

Stat 218 - Day 27
Chi-Square Test for Larger Contingency Tables

The chi-square test that we have studied for a 2×2 table can be easily extended to a largertable. [Note: A 2×2 table means that the row and column variables are both binary and so have only two possible categories. An $r \times k$ table means that the row variable has r possible categories and the column variable has k possible categories.]

- The null hypothesis says that the distribution of the row variable is the same for every category of the column variable.
- Expected counts are calculated in the same manner: $E = \frac{(\text{row total}) \times (\text{column total})}{\text{grand total}}$.
- The test statistic is calculated in the same manner: $\chi_s^2 = \sum \frac{(O - E)^2}{E}$.
- The P -value is found from the chi-square distribution (Table 9) with degrees of freedom equal to $(r-1)(k-1)$.
- Only non-directional hypotheses are possible unless the table is 2×2.

Example: Newspaper reading habits

A sample of 811 men and 1059 women were asked how often they read a newspaper. The results are summarized in the table:

	Men	Women	Total
Every day	375 ()	430 ()	805
A few times per week	182 ()	238 ()	420
Once per week	121 ()	173 ()	294
Less than once per week	133 ()	218 ()	351
Total	811	1059	1870

(a) Identify the two categorical variables in this study. What are the values of r and k for this $r \times k$ table?

(b) State the null and alternative hypotheses for testing whether these sample data provide evidence that newspaper reading frequency differs between men and women in the population.

(c) Calculate the expected counts and record them inside the parentheses in the table.

(d) Calculate the value of the chi-square test statistic.

(e) Determine the P -value of the test as accurately as possible from Table 9.

(f) Do the distributions of newspaper reading frequency differ significantly between men and women at the $\alpha = .05$ significance level? How about at the $\alpha = .01$ level?

(g) Look at the eight pieces of the calculation of the chi-square test statistic. Which cells of the table do the two largest pieces come from? Are the observed counts above or below the expected counts in those cells?

(h) Summarize your conclusions about whether men and women differ in their newspaper reading habits.

(i) Enter the table of counts into Minitab and use it to confirm your analysis (MTB> chis c1-c2).

Example: Federal spending on space program

A sample of 1235 people were asked to identify their political views as liberal, moderate, or conservative. They were also asked to indicate whether they believe that the federal government spends too little, too much, or the right amount on the space program. The responses are summarized in the following table:

	Liberal	Moderate	Conservative	Total
Too little	50	47	36	133
Just right	164	214	212	590
Too much	162	174	176	512
Total	376	435	424	1235

(a) Identify the two categorical variables in this study.

(b) What are the values of r and k for this $r \times k$ table?

In the “newspaper reading” example, researchers took one random sample of men and an independent random sample of women, classifying the people according to their newspaper reading frequency. But in this “government spending” example, researchers took one random sample of people and classified them according to two variables: political viewpoint and opinion about federal spending on the space program. The chi-square test in this context tests the null hypotheses that these two variables are **independent**.

(c) Conduct the chi-square test. (Feel free to use Minitab.) Report the hypotheses, test statistic, and P -value. Also state your conclusion in context and explain the reasoning process by which it follows from your analysis.

Example: More on federal spending

A sample of 1264 people were also asked to give their opinion on whether the federal government spends too little, too much, or the right amount on the *environment*. Results are summarized in the following table:

	Liberal	Moderate	Conservative	Total
Too little	292	271	239	802
Just right	85	142	133	360
Too much	11	27	64	102
Total	388	440	436	1264

Conduct a chi-square test on these data. (Feel free to use Minitab.) Report the following:

(a) expected counts

(b) test statistic

(c) P -value (as accurately as possible)

(d) test decision at the $\alpha=.01$ level

(e) summary of what your analysis reveals, in context

