Section 9.2 - Inference for Two-way Tables
Section 9.3 - Formulas and Models for Two-way Tables

Statistics 104

Autumn 2004

Inference for Two-way Tables

Interested in whether there is an association between 2 categorical variables.

For example, is there a relationship between treatment and survival or major applied to and acceptance.

The null hypothesis can be thought of in terms of independence or conditional distributions.

Example: Mail survey response rates

- A survey was mailed to 841 skydivers
- 427 received a survey with a plain cover
- 414 received a survey with a cover graphic of a skydiver
- Interest in whether the response times were different for the different covers

- The response times were based on the postmark for the returned surveys

	Response Time (days)					
Cover	1-7	8-14	15-31	32-60	Not Returned	Total
Graphic	70	76	51	19	198	414
Plain	84	53	51	32	207	427
Total	154	129	102	51	405	841

The hypotheses to be examined are

H_0: the response time category probabilities are the same for the two cover designs ($p_{G1} = p_{P1}, \ldots, p_{G5} = p_{P5}$), i.e. response time and cover are independent.

H_A: the proportions are not the same for all response time categories for the two cover designs ($p_{Gi} \neq p_{Pi}$ for some response time category i), i.e. response time and cover are dependent.

Conditional distribution of response times, given cover type

	Response Time (days)					
Cover	1-7	8-14	15-31	32-60	Not Returned	Total
Graphic	0.1691	0.1836	0.1232	0.0459	0.4783	1
Plain	0.1967	0.1242	0.1194	0.0749	0.4848	1
Total	0.1831	0.1534	0.1213	0.0606	0.4816	1

If H_0 is true, the row conditionals should be similar to the marginal distribution of response times.

(If two events are independent, $P[A|B] = P[A]$.)

This roughly seems to be the case.

However, instead of comparing the conditional distributions directly, the usual approach is to compare the observed counts in each cell with what would be expected under the null hypothesis.

These are calculated by

$$\text{Expected count} = \frac{\text{Row total} \times \text{Column total}}{n}$$

Note that this can be rewritten as

$$\text{Expected count} = \frac{\text{Row total}}{n} \times \text{Column total} \tag{1}$$

or

$$\text{Expected count} = \frac{\text{Column total}}{n} \times \text{Row total} \tag{2}$$

The set of $\frac{\text{Row total}}{n}$ gives the row marginal distribution and the set of $\frac{\text{Column total}}{n}$ gives the column marginal distribution.

Case (2), which is the one we are interested in here, says to take the number of observations for each cover time (Row total) and distribute them relative to the marginal distribution of response time types.

The expected counts for the table are

	Response Time (days)					
Cover	1-7	8-14	15-31	32-60	Not Returned	Total
Graphic	75.8	63.5	50.2	25.1	199.4	414
Plain	78.2	65.5	51.8	25.9	205.6	427
Total	154	129	102	51	405	841

For example, the expected count for Plain cover and a response time of 15-31 days is

$$\frac{427 \times 102}{841} = 51.79$$

The table of expected counts has the same marginal counts as the observed data.

Chi-square Statistic

The common test statistic for examining whether there is an association between two variables in an $r \times c$ table is

$$X^2 = \sum_{\text{all cells}} \frac{(\text{Obs count} - \text{Exp count})^2}{\text{Exp count}}$$

If the null hypothesis isn't true, then there should be some cells in the table where the observed and expected counts are very different. This would lead to a large value for for X^2.

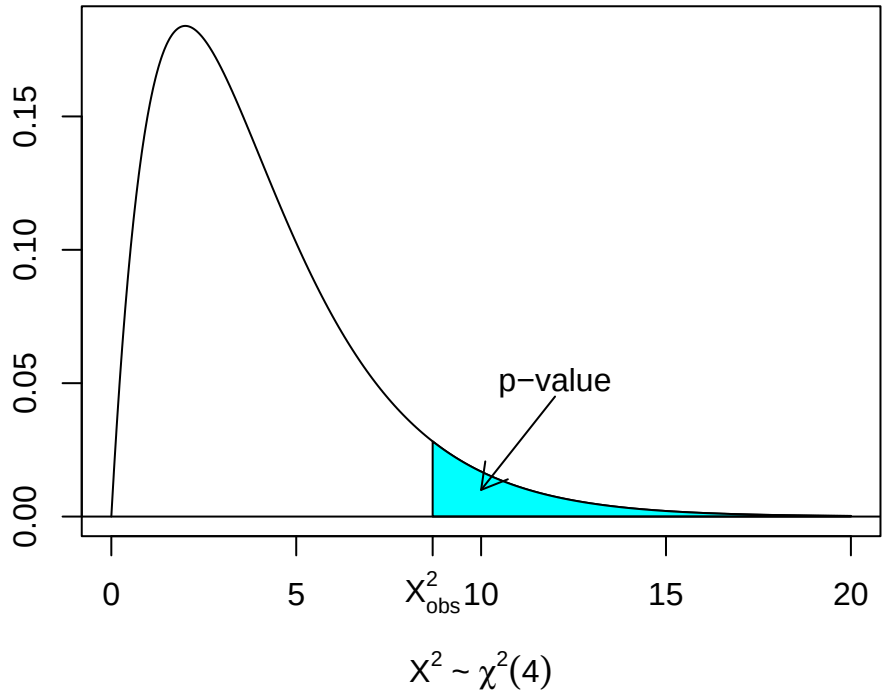
If the number of observations is large, the sampling distribution of X^2 is approximately Chi-squared (χ^2) with $(r-1) \times (c-1)$ degrees of freedom.

The p-value for the Chi-square test is

$$p - \text{value} = P[X^2 \geq X^2_{obs}]$$

where $X^2 \sim \chi^2(df)$.

Note that for the Chi-square test on a two-way table, usually only the upper tail is used in calculating the p-value. The lower tail (values close to 0) are highly consistent with the null hypothesis, so shouldn't be included when calculating the p-value.



If a fixed level α hypothesis test is desired, the rejection region is of the form

$$\text{Reject } H_0 \text{ if } X^2_{obs} \geq \chi^{2*}$$

These critical values for the Chi-square distributions are given in Table F.

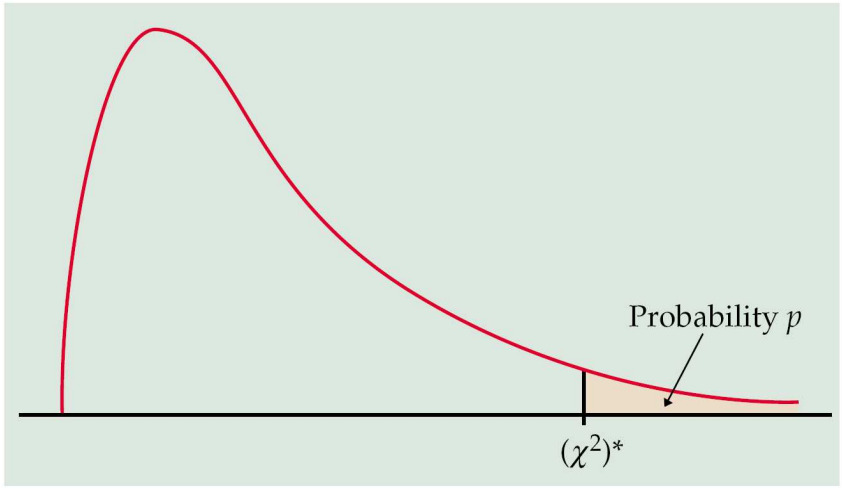


Table entry for p is the critical value $(\chi^2)^*$ with probability p lying to its right.

TABLE F χ^2 distribution critical values

	Tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.32	1.64	2.07	2.71	3.84	5.02	5.41	6.63	7.88	9.14	10.83	12.12
2	2.77	3.22	3.79	4.61	5.99	7.38	7.82	9.21	10.60	11.98	13.82	15.20
3	4.11	4.64	5.32	6.25	7.81	9.35	9.84	11.34	12.84	14.32	16.27	17.73
4	5.39	5.99	6.74	7.78	9.49	11.14	11.67	13.28	14.86	16.42	18.47	20.00
5	6.63	7.29	8.12	9.24	11.07	12.83	13.39	15.09	16.75	18.39	20.51	22.11
6	7.84	8.56	9.45	10.64	12.59	14.45	15.03	16.81	18.55	20.25	22.46	24.10
7	9.04	9.80	10.75	12.02	14.07	16.01	16.62	18.48	20.28	22.04	24.32	26.02
8	10.22	11.03	12.03	13.36	15.51	17.53	18.17	20.09	21.95	23.77	26.12	27.87
9	11.39	12.24	13.29	14.68	16.92	19.02	19.68	21.67	23.59	25.46	27.88	29.67
10	12.55	13.44	14.53	15.99	18.31	20.48	21.16	23.21	25.19	27.11	29.59	31.42
11	13.70	14.63	15.77	17.28	19.68	21.92	22.62	24.72	26.76	28.73	31.26	33.14
12	14.85	15.81	16.99	18.55	21.03	23.34	24.05	26.22	28.30	30.32	32.91	34.82
13	15.98	16.98	18.20	19.81	22.36	24.74	25.47	27.69	29.82	31.88	34.53	36.48
14	17.12	18.15	19.41	21.06	23.68	26.12	26.87	29.14	31.32	33.43	36.12	38.11

Observed counts:

	Response Time (days)				
Cover	1-7	8-14	15-31	32-60	Not Returned
Graphic	70	76	51	19	198
Plain	84	53	51	32	207

Expected counts:

	Response Time (days)				
Cover	1-7	8-14	15-31	32-60	Not Returned
Graphic	75.8	63.5	50.2	25.1	199.4
Plain	78.2	65.5	51.8	25.9	205.6

$$X^2 = \frac{(70-75.8)^2}{75.8} + \frac{(76-63.5)^2}{63.5} + \ldots + \frac{(198-199.4)^2}{199.4}$$
$$+\frac{(84-78.2)^2}{78.2} + \frac{(53-65.5)^2}{65.5} + \ldots + \frac{(207-205.6)^2}{205.6} = 8.6884$$

$$r = 2, c = 5 \Rightarrow df = (2-1)(5-1) = 4$$

$$p - \text{value} = P[X^2 \geq 8.6884] = 0.069$$

To perform the Chi-square test in Stata, the following can be done

```
. tabulate Cover Response_Time [fweight=Count], chi2

           |                    Response_Time
     Cover |     01-07      08-14      15-31      32-60  Not retur |     Total
-----------+-------------------------------------------------------+----------
   Graphic |        70         76         51         19        198 |       414
     Plain |        84         53         51         32        207 |       427
-----------+-------------------------------------------------------+----------
     Total |       154        129        102         51        405 |       841

          Pearson chi2(4) =   8.6884   Pr = 0.069
```

Notes

1. This assumes that the data is entered similarly to

```
. list

     +---------------------------------+
     | Response_Time     Cover   Count |
     |---------------------------------|
  1. |         01-07   Graphic      70 |
  2. |         08-14   Graphic      76 |
  3. |         15-31   Graphic      51 |
  4. |         32-60   Graphic      19 |
  5. |  Not returned   Graphic     198 |
     |---------------------------------|
  6. |         01-07     Plain      84 |
  7. |         08-14     Plain      53 |
  8. |         15-31     Plain      51 |
  9. |         32-60     Plain      32 |
 10. |  Not returned     Plain     207 |
     +---------------------------------+
```

2. There are actually 2 different Chi-square tests for examining two-way tables. The one discussed is Pearson's Chi-square. The likelihood ratio Chi-square usually gives similar answers.

As mentioned earlier, this test has a sampling distribution with an approximate Chi-square distribution. The approximation should be reasonable if the expected cell counts in every cell are at least 5.

Chi-square test for 2 $\times$ 2 tables

Instead of doing a z test examining $H_0 : p_1 = p_2$, we could do a Chi-square test instead. Note that the same hypothesis is being examined with both tests.

For the Aspirin example

```
. tabulate treatment stroke [fweight=count], chi2 expected
           |        stroke
 treatment |        No        Yes |     Total
-----------+----------------------+----------
   Aspirin |        63         15 |        78
           |      53.3       24.7 |      78.0
-----------+----------------------+----------
   Placebo |        43         34 |        77
           |      52.7       24.3 |      77.0
-----------+----------------------+----------
     Total |       106         49 |       155
           |     106.0       49.0 |     155.0
      Pearson chi2(1) =  11.1349   Pr = 0.001
```

```
| Key                 |
|---------------------|
|     frequency       |
| expected frequency  |
+---------------------+
```

```
. prtesti 78 63 77 43, count

Two-sample test of proportion                     x: Number of obs = 78
                                                  y: Number of obs = 77

------------------------------------------------------------------------------
Variable |       Mean   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------+--------------------------------------------------------------------
       x |   .8076923   .0446246                      .7202298    .8951548
       y |   .5584416   .0565897                      .4475277    .6693554
---------+--------------------------------------------------------------------
    diff |   .2492507   .0720677                      .1080007    .3905008
         |  under Ho:   .0746952     3.34   0.001
------------------------------------------------------------------------------

               Ho: proportion(x) - proportion(y) = diff = 0

      Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
       z =  3.337                   z =  3.337                   z =  3.337
   P < z =  0.9996          P > |z| =  0.0008          P > z =  0.0004
```

These two tests seem to be giving similar answers. In fact, they are giving the same answer, assuming that the alternative hypothesis is $H_A : p_1 \neq p_2$.

1. $z^2 = 3.337^2 = 11.1349 = X^2$

2. p-value(z test) = p-value(χ^2 test)

3. $(z^*)^2 = \chi^{2*}$

1) can be shown for any 2 $\times$ 2 table. The result isn't special for this example.

The last two facts come from the probability result that if $z \sim N(0, 1)$, then $Y = z^2 \sim \chi^2(1)$

The difference in the reported p-values is just due to the number of significant digits Stata wants to use in the output for the two routines.

What to do if the sample sizes are small?

Similar to the one sample binomial problems, there is an exact procedure for two-way tables that can be used when the normal approximation doesn't hold. In the 2×2 table case, its known as Fisher's exact test. It can be easily performed in Stata.

The test statistic reported is a p-value, the probability of generating a more extreme table than that observed that has the same marginal counts.

Example: Risk of Urethritis in Seminal Super Shedding (SSS) is HIV-1

Is there a different risk of urethritis in two different groups of HIV-1 patients

```
. tabulate urethritis SSS [fweight=count], chi2 exact expected

           |          SSS
urethritis |       SSS     nonSSS |     Total
-----------+----------------------+----------
        No |         6         60 |        66
           |       8.3       57.8 |      66.0
-----------+----------------------+----------
       Yes |         3          3 |         6
           |       0.8        5.3 |       6.0
-----------+----------------------+----------
     Total |         9         63 |        72
           |       9.0       63.0 |      72.0

          Pearson chi2(1) =   8.4156   Pr = 0.004
           Fisher's exact =                 0.023
   1-sided Fisher's exact =                 0.023
```

While both tests are significant at $\alpha = 0.05$, you will make a different conclusion for $\alpha = 0.01$. The p-value from the Pearson Chi-square is off by a factor between 5 and 6.

When the sample sizes are large, the p-values from the Pearson χ^2 test are a good approximation to the exact p-values. However when the sample sizes are small, you may get a different answer.

```
 . tabulate treatment stroke
[fweight=count], chi2 expected exact

           |        stroke
 treatment |        No        Yes |     Total
-----------+----------------------+----------
   Aspirin |        63         15 |        78
           |      53.3       24.7 |      78.0
-----------+----------------------+----------
   Placebo |        43         34 |        77
           |      52.7       24.3 |      77.0
-----------+----------------------+----------
     Total |       106         49 |       155
           |     106.0       49.0 |     155.0

          Pearson chi2(1) =  11.1349   Pr = 0.001  (0.0008 in R)
           Fisher's exact =                 0.001  (0.0010 in R)
```