significance level, what conclusion should he reach? ($\chi^2 = 13.712$)