The caution extended by SAS Institute Inc. to users of their categorical model procedure, CATMOD, does not clearly identify the risks involved when analyzing models with small sample size. A SAS program was written and executed to assess these risks by variation of three binomial parameter values under the null hypothesis (P = 0.1, 0.3, 0.5) with constant sample sizes of 2 to 120. For sample sizes less than twenty, the 4-dimensional sample space for a 2x2 categorical model was formed and the coordinates of each point used to construct cell frequencies for a categorical table with two factors and one binomial response. The significance of the interaction term in these tables was determined by use of CATMOD. For each table that yielded a \( \chi^2 \) value of 3.84 or greater, the binomial probability function of SAS (PROBNML) was used to calculate the probability associated with point in the sample space. The cumulative probability of the significant tables represents the true size of the rejection region. For sample sizes greater than twenty, a simulation method was used for each of three null hypotheses to estimate the size of the critical region as the proportion of the sampled 2000 tables which yielded \( \chi^2 \) values of 3.84 or greater. The results demonstrated that the Type I error rate is less than the nominal \( \alpha \) of 0.05 for sample sizes less than 20 when using the default options of SAS. Conversely, the joint response function produced a critical region that was more often greater than 0.05. Different additions to empty cells had the greatest effect when sample sizes were less than twenty.

## Introduction

Categorical data is information that is classified into discrete categories of nominal or ordinal scales. Observations of subjects under one or more categorical variables can be represented by contingency tables of one or more dimensions. When more than two categorical variables are present, a distinction can be made between independent and dependent variables. Profiles based on independent variables are called population profiles and those based on the dependent variables are called response profiles. Any one subject can then be classified into one population profile and one response profile. This is the type of arrangement that is used to analyze categorical data as a linear model.

The use of linear modeling for the analysis of categorical data is a recent development that is gaining popularity among researchers. The theoretical justification for the linear model approach was developed by Wald (1943), Neyman (1949), and Grizzle, Starmer, and Koch (1969). This increase in use of the linear model approach is part due to the availability of computer software designed for such analysis. One such computer program is the SAS procedure CATMOD (1985).

Hypotheses about population parameters in a categorical model are tested with the generalized Wald statistic, which is asymptotically distributed as a chi-square (\( \chi^2 \)). This asymptotic property can present problems to researchers, since the behavior of the statistic is not known when computed for small sample sizes. Computational difficulties occur when transformations, such as the logit or the cell proportion, are performed on the response probabilities in the presence of one or more empty cells. Furthermore, the effect that addition of a small value to empty cells has on the Type I error rate has not been investigated.

The purpose of this paper is to determine the Type I error rate of CATMOD when the expected size of the critical region is equal to 0.05 for sample sizes of two to one hundred twenty observations. For sample sizes less than twenty, the assessment is made by summing the probabilities of all points of the sample space that fall into the rejection region. For larger sample sizes, a simulation study is conducted to estimate the Type I error rate. The effect that the addition of various small values to empty cells has on the Type I error rate is also investigated and a comparison is made between the default logit response function and the joint response function.

## Methods

A categorical model with two independent factor variables and one binomial response variable was used for this investigation. The factor variables, designated GROUP and TREATMENT, each assume values of 1 or 2, while the dependent response variable assumes values of success or failure. This model can be represented as four populations with two response categories. The row margins corresponding to the population sizes were held constant within a sample space so that \( n_1 - n_2 - n_3 - n_4 = n \), where \( 1 < n < 120 \). The frequency table for this model is given by the following:

<table>
<thead>
<tr>
<th>POPULATIONS (S)</th>
<th>RESPONSE CATEGORY</th>
</tr>
</thead>
<tbody>
<tr>
<td>GROUP TREATMENT</td>
<td>FAILURE SUCCESS</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>4</td>
</tr>
</tbody>
</table>

The frequencies in such a table are used to calculate estimates of the population parameters \( \pi_j \), the proportion of subjects in the \( j \)th population exhibiting the \( k \)th level of response. The estimate of \( \pi_j \) is \( \hat{\pi}_j = \frac{n_{jk}}{n} \), and \( n_j = n - \sum_{k} n_{jk} \).

As an illustration, consider the 4x4 sample space of contingency tables within GROUP 1 having fixed sample sizes of three observations each. This sample space is presented in the following display:
The probability of any point of this sample space is the product of independent binomial probabilities. Under a null hypothesis, eg., \( H_0: \pi_{12} = \pi_{22} = \pi_{32} = \pi_{42} = 0.3 \), the designated point (2,1) of the GROUP 1 sample space will occur with a probability given by the following:

\[
\text{Prob}(2,1) = (0.3) (0.7)^{2-2} \times (0.3) (0.7)^{2-1} = 0.083349
\]

Given the above sample point for GROUP 1, there exists a 4×4 companion sample space for GROUP 2 which is presented in the following display:

The complete model, the designated points (2,1) and (3,1) in the GROUP 1 and 2 sample spaces, respectively, define a single point in the four dimensional sample space of 256 possible points. These two points are combined to produce the following contingency table:

<table>
<thead>
<tr>
<th>POPULATIONS (S)</th>
<th>RESPONSE CATEGORY</th>
</tr>
</thead>
<tbody>
<tr>
<td>GROUP TREATMENT</td>
<td>FAILURE</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

Under \( H_0: \pi_{12} = \pi_{22} = \pi_{32} = \pi_{42} = 0.3 \), the probability of the complete model frequency table is the product of the within GROUP probabilities. For the above table, this probability is given by the following:

\[
\text{Prob}(2,1,3,1) = \text{Prob}(2,1) \times \text{Prob}(3,1) = 0.000992436
\]

With this enumeration approach, all possible contingency tables were generated for each constant sample size \( n \) of the populations. Each table was analyzed using the CATMOD procedure with the following model specification statement:

\[
\text{PROC CATMOD;}
\text{MODEL RESPONSE = GROUP TREATMENT;}
\text{RESPONSE LOGIT;}
\text{RESPONSE JOINT;}
\]

As SAS processed each possible table from the sample space, the calculated \( \chi^2 \) values were read from the output. For each table that produced a \( \chi^2 \) value greater than \( 3.84 \) for the interaction effect, the SAS function PROBBNML was called to calculate the probability of that table under each of three null hypotheses: \( H_0: \pi_{12} = \pi_{22} = \pi_{32} = \pi_{42} = 0.1, 0.3, \) or \( 0.5 \). These probabilities were accumulated over the sample space and thus represent the actual Type I error rate for the test of interaction.

For samples sizes greater than twenty, a simulation method was used to select a random sample of 2000 points from the sample space under the three null hypotheses above. The 2000 points were selected using the UNIFORM function to sample from an array that had element values in proportion to their expected binomial frequency.

The 2000 sample space points were used to construct frequencies for the categorical tables. These tables were processed using the same model statement as above. In this manner, the same points were analyzed with both the joint and the logit response functions. The proportion of tables that produced significant values was used as the estimate of the size of the critical region. Three independent estimates of the critical region were obtained and averaged to produce an accurate estimate.

For both enumeration and simulation methods, the Type I error rate obtained from weighted least squares (WLS) estimation with the logit response function, was compared with that obtained from the joint response function. Additional analyses were conducted that involved addition of small values of 0.01 or 0.5 to empty cells prior to analysis.

RESULTS

With the enumeration approach and the default logit response function, in which SAS adds 0.5 to empty cells, the size of the rejection region was below 0.05 for all sample sizes when the null parameter was 0.3 or 0.1 (Figure 1.). With a null parameter of 0.5, CATMOD approached an alpha size test when the sample size was greater than nine yet the size of the region was not consistent with increasing sample sizes.

The addition of 0.01 to empty cells when using the logit response function, resulted in a critical region that was smaller than that of the default addition of 0.5 (Figure 2). The addition of different small values to empty cells had the greatest effect when sample sizes were less than nine for null parameter of 0.5 and less than fifteen for null parameter of 0.3.
In contrast to the logit function, the use of the joint response function with 0.5 added to empty cells yielded a critical region that exceeded the expected size of 0.05 for most values of the null parameter and sample size (Figure 3). Only for the parameter value of 0.1 did the critical region fail to achieve a level of 0.05.

The choice of a constant for addition to empty cells produced variable effects on the Type I error rate when the joint response function was requested. For null parameter of 0.5, the addition of 0.01 produced a critical region size that was much larger than that associated with the addition of 0.5 for sample sizes greater than seven (Figure 4). The greatest effect of the different additions is seen when the population parameter is 0.1 (Figure 5). For all null parameters, the addition of 0.01 produced a Type I error rate that was generally greater than that for the addition of 0.5.

The results of the simulation approach indicate that when using the logit function, the size of the critical region approached the nominal 0.05 when sample sizes were greater than forty and the null parameter value was 0.3 or 0.5 (Figure 6). An alpha size test is approximated when sample sizes approached 120 with a null parameter of 0.1. For the joint response function, the critical region remained greater than the expected size of 0.05 until sample sizes reached eighty (Figure 7). Since there was little difference in the size of the critical region associated with the addition of the two small values to empty cells for either response function at these sample sizes, no illustration is given.

DISCUSSION

Our results indicate that for the categorical model studied, the CATMOD procedure is conservative for the test of an interaction effect when using the default logit response function with small sample sizes. For sample sizes less than twenty, the Type I error rate was below the expected five percent for most combinations of sample size and null parameter value. The procedure is especially conservative for small values of the null parameter. The value of 0.01 added to empty cells when using the logit response function produced a slightly reduced critical region size compared to that of the SAS default addition of 0.5. The most notable difference between the addition of the two values occurred when sample sizes of the populations were less than nine.

Specification of the joint response function produces critical regions not only larger than those associated with the logit but often greater than the nominal value of 0.05. This would indicate an overly liberal test. Furthermore, the Type I error rate increases as the value of the constant added to empty cells is decreased. The effect of the value added to empty cells becomes more profound when the null parameter is less than 0.5.

Sample size needs to be twenty or greater to achieve an alpha size test when using the logit response function with null parameter of 0.3 or 0.5. With null parameters of 0.3 or 0.5, the joint response function consistently produced a critical region size greater than 0.05 until sample size exceeded sixty. With a small population proportion, such as 0.1, the sample sizes needs to be at least sixty, that is 240 total observations, to achieve an alpha size critical region when using the logit response function; whereas the sample size needs to be forty, that is 160 total observations, when using the joint response function.

In the CATMOD chapter of the SAS Statistics manual, SAS extends a caution to users that leaves the reader with the impression that all CATMOD results are invalid for small sample sizes. This caution may be misleading to researchers. Our results indicate that caution should be used in interpreting nonsignificant results when using the default logit response function, but that significant results should not be dismissed as invalid. Conversely, significant results obtained when using the joint response function should be viewed with caution, since the size of the critical region is much larger than the nominal value of 0.05 for sample sizes less than sixty.

REFERENCES


Wald, A. [1943]. Tests of statistical hypotheses concerning several parameters when the number of observations is large. Trans. Amer. Math Soc. 54, 426-482.
A COMPARISON BETWEEN TWO VALUES ADDED TO EMPTY CELLS

FIGURE 2. A COMPARISON BETWEEN TWO VALUES ADDED TO EMPTY CELLS WHEN USING THE LOGIT FUNCTION

FIGURE 3. ENUMERATION METHOD USING THE JOINT RESPONSE FUNCTION AND 0.5 ADDED TO EMPTY CELLS

FIGURE 4. A COMPARISON BETWEEN TWO VALUES ADDED TO EMPTY CELLS WHEN USING THE JOINT RESPONSE FUNCTION

FIGURE 5. A COMPARISON BETWEEN TWO VALUES ADDED TO EMPTY CELLS WHEN USING THE JOINT RESPONSE FUNCTION

FIGURE 6. SIMULATION METHOD USING THE LOGIT RESPONSE FUNCTION WITH SAS DEFAULT OF 0.5 ADDED TO EMPTY CELLS

FIGURE 7. SIMULATION METHOD USING THE JOINT RESPONSE FUNCTION WITH 0.5 ADDED TO EMPTY CELLS