Conditional Analysis of Accidents at Four-Approach Traffic Circles

A. AL-BAKRI, MARK R. CONAWAY, AND JAMES STONER

A conditional analysis for relating the number of accidents at four-approach traffic circles to the geometric and flow characteristics of the circles is presented. The conditional analysis takes into account the association among the observations taken at the four approaches within a traffic circle. It also allows for the inherent differences in safety among different circles. The main advantage of the conditional approach is one can use it to estimate the effects of geometric and flow variables without having to specify a distribution to represent the variability between circles. The conditional method is applied to a study of traffic circles in Amman, Jordan.

The relationship between the number of accidents occurring at a traffic circle and the geometric characteristics and traffic flows of the circle is examined in this paper. The data were collected at seven traffic circles in Amman, Jordan. Each had four approaches approximately at right angles, relatively large circular central islands, parallel entries, and yield-sign traffic control. The data were collected by two teams of 30 police officers and 6 graduate students in transportation engineering at the Jordan University of Science and Technology. The teams collected information on the geometric characteristics and traffic flows in each approach in each traffic circle and counts of several types of accidents. The data collection involved designing special forms and training the teams to ensure the accuracy of the data collection. Details of the procedure are given by Al-Bakri (*1*).

The method of analysis is similar to that of Maycock and Hall (*2*) in that it is based on the assumption that the accident counts have Poisson distributions, the mean of the Poisson distribution depending on the geometric and flow variables. The geometric and flow variables to be used in this study, given here, are described in Figure 1. A more comprehensive definition of these variables is contained in the Jordanian Department of Transport's departmental standards.

- CE—Entry-path curvature in meters^{-1} is the shortest straight-ahead vehicle path.
- CA—Approach curvature in meters^{-1} is the reciprocal of the minimum radius of the bend nearest to the traffic circle.
- ICD—Inscribed circle diameter in meters; diameter of the largest circle that can be inscribed in the outline of the traffic circle.
- CID—Central island diameter in meters.
- QC—Average daily traffic volume in vehicles per day circulating the circle.
- QE—Average daily traffic volume in vehicles per day entering the circle.
- E—Entry width in meters, measured at a point in the upstream approach.
- V—Approach half-width in meters.
- θ—Angle in degrees between the approach leg and next approach leg clockwise.
- PED—Pedestrian volume per day.
- Entering Accidents—Collisions between an entering vehicle and a vehicle within the right-of-way.
- Approaching Accidents—Collisions between vehicles on the approach to the circle, such as rear-end impacts and lane-changing accidents.
- Single-Vehicle Accidents—Collisions involving a vehicle and the circle layout, signs, lighting columns, and such.
- Other Accidents—Collisions between circulating vehicles, circulating vehicles and vehicles exiting the circle, and exiting vehicles and entering or exiting vehicles.
- Pedestrian Accidents—Collisions involving pedestrians.

It is also assumed that a random effect is associated with each traffic circle, which affects the variability of the accident counts within the circle. Similar assumptions were made by Maycock and Hall (*2*), who also assumed that the random effects have a gamma distribution. They estimated the effects of the geometric and flow variables and the parameters of the assumed gamma distribution. A conditional likelihood approach to the problem is taken, so the effects of the geometric and flow variables can be estimated without a distribution for the random effects being specified. This is an important distinction, because the effect of misspecifying the distribution for the random effects is not known.

METHODS

In the method used to analyze the data, Y_{ij} denotes the number of accidents in approach j of traffic circle i, and μ_{ij} is the mean of Y_{ij} for $j = 1, \ldots, 4$ and $i = 1, \ldots, 7$. The geometric and flow variables associated with the jth approach in circle i will be denoted by $\underline{x}_{ij}$. A regression approach to the problem would use the model $Y_{ij} = \underline{x}'_{iy}\underline{\beta} + \varepsilon_{ij}$, and the following assumptions: (a) the Y_{ij}s are normally distributed with mean $\mu_{ij} = \underline{x}'_{ij}\underline{\beta}$, and (b) the Y_{ij}s are independent of one another. Because the number of accidents at an approach is a count variable, with possible values 0, 1, 2, 3, . . . , the normal assumption does not seem valid. An alternative analysis, pro-

A. Al-Bakri and J. Stoner, Department of Civil and Environmental Engineering, University of Iowa, Iowa City, Iowa 52242. M. R. Conaway, Department of Statistics and Actuarial Science, University of Iowa, Iowa City, Iowa 52242.

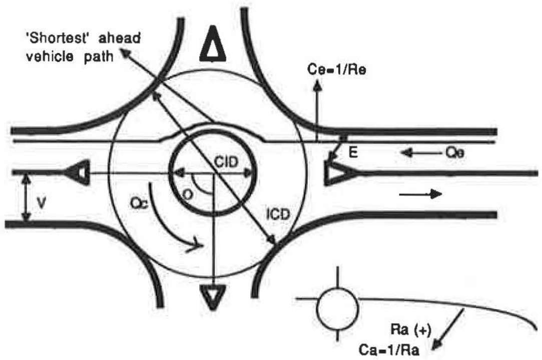

FIGURE 1 Geometric characteristics of traffic circles.

posed by Hauer (*3*), is based on the assumption that the number of accidents at each approach has a Poisson distribution with mean μ_{ij}, with $\ln(\mu_{ij}) = \underline{x}'_{ij}\underline{\beta}$. Hauer fit generalized linear models to estimate $\underline{\beta}$, the effects of the geometric and flow variables. This is based on a more realistic assumption about the distribution of the accident counts than is the regression method, but it ignores the fact that the observations from the same circle are likely to be dependent, even after adjusting for geometric and flow variables. In addition, as noted by Maycock and Hall, one would expect that some traffic circles are inherently more dangerous than others, in the sense that they have more accidents than others with similar geometric and flow characteristics.

Maycock and Hall account for the inherent differences among traffic circles by assuming that the number of accidents has a Poisson distribution with mean q, where q has a gamma distribution with parameters μ and S. The parameter μ depends on the geometric and flow variables through the model $\ln(\mu) = \underline{x}'\underline{\beta}$. Averaging over the assumed gamma distribution, this model states that the number of accidents has a negative binomial distribution with mean μ and variance $\mu(\mu + S)/S$. Maycock and Hall outline a procedure for estimating S then use generalized linear modeling to estimate $\underline{\beta}$.

Several features make the Maycock and Hall method inappropriate for this data. If the differences among traffic circles are assumed to be random, then averaging over the distribution of the random effects induces a dependence among the four observations within a circle. The generalized linear model approach of Maycock and Hall would not take this into account; it would treat all the observations as if they were independent. A second problem that arises in this study, though not in the Maycock and Hall study, is that there are only seven traffic circles to be used in estimating the parameters of the assumed gamma distribution. With such a small sample, it is difficult to check the assumption of a gamma distribution; even if the assumption were correct, it would be difficult to obtain reliable estimates of the parameters of the gamma distribution. In this study, these problems will be solved with a conditional analysis, which will allow inferences to be drawn about the effect of the approach-specific geometric and flow characteristics without requiring the specification of a distribution to represent the inherent differences among traffic circles. This conditional analysis yields valid inferences about the effects of the geometric and flow variables under a wide variety of possible distributions, including common distributions such as the gamma or log-normal distributions. Being able to obtain valid inferences under a variety of distributions may be important, because the effect of misspecifying the distribution—that is, assuming a gamma distribution when the true distribution is not gamma—is not yet known.

The theoretical justification of the conditional approach will be given in the next section. In this section, the basic model underlying the conditional approach will be outlined and the interpretation of the parameters in the model will be discussed. Following the earlier notation, Y_{ij} represents the accident count in approach j of circle i, $j = 1, \ldots, 4$ and $i = 1, \ldots, 7$, and $\underline{x}_{ij}$ represents the geometric and flow variables associated with this approach. To represent the inherent differences among traffic circles, a random quantity (θ_i) is associated with the ith circle. The model is that, given θ_i, the accident counts in the four approaches (Y_{i1}, Y_{i2}, Y_{i3}, Y_{i4}) are independent Poisson variables, with mean $\theta_i\mu_{ij}$. The parameter μ_{ij} depends on the geometric and flow variables through $\ln(\mu_{ij}) = \underline{x}'_{ij}\underline{\beta}$. With this model it is assumed that the effect of the ith circle is to multiply the mean number of accidents in the four approaches by the same factor (θ_i).

Figure 2 illustrates the model for a traffic circle with the associated effect θ_i. If θ_i were averaged over some assumed distribution, the common factor of θ_i in each of the four approaches would induce an association among the accident counts within the same circle, even after adjustments for the geometric and flow variables.

The multiplicative model yields $E(Y_{ij}) = E[E(Y_{ij}|\theta_i)] = \mu_{ij} E(\theta_i)$, and if an intercept term is included in the model for the $\ln(\mu_{ij})$, then the model can be reparametrized so that $E(\theta_i) = 1$. With this reparametrization, $E(Y_{ij}) = \mu_{ij}$. The unconditional variance of Y_{ij} is given by $\mathrm{Var}(Y_{ij}) = \mu_{ij} + \mu_{ij}^2 \mathrm{Var}(\theta_i)$. The covariance between two accident counts within the same circle is $\mathrm{Cov}(Y_{ij}, Y_{ik}) = \mu_{ij}\, \mu_{ik}\, \mathrm{Var}(\theta_i)$. The multiplicative model is the simplest model that allows for additional variability in the Y_{ij} because of inherent differences in the circle, and it allows observations within the same circle to be associated.

The fundamental idea behind the conditional analysis is to consider the conditional distribution of the four accident counts (Y_{i1}, Y_{i2}, Y_{i3}, Y_{i4}), given the total number of accidents at that traffic circle ($Y_{i+} = Y_{i1} + Y_{i2} + Y_{i3} + Y_{i4}$). Analyzing Poisson variables by conditioning on their sum is a standard statistical technique used in a number of applications [compare McCullagh and Nelder (*4*)]. One of the attractive features of the

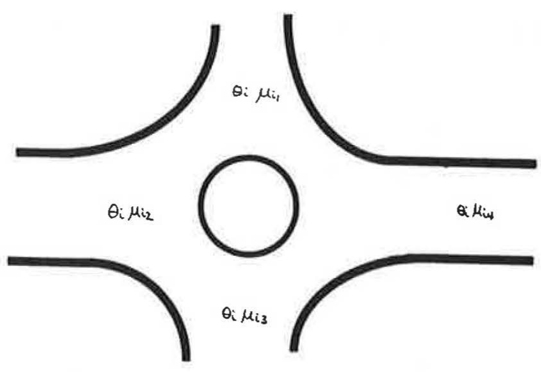

FIGURE 2 Expected number of accidents in each approach.

conditional analysis is that the resulting distribution is multinomial, with parameters y_{i+} and $(p_{i1}, p_{i2}, p_{i3}, p_{i4})$, where

$$p_{ij} = \frac{\exp(\underline{x}_{ij}'\underline{\beta})}{\sum_k \exp(\underline{x}_{ik}'\underline{\beta})} \tag{1}$$

This has the form of a multinomial logit model [Agresti (5)], which facilitates the interpretation of the parameters and allows for the computation of the conditional estimate of $\underline{\beta}$ in a standard statistical package such as GLIM (6). To illustrate the interpretation of the parameters in the conditional analysis, consider Approaches 1 and 2 in a particular traffic circle and suppose that the approaches are identical in all the geometric and flow characteristics except the one measured by the predictor x_{ijm}. In the conditional distribution, $p_{i1}/p_{i2} = \exp[\beta_m(x_{i1m} - x_{i2m})]$, so that the expected number of accidents in Approach 1 would be $\exp[\beta_m(x_{i1m} - x_{i2m})]$ more than expected in Approach 2.

The conditional estimate of $\underline{\beta}$ can be computed in GLIM by specifying that the Y_{ij} have Poisson distributions and using the model $\ln(\mu_{ij}) = \underline{x}_{ij}'\underline{\beta} + \Sigma_i \delta_i I_i$, where I_i is an indicator variable for the ith circle. Note that the model for the mean, μ_{ij}, has the form of a parallel regressions model. This illustrates one of the drawbacks of the conditional analysis. The effect on accidents of changing the geometric and flow variables can be estimated, but the actual value for the mean number of accidents that would occur at a traffic circle with given geometric and flow characteristics cannot be predicted.

DETAILS OF CONDITIONAL ANALYSIS

The theoretical justification for using the conditional estimates will be outlined. As before, let Y_{ij} be the number of accidents in approach j, $j = 1, \ldots, 4$, of circle i, and let Y_{i+} be the total number of accidents at circle i, $i = 1, \ldots, 7$. The random quantity θ_i is associated with the ith circle, and the conditional distributions of the accident counts $(Y_{i1}, Y_{i2}, Y_{i3}, Y_{i4})$, given θ_i, are assumed to be independent Poisson random variables with means $\theta_i\mu_{ij}$, $j = 1, \ldots, 4$. In addition, assume that θ_i is sampled from a population with density g.

As in Maycock and Hall (2), the distribution of the accident counts in the ith circle is considered averaging over the distribution of the random effect θ_i. Let $L_i(y_{i1}, y_{i2}, y_{i3}, y_{i4}; \underline{\beta}) = P(Y_{i1} = y_{i1}, Y_{i2} = y_{i2}, Y_{i3} = y_{i3}, Y_{i4} = y_{i4} \mid \underline{\beta})$ be the likelihood from circle i, averaging over the distribution of the θ_i. The likelihood is

$$L_i(\underline{y}_i; \underline{\beta}) = \int P(Y_{i1} = y_{i1}, Y_{i2} = y_{i2}, Y_{i3} = y_{i3}, Y_{i4} = y_{i4} \mid \theta_i) g(\theta_i)\, d\theta_i \tag{2}$$

$$= \int_{y_{i+}} \sum P(Y_{i1} = y_{i1}, \ldots, Y_{i4} = y_{i4} \mid Y_{i+}, \theta_i) h(Y_{i+} \mid \theta_i) g(\theta_i)\, d\theta_i \tag{3}$$

Because Y_{i+} is a sufficient statistic for θ_i, the conditional distribution of $(Y_{i1}, Y_{i2}, Y_{i3}, Y_{i4})$ given Y_{i+} is free from θ_i. This conditional distribution is given by

$$P(Y_{i1} = y_{i1}, Y_{i2} = y_{i2}, Y_{i3} = y_{i3}, Y_{i4} = y_{i4} \mid Y_{i+}, \theta_i)$$

$$= \prod_j \left[\frac{\exp(\underline{x}_{ij}'\underline{\beta})}{\sum_k \exp(\underline{x}_{ik}'\underline{\beta})}\right]^{y_{ij}} \quad \text{if } \sum_j Y_{ij} = y_{i+}$$

$$= 0 \quad \text{if } \sum_j Y_{ij} \neq y_{i+} \tag{4}$$

Although Equation 3 is written in terms of a sum over the values of Y_{i+}, Equation 4 indicates that only one term in the sum is nonzero. From this expression for the conditional distribution of $(Y_{i1}, Y_{i2}, Y_{i3}, Y_{i4})$ given Y_{i+}, the following expression can be written:

$$L_i(\underline{y}_i; \underline{\beta}) = \prod_j \left[\frac{\exp(\underline{x}_{ij}'\underline{\beta})}{\sum_k \exp(\underline{x}_{ik}'\underline{\beta})}\right]^{y_{ij}} * \int h(Y_{i+} \mid \theta_i) g(\theta_i)\, d\theta_i \tag{5}$$

Because the observations at different traffic circles are assumed to be independent, the likelihood from the sample is the product of the likelihoods from each circle and can be written as

$$L = \prod_i \prod_j \left[\frac{\exp(\underline{x}_{ij}'\underline{\beta})}{\sum_k \exp(\underline{x}_{ik}'\underline{\beta})}\right]^{y_{ij}} * \prod_i \int h(Y_{i+} \mid \theta_i) g(\theta_i)\, d\theta_i \tag{6}$$

The distribution of the random effect enters only through the second factor of the likelihood, and an estimate of $\underline{\beta}$ can be computed without a distribution for g being specified, by maximizing the first factor only. The resulting estimates of $\underline{\beta}$ are known as conditional maximum likelihood estimates and possess a number of desirable statistical properties [Andersen (7)]. These properties depend on the size of the Y_{i+}, not on the number of circles in the sample. This makes the conditional approach particularly well suited for these data. There are few traffic circles in the sample, but a fairly large number of accidents were observed at each one.

It should be noted that there is some loss of information about $\underline{\beta}$ in doing the conditional procedure. Because the distribution of Y_{i+} depends on $\underline{\beta}$, ignoring the second factor of the likelihood ignores some of the information about $\underline{\beta}$. Computing the distribution of Y_{i+}, however, reveals that it depends on $\underline{\beta}$ only through the quantity $\mu_{i+} = \Sigma_j \exp(\underline{x}_{ij}'\underline{\beta})$. In trying to estimate "within circle" effects, this should contain little information about the amount of information about $\underline{\beta}$ in the conditional distribution, so that the loss of information from ignoring the second factor should not be large. Precisely how much is lost in the conditioning is difficult to answer, because the amount of information lost depends on the true distribution of the θ_i. A more important feature to note is that, with the conditional procedure, the effects of factors common to all the approaches of a traffic circle cannot be estimated. These disadvantages of the conditional procedure must be balanced against the gains made in estimating the effects of approach-specific factors and in guarding against biases that can result from misspecifying the distribution for the θ_i.

RESULTS OF ANALYSIS

Before the results of the conditional analysis are presented, a brief description of how one can select a model using generalized linear models will be given. A comprehensive treatment of this topic is found in McCullagh and Nelder (*4*). A description of fitting generalized linear models for analyzing traffic data is given in Maycock and Hall (*2*).

Besides providing estimates for the coefficients of the predictors in the model, GLIM also provides a way of checking the fit of the model and of checking whether or not a predictor is a significant addition to a model. These measures are based on a quantity known as the deviance, which, for the Poisson models, equals

$$2 * \sum_{ij} y_{ij} \ln(y_{ij}/\hat{m}_{ij})$$

where $\hat{m}_{ij}$ is an estimate, based on the model being fit, of the expected number of accidents in approach j of circle i. Associated with the deviance is the number of degrees of freedom, which equals the number of observations minus the number of parameters being estimated. A model that describes the data should have a deviance approximately equal to the number of degrees of freedom.

To check if a predictor is a significant addition to a model, compare the deviances from two models: one without the predictor and one after the predictor has been added to the model. For example, suppose a model includes only the logarithm of the entering volume (LNQE) as a predictor. Let d_1 represent the deviance found by fitting this model. To see if another predictor—say, CA—significantly improves the model, obtain the deviance (d_2) that results from using both LNQE and CA as predictors. If the difference ($d_1 - d_2$) is large, compared with a chi-squared distribution with 1 degree of freedom, then it would be concluded that CA is a significant addition to the model. This procedure can be thought of as the generalized linear model version of doing an *F*-test to determine whether a predictor is a significant addition to a regression model.

One of the analyses was performed with total accidents as the dependent variable, where total accidents include entering, approaching, single-vehicle, pedestrian, and other accidents. The most parsimonious model that fit the data included terms for LNQE, CA, and CE. None of the other available predictors significantly improves the model. Deleting the flow or either of the curvature variables results in a model that fits significantly worse than the chosen model. The conditional estimates from the model that includes the predictors CA, CE, and LNQE are given in Table 1. The estimates were computed with and without an outlying observation.

The deviance associated with this model is 26.2 on 18 degrees of freedom (p = .1). The fit of this model is adequate, but not particularly good, primarily because of one large residual, corresponding to Arm 3 in the traffic circle R6. This arm has one of the smaller values of flow in the data set, and one would expect fewer accidents here than the 15 that were

TABLE 1 ESTIMATES FOR FINAL MODEL

With case 12			Without case 12		
Estimate	Std. Error	Predictor	Estimate	Std. Error	Predictor
-3.67	1.26	intercept	-5.261	.46	intercept
0.73	0.15	LNQE	.89	.16	LNQE
45.04	17.4	CA	55.86	18.2	CA
159.8	36.9	CE	150.0	37.9	CE

observed. This arm has six single-vehicle accidents, an unusually large number that contributes to the large total number of accidents. No other arm has more than two single-vehicle accidents. Deleting this observation and refitting the model yields a deviance of 17.5 on 17 degrees of freedom.

One way to see how to interpret these estimates is to consider two approaches in the same traffic circle and suppose that Approach 1 has a log of entering flow of LNQE_1, an entry curvature of CE_1, and an approach curvature of CA_1. Similarly, Approach 2 has values of LNQE_2, CE_2, and CA_2 for these characteristics. Using the estimates from the final model (without Case 12), we have $\hat{p}_1/\hat{p}_2 = \exp[.89 * (\text{LNQE}_1 - \text{LNQE}_2) + 150.0 * (\text{CE}_1 - \text{CE}_2) + 55.86 * (\text{CA}_1 - \text{CA}_2)]$.

To interpret the estimated coefficient (.89) for LNQE, suppose that the two approaches have the same approach curvature and entry curvature, but Approach 1 has an entering flow of α ($\alpha \geq 1$) times the entering flow of the Approach 2. In other words, $\text{LNQE}_1 = \ln(\alpha) + \text{LNQE}_2$, $\text{CE}_1 = \text{CE}_2$, and $\text{CA}_1 = \text{CA}_2$. For these approaches,

$$\hat{p}_1/\hat{p}_2 = \exp[.89 * \ln(\alpha)] = \alpha^{.89} \quad (7)$$

so that the estimate of accidents in Approach 1 would be $\alpha^{.89}$ times the number of accidents in Approach 2. Note that the standard error of the estimate for the coefficient of LNQE is .16, so that the estimate of .89 is not inconsistent with the hypothesis that the parameter equals 1. A value of 1 for the parameter is intuitively appealing, because this means that on average the number of accidents, adjusting for the other factors, changes in direct proportion to the traffic volume.

Similar calculations can be done to interpret the coefficients of the other predictors in the model. CE is measured in units of .001 and has an estimated coefficient of 150.0. This indicates that for two arms with identical flows and approach curvatures, 1.16 times the number of accidents would be expected in an arm with entry curvature CE + .001 as would be expected in an arm with entry curvature CE. CA is also measured in units of .001; its estimated coefficient is 55.86. From this, 1.06 times the number of accidents would be expected in an arm with approach curvature CA + .001 as would be expected in an arm in the same circle with approach curvature CA, if the flow and entry curvature were held constant.

The outlier primarily affects the estimate of the coefficient of LNQE. This might be expected given that Case 12 has a relatively small flow but an unusually large number of accidents. From the estimates that include Case 12, 1.17 times the number of accidents would be expected for an arm with entry curvature CE + .001 as would be expected for an arm with curvature CE. Similarly, 1.05 times the number of accidents would be expected in an arm with approach curvature of CA + .001 as for CA. These estimates are close to those based on deleting Case 12.

These findings are consistent with those of Maycock and Hall (*2*), who also found that CA, CE, and LNQE were important predictors of total accidents.

CONCLUSIONS

In this paper, a conditional analysis was used to relate the number of traffic accidents at traffic circles to the geometric and flow characteristics of the circles. The conditional analysis is based on a Poisson distribution for the accident counts and is particularly suited for these data. It can give valid estimates of the effects without a large number of traffic circles in the sample and without a specified particular distribution for the variability among the circles. This can protect against biases that might result from misspecifying this distribution. All the computations can be done with readily available statistical software, which also provides methods for checking the adequacy of the model and the effect of adding predictors to the model. Applying the method to data collected in Amman, Jordan, it was found that, besides traffic volume, the entry and approach curvatures were important factors in determining the number of accidents.

ACKNOWLEDGMENTS

Part of this work was completed while Mark Conaway was visiting assistant professor at the department of biostatistics at the Harvard School of Public Health and is partially supported by a grant from the National Institutes of Health. Ala Al-Bakri was partially supported by a quarter-time research and teaching assistantship at the University of Iowa and by the Jordanian Royal Endowment for Culture and Education, chaired by Queen Noor Al Hussein, and the Public Security Directorate.

REFERENCES

1. A. Al-Bakri. *Investigation of the Effectiveness of Selected Intersection Transportation System Management Techniques and Traffic Enforcement Levels in a Developing Country: Jordan Case Study.* Ph.D. dissertation. Department of Civil and Environmental Engineering, University of Iowa, Iowa City, 1990.
2. G. Maycock and R. D. Hall. *Accidents at 4-Arm Roundabouts.* Report 1120. U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1984.
3. E. Hauer, J. Ng, and J. Lovell. Estimation of Safety at Signalized Intersections. In *Transportation Research Record 1185*, TRB, National Research Council, Washington, D.C., 1988.
4. P. McCullagh and J. A. Nelder. *Generalized Linear Models*, 2nd ed. Chapman and Hall, London, England, 1989.
5. *GLIM 3.77.* Numerical Algorithms Group, Inc., Downers Grove, Ill.
6. A. Agresti. *Categorical Data Analysis.* John Wiley & Sons, New York, N.Y., 1990.
7. E. B. Andersen. Asymptotic Properties of Conditional Maximum Likelihood Estimators. *Journal of the Royal Statistical Society*, Series B, Vol. 32, 1973, pp. 283–301.

Publication of this paper sponsored by Committee on Methodology for Evaluating Highway Improvements.