2

Two-Dimensional Tables and Simple Logistic Regression

At this point, it is not our primary intention to provide a rigorous account of logistic regression and log-linear model theory. Such a treatment demands extensive use of advanced calculus and asymptotic theory. On the other hand, some knowledge of the basic issues is necessary for a correct understanding of *applications of logistic regression and log-linear models*. In this chapter, we address these basic issues for the simple case of two-dimensional tables and simple logistic regression. For a more elementary discussion of two-dimensional tables and simple logistic regression including substantial data analysis, see Christensen (1996a, Chapter 8). In fact, we assume that the reader is familiar with such analyses and use the topics in this chapter primarily to introduce key theoretical ideas.

2.1 Two Independent Binomials

Consider two binomials arranged in a 2×2 table. Our interest is in examining possible differences between the two binomials.

EXAMPLE 2.1.1. A survey was conducted to examine the relative attitudes of males and females about abortion. Of 500 females, 309 supported legalized abortion. Of 600 males, 319 supported legalized abortion. The data can be summarized in tabular form:

OBSERVED VALUES

	Support	Do Not Support	Total
Female	309	191	500
Male	319	281	600
Total	628	472	1,100

Note that the totals on the right-hand side of the table (500 and 600) are fixed by the design of the study. The totals along the bottom of the table are observed random variables. It is assumed that for each sex, the numbers of supporters and nonsupporters form a binomial random vector (ordered pair). We are interested in whether these numbers indicate that a person's sex affects their attitude toward legalized abortion. Note that the categories are Support and Do Not Support legalized abortion. Not supporting legalized abortion is distinct from opposing it. Anyone who is indifferent neither supports nor opposes legalized abortion.

We now introduce the notation that will be used for tables of counts in this book. For a 2×2 table, the observed values are denoted by n_{ij}, $i = 1, 2$ and $j = 1, 2$. Marginal totals are written $n_{i\cdot} \equiv n_{i1} + n_{i2}$ and $n_{\cdot j} \equiv n_{1j} + n_{2j}$. The total of all observations is $n_{\cdot\cdot} \equiv n_{11} + n_{12} + n_{21} + n_{22}$. The probability of having an observation fall in the ith row and jth column of the table is denoted p_{ij}. The number of observations that one would expect to see in the ith row and jth column (based on some statistical model) is denoted m_{ij}. For independent binomial rows, $m_{ij} = n_{i\cdot} p_{ij}$. Marginal totals $p_{i\cdot}$, $p_{\cdot j}$, $m_{i\cdot}$, and $m_{\cdot j}$ are defined like $n_{i\cdot}$ and $n_{\cdot j}$.

All of this notation can be summarized in tabular form.

OBSERVED VALUES

		Columns 1	2	Totals
Rows	1	n_{11}	n_{12}	$n_{1\cdot}$
	2	n_{21}	n_{22}	$n_{2\cdot}$
	Totals	$n_{\cdot 1}$	$n_{\cdot 2}$	$n_{\cdot\cdot}$

PROBABILITIES

		Columns 1	2	Totals
Rows	1	p_{11}	p_{12}	$p_{1\cdot}$
	2	p_{21}	p_{22}	$p_{2\cdot}$
	Totals	$p_{\cdot 1}$	$p_{\cdot 2}$	$p_{\cdot\cdot}$

EXPECTED VALUES

		Columns		
		1	2	Totals
Rows	1	m_{11}	m_{12}	$m_{1\cdot}$
	2	m_{21}	m_{22}	$m_{2\cdot}$
	Totals	$m_{\cdot 1}$	$m_{\cdot 2}$	$m_{\cdot\cdot}$

Our interest is in finding estimates of the p_{ij}'s, developing models for the p_{ij}'s, and performing tests on the p_{ij}'s. Equivalently, we can concern ourselves with estimates, models, and tests for the m_{ij}'s.

In Example 2.1.1, our interest is in whether sex is related to support for legalized abortion. Note that $p_{11}+p_{12}=1$ and $p_{21}+p_{22}=1$. (Equivalently $m_{11}+m_{12}=500$ and $m_{21}+m_{22}=600$.) If sex has no effect on opinion, the distribution of support versus nonsupport should be the same for both sexes. In particular, it is of interest to test the null hypothesis (model)

$$H_0 : p_{11} = p_{21} \text{ and } p_{12} = p_{22}\,.$$

With $p_{i1}+p_{i2} = 1$, the equality $p_{11} = p_{21}$ holds if and only if $p_{12} = p_{22}$ holds. In other words, females and males have the same probability of "support" if and only if they have the same probability for "do not support." It suffices to test that the probability of support is the same for both sexes, i.e.,

$$H_0 : p_{11} = p_{21}$$

or, equivalently,

$$H_0 : p_{11} - p_{21} = 0\,.$$

To test this hypothesis, we need an estimate of $p_{11}-p_{21}$ and the standard error (SE) of the estimate. Each row is binomial with sample size $n_{i\cdot}$, so a natural estimate of p_{ij} is the proportion of observations falling in cell ij relative to the total number of observations in the ith row, i.e.,

$$\hat{p}_{ij} = n_{ij}/n_{i\cdot}\,.$$

For the abortion example, $\hat{p}_{11} = 309/500$ and $\hat{p}_{21} = 319/600$. The estimate of $p_{11}-p_{21}$ is

$$\hat{p}_{11} - \hat{p}_{21} = (n_{11}/n_{1\cdot}) - (n_{21}/n_{2\cdot})\,.$$

The two rows of the table were sampled independently so the variance of $\hat{p}_{11} - \hat{p}_{21}$ is

$$\begin{aligned}\mathrm{Var}(\hat{p}_{11} - \hat{p}_{21}) &= \mathrm{Var}(\hat{p}_{11}) + \mathrm{Var}(\hat{p}_{21}) \\ &= p_{11}p_{12}/n_{1\cdot} + p_{21}p_{22}/n_{2\cdot}\,,\end{aligned}$$

cf. Exercise 1.6.7. Finally,

$$\mathrm{SE}(\hat{p}_{11} - \hat{p}_{21}) = \sqrt{\hat{p}_{11}\hat{p}_{12}/n_{1\cdot} + \hat{p}_{21}\hat{p}_{22}/n_{2\cdot}}\,.$$

For the abortion example,

$$\mathrm{SE}(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{(309/500)(191/500)}{500} + \frac{(319/600)(281/600)}{600}} = .0298\,.$$

One other thing is required before we can perform a test. We need to know the distribution of $[(\hat{p}_{11} - \hat{p}_{21}) - (p_{11} - p_{21})]/\mathrm{SE}(\hat{p}_{11} - \hat{p}_{21})$. By appealing to the Central Limit Theorem and the Law of Large Numbers (cf. Lindgren, 1993), if $n_{1\cdot}$ and $n_{2\cdot}$ are large, we can use the approximate distribution

$$\frac{(\hat{p}_{11} - \hat{p}_{21}) - (p_{11} - p_{21})}{\mathrm{SE}(\hat{p}_{11} - \hat{p}_{21})} \sim N(0,1)\,.$$

To perform a test of

$$H_0 : p_{11} - p_{21} = 0$$

versus

$$H_A : p_{11} - p_{21} \neq 0,$$

assume that H_0 is true and look for evidence against H_0. If H_0 is true, the approximate distribution is

$$\frac{(\hat{p}_{11} - \hat{p}_{21}) - 0}{\mathrm{SE}(\hat{p}_{11} - \hat{p}_{21})} \sim N(0,1)\,.$$

If the alternative hypothesis is true, $\hat{p}_{11} - \hat{p}_{21}$ still estimates $p_{11} - p_{21}$ so the test statistic

$$\frac{(\hat{p}_{11} - \hat{p}_{21}) - 0}{\mathrm{SE}(\hat{p}_{11} - \hat{p}_{21})}$$

tends to be either a large positive value if $p_{11} - p_{21} > 0$ or a large negative value if $p_{11} - p_{21} < 0$. An $\alpha = .05$ level test rejects H_0 if

$$\frac{(\hat{p}_{11} - \hat{p}_{21}) - 0}{\mathrm{SE}(\hat{p}_{11} - \hat{p}_{21})} > 1.96$$

or if

$$\frac{(\hat{p}_{11} - \hat{p}_{21}) - 0}{\mathrm{SE}(\hat{p}_{11} - \hat{p}_{21})} < -1.96\,.$$

The values -1.96 and 1.96 cut off the probability .025 from the bottom and top of a $N(0,1)$ distribution, respectively. Thus, the total probability of rejecting H_0 when H_0 is true is $.025 + .025 = .05$. Recall that this test is based on a large sample approximation to the distribution of the test statistic.

For the abortion example, the test statistic is

$$\frac{(309/500) - (319/600)}{.0298} = 2.90\,.$$

Because $2.90 > 1.96$, H_0 is rejected at the $\alpha = .05$ level. There is evidence of a relationship between sex and attitudes about legalized abortion. These data indicate that females are more likely to support legalized abortion.

Before leaving this test procedure, we mention an alternative method for computing $\mathrm{SE}(\hat{p}_{11} - \hat{p}_{21})$. Recall that

$$\mathrm{Var}(\hat{p}_{11} - \hat{p}_{21}) = p_{11}p_{12}/n_{1\cdot} + p_{21}p_{22}/n_{2\cdot}\,.$$

If H_0 is true, $p_{11} = p_{21}$ and $p_{12} = p_{22}$. These facts can be used in estimating the variance of $\hat{p}_{11} - \hat{p}_{21}$. A pooled estimate of $p \equiv p_{11} = p_{21}$ is $(n_{11} + n_{21})/(n_{1\cdot} + n_{2\cdot}) = n_{\cdot 1}/n_{\cdot\cdot} = 628/1100$. A pooled estimate of $(1 - p) \equiv p_{12} = p_{22}$ is $n_{\cdot 2}/n_{\cdot\cdot} = 472/1100$. This yields

$$\mathrm{Var}(\hat{p}_{11} - \hat{p}_{21}) = p(1 - p)(1/n_{1\cdot} + 1/n_{2\cdot})$$

and

$$\begin{aligned}\mathrm{SE}(\hat{p}_{11} - \hat{p}_{21}) &= \sqrt{(628/1100)(472/1100)[(1/500) + (1/600)]} \\ &= .0300\,.\end{aligned}$$

The test statistic computed with the new standard error is

$$\frac{(309/500) - (319/600)}{.0300} = 2.87777\,.$$

For these data, the results are essentially the same.

The test procedures discussed above work nicely for two independent binomials, but, unfortunately, they do not generalize to more than two binomials or to situations in which there are more than two possible outcomes (e.g., support, oppose, no opinion). An alternative test procedure is based on what is known as the *Pearson chi-square test statistic*. This test is equivalent to the test given above using the pooled estimate of the standard error. Moreover, Pearson's chi-square is applicable in more general problems. The Pearson test statistic is based on comparing the observed table values in the 2×2 table with estimates of the expected values that are obtained assuming that H_0 is true.

In the abortion example, if H_0 is true, then $p = p_{11} = p_{21}$ and $\hat{p} = 628/1100$. Similarly, $(1-p) = p_{12} = p_{22}$ and $(1-\hat{p}) = 472/1100$. As before, the expected values are $m_{ij} = n_{i\cdot}p_{ij}$. The estimated expected values under H_0 are $\hat{m}_{ij}^{(0)} = n_{i\cdot}\hat{p}_{ij}$, where $\hat{p}_{ij}$ is $\hat{p}$ if $j = 1$ and $(1 - \hat{p})$ if $j = 2$. More generally,

$$\hat{m}_{ij}^{(0)} = n_{i\cdot}(n_{\cdot j}/n_{\cdot\cdot})\,. \tag{1}$$

The Pearson chi-square statistic is defined as

$$X^2 = \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{\left(n_{ij} - \hat{m}_{ij}^{(0)}\right)^2}{\hat{m}_{ij}^{(0)}}\,.$$

If H_0 is true, then n_{ij} and $\hat{m}_{ij}^{(0)}$ should be near each other, and the terms $(n_{ij} - \hat{m}_{ij}^{(0)})^2$ should be reasonably small. If H_0 is not true, then the $\hat{m}_{ij}^{(0)}$'s, which are estimates based on the assumption that H_0 is true, should do a poor job of predicting the n_{ij}'s. The terms $(n_{ij} - \hat{m}_{ij}^{(0)})^2$ should be larger when H_0 is not true.

Note that a prediction $\hat{m}_{ij}^{(0)}$ that is, say, three away from the observed value n_{ij}, can be either a good prediction or a bad prediction depending on how large the value in the cell should be. If $n_{ij} = 4$ and $\hat{m}_{ij}^{(0)} = 1$, the prediction is poor. If $n_{ij} = 104$ and $\hat{m}_{ij}^{(0)} = 101$, the prediction is good. The $\hat{m}_{ij}^{(0)}$ in the denominator of each term of X^2 is a scale factor that corrects for this problem.

The hypothesis $H_0 : p_{11} = p_{21}$ and $p_{12} = p_{22}$ is rejected at the $\alpha = .05$ level if

$$X^2 \geq \chi^2(.95, 1)\,.$$

The test is based on the fact that if H_0 is true, then as $n_{1\cdot}$ and $n_{2\cdot}$ get large, X^2 has approximately a $\chi^2(1)$ distribution. This is a consequence of the Central Limit Theorem and the Law of Large Numbers, cf. Exercise 2.1.

For the abortion example

$\hat{m}_{ij}^{(0)}$	Support	Do Not Support	Totals
Female	285.5	214.5	500
Male	342.5	257.5	600
Totals	628	472	1100

$$X^2 = 8.2816,$$

$$\chi^2(.95, 1) = 3.84\,.$$

Because $8.2816 > 3.84$, the $\alpha = .05$ test rejects H_0.

Note that $8.2816 = (2.8777)^2$ and that $3.84 = (1.96)^2$. For 2×2 tables, the results of Pearson chi-square tests are exactly equivalent to the results of normal theory tests using the pooled estimate in the standard error. By definition, $\chi^2(1-\alpha, 1) = [z(1-\frac{\alpha}{2})]^2$ for $\alpha \in (0, .5]$. Also,

$$X^2 = \frac{(\hat{p}_{11} - \hat{p}_{21})^2}{\hat{p}(1-\hat{p})\left(\frac{1}{n_{1\cdot}} + \frac{1}{n_{2\cdot}}\right)}\,. \tag{2}$$

EXERCISE 2.1. Prove equation (2).

By comparing the n_{ij}'s to the $\hat{m}_{ij}^{(0)}$'s, we can examine the nature of the differences in the two binomials. One simple way to do this comparison is to examine a table of *residuals*, i.e., the $(n_{ij} - \hat{m}_{ij}^{(0)})$'s. In order to make accurate evaluations of how well $\hat{m}_{ij}^{(0)}$ is predicting n_{ij}, the residuals need to be rescaled or standardized. Define the *Pearson residuals* as

$$\tilde{r}_{ij} = \frac{n_{ij} - \hat{m}_{ij}^{(0)}}{\sqrt{\hat{m}_{ij}^{(0)}}},$$

$i = 1, 2$, $j = 1, 2$. Note that $X^2 = \sum_{ij} \tilde{r}_{ij}^2$. The Pearson residuals for the abortion data are

$\tilde{r}_{ij}$	Support	Do Not Support
Female	1.39	−1.60
Male	−1.27	1.46

The positive residual 1.39 indicates that more females support legalized abortion than would be expected under H_0. The negative residual −1.27 indicates that fewer males support abortion than would be expected under H_0. Together, the values 1.39 and −1.27 indicate that proportionately more females support legalized abortion than males. Equivalently, proportionately more males do not support legalized abortion than females.

2.1.1 The Odds Ratio

A commonly used technique in the analysis of count data is the examination of odds ratios. In the abortion example, the odds of females supporting legalized abortion is p_{11}/p_{12}. The odds of males supporting legalized abortion is p_{21}/p_{22}. The odds ratio is

$$\frac{(p_{11}/p_{12})}{(p_{21}/p_{22})} = \frac{p_{11}p_{22}}{p_{12}p_{21}}.$$

Note that if the two binomials are identical, then $p_{11} = p_{21}$ and $p_{12} = p_{22}$, so

$$\frac{p_{11}p_{22}}{p_{12}p_{21}} = 1.$$

An alternative to using Pearson's chi-square for examining whether two binomials are the same is to examine the estimated odds ratio. Using $\hat{p}_{ij} = n_{ij}/n_{i\cdot}$ gives

$$\frac{\hat{p}_{11}\hat{p}_{22}}{\hat{p}_{12}\hat{p}_{21}} = \frac{n_{11}n_{22}}{n_{12}n_{21}}.$$

For the abortion example, the estimate is

$$\frac{(309)(281)}{(191)(319)} = 1.425.$$

This is only an estimate of the population odds ratio, but it is fairly far from the target value of 1. In particular, we have estimated that the odds of a female supporting legalized abortion are about one and a half times as large as the odds of a male supporting legalized abortion.

We may wish to test the hypothesis that the odds ratio equals 1. Equivalently, we can test whether the log of the odds ratio equals 0. The log odds ratio is $\log(1.425) = .354$. The asymptotic (large sample) standard error of the log odds ratio is

$$\begin{aligned}\sqrt{\frac{1}{n_{11}}+\frac{1}{n_{12}}+\frac{1}{n_{21}}+\frac{1}{n_{22}}} &= \sqrt{\frac{1}{309}+\frac{1}{191}+\frac{1}{319}+\frac{1}{281}} \\ &= .123\,.\end{aligned}$$

The estimate minus the hypothesized value over the standard error is

$$\frac{.354-0}{.123} = 2.88\,.$$

Comparing this to a $N(0,1)$ distribution indicates that the log-odds ratio is greater than zero, thus the odds ratio is greater than 1. Note that this test is not equivalent to the other tests considered, even though the numerical value of the test statistic is similar to the other normal theory tests.

A 95% confidence interval for the log odds ratio has the end points $.354 \pm 1.96(.123)$. This gives an interval $(.113, .595)$. The log odds ratio is in the interval $(.113, .595)$ if and only if the odds ratio is in the interval $(e^{.113}, e^{.595})$. Thus, a 95% confidence interval for the odds ratio is $(e^{.113}, e^{.595})$, which simplifies to $(1.12, 1.81)$. We are 95% confident that the true odds of women supporting legalized abortion is between 1.12 and 1.81 times greater than the odds of men supporting legalized abortion.

2.2 Testing Independence in a 2×2 Table

In Section 1, we obtained a 2×2 table by looking at two populations, each divided into two categories. In this section, we consider only one population but divide it into two categories in each of two different ways. The two different ways of dividing the population will be referred to as factors.

In Section 1, we examined differences between the two populations. In this section, we examine the nature of the one population being sampled. In particular, we examine whether the two factors affect the population independently or whether they interact to determine the nature of the population.

EXAMPLE 2.2.1. As part of a longitudinal study, a sample of 3182 people without cardiovascular disease were cross-classified by two factors: personality type and exercise. Personality type was categorized as type A or type B. Type A persons show signs of stress, uneasiness, and hyperactivity. Type

B persons are relaxed, easygoing, and normally active. Exercise is categorized as persons who exercise regularly and those who do not. The data are given in the following table:

		Personality		
	n_{ij}	A	B	Totals
Exercise	Regular	483	477	960
	Other	1101	1121	2222
	Totals	1584	1598	3182

Although notations for observations (n_{ij}'s), probabilities (p_{ij}'s), and expected values (m_{ij}'s) are identical to those in Section 1, the meaning of these quantities has changed. In Section 1, the rows were two independent binomials, so $p_{11}+p_{12} = 1 = p_{21}+p_{22}$. In this section, there is only one population, so the constraint on the probabilities is that $p_{11}+p_{12}+p_{21}+p_{22} = 1$.

In this section, our primary interest is in determining whether the row factor is independent of the column factor and if not, how the factors deviate from independence. The probability of an observation falling in the ith row and jth column of the table is p_{ij}. The probability of an observation falling in the ith row is $p_{i\cdot}$. The probability of the jth column is $p_{\cdot j}$. Rows and columns are independent if and only if for all i and j

$$p_{ij} = p_{i\cdot}p_{\cdot j} \, . \tag{1}$$

The sample size is $n_{\cdot\cdot}$, so the expected counts in the table are

$$m_{ij} = n_{\cdot\cdot}p_{ij} \, .$$

If rows and columns are independent, this becomes

$$m_{ij} = n_{\cdot\cdot}p_{i\cdot}p_{\cdot j} \, . \tag{2}$$

It is easily seen that condition (1) for independence is equivalent to

$$m_{ij} = m_{i\cdot}m_{\cdot j}/n_{\cdot\cdot} \, . \tag{3}$$

Pearson's chi-square can be used to test independence. Pearson's statistic is again

$$X^2 = \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{\left(n_{ij} - \hat{m}_{ij}^{(0)}\right)^2}{\hat{m}_{ij}^{(0)}}$$

where $\hat{m}_{ij}^{(0)}$ is an estimate of m_{ij} based on the assumption that rows and columns are independent. If we take $\hat{m}_{i\cdot} = n_{i\cdot}$ and $\hat{m}_{\cdot j} = n_{\cdot j}$, then equation (3) leads to

$$\hat{m}_{ij}^{(0)} = n_{i\cdot}n_{\cdot j}/n_{\cdot\cdot} \, . \tag{4}$$

Equation (4) can also be arrived at via equation (2). An obvious estimate of $p_{i\cdot}$ is

$$\hat{p}_{i\cdot} = n_{i\cdot}/n_{\cdot\cdot} \,.$$

Similarly,

$$\hat{p}_{\cdot j} = n_{\cdot j}/n_{\cdot\cdot} \,.$$

Substitution into equation (2) leads to equation (4). It is interesting to note that equation (4) is numerically identical to formula (2.1.1), which gives $\hat{m}_{ij}$ for two independent binomials. Just as in Section 1, for the purposes of testing, X^2 is compared to a $\chi^2(1)$ distribution. Pearson residuals are again defined as

$$\tilde{r}_{ij} = \frac{n_{ij} - \hat{m}_{ij}^{(0)}}{\sqrt{\hat{m}_{ij}^{(0)}}} \,.$$

For the personality-exercise data, we get

	$\hat{m}_{ij}^{(0)}$	Personality A	Personality B	Totals
Exercise	Regular	477.9	482.1	960
	Other	1106.1	1115.9	2222
	Totals	1584	1598	3182

$$X^2 = .156$$

The test is not significant for any reasonable α level. (The *P value* is greater than .5.) There is no significant evidence against independence of exercise and personality type. In other words, the data are consistent with the interpretation that knowledge of personality type gives no new information about exercise habits or, equivalently, knowledge of exercise habits gives no new information about personality type.

2.2.1 The Odds Ratio

Just as in examining the equality of two binomials, the odds ratio can be used to examine the independence of two factors in a multinomial sample. In the personality-exercise data, the odds that a person exercises regularly are $p_{1\cdot}/p_{2\cdot}$. In addition, the odds of exercising regularly can be examined separately for each personality type. For type A personalities, the odds are p_{11}/p_{21}, and for type B personalities, the odds are p_{12}/p_{22}. Intuitively, if exercise and personality types are independent, then the odds of regular exercise should be the same for both personality types. In particular, the ratio of the two sets of odds should be one.

Proposition 2.2.2. If rows and columns are independent, then the odds ratio

$$\frac{(p_{11}/p_{21})}{(p_{12}/p_{22})} = \frac{p_{11}p_{22}}{p_{12}p_{21}}$$

equals one.

Proof. By equation (1)

$$\frac{p_{11}p_{22}}{p_{12}p_{21}} = \frac{p_{1\cdot}p_{\cdot 1}p_{2\cdot}p_{\cdot 2}}{p_{1\cdot}p_{\cdot 2}p_{2\cdot}p_{\cdot 1}} = 1.$$

□

If the odds ratio is estimated under the assumption of independence, $\hat{p}_{ij} = \hat{p}_{i\cdot}\hat{p}_{\cdot j} = n_{i\cdot}n_{\cdot j}/(n_{\cdot\cdot})^2$; so the estimated odds ratio is always one. A more interesting approach is to estimate the odds ratio without assuming independence and then see how close the estimated odds ratio is to one. With this approach, $\hat{p}_{ij} = n_{ij}/n_{\cdot\cdot}$ and

$$\frac{\hat{p}_{11}\hat{p}_{22}}{\hat{p}_{12}\hat{p}_{22}} = \frac{n_{11}n_{22}}{n_{12}n_{21}}\,.$$

In the personality-exercise example, the estimated odds ratio is

$$\frac{(483)(1121)}{(477)(1101)} = 1.03$$

which is very close to one. The log odds are .0305, the asymptotic standard error is $[1/483+1/477+1/1101+1/1121]^{1/2} = .0772$, and the test statistic for H_0 that the log odds equal 0 is $.0305/.0772 = .395$. Again, there is no evidence against independence.

EXERCISE 2.2. Give a 95% confidence interval for the odds ratio. Explain what the confidence interval means.

2.3 $I \times J$ Tables

The situation examined in Section 1 can be generalized to consider samples from I different populations, each of which is divided into J categories. We assume that the samples from different populations are independent and that each sample follows a multinomial distribution. This is *product-multinomial sampling.*

Similarly, a sample from one population that is categorized by two factors can be generalized beyond the case considered in Section 2. We allow one factor to have I categories and the other factor to have J categories. Between the two factors, the population is divided into a total of IJ categories. The distribution of counts within the IJ categories is assumed to have a multinomial distribution. Consequently, this sampling scheme is multinomial sampling.

An $I \times J$ table of the observations n_{ij}, $i = 1, \ldots, I$, $j = 1, \ldots, J$, can be written

	n_{ij}	Factor 2 (Categories) 1	2	$\ldots$	J	Totals
	1	n_{11}	n_{12}	$\ldots$	n_{1J}	$n_{1\cdot}$
Factor 1	2	n_{21}	n_{22}	$\ldots$	n_{2J}	$n_{2\cdot}$
(Populations)	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
	I	n_{I1}	n_{I2}	$\ldots$	n_{IJ}	$n_{I\cdot}$
	Totals	$n_{\cdot 1}$	$n_{\cdot 2}$	$\ldots$	$n_{\cdot J}$	$n_{\cdot\cdot}$

with similar tables for the probabilities p_{ij} and the expected values m_{ij}.

The analysis of product-multinomial sampling begins by testing whether all of the I multinomial populations are identical. In other words, we wish to test the model

$$H_0: \; p_{1j} = p_{2j} = \cdots = p_{Ij} \text{ for all } j = 1, \ldots, J\,. \tag{1}$$

against the alternative

$$H_A: \; \text{model (1) is not true}\,.$$

This is described as testing for *homogeneity of proportions.*

We continue to use Pearson's chi-square test statistic to evaluate the appropriateness of the null hypothesis model. Pearson's chi-square requires estimates of the expected values m_{ij}. Each sample i has a multinomial distribution with $n_{i\cdot}$ trials, so

$$m_{ij} = n_{i\cdot} p_{ij}\,.$$

If H_0 is true, p_{ij} is the same for all values of i. A pooled estimate of the common value of the p_{ij}'s is

$$\hat{p}_{ij}^{(0)} = n_{\cdot j}/n_{\cdot\cdot}\,.$$

In other words, if all the populations have the same probability for category j, an estimate of this common probability is the total number of observations in category j divided by the overall total number of observations. From this probability estimate we obtain

$$\hat{m}_{ij}^{(0)} = n_{i\cdot}(n_{\cdot j}/n_{\cdot\cdot})\,.$$

In both $\hat{p}_{ij}^{(0)}$ and $\hat{m}_{ij}^{(0)}$, the superscript (0) is used to indicate that the estimate was obtained under the assumption that H_0 was true. Pearson's chi-square test statistic is

$$X^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{\left(n_{ij} - \hat{m}_{ij}^{(0)}\right)^2}{\hat{m}_{ij}^{(0)}} .$$

For large samples, if H_0 is true, the approximation

$$X^2 \sim \chi^2((I-1)(J-1))$$

is valid. H_0 is rejected in an α level test if

$$X^2 > \chi^2(1-\alpha, (I-1)(J-1)) .$$

Note that if $I = J = 2$, these are precisely the results discussed in Section 1.

The analysis of a multinomial sample begins by testing for independence of the two factors. In particular, we wish to test the model

$$H_0 : p_{ij} = p_{i\cdot}p_{\cdot j}, \quad i = 1, \ldots, I, \ j = 1, \ldots, J . \tag{2}$$

We again use Pearson's chi-square. The marginal probabilities are estimated as

$$\hat{p}_{i\cdot} = n_{i\cdot}/n_{\cdot\cdot}$$

and

$$\hat{p}_{\cdot j} = n_{\cdot j}/n_{\cdot\cdot} .$$

Because $m_{ij} = n_{\cdot\cdot}p_{ij}$, if the model in (2) is true, we can estimate m_{ij} with

$$\begin{aligned} \hat{m}_{ij}^{(0)} &= n_{\cdot\cdot}\hat{p}_{i\cdot}\hat{p}_{\cdot j} \\ &= n_{\cdot\cdot}(n_{i\cdot}/n_{\cdot\cdot})(n_{\cdot j}/n_{\cdot\cdot}) \\ &= n_{i\cdot}n_{\cdot j}/n_{\cdot\cdot} \end{aligned}$$

where the (0) in $\hat{m}_{ij}^{(0)}$ indicates that the estimate is obtained assuming that (2) holds. The Pearson chi-square test statistic is

$$X^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{\left(n_{ij} - \hat{m}_{ij}^{(0)}\right)^2}{\hat{m}_{ij}^{(0)}}$$

which, if (2) is true and the sample size is large, is approximately distributed as a $\chi^2\left((I-1)(J-1)\right)$. H_0 is rejected at the α level if

$$X^2 > \chi^2\left(1-\alpha, (I-1)(J-1)\right) .$$

Once again, if $I = J = 2$, we obtain the previous results given for 2×2 tables. Moreover, the test procedures for product-multinomial sampling and for multinomial sampling are numerically identical. Only the interpretations of the tests differ.

EXAMPLE 2.3.1. Fifty-two males between the ages of 11 and 30 were operated on for knee injuries using arthroscopic surgery. The patients were classified by type of injury: twisted knee, direct blow, or both. The results of the surgery were classified as excellent (E), good (G), and fair or poor (F-P). These data can reasonably be viewed as either multinomial or product-multinomial. As a multinomial, we have 52 people cross-classified by type of injury and result of surgery. However, we can also think of the three types of injuries as defining different populations. Each person sampled from a population is given arthroscopic surgery and then the result is classified. Because our primary interest is in the result of surgery, we choose to *think* of the sampling as product-multinomial. The form of the analysis is identical for both sampling schemes. The data are

		Result			
	n_{ij}	E	G	F-P	Totals
	Twist	21	11	4	36
Injury	Direct	3	2	2	7
	Both	7	1	1	9
	Totals	31	14	7	52

The estimated expected counts under H_0 are

		Result			
	$\hat{m}_{ij}^{(0)}$	E	G	F-P	Totals
	Twist	21.5	9.7	4.8	36
Injury	Direct	4.2	1.9	.9	7
	Both	5.4	2.4	1.2	9
	Totals	31	14	7	52

with

$$X^2 = 3.229$$

and

$$df = (3-1)(3-1) = 4\,.$$

If the sample size is large, X^2 can be compared to a χ^2 distribution with four degrees of freedom. If we do this, the *P value* for the test is .52. Unfortunately, it is quite obvious that the sample size is not large. The number of observations in many of the cells of the table is small. This is a serious problem and aspects of the problem are discussed in Section 4, the subsection of Section 3.5 on Other Sampling Models, and Chapter 8. However, to the extent that this book focuses on distribution theory, it focuses on asymptotic distributions. For now, we merely state that in this example, the n_{ij}'s and $\hat{m}_{ij}^{(0)}$'s are such that, when taken together with the very large *P value*, we feel safe in concluding that these data provide no evidence of different surgical results for the three types of injuries. (This conclusion is borne out by the fact that an exact small sample test yields a similar *P value*, cf. Section 3.5.)

2.3.1 Response Factors

In Example 2.3.1, the result of surgery can be thought of as a response, whereas the type of injury is used to explain the response. Similarly, in Example 2.1.1, opinions on abortions can be considered as a response and sex can be considered as an explanatory factor.

The existence of response factors is often closely tied to the sampling scheme. Product-multinomial sampling is commonly used with an independent multinomial sample taken for every combination of the explanatory factors and the categories of the multinomials being the categories of the response factors. This is illustrated in Example 2.1.1 where there are two independent multinomials (binomials), one for males and one for females. The categories for each multinomial are Support and Do Not Support legalized abortion. Example 3.5.2 in the next chapter involves two explanatory factors, Sex and Socioeconomic Status, and one response factor, Opinion on Legalized Abortion. Each of the four combinations obtained from the two sexes and the two statuses define an independent multinomial. In other words, there is a separate multinomial sample for each combination of sex and socioeconomic status. The categories of the response factor, Support and Do Not Support legalized abortion, are the categories of the multinomials.

More generally, the categories of a response factor can be cross-classified with other response factors or explanatory factors to yield the categories in a series of independent multinomials. This situation is of most interest when there are several factors involved. Some factors can be cross-classified to define the multinomial populations while other factors can be cross-classified with the response factors to define the categories of the multinomials. Example 2.3.1 illustrates the simplest case in which there is one explanatory factor, Injury, crossed with one response factor, Result, to define the categories of the multinomial. Both Injury and Result have three levels so the multinomial has a total of nine categories. With only two factors in the table, there can be only one multinomial sample because there are no other factors available to define various multinomial populations. Example 3.5.1 is more general in that it has two independent multinomials, one for each sex. Each multinomial has six categories. The categories are obtained by cross-classifying the explanatory factor, Political Party, having three levels, with the response factor, Abortion Opinion, having two levels.

In this more general sampling scheme, one often conditions on all factors other than the response so that the analysis is reduced to that of the original sampling scheme in which every combination of explanatory factors defines an independent multinomial. Again, this is illustrated in Example 2.3.1. While the sampling was multinomial, we treated the sampling as product-multinomial with an independent multinomial for each level of the Injury. The justification for treating the data as product-multinomial is that we conditioned on the Injury.

While the sampling techniques described above are probably the most commonly used, there are alternatives that are also commonly used. For example, in medicine a response factor is often some disease with levels that are various states of the disease. If the disease is at all rare, it may be impractical to sample different populations and see how many people fall into the various levels of the disease. In this case, one may need to take the disease levels as populations, sample from these populations, and investigate various characteristics of the populations. Thus the "explanatory" factors discussed above would be considered descriptive factors here. This sampling scheme is often called *retrospective* for obvious reasons. The other schemes discussed above are called *prospective*. These issues are discussed in more detail in the introduction to Chapter 4 and in Sections 4.7 and 11.7.

2.3.2 Odds Ratios

The null hypotheses (1) and (2) can be rewritten in terms of odds ratios.

Proposition 2.3.2. Under product-multinomial sampling $p_{1j} = \cdots = p_{Ij} > 0$ for all $j = 1, \ldots, J$ if and only if

$$\frac{p_{ij}p_{i'j'}}{p_{ij'}p_{i'j}} = 1$$

for all i, $i' = 1, \ldots, I$ and j, $j' = 1, \ldots, J$.

Proof. a) *Equality of probabilities across rows implies that the odds ratios equal one.* By substitution,

$$\frac{p_{ij}p_{i'j'}}{p_{ij'}p_{i'j}} = \frac{p_{ij}p_{ij'}}{p_{ij'}p_{ij}} = 1.$$

b) *All odds ratios equal to one implies equality of probabilities across rows.* Recall that $p_{i\cdot} = 1$ for all $i = 1, \ldots, I$, so that $p_{\cdot\cdot} = I$. In addition, $p_{ij}p_{i'j'}/p_{ij'}p_{i'j} = 1$ implies $p_{ij}p_{i'j'} = p_{ij'}p_{i'j}$. Note that

$$p_{ij} = p_{ij}p_{\cdot\cdot}/I \quad = \quad \frac{1}{I}\sum_{i'=1}^{I}\sum_{j'=1}^{J} p_{ij}p_{i'j'}$$

$$
\begin{aligned}
&= \frac{1}{I}\sum_{i'=1}^{I}\sum_{j'=1}^{J} p_{ij'}p_{i'j} \\
&= \frac{1}{I}\sum_{j'=1}^{J} p_{ij'}\sum_{i'=1}^{I} p_{i'j} \\
&= \frac{1}{I}\sum_{j'=1}^{J} p_{ij'}p_{\cdot j} \\
&= \frac{1}{I}p_{\cdot j}\sum_{j'=1}^{J} p_{ij'} \\
&= \frac{1}{I}p_{\cdot j}p_{i\cdot} \\
&= p_{\cdot j}/I\,.
\end{aligned}
$$

Because this holds for any i and j, $p_{\cdot j}/I = p_{1j} = p_{2j} = \cdots = p_{Ij}$ for $j = 1, \ldots, J$. □

Proposition 2.3.3. Under multinomial sampling, $0 < p_{ij} = p_{i\cdot}p_{\cdot j}$ for all $i = 1, \ldots, I$, $j = 1, \ldots, J$ if and only if

$$\frac{p_{ij}p_{i'j'}}{p_{ij'}p_{i'j}} = 1$$

for all $i, i' = 1, \ldots, I$ and $j, j' = 1, \ldots, J$.

Proof. a) *Independence implies that the odds ratios equal one.*

$$\frac{p_{ij}p_{i'j'}}{p_{ij'}p_{i'j}} = \frac{p_{i\cdot}p_{\cdot j}p_{i'\cdot}p_{\cdot j'}}{p_{i\cdot}p_{\cdot j'}p_{i'\cdot}p_{\cdot j}} = 1.$$

b) *All odds ratios equal to one implies independence.* If $p_{ij}p_{i'j'}/p_{ij'}p_{i'j} = 1$ for all i, i', j, and j', then $p_{ij}p_{i'j'} = p_{ij'}p_{i'j}$. Moreover, because $p_{\cdot\cdot} = 1$,

$$
\begin{aligned}
p_{ij} = p_{ij}p_{\cdot\cdot} = \sum_{i'=1}^{I}\sum_{j'=1}^{J} p_{ij}p_{i'j'} &= \sum_{i'=1}^{I}\sum_{j'=1}^{J} p_{ij'}p_{i'j} \\
&= \sum_{i'=1}^{I} p_{i'j}\sum_{j'=1}^{J} p_{ij'} \\
&= \sum_{i'=1}^{I} p_{i'j}p_{i\cdot} \\
&= p_{i\cdot}\sum_{i'=1}^{I} p_{i'j}
\end{aligned}
$$

$$= \quad p_{i\cdot}p_{\cdot j} \qquad \square$$

There is a great deal of redundancy in specifying that

$$\frac{p_{ij}p_{i'j'}}{p_{ij'}p_{i'j}} = 1$$

for all i, i', j, and j'. For example, if $i = i'$, then $p_{ij}p_{i'j'}/p_{ij'}p_{i'j} = p_{ij}p_{ij'}/p_{ij'}p_{ij} = 1$ and no real restriction has been placed on the p_{ij}'s. A similar result occurs if $j = j'$. More significantly, if

$$p_{12}p_{23}/p_{13}p_{22} = 1$$

and

$$p_{12}p_{24}/p_{14}p_{22} = 1,$$

then

$$\begin{aligned} 1 &= (p_{12}p_{23}/p_{13}p_{22})(p_{14}p_{22}/p_{12}p_{24}) \\ &= p_{14}p_{23}/p_{13}p_{24} \,. \end{aligned}$$

In other words, the fact that two of the odds ratios equal one implies that a third odds ratio equals one. It turns out that the condition

$$\frac{p_{11}p_{ij}}{p_{1j}p_{i1}} = 1$$

for $i = 2, \ldots, I$ and $j = 2, \ldots, J$ is equivalent to the condition that all odds ratios equal one.

Proposition 2.3.4. $p_{ij}p_{i'j'}/p_{ij'}p_{i'j} = 1$ for all i, i', j and j' if and only if $p_{11}p_{ij}/p_{1j}p_{i1} = 1$ for all $i \neq 1$, $j \neq 1$.

Proof. Clearly, if the odds ratios are one for all i, i', j, and j', then $p_{11}p_{ij}/p_{1j}p_{i1} = 1$ for all i and j. Conversely,

$$\begin{aligned} 1 &= \left(\frac{p_{11}p_{ij}}{p_{1j}p_{i1}}\right)\left(\frac{p_{11}p_{i'j'}}{p_{1j'}p_{i'1}}\right) \Big/ \left(\frac{p_{11}p_{ij'}}{p_{1j'}p_{i1}}\right)\left(\frac{p_{11}p_{i'j}}{p_{1j}p_{i'1}}\right) \\ &= \frac{p_{ij}p_{i'j'}}{p_{ij'}p_{i'j}} \end{aligned}$$

$\square$

Of course, in practice the p_{ij}'s are never known. We can investigate independence by examining the estimated odds ratios

$$\hat{p}_{ij}\hat{p}_{i'j'}/\hat{p}_{ij'}\hat{p}_{i'j} = n_{ij}n_{i'j'}/n_{ij'}n_{i'j}$$

or, equivalently, we can look at the log of this. For large samples, the log of the estimated odds ratio is normally distributed with standard error

$$\mathrm{SE} = \sqrt{\frac{1}{n_{ij}} + \frac{1}{n_{ij'}} + \frac{1}{n_{i'j}} + \frac{1}{n_{i'j'}}}\,.$$

This allows the construction of asymptotic tests and confidence intervals for the log odds ratio. Of particular interest is the hypothesis

$$H_0 : p_{ij}p_{i'j'}/p_{ij'}p_{i'j} = 1\,.$$

After taking logs, this becomes

$$H_0 : \log(p_{ij}p_{i'j'}/p_{ij'}p_{i'j}) = 0\,.$$

EXAMPLE 2.3.5. We continue with the knee injury data of Example 2.3.1. From Proposition 2.3.4, it is sufficient to examine

$$\begin{aligned}
\frac{n_{11}n_{22}}{n_{12}n_{21}} = 21(2)/11(3) &= 1.27\,,\\
\frac{n_{11}n_{23}}{n_{13}n_{21}} = 21(2)/4(3) &= 3.5\,,\\
\frac{n_{11}n_{32}}{n_{12}n_{31}} = 21(1)/11(7) &= .27\,,\\
\frac{n_{11}n_{33}}{n_{13}n_{31}} = 21(1)/4(7) &= .75\,.
\end{aligned}$$

Although the X^2 test indicated no difference in the populations (populations = injury types), *at least* two of these estimated odds ratios *seem* substantially different from 1. In particular, relative to having an F-P result, the odds of an excellent result are about 3.5 times larger with twist injuries than with direct blows. Also, relative to having a good result, the odds of an excellent result from a twisted knee are only .27 of the odds of an excellent result with both types of injury. These numbers seem quite substantial, but they are difficult to evaluate without some idea of the error to which the estimates are subject. To this end, we use the large sample standard errors for the log odds ratios. Testing whether the log odds ratios are different from zero, we get

odds ratio	log (odds ratio)	SE	z
1.27	0.2412	0.9858	0.24
3.5	1.2528	1.0635	1.18
.27	−1.2993	1.1320	−1.15
.75	−0.2877	1.2002	−0.24

The large standard errors and small z values are consistent with the result of the X^2 test. None of the odds ratios appear to be substantially different from 1. Of course, it should not be overlooked that the standard errors are really only valid for large samples and we do not have large samples. Thus, all of our conclusions about the individual odds ratios must remain tentative.

2.4 Maximum Likelihood Theory for Two-Dimensional Tables

In this section, we introduce the *likelihood function*, *maximum likelihood estimates*, and *(generalized) likelihood ratio tests.* A valuable result for maximum likelihood estimation is given below without proof.

Lemma 2.4.1. Let $f(p_1, \ldots, p_r) = \sum_{i=1}^r n_i \log p_i$. If $n_i > 0$ for $i = 1, \ldots, r$, then, subject to the conditions $0 < p_i < 1$ and $p. = 1$, the maximum of $f(p_1, \ldots, p_r)$ is achieved at the point $(p_1, \ldots, p_r) = (\hat{p}_1, \ldots, \hat{p}_r)$ where $\hat{p}_i = n_i / n.$.

In this section, we consider product-multinomial sampling of I populations, with each population divided into the same J categories. The I populations will form the rows of an $I \times J$ table. No results will be presented for multinomial sampling in an $I \times J$ table. The derivations of such results are similar to those presented here and are left as an exercise.

The probability of obtaining the data $n_{i1}, \ldots, n_{iJ}$ from the ith multinomial sample is

$$\frac{n_{i\cdot}!}{\prod_{j=1}^J n_{ij}!} \prod_{j=1}^J p_{ij}^{n_{ij}} .$$

Because the I multinomials are independent, the probability of obtaining all of the values n_{ij}, $i = 1, \ldots, I$, $j = 1, \ldots, J$, is

$$\prod_{i=1}^I \left[\frac{n_{i\cdot}!}{\prod_{j=1}^J n_{ij}!} \prod_{j=1}^J p_{ij}^{n_{ij}} \right] . \tag{1}$$

Thus, if we know the p_{ij}'s, we can find the probability of obtaining any set of n_{ij}'s. In point of fact, we are in precisely the opposite position. We do not know the p_{ij}'s, but we do know the n_{ij}'s. The n_{ij}'s have been observed. If we think of (1) as a function of the p_{ij}'s, we can write

$$L(p) = \prod_{i=1}^I \left[\frac{n_{i\cdot}!}{\prod_{j=1}^J n_{ij}!} \prod_{J=1}^J p_{ij}^{n_{ij}} \right] \tag{2}$$

where $p = (p_{11}, p_{12}, \ldots, p_{IJ})$. $L(p)$ is called the likelihood function for p. Some values of p give a very small probability of observing the n_{ij}'s that