

Likelihood Analysis of Three-Way Contingency Tables

J. K. LINDSEY AND F. W. NASH

*Fisheries Research Board of Canada
Biological Station, Nanaimo, B.C.*

LINDSEY, J. K., AND F. W. NASH. 1972. Likelihood analysis of three-way contingency tables. *J. Fish. Res. Bd. Canada* 29: 590-591.

The three-way contingency table is considered as an analysis of variance problem. Since the appropriate distribution is multinomial rather than normal, the log likelihood ratios are used to make inferences instead of the variance ratios.

LINDSEY, J. K., AND F. W. NASH. 1972. Likelihood analysis of three-way contingency tables. *J. Fish. Res. Bd. Canada* 29: 590-591.

Nous considérons les tableaux de contingence à triple entrée pour l'étude des problèmes d'analyse de la variance. Puisque la distribution appropriée est polynomiale plutôt que normale, les inférences sont basées sur les rapports de la vraisemblance plutôt que sur les rapports de la variance.

Received September 27, 1971

CONSIDER a three-way contingency table of size $I \times J \times K$ where individual cells contain counts, N_{ijk} , of individuals. Suppose the rows one way follow a multinomial distribution. Then, the table may be analyzed as an analysis of variance problem with multinomial "error" and, say, $I \times J$ cells. Then, in the (i, j) th cell, the probability of the k th type of individual occurring will be p_{ijk} where $\sum_k p_{ijk} = 1$. The maximum likelihood estimates will be

$$\hat{p}_{ijk} = N_{ijk} / \sum_k N_{ijk}$$

and the maximized log likelihood function

$$\log L_{IJ} = \sum_{i,j,k} [N_{ijk} \log (N_{ijk} / \sum_k N_{ijk})].$$

Suppose now that all I rows have the same multinomial distribution, i.e., there is no effect of difference in condition among rows. Then, in the j 'th cell of all rows, the probability of the k 'th type of individual occurring will be p_{jk} where $\sum_k p_{jk} = 1$ and $p_{jk} = \sum_i N_{ijk} / \sum_i N_{ijk}$, so that the maximized log likelihood function is

$$\log L_J = \sum_{j,k} [N_{ijk} \log (\sum_i N_{ijk} / \sum_i N_{ijk})].$$

Similarly, if all J columns have the same distribution, the probabilities will be p_{ik} where $\sum_k p_{ik} = 1$

and $p_{ik} = \sum_j N_{ijk} / \sum_j N_{ijk}$ so that the maximized log likelihood function is

$$\log L_I = \sum_{i,j,k} [N_{ijk} \log (\sum_j N_{ijk} / \sum_j N_{ijk})].$$

Then, the greater the difference between $\log L_J$ and $\log L_{IJ}$, the more likely is the effect of treatment differences among rows.

If no treatment effect is assumed among either rows or columns, the probabilities will be p_k where $\sum_k p_k = 1$ and $\hat{p}_k = \sum_{i,j} N_{ijk} / \sum_{i,j,k} N_{ijk}$ so that the maximized log likelihood function is

$$\log L = \sum_{i,j,k} [N_{ijk} \log (\sum_{i,j} N_{ijk} / \sum_{i,j,k} N_{ijk})].$$

A measure of interaction between rows and columns, by analogy with that used in normal theory, is given by $\log L_I + \log L_J - \log L - \log L_{IJ}$. Thus, a table may be constructed, as in normal theory analysis of variance, to analyze the contingency table:

Overall effect	$\log L - \log L_{IJ}$
Between rows	$\log L_J - \log L_{IJ}$
Between columns	$\log L_I - \log L_{IJ}$
Interaction	$\log L_I + \log L_J - \log L - \log L_{IJ}$

Note that the three effects, between rows, between columns, and interaction, are all approximate. For example, the effects between rows and between columns are exact only if there is no interaction

effect. The exact estimates of these three effects may only be obtained by a very time-consuming iterative process, usually involving problems of overflow with most computers. In numerical examples involving the binomial distribution, little difference was found between the approximate and the exact estimates. If the interaction effect is plausible, $\log L - \log L_I$ and $\log L - \log L_J$ are more appropriate for the row and column effects, respectively. But these are also approximate in the same way as the interaction effect: the likelihood equations are nonlinear, requiring iterative solution.

Example 1: An experiment was performed to compare three methods (J) of sampling organisms on the bottom of a stream. Twenty localities (I) were selected at 3-ft intervals along a stream bed. At each location, each method was used in an adjacent area to count the organisms of each of twenty-six species (K) living there. In the analysis, the three sampling methods were compared in pairs:

Methods compared	1 vs 2	2 vs 3	1 vs 3
Overall effect	-5931.76	-4522.65	-3054.20
Between locations	-2183.26	-1852.48	-2596.27
Between methods	-4501.35	-2681.75	-1580.00
Interaction	-752.84	-11.58	-1122.08

We know, from knowledge of the stream, that the distribution of organisms varies with location, i.e., that the effect of location is significant. In making a comparison between effects of location and of method, the numbers of degrees of freedom should be considered: 475 and 25, respectively. Hence, the difference in log likelihood for locations would be expected to be greater than that for methods for the same level of plausibility of effect. These considerations imply that differences between methods 1 and 2 and, to a slightly lesser extent, between 2 and 3 are significant, whereas the difference between 1 and 3, if significant, is much less so.

Example 2: Length-age frequency tables were available for a given species of fish from a number of samples taken in 1 week. We wished to determine if the tables may be combined. At a given age I (=10), the length K (=48) has a multinomial distribution within a given sample J (=6). Then, the analysis is given in the following table:

Overall effect	-301.65
Between ages	-278.83
Between samples	-38.22
Interaction	-15.40

We used the same reasoning as in the previous example. As expected, the length distribution differs among ages, but, since the effect between samples may be considered to be relatively implausible, the samples may be combined with little loss of information about the differences in length distribution at different ages.

As can be seen from the examples, the inferences to be drawn using the likelihood function are "weaker" than those possible when a test of significance is applicable [see, for example, the references provided by Lindsey (1970)]. Differences in maximized log likelihoods are compared to give an indication of the relative size of the various effects, as F-ratios are occasionally compared even although the distribution is known to be non-normal. The comparison using likelihoods will be more accurate than in such cases using the F-ratio, since the additional assumption of normality is not introduced. This remark also applies to considering, for example, $-2(\log L - \log L_{IJ})$ as a χ^2 variate, since the asymptotic normal assumption, being not strictly applicable, has not been introduced.

Of course, the procedure outlined above may be used with any theoretical distribution by replacing the multinomial likelihood function by that for the desired distribution, e.g., the Poisson. When this is done using the normal distribution, the differences in log likelihood corresponding to the various F-tests of analysis of variance result. A simple example, involving inferences using the likelihood function, has been outlined by Lindsey (1970).

An IBM 1130 computer program for the analysis given in this paper is available from the Biological Station, Nanaimo, B.C.

Acknowledgments — The authors would like to express their appreciation to J. C. Mason and A. S. Hourston for bringing the problems in the examples to their attention.

LINDSEY, J. K. 1970. Exact inferences about the parameter for an exponential growth curve following a Poisson distribution. *J. Fish. Res. Bd. Canada* 27: 172-174.