Table 2.6 Adjusted Residuals (in Parentheses) for Testing Independence in Table 2.5

	Party Identification		
Gender	Democrat	Independent	Republican
Females	279	73	225
	(2.29)	(0.46)	(−2.62)
Males	165	47	191
	(−2.29)	(−0.46)	(2.62)

indicates lack of fit of H_0 in that cell. Table 2.6 shows the adjusted residuals for testing independence in Table 2.5. For the first cell, for instance, $n_{11} = 279$ and $\hat{\mu}_{11} = 261.4$. The first row and first column marginal proportions equal $p_{1+} = 577/980 = .589$ and $p_{+1} = 444/980 = .453$. Substituting into (2.4.4), the adjusted residual for this cell equals

$$\frac{279 - 261.4}{\sqrt{261.4(1 - .589)(1 - .453)}} = 2.29.$$

This cell shows a greater discrepancy between n_{11} and $\hat{\mu}_{11}$ than one would expect if the variables were truly independent.

Table 2.6 shows large positive residuals for female Democrats and male Republicans, and large negative residuals for female Republicans and male Democrats. Thus, there were significantly more female Democrats and male Republicans and fewer female Republicans and male Democrats than the hypothesis of independence predicts. An odds ratio describes this evidence of a gender gap. The 2×2 table of Democrat and Republican identifiers has a sample odds ratio of $(279)(191)/(225)(165) = 1.44$. Of those subjects identifying with one of the two parties, the estimated odds of identifying with the Democrats rather than the Republicans were 44% higher for females than males.

For each party, Table 2.6 shows that there is only one nonredundant residual; the one for females is the negative of the one for males. The observed counts and the estimated expected frequencies have the same row and column totals. Thus, in a given column, if $n_{ij} > \hat{\mu}_{ij}$ in one cell, the reverse must happen in the other cell. The differences $n_{1j} - \hat{\mu}_{1j}$ and $n_{2j} - \hat{\mu}_{2j}$ have the same magnitude but different signs, implying the same pattern for their adjusted residuals.

2.4.6 Partitioning Chi-Squared

Chi-squared statistics have a reproductive property. If one chi-squared statistic has $df = df_1$ and a separate, independent, chi-squared statistic has $df = df_2$, then their sum is chi-squared with $df = df_1 + df_2$. For instance, if we had a table of form Table 2.5 for college-educated subjects and a separate one for subjects not having a college education, the sum of the X^2 values or the sum of the G^2 values from the two tables would be a chi-squared statistic with $df = 2 + 2 = 4$.

Similarly, chi-squared statistics having $df > 1$ can be broken into components with fewer degrees of freedom. For instance, a statistic having $df = 2$ can be partitioned into two independent components each having $df = 1$. Another supplement to a test of independence partitions its chi-squared test statistic so that the components represent certain aspects of the association. A partitioning may show that an association primarily reflects differences between certain categories or groupings of categories.

We illustrate with a partitioning of G^2 for testing independence in $2 \times J$ tables. The test statistic then has $df = (J - 1)$, and we partition it into $J - 1$ components. The jth component is G^2 for testing independence in a 2×2 table, where the first column combines columns 1 through j of the original table, and the second column uses column $j + 1$ of the original table. That is, G^2 for testing independence in a $2 \times J$ table equals the sum of a G^2 statistic that compares the first two columns, plus a G^2 statistic for the 2×2 table that combines the first two columns and compares them to the third column, and so on, up to a G^2 statistic for the 2×2 table that combines the first $J - 1$ columns and compares them to the last column. Each component G^2 statistic has $df = 1$.

Consider again Table 2.5. The first two columns of this table form a 2×2 table with cell counts, by row, of $(279, 73/165, 47)$. For this component table, $G^2 = 0.16$, with $df = 1$. Of those subjects who identify either as Democrats or Independents, there is little evidence of a difference between females and males in the relative numbers in the two categories. We form the second 2×2 table by combining these columns and comparing them to the Republican column, giving the table with rows $(279 + 73, 225/165 + 47, 191) = (352, 225/212, 191)$. This table has $G^2 = 6.84$, based on $df = 1$. There is strong evidence of a difference between females and males in the relative numbers identifying as Republican instead of Democrat or Independent. Note that $0.16 + 6.84 = 7.00$; that is, the sum of these G^2 components equals G^2 for the test of independence for the complete 2×3 table. This overall statistic primarily reflects differences between genders in choosing between Republicans and Democrats/Independents.

It might seem more natural to compute G^2 for separate 2×2 tables that pair each column with a particular one, say the last. Though this is a reasonable way to investigate association in many data sets, these component statistics are not independent and do not sum to G^2 for the complete table. Certain rules determine ways of forming tables so that chi-squared partitions, but they are beyond the scope of this text (see, e.g., Agresti (1990), p. 53, for rules and references). A necessary condition is that the G^2 values for the component tables sum to G^2 for the original table.

The G^2 statistic has exact partitionings. The overall Pearson X^2 statistic does not equal the sum of the X^2 values for the separate tables in a partition. However, it is valid to use the X^2 statistics for the separate tables in the partition; they simply do not provide an exact algebraic partitioning of the X^2 statistic for the overall table.

2.4.7 Comments on Chi-Squared Tests

Chi-squared tests of independence, like any significance tests, have serious limitations. They simply indicate the degree of evidence for an association. They are rarely adequate for answering all questions we have about a data set. Rather than relying solely on results of these tests, one should study the nature of the association. It is sensible to decompose chi-squared into components, study residuals, and estimate parameters such as odds ratios that describe the strength of association.

The X^2 and G^2 chi-squared tests also have limitations in the types of data sets for which they are applicable. For instance, they require large samples. The sampling distributions of X^2 and G^2 get closer to chi-squared as the sample size n increases, relative to the number of cells IJ. The convergence is quicker for X^2 than G^2. The chi-squared approximation is often poor for G^2 when $n/IJ < 5$. When I or J is large, it can be decent for X^2 when some expected frequencies are as small as 1. Section 7.4.3 provides further guidelines, but these are not crucial since small-sample procedures are available whenever we question whether n is sufficiently large. Section 2.6 discusses these.

The $\{\hat{\mu}_{ij} = n_{i+}n_{+j}/n\}$ used in X^2 and G^2 depend on the row and column marginal totals, but not on the order in which the rows and columns are listed. Thus, X^2 and G^2 do not change value with arbitrary reorderings of rows or of columns. This means that these tests treat both classifications as nominal. We ignore some information when we use them to test independence between ordinal classifications. When at least one variable is ordinal, more powerful tests of independence usually exist. The next section presents such a test.

2.5 TESTING INDEPENDENCE FOR ORDINAL DATA

The chi-squared test of independence using test statistic X^2 or G^2 treats both classifications as nominal. When the rows and/or the columns are ordinal, test statistics that utilize the ordinality are usually more appropriate.

2.5.1 Linear Trend Alternative to Independence

When the row variable X and the column variable Y are ordinal, a "trend" association is quite common. As the level of X increases, responses on Y tend to increase toward higher levels, or responses on Y tend to decrease toward lower levels. One can use a single parameter to describe such an ordinal trend association. The most common analysis assigns scores to categories and measures the degree of *linear trend* or correlation.

We next present a test statistic that is sensitive to positive or negative linear trends in the relationship between X and Y. It utilizes correlation information in the data. Let $u_1 \leq u_2 \leq \cdots \leq u_I$ denote scores for the rows, and let $v_1 \leq v_2 \leq \cdots \leq v_J$ denote scores for the columns. The scores have the same ordering as the category levels and are said to be *monotone*. The scores reflect distances between categories, with greater distances between categories treated as farther apart.

The sum $\sum_{i,j} u_i v_j n_{ij}$, which weights cross-products of scores by the frequency of their occurrence, relates to the covariation of X and Y. For the chosen scores, the Pearson product-moment correlation between X and Y equals the standardization of