Revista Mexicana de Análisis de la Conducta
Núm. Monográfico, Vol. 17, No. 3, 1991, pp. 65-83

From Lags to Logs : Advances in Sequential Analysis

De Lags a Logs : Avances en Análisis Secuencial

Roger Bakeman
Georgia State University

RESUMEN

El análisis secuencial no es un tópico estadístico unificado y simple como el análisis de variancia o la regresión múltiple. Es la aplicación de diversas técnicas a los datos secuenciales de tipo categórico. Los primeros artículos sobre el tema daban especial importancia a la prueba binomial *z* y al análisis lag-secuencial. Los artículos posteriores contribuyeron a la formulación del puntaje *z* corregido y al surgimiento del análisis log-lineal. A pesar de su aparente precisión, el análisis log-lineal rara vez ha sido utilizado para el análisis secuencial. Este trabajo constituye una breve introducción al análisis log-lineal, y enfatiza la forma en que las técnicas de log-lineal se pueden utilizar para el análisis secuencial. Como conclusión, se señala la disponibilidad de un formato estándar para los datos secuenciales y la utilidad de un programa de computadora genérico que pueda reducir los datos secuenciales a tablas de contingencia para el análisis log-lineal.
Palabras clave: Análisis secuencial, análisis lag-secuencial, análisis log-lineal.

ABSTRACT

Sequential analysis is not a single unified statistical topic, like the analysis of variance or multiple regression. Instead it is the application of a number of existing techniques to sequential categorical data. Early articles stressed the binomial test z score and lag-sequential analysis; subsequent articles provided the correct z score formula and suggested log-lineal analysis. But in spite of its apparent appropriateness, log-linear analysis has seldom been used for sequential analysis. The presente paper constitutes a brief introduction to log-linear analysis and emphasizes how log-linear techniques can be used for sequential analysis. It concludes by noting the availability of a standard format for sequential data and the usefulness of a general purpose computer program that would reduce sequential data into contingency tables suitable for log-linear analysis.
Key words: sequential analysis, lag-sequential analysis, log-sequential analysis.

FROM LAGS TO LOGS: ADVANCES IN SEQUENTIAL ANALYSIS

Sequential analysis (Bakeman & Gottman, 1986; Bakeman & Gottman, 1987; Sackett, 1987; Anguera, 1991) is not a single unified statistical topic, like the analysis of variance or multiple regression. Instead it is the application of a number of existing techniques to sequential categorical data. Nonetheless our understanding of which techniques are relevant and how they should be applied has developed dramatically during the past 20 years.

Early articles stressed the binomial test *z* score and lag-sequential analysis (e.g., Bakeman, 1978; Bakeman & Dabbs, 1976; Sackett, 1979). Allison and Liker (1982) agreed that a *z* score could be used to test sequential dependency, but noted that the binomial computation recommended in the early articles was technically incorrect and that a slightly different computation should be used instead (because expected cell frequencies were estimated from the marginals, not determined by theory; see Bakeman & Gottman, 1986). They also suggested that log-linear modeling might be used but did not develop that idea in any detail.

Observational data of the sort appropriate for sequential analyses can almost always be presented in the form of multidimensional contingency tables. Castellan (1979) may have been the first to point this out, and Allison and Liker (1982) explicitly suggested log-linear analyses in their much-cited paper. Yet for much of the 1980's researchers interested in sequential analysis rarely used log-linear analyses (exceptions are Cohn & Tronick, 1987; and Stevenson, Ver Hoeve, Roach, & Leavitt, 1986), even though such analyses are probably the most appropriate ones for sequential data.

The present paper constitutes a brief introduction to log-linear analysis (see also Bakeman, Adamson, & Strisik, 1989; Bakeman, 1991) and emphasizes how log-linear techniques can be used for sequential analysis. Given similar data and questions, a log-linear and a lag-sequential analysis yield similar results, as M.T. Anguera, A. Blanco, and I intend to demonstrate in a forthcoming paper. Log-linear techniques, however, can answer questions that lag-sequential analysis cannot and, moreover, are much more established statistically (the standard reference is Bishop, Fienberg, & Holland, 1975; see also Fienberg, 1980; useful introductions are provided by Kennedy, 1983; Knoke & Burke, 1980; and Chapter 7 in Tabachnick & Fidell, 1989).

When learning about a new topic, it is often helpful to begin with what we know. Most readers of this paper understand how to analyze two-dimensional contingency tables using simple chi-square techniques, thus analysis of such tables serves as a useful starting point. In this paper I will begin by showing how to analyze a simple 2 x 2 table, first using a chi-square and then a log-linear analysis. In the course of this exposition, I will also demonstrate some useful and basic descriptive statistics for 2 x 2 tables. The level of exposition is basic and straightforward and is intended to appeal to the sort of social science researcher who is somewhat suspicious of and occasionally intimidated by mathematical statistics.

The 2 x 2 Table: Basic Descriptives and Chi-Squere Analysis

Imagine that mother-child conversation is recorded. Each turn of talk is coded as either responsive to the partner's previous turn (coded *Yes*) or not (coded *No*). The corpus is then scanned for mother-to-child transitions. There are four possible kinds of transitions:

1. Mother *Yes* to Child *Yes*
2. Mother *Yes* to Child *No*
3. Mother *No* to Child *Yes*
4. Mother *No* to Child *No*

The tallies for such transitions are usually displayed as a 2 x 2 table like Table 1.

Mother-Child Interaction: Observed Frequencies, Simple Probabilities, and Transitional Probabilities

Mom Code / Statistic	Child Code: Yes	Child Code: No	Statistic: Sum	Statistic: Prob.
Yes				
Obs. Freq.	27	4	31	.52
Trans. Prob.	.87	.13		
No				
Obs. Freq.	18	11	29	.48
Trans. Prob.	.62	.38		
Sum	45	15	60	
Probability	.75	.25		1.00

Simple and conditional frequencies and probabilities. In addition to the raw counts or *observed frequencies* (symbolized as *obs* or *f*), simple or *unconditional probabilities* (symbolized as *p*) are one of the most basic descriptive statistics computed for data like those in Table 1. The total number of tallies is symbolized with *N*. Thus:

$$p(Mom=Yes) = f(\text{Mom}=\text{Yes}) / N = 31/60 = .52$$

is the probability that the transitions tallied began with a mother code of *Yes*. Similarly:

$$p(\text{Child}=\text{Yes}) = f(\text{Child}=\text{Yes}) / N = 45/60 = .75$$

is the probability that the transitions tallied ended with a child code of *Yes*. For these (manufactured) data, children had an overall tendency to be responsive, even more so than their mothers.

Usually data are organized in a 2 X 2 table in order to determine whether the row variable (in this case, the mother's turn of talk) is associated with or affects the column variable (in this case, the child's turn of talk). For example, we might want to know whether the child is more likely to be responsive if the mother begins by being responsive. The conditional or *transitional probability* lets us describe the state of affairs observed (again see Table 1). For example:

$$p(\text{Child}=\text{Yes}/\text{Mom}=\text{Yes}) = f(\text{Mom}=\text{Yes} \ \& \ \text{Child}=\text{Yes}) / f(\text{Mom}=\text{Yes}) = 27/31 = .87$$

is the probability that the child's code will be *Yes* given that the mother's code was *Yes* (the slash is read as "given").
Similarly:

$$p(\text{Child}=\text{Yes}/\text{Mom}=\text{No}) = f(\text{Mom}=\text{No} \ \& \ \text{Child}=\text{Yes}) / f(\text{Mom}=\text{No}) = 18/29 = .62$$

is the probability that the child's code will be *No* given that the mother's code was *Yes*.

Expected frequencies and raw residuals. It appears that children were more likely to be responsive if their mother had begun by being responsive, but is the diference between .87 and .62 statistically significant? In order to make this determination, *expected frequencies* (symbolized as *exp*) and residuals are needed (see Table 2). Expected frequencies for the cells are computed from the totals in the margins (which for that reason are called marginals). For example:

$$exp(\text{Mom}=\text{Yes} \ \& \ \text{Child}=\text{Yes}) = f(\text{Mom}=\text{Yes}) \text{ X } p(\text{Child}=\text{Yes}) = f(\text{Mom}=\text{Yes}) \text{ X } f(\text{Child}=\text{Yes}) / \text{N} = 31 \text{ X } 45/60 = 23.35.$$

The *raw residuals* for the cells are then simply the differences between observed and expected frequencies. For example:

$$\text{Residual (Mom}=\text{Yes} \ \& \ \text{Child}=\text{Yes}) = obs(\text{Mom}=\text{Yes} \ \& \ \text{Child}=\text{Yes}) - exp(\text{Mom}=\text{Yes} \ \& \ \text{Child}=\text{Yes}) = 27 - 23.35 = 3.75$$

Pearson Chi-square. Most readers learned long ago to use expected frequencies and raw, residuals to compute a statistic called the *Pearson chi-square*. It is

TABLA 2 Independence Model: Observed and Expected Frequencies, Residuals, and Z-Scores

Mom Code / Statistic	Child Code: Yes	Child Code: No
Yes		
Obs. Freq.	27	4
Exp. Freq.	23.25	7.75
Residual	3.75	-3.75
Z score	2.24	-2.24
No		
Obs. Freq.	18	11
Exp. Freq.	21.75	7.25
Residual	-3.75	3.75
Z score	-2.24	2.24

distributed approximately as chi-square and is symbolized here as X^2. Letting R, C, r, and c represent the number of rows, the number of columns, a particular row, and a particular column, respectively, then the Pearson chi-square, summed over all cells of an R X C table, is:

$$X^2 = \sum \frac{(obs_{rc} - exp_{rc})^2}{exp_{rc}}. \quad (1)$$

For the present example:

$$X^2 = (27 - 23.25)^2/23.25 + (4 - 7.75)^2/7.75 + (18 - 21.75)^2/21.75 + (11 - 7.25)^2/7.25$$
$$= 0.60 + 1.81 + 0.65 + 1.94 = 5.00.$$

Degrees of freedom for a 2 X 2 table are 1 and the critical value for chi-square with 1 degree of freedom, alpha = .05, is 3.84, thus for the present example rows and columns are not independent. Apparently the child's turn of talk is affected by the mother's preceding turn.

Z scores and adjusted residuals. For a 2 X 2 table, as for a one-way analysis of variance with two groups, no further tests are necessary. For other two-dimensional tables, however, as for one-way analyses of variance with more than two groups, some sort of post hoc test is necessary in order to determine how the cells vary among themselves. For example, imagine that the child's consequent turn was categorized as *Yes*, *No*, or *Maybe*. Then a significant chi-square for the resulting 2 X 3 table could mean that the mother's antecedent turn affected the distribution of the child's *Yes*'s, or the child's *No*'s, or the child's *Maybe*'s, or all three, or only some combination of two of the child's possible responses.

The residuals provide a descriptive sense of the effect. For a 2 X 2 table, although the absolute magnitude of the four residuals is the same, the sign is informative. For the present example, we learn that the child was more likely to be coded *Yes* if the mother was coded *Yes*, and more likely to be coded *No* if the mother was coded *No* (the residuals are positive; see Table 2). For a 2 x 3 table (or any table other than a 2 X 2), the magnitude of the residuals can be informative as well. But what is a significant magnitude?

Residuals can be standardized, that is, transformed into a score that is distributed approximately normally. Then any standardized residuals larger than 1.96 (p . < 05) or 2.58 (p <.01) can be regarded as worthy of attention and interpretation. In the chi-square and sequential analysis literature, standardized residuals are usually called z scores. Z scores are computed for each cell of a two-dimensional table. Letting f_{rc} represent the observed frequency for that cell, f_r the frequency for tallies in that row, pc the probability for tallies in that row, the formula is:

$$z = \frac{f_{rc} - f_r p_c}{\sqrt{f_r p_c (1-p_c)\,(1-p_r)}} \qquad (2)$$

For the present example:

$$z = \frac{27 - 31 \text{ X } .75}{\sqrt{31 \text{ X } .75 \text{ X } .25 \text{ X } .48}} = 2.24.$$

In the sequential analysis literature, this is sometimes called the Allison and Liker z score (Allison & Liker, 1982; see also Bakeman & Gottman, 1986).

Early in the log-linear literature, a statistic was defined and called the *standardized residual* (it is the residual divided by the square root of the expected frequency). But this was premature. Haberman (1973, 1978) showed that another statistic, called the *adjusted residual*, not only was a better approximation to a normal distribution, but also provided identical absolute values for the four the cells of a 2 X 2 table, wich the standardized residual does not. Moreover, the adjusted residual is the same as the z score defined in the previous paragraph. For all these reasons, I believe it makes sense to use the adjusted and not the standardized residual (some computer programs like SPSSX'S LOGLINEAR and CROSSTABS provide both, others like SPSSX'S HILOGLINEAR provide only the standardized residual).

Two-dimensional chi-square analyses. The statistical education of many social science researchers does not go much beyond the material presented in this section, at least with respect to the chi-square goodness-of-fit test (observed frequencies, organized as a row of K cells, are compared with theoretically expected ones, as in the binomial or sign test). And most know how to generalize from a 2 X 2 table to other two-dimensional tables and know that an R X C table has (R - 1) (C - 1) degrees of freedom. Many even know that the chi-square test is called a *test of independence* of rows and columns if the units (susbjects, transitions, etc.) tallied are free to vary along both rows and columns (e.g., subjects are selected

and then are categorized a male or female and a old or young) and a *test of homogeneity* of proportions if the number of units in each row is predetermined (e.g., samples of males and females are selected and then categorized as old or young).

Traditional chi-square analysis does not allow for the ready analysis of tables with more than two dimensions, which is its major limitation. For example, imagine that we had two kinds of mother-child dyads: dyads containing a depressed and a normal mother, or dyads subjected to two kinds of treatment. The resulting data would be organized as a 2 X 2 X 2 table: dimension or factor A would be dyad (Depressed/Normal), factor B would be responsiveness of the mother's antecedent turn /Yes/No), and factor C would be responsiveness of the child's consequent turn (Yes/No). Such tables are readily analyzed with log-linear techniques, which allow us to answer questions like. Is there a relation between antecedent and consequent turns? Is that relation different for depressed and normal dyads? Before showing how such question can be answered, however, first I will demonstrate how a log-linear analysis of the simple 2 X 2 table just presented would proceed. This provides a simple and familiar context in which to present terms and ideas that will be elaborated later.

The 2 x 2 table: Log-Linear Analysis

The "log" in log-linear analysis stands for logarithm, a topic that most of us encountered and probably forgot long ago. Briefly, a *logarithm* is an *exponent*. Consider the following:

$$\begin{aligned}
2^0 &= 1, & \log_2 1 &= 0\\
2^1 &= 2, & \log_2 2 &= 1\\
2^2 &= 4, & \log_2 4 &= 2\\
2^3 &= 8, & \log_2 8 &= 3\\
2^4 &= 16, & \log_2 16 &= 4\\
2^5 &= 32, & \log_2 32 &= 5\\
2^6 &= 64, & \log_2 64 &= 6\\
2^7 &= 128, & \log_2 128 &= 7.
\end{aligned}$$

This series serves to remind us that, for example, the log of 8 (base 2) is 3 because 2 raised to the power 3 is 8. Similarly, the log (base 2) of 128 is 7 because 2 raised to the power 7 is 128. In theory, any base could be used, although 2, 10, and *e* are the most common.

Natural logs. Recall that *e* is a constant whose value is approximately 2.718 and that logarithms that use the base *e* are called natural logs. For example:

$$\begin{aligned}
\ln_e \quad 1 &= 0\\
\ln_e \quad 3 &= 1.10\\
\ln_e \quad 8 &= 2.08\\
\ln_e \quad 21 &= 3.04\\
\ln_e \quad 55 &= 4.01\\
\ln_e \quad 150 &= 5.01\\
\ln_e \quad 404 &= 6.00\\
\ln_e \quad 1100 &= 7.00
\end{aligned}$$

where *ln* indicates a natural log. As you might guess, log-linear analyses are based on natural logarithms, although a user could remain innocent of this fact and still perform competent analyses.

Log expected frequencies. When testing for the independence of rows and columns (or the homogeneity of proportions in the different rows), the formula for an expected frequency for the cell in row *r* and column c is:

$$\underline{\exp}_{rc} = \underline{f}_r \underline{p}_c = \underline{f}_r \text{ X } \underline{f}_c / \underline{N} .$$

When numbers are multiplied, their logarithms are added, and when numbers are divided, the logarithm of the divisor is subtracted from the logarithm of the dividend. Therefore:

$$\ln (\underline{\exp}_{rc}) = \ln (\underline{f}_r) + \ln (\underline{f}_c) - \ln (\underline{N}) .$$

If we let:

$$u_r = \ln (\underline{f}_r)$$
$$u_c = + \ln (\underline{f}_c)$$
$$u = - \ln (\underline{N})$$

then, the formula for the logarithms of the cell frequencies can be expressed as:

$$\ln (\exp_{rc}) = u + u_r + u_c. \qquad (3)$$

This is intriguing. The log-linear expression (*log* because logarithms of observed frequencies are predicted or modeled; *linear* because the prediction equation consists of additive effects) looks remarkably like the model for a two-way analysis of variance (ANOVA) that postulates two main effects but no interaction.:

$$\underline{Y}_{ijk} = \mu + \alpha_i + \beta_j + \in_{ijk} . \qquad (4)$$

The two-main-effects ANOVA model and the independence log-linear model are alike in that both postulate a row and a column effect, but no row X column interaction. More generally, they are alike in that both are linear models, and statisticians have found linear models tractable and immensely useful, as the broad use of analysis of variance and multiple regression attests.

The ANOVA and log-linear models differ in a number of ways. The ANOVA model predicts individual scores and includes an error term, whereas the log-linear model predicts cell frequencies and does not include an error term. Still, the fundamental insight remains: if logarithms of expected frequencies are modeled, then linear models for the cells of contingency tables can be formed. These models have all the advantages of linear models generally, including the partitioning of effects into main effects and various interactions among them. Thus log-linear analyses allow those social science investigators whose data is most naturally organized into contingency tables ready and direct answers to the research questions that concern them.

Log likelihood ratio chi-square. The log-linear model of independence (row and column effects only, row and column do not interact) generates the log expected frequencies shown in Table 3. Using Equation 3 and the values given in the table, the arithmetic is as follows (computed before rounding, thus numbers shown here may not add exactly):

$$\ln(\underline{\exp}_{rc}) = u + u_r + u_c$$

$$\ln(\underline{exp}_{rc}) = -\ \ln(\underline{N}) + \ln(\underline{f}_r) + \ln(\underline{f}_c)$$
$$\ln(\underline{exp}_{11}) = -4{,}09 + 3.43 + 3.81 = 3.15$$
$$\ln(\underline{exp}_{12}) = -4.09 + 3.43 + 2.71 = 2.05$$
$$\ln(\underline{exp}_{21}) = -4.09 + 3.37 + 3.81 = 3.08$$
$$\ln(\underline{exp}_{22}) = -4.09 + 3.37 + 2.71 = 1.98.$$

These are logs for expected values (given the model of independence) but the question remains, how well do these expected values fit the observed ones, that is, how close are they to the logs for the observed counts?

We could work backwards, findings the antilogs for the logs in Table 3 (which are the expected frequencies shown in Table 2) and compute a Pearson chi-square. For a number of technical reasons, however, another statistic is preferred for log-li-

TABLA 3
Independence Model: Expected Frequencies and Log ExpectedFrequencies

Mom Code Statistic	Child Code Yes	Child Code No	Statistic Sum	Statistic Ln
Yes				
Exp. Freq.	23.25	7.75	31	3.43
Ln Exp. Freq.	3.15	2.05		
No				
Exp. Freq.	21.75	7.25	29	3.37
Ln Exp. Freq.	3.08	1.98		
Sum	45	15	60	
Ln	3.81	2.71		4.09

near analyses. It is called the *log likelihood ratio chi-square*, or, simply, the likelihood ratio chi-square, usually symbolized as G^2. It too is distributed approximately as chi-square. Summed over all cells in the R X C table, it is:

$$\underline{G}^2 = 2 \sum obs_{rc} [\ln(obs_{rc}) - \ln(exp_{rc})] \qquad (5)$$

or (because $\log a - \log b = \log (a/b)$:

$$\underline{G}^2 = 2 \sum obs_{rc} \ln(obs_{rc}/exp_{rec}).$$

For the present example:

$$\underline{G}^2 = 2 \text{ X } (27 \text{ X } [3.30 - 3.15] + 4 \text{ X } [1.39 - 2.05] + 18 \text{ X } [2.89 - 3.08] + 11 \text{ X } [2.40 - 1.98])$$
$$= 5.14.$$

Again, this result is significant at the .05 level. The model of independence fails to fit the data. As before we conclude that the child's turn of talk was affected by the mother's preceding turn and not independent of it.

Usually, Pearson and likelihood ratio chi-squares yield identical levels of significance. When they do not, it should only serve to remind us that both X^2 and G^2 are distributed *approximately* as chi-square; neither is necessarily distributed *exactly* as the theoretical chi-square. But in any case, although both are printed by most computer programs (e.g., SPSSX's LOGLINEAR and HILOGLINEAR), the likelihood ratio chi-square approximation should be used for log-linear analyses.

Hierarchical Models for 2 x 2 tables

Hierarchical log-linear models, like hierarchical multiple regression models, refer to a series of nested models in which successively more complex models incorporate all less complex ones. Whenever log-linear models are discussed, it is almost always safe to assume that hierarchical models are intended. Nonhierarchical log-linear models are possible, but even experts find them problematic and often recommend against their use. For example, the likelihood ratio chi-square can be partitioned among the various effects for hierarchical but not nonhierarchical models. Moreover, as demonstrated subsequently, there are statistical criteria for choosing among hierarchical but not Nonhierarchical models. For all these reasons, only hierarchical log-linear models are discussed here.

Log-linear models generate predicted values for cell frequencies. Each successive model in the hierarchy includes an additional term--that is, takes an additional factor into account--in effect imposing an additional constraint on the cell frequencies generated by the model. As a result, the cell frequencies predicted by successive models come closer and closer to replicating or fitting the tallies actually observed. For a 2 X 2 table, a series of four models can be defined.

The null model. As a first step, imagine that we knew only that 60 mother-child transitions were tallied, and were asked to "model" the scores as best we could. Given this limited information, our best guess would be that half the transitions began with the mother being responsive, and half ended with the child being responsive, such that the four types of transitions occurred equally often. Because this first model only takes the total number of tallies into account and nothing else, it is often called the *null model.* It is usually symbolized as [0], a zero in brackets. This model necessarily generates equal frequencies for all four cells. For that reason, in addition to the null, it is sometimes called the equiprobable model as well (see Figure 1).

The child only model. The second model assumes knowledge of one factor or dimension only. For example, imagine that we knew only that 45 transitions ended with *Child=Yes* and 15 with *Child=No*, but did not know the distribution

for *Mother=Yes* and *Mother=No. Then we would distribute the 45 Child=Yes* and the 15 *Child=No* tallies evenly between the two rows representing the mother, guessing that half of her responses were *Yes* and half *No*. Because this model takes into account just the child tallies (i.e., the column totals) presented in the margins of the 2 X 2 table, it is called here the *child only model* and is represented symbolically with [C] (again, see Figure 1).

The child plus mother model. The third model assumes knowledge of both factors or dimensions. It assumes that both row and column marginal totals are known; that is, in addition to knowing that 45 transitions ended with *Child=Yes* and 15 with *Child=No*, we also know that 31 began with *Mother=Yes* and 29 with *Mother=No*. This additional information allows us to refine our guesses. Taking both row and column marginal totals into account, we would expect the 31 marginal tallies in row 1 and the 29 in row 2 to be distributed into the cells like the column marginal totals, that is, with a 45:15 or a 3:1 ratio. These considerations generate the expected cell frequencies shown in Figure 1. Because this model takes into account marginal totals for both mother and child categories, it is called here the *mother plus child model* and is represented symbolically with [M] [C]. Then [M] [C] model is, of course, the independence model discussed in previous sections.

[0], Null Model:

Mom Code	Child Code: Yes	Child Code: No	
Yes	15	15	
No	15	15	
			60

[C], Child Only Model:

Mom Code	Child Code: Yes	Child Code: No	
Yes	22.5	7.5	
No	22.5	7.5	
	45	15	60

[M][C], Mom + Child Model:

Mom Code	Child Code: Yes	Child Code: No	
Yes	23.25	7.75	31
No	21.75	7.25	29
	45	15	60

[MC], Mother X Child Model:

Mom Code	Child Code: Yes	Child Code: No	
Yes	27	4	31
No	28	22	29
	45	15	60

Figure 1. Expected cell frequencies generated by four hierarchical log-linear models for a 2 X 2 table. Each successive model imposes an additional constraint on the cell frequencies.

The saturated model. The fourth and final model assumes knowledge, not just of the row and column totals, but also knowledge of the frequencies in the four cells. For a 2 X 2 table, this fourth model generates cell frequencies identical to those observed (again, see Figure 1). Such a model is called a *saturated model* and necessarily the scores generated by it fit the observed scores perfectly. Because this model, in addition to marginal totals for both the mother and child categories, takes the cross-classification of the cells by the these two factors into account, it is called here the *mother times child model* and is represented symbolically with [M] [C] [MC].

Because each successive model after the null adds a term--first [C], then [M] [C], then [M] [C] [MC]--it makes sense to represent the fourth model is usually represented simply as [MC]. Scores that satisfy the [MC] constraint--scores that reflect the actual mother turn X child turn cross-classification--necessarily satisfy the [M] (row or mother category) and the [C] (column or child category) marginal constraints as well. Thus, once the fourth model has been represented with [MC], it is unnecessary to add [M] and [C] to it. For that reason, and in the interest of simplicity, [MC] is usually presented by itself; the [M] and [C] terms are understood implicitly.

Backwards elimination. For ease of exposition, the four models were just presented beginning with the simplest or null model and progressing to the most complex or saturated model. Usually, however, a log-linear analysis precedes in the opposite direction. We begin with the saturated model, which by definition fits the data perfectly, and proceed to eliminate terms (which is called *backwards elimination*). The goal is to find the simplest model that still fits the data, that is, whose chi-square remains nonsignificant. The likelihood ratio chi-sqare assesses how well the expected scores for the models defined at each successive steps fit the observed data (see Table 4). And the change in chi-square from step to step indicates how important the term, or effect, or constraint deleted at that step is (see Table 5).

For the sake of completeness, and to illustrate how a log-linear analysis migth proceed, all four steps for the present 2 X 2 example are shown in Tables 4 and 5. But in practice, an investigator faced with a two-dimensional contingency table would probably only compute chi-square for the independence or homogeneity of proportions model (here the [M] [C] model). If this model fails --that is, if it generates scores quite discrepant from those actually observed as reflected in a chi-square significantly different from zero--then we reject this model and accept the only one that could fit, the full or saturated model, and we conclude that the two dimensions of the contingency table are not independent (or that the proportions reflected in the rows are not homogeneous). This, in fact, is the substantive result usually desired by researchers.

Table 4

Hierarchical Models: Expected Frequencies , Lag Expected Frequencies, and G^2

Model Statistic	G^2	df	Mom Code			
			Yes		No	
			Child Code		Child Code	
			Yes	No	Yes	No
[MC], $u + u_m + u_c + u_{mc}$ (saturated model)	0.0	0				
Exp. Freq.			27	4	18	11
Ln Exp. Freq			3.30	1.39	2.89	2.40
[M][C], $u + u_m + u_c$ (mom + child model)	5.14	1				
Exp. Freq.			23.25	7.75	21.75	7.25
Ln Exp. Freq.			3.15	2.05	3.08	1.98
[C], $u + u_c$ (child only model)	5.21	2				
Exp. Freq.			22.5	7.5	22.5	7.5
Ln Exp. Freq.			3.11	2.01	3.11	2.01
[0], u (null model)	20.91	3				
Exp. Freq.			15	15	15	15
Ln Exp. Freq.			2.71	2.71	2.71	2.71

Table 5

Hierarchical Models for a 2 X 2 Table: G^2 and Partial G^2

Step	Model	G^2	df	Deleted	ΔG^2	Δdf
1	[MC], saturated	0.0	0			
2	[M][C], mom + child	5.14	1	[MC]	5.14	1
3	[C], child only	5.21	2	[M]	0.07	1
4	[0], null model	20.91	3	[C]	15.70	1

Note. M represents the mother or row dimension, C represents the child or column dimension.

Describing results. For the present example, we now know that the mother's antecedent turn has a significant effect on her child's subsequent one. The model that includes the saturated or 2 way interaction term fits the data (likelihood chi-square [0] = 0.0, p = 1) whereas the model that deletes the [MC] term does not (likelihood chi-square[1] = 5.14, p .<05). The probability that the child's turn will be coded responsive is .75 overall. This becomes .87 when the mother's turn was coded responsive and. 62 when the mother's turn was code unresponsive. Alternatively, we could note that the odds that the child will be responsive is 6.75 to 1 if the mother was responsive and is only 1.64 to 1 if she was not (the ratio of these two ods, called the odds ratio, is 4.13, and is a common descriptive statistic used in epidemiological studies). In any case, we now know from the log-linear analysys that the difference between these two conditional probabilities, or two ods, is statistically significant.

Degrees of freedom: Multiple regression and log-linear approaches compared. We have yet to discuss how degrees of freedom for the chi-square statistics computed in Tables 4 and 5 are determined. This is an important topic for two reasons. First, we need to know degrees of freedom before we can assign significance to a particular value for chi-square. Chi-square, like the F distribution, is actually a family of distributions, and its exact shape--and hence the critical value demarcating 5% or 1% of the area under the curve--depends on the degrees of freedom. Second, understanding how degrees of freedom are determined for contingency tables highlights some of the key differences between the multiple regression or ANOVA and log-linear approaches.

Both hierarchical multiple regression and log-linear approaches consist of a series of steps. A model is associated with each step and each successive model takes additional information into account. For multiple regression analyses, the models predict scores for individuals. The starting point is total variance, which reflects differences between the observed scores and the mean. When computing total variance, only one constraint is imposed on the individual scores--that is, the only parameter specified is the mean--and so we claim N - 1 degrees of freedom total. Successive steps include more factors and make successively more accurate predictions concerning individuals' scores and thereby reduce residual or error variance. This has the effect of increasing the R^2 accounted for by the multiple regression model specified at each successive step.

For log-linear analyses, on the other hand, the models predict scores for the cells of a contingency or cross-classification table, not for individuals. For the present 2 X 2 example, the starting point is the empty table with its four cells. The null model imposes only one constraint on these cells, namely that the four cell frequencies sum to the total number of tallies. Because there are four scores initially constrained by one parameter, N or the total number of tallies, the degrees of freedom for the null model is 4 -1 or 3. In other words, three cell frequencies are free to vary, but once three have been entered, the fourth is determined by the requirement that the four sum to the total. If K is the number

of-cells, then degrees of freedom total is K - 1, which for the present example equals 3.

Similar reasoning applies to the column only model (here the [C] model), except now two parameters constrain the cell frequencies, one representing the grand total and one the column or child category totals, which results in K - 2 or, for the present example, 2 degrees of freedom. Similarly, the row plus column or independence model (here the [M] [C] model) is constrained by three parameters, which results in K - 3 or, for the present example, 1 degree of freedom. In other words, the four cells for the 2 X 2 independence or homogeneity of proportions model are constrained by both row and column marginals. Once a score is entered in one cell, values for other cells are determined. Because only one cell is free to vary, there is only one degree of freedom associated with this model.

The present example consists of a 2 X 2 table, but two-dimensional tables with more than two levels per dimension are possible. Speaking generally, if R represents the number of levels for the row dimension and C represents the number of levels for the column dimension, then the degrees of freedom for an R X C table are (R - 1) times (C - 1), as noted earlier. Note that degrees of freedom for contingency tables do not take the number of tallies into account, only the number of dimensions and the number of levels for each dimension. This contrasts with the degrees of freedom computations for multiple regresion analyses, which take the total number of subjects into account and, as noted earlier, reflects a key difference between multiple regression models, which predict individual scores, and log-linear models, which predict cell frequencies instead.

Partial chi-square. One additional detail In Table 5 requires clarification. Each step in a hierarchical log-linear analysis is associated with a total chi-square, which represents how well the scores generated by the present model fit the cell frequencies actually observed, and also with a change in chi-square, which represents the contribution of the term deleted at that step and which is sometimes called a *partial chi-square* (symbolized with a Δ or delta in Table 5). In order to evaluate the significance of the change in chi-square we need to know its degrees of freedom, which is the difference between the degrees of reedom for the present and previous model. For the present example, the change in chi-square from the third to the fouth step is 20.91 minus 5.21, which equals 15.70. The corresponding change in degrees of freedom is 3 minus 2 which equals 1.

Hierarchical orders. Tables 4 and 5 show only one of the two possible ways in which terms for a 2 X 2 table could be deleted. The 2-way saturated term, [MC], must be deleted first, but there are two 1-way terms, [M] and [C], and either [M] could be deleted second and [C] third as in Tables 4 and 5, or [C] could be deleted second and [M] third. There is no intrinsically correct ordering among terms that occupy the same level (i.e., all 1-way terms, all 2-way terms, all 3-way terms, etc.). Instead, investigators must decide on a hierarchical ordering that makes sense, given their substantive and research concerns.

For the present example, the order of the 1-way terms makes little difference. The partial chi-square associated with the mother term is small and insignifi-

cant, which only means that about half of the mother's turns were coded *Yes* and about half *No*. Similarly, the large and significant partial chi-square associated with the child's turn only means that codes for the child's turns were not distributed evenly. But the significance or insignificance of the 1-way terms is placed in shadow by the significant 2-way or saturated term. For this example, only the saturated model fits the data, thus interpretation needs to take into account the cross-classification of, or interaction between, the row and column (i.e, mother and child) factors.

Beyond Two Dimensions

The previous discussion has emphasized not just two-dimensional tables, but the simplest exemplar of a two-dimensional table, a 2 X 2. The purpose was to present terms and issues basic to log-linear analysis in the simplest context possible. Needless to say, interesting applications of log-linear analysis often involve more than two dimensions and so it is worthwhile to consider briefly how such analyses would proceed.

Imagine, for example, that mother-child interaction was coded for two groups of mothers, those who were depressed and those who were not (or those who had received a particular treatment and those who had not) and that their interaction was again coded as for our previous example. A 2 X 2 X 2, group (*Depressed/Not depressed*) X mother's antecedent turn (*Yes/No*) X child's consequent turn (*Yes/No*) contingency table would result. Let *G* represent group. This table could be analyzed with the hierarchical series shown in Table 6.

Table 6

Hierarchical Models: First Five Steps for a 2 X 2 X 2 Table

Step	Model	G^2	df	Deleted	ΔG^2	Δdf
1	[CMG]	0.0	0			
2	[CM][CG][MG]		1	[CMG]		1
3	[CG][MG]		2	[CM]		1
4	[MG]		3	[CG]		1
5	[C][M][G]		4	[MG]		1

Note. M represents the mother dimension, C represents the child dimension, and G represents the group dimension.

If only the saturated model, [CMG], fit the data (i.e., if the chi-square associated with the [CM] [CG] [MG] model was significantly different from zero), then we would conclude that child's consequent turns were affected by mother's

antecedent ones in different ways for depressed and non-depressed mothers. In analysis of variance terms, we would say that group interacted with mother's antecedent turn in accounting for the child's consequent turn. However, if the [CM] [CG] [MG] model fit the data, but the [CG] [MG] model did not, we would conclude that child's consequent turns were affected by mother's antecedent ones (removing the [CM] turn resulted in loss of fit as indicated by a significant increase in chi-square), and in a similar way for both depressed and non-depressed mothers. Finally, if the [CG] [MG] model fit the data, we would conclude that mother's turn did not affect their child's subsequent ones, although the distribution of the mother's codes (as signalled by the [MG] term), and the child's codes as well (as signalled by the [CG] term), was different for the depressed and non-depressed groups.

Computing log-linear statistics. This relatively brief paper can provide only the simplest of introductions to a subtle and far-ranging topic. Ultimately, we learn best by doing and interested readers are encouraged to perform log-linear analyses of their own data. Two programs in SPSSX are especially helpful (see Norusis, 1985). First, use HILOGLINEAR to compute partial chi-squares (specify PRINT ASSOCIATION and use the default DESIGN, which is the saturated model). Then, because HILOGLINEAR only gives standardized residuals, use CROSSTABS to compute adjusted residuals (specify OPTIONS 17) for the tables broken down as indicated by the log-linear analysis.

Reducing sequential data to contingency tables. Most major statistical packages (BMDP, SAS, SPSS, SYSTAT) have routines for log-linear analysis of contingency tables. However, there are no widely-used general purpose computer programs that produce contingency tables from sequential data. Nor is there an accepted standard way to represent sequential data, a state of affairs which I believe has impeded progress. In an attempt to remedy this situation, V. Quera and I have defined a sequential data interchange standard, which we call SDIS (Bakeman & Quera, 1991), and are beginning to define a general purpose computer program that will analyze SDIS data. Such a program, coupled with log-linear analytic techniques, should greatly advance the use of sequential analysis.

REFERENCES

Allison, P.D., & Liker, J.K. (1982). Analyzing sequential categorical data on dyadic interaction: A comment on Gottman. *Psychological Bulletin, 91*, 393-403.

Anguera, M.T. (Ed.). (1991). *Metodología observacional en la investigación psicológica*. Barcelona (Spain): University of Barcelona Press.

Bakeman, R. (1978). Untangling streams of behavior: Sequential analyses of observation data. In G.P. Sackett (Ed.), *Observing behavior* (Vol. 2, *Data collection and analysis methods*, pp. 63-78). Baltimore: University Park Press.

Bakeman, R. (1991) Counts and codes: Analyzing categorical data (pp.255-274). In B.M. Montgomery & S. Duck (Eds.), *Studying interpersonal interaction*. New York: Guilford Publications.

Bakeman, R., Adamson, L.A., & Strisik, P. (1989). Lags and logs: Statistical approaches to interaction. In M.H. Bornstein & J. Bruner (Eds.), *Interaction in human development* (pp.241-260). Hillsdale, NJ: Erlbaum.

Bakeman, R., & Dabbs, J.M. Jr. (1976). Social interaction observed: Some approaches to the analysis of behavior streams. *Personality and Social Psychology Bulliten, 2*, 335-345.

Bakeman, R., & Gottman, J.M. (1986). *Observing interaction: An introduction to sequential analysis*. New York: Cambridge University Press.

Bakeman, R., & Gottman, J.M. (1987). Applying observational methods: A systematic view. In J. Osofsky (Ed.), *Handbook of infant development* (2nd ed., pp. 818-854). New York; Wiley.

Bakeman, R. & Quera, V. (1991). SDIS: A sequential data interchange standard. Atlanta, GA: Developmental Laboratory, Georgia State University.

Bishop, Y.M.M., Fienberg, S.R., & Holland, P.W. (1975). *Discrete multivariate analysis: Theory and practice*. Cambridge, MA: MIT Press.

Castellan, N. J., Jr. (1979) The analysis of behavior sequences. In R.B. Cairns (Ed.), *The analysis of social interactions: Methods, issues, and illustrations* (pp. 81-116)- Hillsdale, NJ: Erlbaum.

Cohn, J.F., & Tronick, E.Z. (1987). Mother-infant face-to-face interaction: The sequence of dyadic states at 3, 6, and 9 months. *Developmental Psychology, 23*, 68-77.

Fienberg, S.E. (1980). *The analysis of cross-classified categorical data* (2nd ed.). Cambridge, MA: MIT Press.

Haberman, S.J. (1973). The analysis of residuals in cross-classified tables. *Biometrics, 29*, 205-220.

Haberman, S.J. (1978). *Analysis of qualitative data* (Vol *1*). New York: Academic Press.

Kennedy, J. J. (1983). *Analyzing qualitative data: Introductory log-linear analysis for behavioral research*. New York: Praeger.

Knoke, D., and Burke, P.J. (1980). *Log-linear models*. Newbury Park, CA.: Sage.

Norusis, M.J. (1985). *SPSSX Advanced Statistics Guide*. New York: McGraw-Hill.

Sackett, G.P. (1979). The lag sequential analysis of contingency and cyclicity in behavioral interaction research. In J. Osofsky (Ed.), *Handbook of Infant Development*, (1st ed., pp. 623-649). New York: Wiley.

Sackett, G.P. (1987). Analysis of sequential social interaction data: Some issues, recent developments, and a casual inference model. In J. Osofsky (Ed.), *Handbook of Infant Development*, (2nd ed., pp. 855-878). New York: Wiley.

Stevenson, M.B., Ver Hoeve, J.N. Roach, M.A., & Leavitt, L.A. (1986). The beginning of conversation: Early patterns of mother-infant vocal responsiveness. *Infant Behavior and Development*, *9*, 423-440.

Tabachnick, B.G., & Fidell, L.S. (1989). *Using multivariate statistics*. New York: Harper & Row.