STA258
Chi-Squared Tests

Al Nosedal.
University of Toronto.

Winter 2017

Example

Suppose that customers can purchase one of the three brands of milk at a supermarket. In a study to determine whether one brand is preferred over another, a record is made of a sample of $n = 300$ milk purchases. The data are shown below. Do the data provide sufficient evidence to indicate a preference for one or more brands?

Brand 1	Brand 2	Brand 3	Total
78	117	105	300

Step 1. State Hypotheses.

If all the brands are **equally** preferred, then the probability that a purchaser will choose any one brand is the same as the probability of choosing any other - that is, $p_1 = p_2 = p_3 = 1/3$. Therefore, the null hypothesis of "no preference" is

$$H_0 : p_1 = p_2 = p_3 = 1/3$$

If p_1, p_2, and p_3 are not all equal, the brands are not equally preferred; in other words, the purchasers must have a preference for one (or possibly) two brands. The alternative hypothesis is

$$H_a : p_1, p_2, \text{ and } p_3 \text{ are not all equal}$$

Therefore, we seek a test statistic that will detect a **lack of fit** of the observed **cell counts** to our hypothesized (null) expected cell counts based on the hypothesized cell probabilities.
These expected values are:
$E(n_1) = np_1 = (300)\left(\frac{1}{3}\right) = 100$
$E(n_2) = np_2 = (300)\left(\frac{1}{3}\right) = 100$
$E(n_3) = np_3 = (300)\left(\frac{1}{3}\right) = 100$

Table of Expected Counts

Brand 1	Brand 2	Brand 3	Total
100	100	100	300

The test statistic for comparing the observed and expected cell counts (and, consequently, testing $H_0 : p_1 = p_2 = p_3 = 1/3$ is the $\mathbf{X^2}$ **statistic**:

$$X^2 = \sum_{all\ \ cells} \frac{(\text{observed - expected})^2}{\text{expected}} = \sum_{all\ \ cells} \frac{(n_i - E(n_i))^2}{E(n_i)}$$

$$X^2 = \frac{(78-100)^2}{100} + \frac{(117-100)^2}{100} + \frac{(105-100)^2}{100}$$

$$X^2 = 4.84 + 2.89 + 0.25 = 7.98$$

Step 3. Finding P-value

To find the P-value, compare X^2 with critical values from the chi-square distribution with degrees of freedom one fewer than the number of values the brand can take. That's $3 - 1 = 2$ degrees of freedom. From Table, we see that $X^2 = 7.98$ falls between 0.02 and 0.01 critical values of the chi-square distribution with 2 degrees of freedom. So the P-value of $X^2 = 7.98$ is between 0.01 and 0.02 $(0.01 < P - value < 0.02)$.

Step 4. Conclusion

If we used $\alpha = 0.05$, since our $P - value < \alpha = 0.05$, we could reject H_0 at the 5% significance level. We would conclude that the three brands of milk are **not** equally preferred.

Example

Raymond Weil is about to come out with a new watch and wants to find out whether people have special preferences of the color of the watchband, or whether all four colors under consideration are equally preferred. A random sample of 80 prospective watch buyers is selected. Each person is shown the watch with four different band colors and asked to state his or her preference. The results (observed counts) are given below.

Tan	Brown	Maroon	Black	Total
12	40	8	20	80

Use $\alpha = 0.01$.

Step 1. State Hypotheses.

If all the brands are **equally** preferred, then the probability that a purchaser will choose any one color is the same as the probability of choosing any other - that is, $p_1 = p_2 = p_3 = p_4 = 1/4$. Therefore, the null hypothesis of "no preference" is

$$H_0 : p_1 = p_2 = p_3 = p_4 = 1/4$$

If p_1, p_2, p_3 and p_4 are not all equal, the colors are not equally preferred. The alternative hypothesis is

$$H_a : p_1, p_2, p_3 \text{ and } p_4 \text{ are not all equal}$$

Therefore, we seek a test statistic that will detect a **lack of fit** of the observed **cell counts** to our hypothesized (null) expected cell counts based on the hypothesized cell probabilities.

These expected values are:

$E(n_1) = np_1 = (80)\left(\frac{1}{4}\right) = 20$

$E(n_2) = np_2 = (80)\left(\frac{1}{4}\right) = 20$

$E(n_3) = np_3 = (80)\left(\frac{1}{4}\right) = 20$

$E(n_4) = np_4 = (80)\left(\frac{1}{4}\right) = 20$

Table of Expected Counts

Tan	Brown	Maroon	Black	Total
20	20	20	20	80

Step 2. Computing test statistic

The test statistic for comparing the observed and expected cell counts (and, consequently, testing $H_0 : p_1 = p_2 = p_3 = p_4 = 1/4$ is the $\mathbf{X^2}$ **statistic**:

$$X^2 = \sum_{all\ \ cells} \frac{(\text{observed - expected})^2}{\text{expected}} = \sum_{all\ \ cells} \frac{(n_i - E(n_i))^2}{E(n_i)}$$

$$X^2 = \frac{(12-20)^2}{20} + \frac{(40-20)^2}{20} + \frac{(8-20)^2}{20} + \frac{(2-20)^2}{20}$$

$$X^2 = 64/20 + 400/20 + 144/20 + 0 = 30.4$$

Step 3. Finding P-value

To find the P-value, compare X^2 with critical values from the chi-square distribution with degrees of freedom one fewer than the number of values the color can take. That's $4 - 1 = 3$ degrees of freedom. From Table, we see that $X^2 = 30.4$ is greater than the greatest entry in the df $= 3$ row, which is the critical value for tail area 0.0005. The P-value is therefore smaller than 0.0005.

Step 4. Conclusion

Since our $P - value < \alpha = 0.01$, we conclude that there is evidence to reject the null hypothesis that all four colors are equally likely to be chosen. Some colors are probably preferable to others. Our P-value is very small.

R Code

```r
x = c(12, 40, 8, 20);

chisq.test(x);

##
##  Chi-squared test for given probabilities
##
## data:  x
## X-squared = 30.4, df = 3, p-value = 1.137e-06

# chisq.test(x) gives you test statistic;
# degrees of freedom and P-value;
```

R Code

```
chisq.test(x)$expected;

## [1] 20 20 20 20

# gives you expected counts;
```

Exercise

Consider a multinomial experiment involving $n = 150$ trials and $k = 5$ cells. The observed frequencies resulting from the experiment are shown in the accompanying table, and the null hypothesis to be tested is as follows:

$$H_0 : p_1 = 0.1, \ \ p_2 = 0.2, \ \ p_3 = 0.3, \ \ p_4 = 0.2, \ \ p_5 = 0.2$$

Test the hypothesis at the 1% significance level.

Cell	1	2	3	4	5
Frequency	12	32	42	36	28

Therefore, we seek a test statistic that will detect a **lack of fit** of the observed **cell counts** to our hypothesized (null) expected cell counts based on the hypothesized cell probabilities.

These expected values are:

$$E(n_1) = np_1 = (150)\left(\frac{1}{10}\right) = 15$$
$$E(n_2) = np_2 = (150)\left(\frac{2}{10}\right) = 30$$
$$E(n_3) = np_3 = (150)\left(\frac{3}{10}\right) = 45$$
$$E(n_4) = np_4 = (150)\left(\frac{2}{10}\right) = 30$$
$$E(n_5) = np_5 = (150)\left(\frac{2}{10}\right) = 30$$

Table of Expected Counts

Cell	1	2	3	4	5
Frequency	15	30	45	30	30

Step 2. Computing test statistic

$$X^2 = \sum_{all\ \ cells} \frac{(\text{observed - expected})^2}{\text{expected}} = \sum_{all\ \ cells} \frac{(n_i - E(n_i))^2}{E(n_i)}$$

$$X^2 = \frac{(12-15)^2}{15} + \frac{(32-30)^2}{30} + \frac{(42-45)^2}{45} + \frac{(36-30)^2}{30} + \frac{(28-30)^2}{30}$$

$$X^2 = 9/15 + 4/30 + 9/45 + 36/30 + 4/30 = 2.2667$$

To find the P-value, compare X^2 with critical values from the chi-square distribution with degrees of freedom one fewer than the number of "columns". That's $5 - 1 = 4$ degrees of freedom. From Table, we see that $X^2 = 2.2667$ is smaller than the smallest entry (5.39) in the df $= 4$ row, which is the critical value for tail area 0.25. The P-value is therefore greater than 0.25.

Step 4. Conclusion

Since our $P - value > 0.25 > \alpha = 0.01$, we conclude that there is NOT enough evidence to reject the null hypothesis H_0. There is not enough evidence to infer that at least one p_i is not equal to its specified value.

The Chi-square Test for Goodness of fit

A categorical variable has k possible outcomes, with probabilities $p_1, p_2, ..., p_k$. That is , p_i is the probability of the ith outcome. We have n independent observations from this categorical variable.
To test the null hypothesis that the probabilities have specified values

$$H_0 : p_1 = p_{10}, p_2 = p_{20}, ..., p_k = p_{k0}$$

use the **chi-square statistic**

$$X^2 = \sum \frac{(\text{observed count - expected count})^2}{\text{expected count}}$$

(expected counts are equal to np_i).
The P-value is the area to the right of X^2 under the density curve of the chi-square distribution with $k - 1$ degrees of freedom.

Example

An article in Business Week reports profits and losses of firms by industry. A random sample of 100 firms is selected, and for each firm in the sample, we record whether the company made money or lost money, and whether or not the firm is a service company. The data are summarized in the 2×2 contingency table. Using the information in the table, determine whether or not you believe that the two events "the company made a profit this year" and "the company is in the service industry" are independent. Use $\alpha = 0.01$

Data

	Industy type		
	Service	Nonservice	Total
Profit	42	18	60
Loss	6	34	40
Total	48	52	100

Step 1. State Hypotheses

One way to solve the problem is to consider that there are two variables: industry type and profit/loss.

$$H_0 : \text{The two variables are independent}$$

vs

$$H_a : \text{The two variables are dependent}$$

Step 2. Computing test statistic

The **chi-square statistic** is a measure of how far the observed counts in a two-way table are from the expected counts. The formula for the statistic is

$$X^2 = \sum_{all\ \ cells} \frac{(\text{observed - expected})^2}{\text{expected}}$$

The **expected count** in any cell of a two-way table when H_0 is true is

$$\text{expected count} = \frac{\text{row total } \times \text{column total}}{\text{table total}}$$

Expected counts

$$E_{11} = \frac{(60)(48)}{100} = 28.8$$
$$E_{12} = \frac{(60)(52)}{100} = 31.2$$
$$E_{21} = \frac{(40)(48)}{100} = 19.2$$
$$E_{22} = \frac{(40)(52)}{100} = 20.8$$

$$X^2 = \frac{(42-28.8)^2}{28.8} + \frac{(18-31.2)^2}{31.2} + \frac{(6-19.2)^2}{19.2} + \frac{(34-20.8)^2}{20.8} = 29.09$$

Step 3. Finding P-value

To find the P-value, compare X^2 with critical values from the chi-square distribution with $(r-1)\times(c-1) = (2-1)\times(2-1) = 1$ degree of freedom. From Table, we see that $X^2 = 29.09$ is greater than the greatest entry in the df $= 1$ row, which is the critical value for tail area 0.0005. The P-value is therefore smaller than 0.0005.

Step 4. Conclusion

Since our $P - value < \alpha = 0.01$, we conclude that there is evidence to reject the null hypothesis and conclude that the two qualities (variables), profit/loss and industry type, are probably not independent.

Exercise

The operations manager of a company that manufactures shirts wants to determine whether there are differences in the quality of workmanship among the three daily shifts. She randomly selects 600 recently made shirts and carefully inspects them. Each shirt is classified as either perfect or flawed, and the shift that produced it is also recorded. The accompanying table summarizes the number of shirts that fell into each cell. Do these data provide sufficient evidence to infer that there are differences in quality between the three shifts? Use $\alpha = 0.05$.

Data

	Shift		
	1	2	3
Perfect	240	191	139
Flawed	10	9	11

Step 1. State Hypotheses

There are two variables: shift and quality.

$$H_0 : \text{The two variables are independent}$$

vs

$$H_a : \text{The two variables are dependent}$$

Step 2. Computing test statistic

The **chi-square statistic** is a measure of how far the observed counts in a two-way table are from the expected counts. The formula for the statistic is

$$X^2 = \sum_{all \ \ cells} \frac{\text{(observed - expected)}^2}{\text{expected}}$$

The **expected count** in any cell of a two-way table when H_0 is true is

$$\text{expected count} = \frac{\text{row total} \ \times \text{column total}}{\text{table total}}$$

Table of Expected Counts

	Shift		
	1	2	3
Perfect	237.5	190	142.5
Flawed	12.5	10	7.5

Test statistic

$$X^2 = \frac{(240-237.5)^2}{237.5} + \frac{(191-190)^2}{190} + \frac{(139-142.5)^2}{142.5} + \frac{(10-12.5)^2}{12.5} + \frac{(9-10)^2}{10} + \frac{(11-7)^2}{7} = 2.3509$$

Step 3. Finding P-value

To find the P-value, compare X^2 with critical values from the chi-square distribution with $(r-1)\times(c-1) = (2-1)\times(3-1) = 2$ degrees of freedom. From Table, we see that $X^2 = 2.3509$ is smaller than the smallest entry in the df $= 2$ row, which is the critical value for tail area 0.25. The P-value is therefore greater than 0.25.

Step 4. Conclusion

Since our $P - value > 0.25 > \alpha = 0.05$, we **can't reject** H_0. We conclude that we don't have enough evidence to infer that there are differences in quality among the three shifts.

R Code

```
perfect=c(240, 191, 139);

flawed=c(10 , 9 ,11 );

chisq.test(data.frame(perfect,flawed));

##
##  Pearson's Chi-squared test
##
## data:  data.frame(perfect, flawed)
## X-squared = 2.3509, df = 2, p-value = 0.3087
```

R Code

```
chisq.test(data.frame(perfect,flawed))$expected;

##      perfect flawed
## [1,]   237.5   12.5
## [2,]   190.0   10.0
## [3,]   142.5    7.5

# gives you expected counts;
```

Relationship between demands for desktops and laptops

Big Office, a chain of large office supply stores, sells a variety of Windows and Mac laptops. Company executives want to know whether the demands for these two types of computers are related in any way. They might act as complementary products, where high demand for Windows laptops accompanies high demand for Mac laptops, they might act as substitute products (demand for one takes away demand for the other), or their demands might be unrelated. Because of limitations in its information system, Big Office does not have the exact demands for these products. However, it does have daily information on categories of demand, listed in aggregate (that is, over all stores). These data appear in the next slide. Each day's demand for each type of computer is categorized as Low, Medium Low, Medium High, or High. Based on these data, can Big Office conclude that demands for these two products are independent? Use $\alpha = 0.05$.

Data

		Low	Windows Med Low	Med High	High	Total
Mac	Low	4	17	17	5	43
	Med Low	8	23	22	27	80
	Med High	16	20	14	20	70
	High	10	17	19	11	57
	Total	38	77	72	63	250

The table is based on 250 days, so that the counts add to 250. The individual counts show, for example, that demand was high for both Windows and Mac laptops on 11 of the 250 days.

Solution

Test statistic:
$X^2 = 17.242$

df: 9

P-value:
0.045058

Exercise

Econetics Research Corporation, a well-known Montreal-based consulting firm, wants to test how it can influence the proportion of questionnaires returned from surveys. In the belief that the inclusion of an inducement to respond may be important, the firm sends out 1000 questionnaires: Two hundred promise to send respondents a summary of the survey results, 300 indicate that 20 respondents (selected by lottery) will be awarded gifts, and 500 are accompanied by no inducements. Of these, 80 questionnaires promising a summary, 100 questionnaires offering gifts, and 120 questionnaires offering no inducements are returned. What can you conclude from these results? (Use $\alpha = 0.01$).

Data

	Returned		
	Yes	No	Total
Summary	80	120	200
Gifts	100	200	300
No inducements	120	380	500
Total	300	700	1000

Step 1. State Hypotheses

There are two variables: inducement and return.

$$H_0 : \text{The two variables (inducement and return) are independent}$$

vs

$$H_a : \text{The two variables are dependent}$$

The **chi-square statistic** is a measure of how far the observed counts in a two-way table are from the expected counts. The formula for the statistic is

$$X^2 = \sum_{all\ \ cells} \frac{(\text{observed - expected})^2}{\text{expected}}$$

The **expected count** in any cell of a two-way table when H_0 is true is

$$\text{expected count} = \frac{\text{row total } \times \text{column total}}{\text{table total}}$$

Table of Expected Counts

	Returned		
	Yes	No	Total
Summary	60	140	200
Gifts	90	210	300
No inducements	150	350	500
Total	300	700	1000

Test statistic

$$X^2 = \frac{(80-60)^2}{60} + \frac{(120-140)^2}{140} + \frac{(100-90)^2}{90} + \frac{(200-210)^2}{210} + \frac{(120-150)^2}{150} + \frac{(380-350)^2}{350} = 19.682$$

Step 3. Finding P-value

To find the P-value, compare X^2 with critical values from the chi-square distribution with $(r-1) \times (c-1) = (3-1) \times (2-1) = 2$ degrees of freedom. From Table, we see that $X^2 = 19.682$ is greater than the greatest entry in the df $= 2$ row, which is the critical value for tail area 0.0005. The P-value is therefore smaller than 0.0005. (You can find the exact P-value using R).

Step 4. Conclusion

Since our $P - value < \alpha = 0.01$, we **reject** H_0. There is sufficient evidence to infer that the return rates differ among the different inducements.

APPENDIX

The Chi-Square Test

Each of the n_is have Binomial distributions with parameters n and p_i, and the expected numbers falling into cell i is

$$E(n_i) = np_i, \quad i = 1, 2, ..., k.$$

Now suppose that we hypothesize values for $p_1, p_2, ..., p_k$ and calculate the expected value for each cell. Certainly if our hypothesis is true, the cell counts n_i should not deviate greatly from their expected values np_i, for $i = 1, 2, ..., k$.

In 1900 Karl Pearson proposed the following test statistic, which is a function of the squares of the deviations of the observed counts from their expected values, weighted by the reciprocals of their expected values:

$$X^2 = \sum_{i=1}^{k} \frac{[n_i - E(n_i)]^2}{E(n_i)} = \sum_{i=1}^{k} \frac{[n_i - np_i]^2}{np_i}$$

It can be shown that when n is large, X^2 has an approximate chi-square probability distribution.

We will demonstrate this result for the case $k = 2$. If $k = 2$, then $n_2 = n - n_1$ and $p_1 + p_2 = 1$. Thus,

$$X^2 = \sum_{i=1}^{2} \frac{[n_i - E(n_i)]^2}{E(n_i)} = \frac{(n_1 - np_1)^2}{np_1} + \frac{(n_2 - np_2)^2}{np_2}$$

$$= \frac{(n_1 - np_1)^2}{np_1} + \frac{[(n - n_1) - n(1 - p_1)]^2}{n(1 - p_1)}$$

$$= \frac{(n_1 - np_1)^2}{np_1} + \frac{[np_1 - n_1]^2}{n(1 - p_1)}$$

$$= (n_1 - np_1)^2 \left(\frac{1}{np_1} + \frac{1}{n(1 - p_1)} \right)$$

$$= \frac{(n_1 - np_1)^2}{n(1 - p_1)}$$

We have seen that for large n

$$\frac{n_1 - np_1}{\sqrt{np_1(1-p_1)}}$$

has approximately a standard Normal distribution. Since the square of a standard Normal random variable has a χ^2 distribution, for $k = 2$ and large n, X^2 has an approximate χ^2 distribution with 1 degree of freedom (df).