ANALYSIS OF MARKETING EXPERIMENTS WITH A CATEGORICAL RESPONSE VARIABLE

Richard K. Burdick, Arizona State University
Bruce J. Walker, Arizona State University

Abstract

The analysis of experimental data with a qualitative response variable is a problem confronted in many marketing research studies. This paper illustrates a methodology based on generalized least squares that properly handles a response variable that is discrete. The methodology is illustrated on several sets of marketing data and software packages that can be used for computations are noted.

Introduction

The analysis of categorical data is receiving increased attention in the marketing literature. In particular, recent articles by Green, Carmone, and Wachspress (1977) and Flath and Leonard (1979) have illustrated the usefulness of logit-type models in analyzing qualitative marketing data. The purpose of this paper is to illustrate a methodology developed by Grizzle, Starmer, and Koch (1969) that can be used to analyze categorical data in which a response variable can be clearly defined. This paper illustrates the methodology using data collected in a marketing research study where the response variable is dichotomous. The extension of the methodology to polytomous response variables is illustrated and software packages that can be used for computations are noted.

A Marketing Research Example

A study was conducted by Walker (1981) to determine the effect of various incentives on the response rate in a survey. Customers of a national family steak house chain were given a questionnaire when they paid their bill and were requested to mail it to a given address upon completion. Six different incentive treatments were employed in the experiment: A coupon for a free beverage upon the next visit to the steak house; 25¢ donation to a charity specified by the steak house; 25¢ cash; 50¢ cash; 50¢ promised upon completion and return of the questionnaire; and no incentive. The first four incentives were attached to the questionnaire and were received by the customer even if the questionnaire was not returned. The data for the experiment are shown in **Table 1(a)**.

TABLE 1
RESULTS OF WALKER INCENTIVE EXPERIMENT

(a) Frequency Count

Incentive	Number of Non-Responses	Number of Responses
1	61	26
2	57	37
3	59	34
4	44	52
5	72	24
6	77	18

(b) Estimated Parameters

$$\hat{\beta} = \begin{bmatrix} .6601 \\ .0410 \\ -.0538 \\ -.0257 \\ -.2018 \\ .0899 \end{bmatrix}$$

Incentive	Actual Probability of Non-Response	Predicted Probability of Non-Response
1	.7011	.7011
2	.6063	.6063
3	.6344	.6344
4	.4583	.4583
5	.7500	.7500
6	.8105	.8105

One might wish to analyze this data as a single-factor analysis of variance design and perform an overall F-test to test for equal treatment means. The fact that the response variable (did or did not return questionnaire) is binary, however, violates several assumptions of this analysis. The response variable is a Bernoulli random variable rather than a normal random variable as assumed in the analysis of variance model. In addition, since the variance of a Bernoulli random variable is a function of the mean, the variance of the response variable will be a function of the treatment mean and will therefore be heteroskedastic if treatment means are unequal. Finally, the expected value of the binary response variable is defined to be a probability and must therefore be greater than or equal to zero and less than or equal to one. Neter and Wasserman (1974, pp. 320-335) discuss the consequences of these violations and suggest possible corrective measures. The next section of this paper illustrates an alternative methodology that is appropriate when the response variable is qualitative.

The Grizzle-Starmer-Koch Methodology

A methodology developed by Grizzle, Starmer, and Koch (1969), hereafter referred to as GSK, is based on generalized weighted least squares regression and can be used to analyze contingency tables that have a clearly defined response variable and one or more experimental factors. The GSK methodology is appropriate in cases where the question of interest is whether or not the experimental factors affect the categorical response variable.

To illustrate the methodology, the data are represented in terms of the cell frequencies and the expected cell probabilities. This framework is shown in **Table 2**, where r represents the number of categories for the response variable and s represents the number of treatments.[1] If it is assumed that a random sample of size

[1] For a single-factor design, a treatment is defined to be a factor level. For a multi-factor design, a treatment is defined to be a combination of factor levels.

$n_{i.}$ has been selected from population i where i=1,...,s, then a hypothesis of interest is that of homogeneity among the populations. Expressed in terms of the expected cell probabilities shown in Table 2, the null hypothesis is H_o: $\pi_{1j} = \pi_{2j} = \ldots = \pi_{sj}$, for j=1,...,r-1. That is, the hypothesis of homogeneity states that the probability of response category j is equal for all treatments.

TABLE 2
GENERAL FORMAT FOR DISPLAYING DATA
WITH QUALITATIVE RESPONSE VARIABLE

Cell Frequencies

Population (Treatment)	Response Category 1	2 . . . r	Total
1	n_{11}	n_{12} n_{1r}	$n_{1.}$
2	n_{21}	n_{22} n_{2r}	$n_{2.}$
.	.	. . .	.
.	.	. . .	.
.	.	. . .	.
s	n_{s1}	n_{s2} n_{sr}	$n_{s.}$

Expected Cell Probabilities

Population (Treatment)	Response Category 1	2 . . . r	Total
1	π_{11}	π_{12} π_{1r}	1
2	π_{21}	π_{22} π_{2r}	1
.	.	. . .	.
.	.	. . .	.
.	.	. . .	.
s	π_{s1}	π_{s2} π_{sr}	1

The methodology involves definition of meaningful linear or non-linear functions of the π_{ij} represented as $f_m(\pi)$ for m=1,2,...,u $\leq$ (r-1)s. A model of the form

$$\underline{F}(\pi) = X\,\underline{\beta} \qquad (1)$$

is now fit by replacing π_{ij} with $p_{ij} = n_{ij}/n_{i.}$ and using a weighted least squares procedure where $[\underline{F}(\pi)]' = [f_1(\pi),\ldots,f_u(\pi)]$ is a (1xu) row vector, X is a (uxv) design matrix of rank v$\leq$u, and $\underline{\beta}' = [\beta_0\beta_1\ldots\beta_{v-1}]$ is a (1xv) row vector of unknown parameters. Tests of hypotheses are formulated in terms of the parameters contained in $\underline{\beta}$ by considering the null hypothesis H_o:$C\underline{\beta}$=0, where C is a (dxv) matrix of appropriate constants and of rank d$\leq$v. It should be noted that the form of X is the same as in a univariate analysis of variance model whenever a single function $f_m(\pi)$ is constructed for each of the s populations.

For the Walker data shown in Table 1(a), s=6 and r=2. The hypothesis of homogeneity can be stated as H_o: $\pi_{11} = \pi_{21} = \ldots = \pi_{61}$. In words, this hypothesis states the probability of non-response is equal for all six incentive treatments. A meaningful definition for $f_m(\pi)$ that can be used to test this hypothesis is $f_m(\pi) = \pi_{m1}$ for m=1,...,6 and $[\underline{F}(\pi)]' = [\pi_{11}\ \pi_{21}\ \pi_{31}\ \pi_{41}\ \pi_{51}\ \pi_{61}]$. Notice that this choice is not the only one that can be used to test for homogeneity and other functions such as $f_m(\pi) = \pi_{m2}$ and $f_m(\pi) = \pi_{m1} - \pi_{m2}$ for m=1,...,6 are also appropriate.

Since there is only one response function for each of the six populations, the matrix X is the design matrix used in a single-factor analysis of variance model. If indicator variables are defined in the form

$$X_k = \begin{cases} 1 \text{ if incentive level k,} \\ -1 \text{ if incentive level 6, and} \\ 0 \text{ otherwise} \end{cases} \qquad (2)$$

for k=1,...,5, then

$$X = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & -1 & -1 & -1 & -1 & -1 \end{bmatrix} \qquad (3)$$

with u=6 rows and v=6 columns and $\underline{\beta}' = [\beta_0\ \beta_1\ \beta_2\ \beta_3\ \beta_4\ \beta_5]$. In terms of this model, the six populations are homogeneous if $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$. That is, if $\beta_k = 0$ for k=1,...,5, then $\pi_{i1} = \beta_0$ for all i=1,...,6.

Thus, if C is defined to be the (5x6) matrix

$$C = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \qquad (4)$$

then H_o:$C\underline{\beta}$ = 0 is the hypothesis of homogeneity. This hypothesis is tested using a test statistic which has an asymptotic chi-square null distribution with d degrees of freedom. The calculated chi-square test statistic for the Walker data is 35.18 with d=5 degrees of freedom and the hypothesis of homogeneity is rejected (p<.0001). The estimated parameters in $\underline{\beta}$ are shown in Table 1(b).

As shown in Table 1(b), when u=v the model represents a perfect fit of the cell frequencies. In many instances, however, u>v and the model must be tested for adequate fit. The GSK methodology provides a test for model fit that is based on a test statistic that has an asymptotic chi-square null distribution with u-v degrees of freedom.

For example, in a two-factor experiment, the design matrix may contain only main effects and it may be desired to test whether this non-interaction model provides adequate fit.

As in a single-factor analysis of variance experiment, the rejection of the null hypothesis of equal treatment means requires examination of pairwise treatment differences. For the single-factor design used in the Walker study, each of the 15 [(s)(s-1)/2] pairwise comparisons is represented in the form $C\underline{\beta}$ and tested equal to zero. For example, the probability of non-response is equal for incentive one and incentive two (i.e., $\pi_{11} = \pi_{21}$) if $\beta_1 = \beta_2$, or $\beta_1 - \beta_2 = 0$. This hypothesis can be tested by defining C to be the (1 x 6) vector [0 1 -1 0 0 0], and testing $C\underline{\beta}$ = 0 using the chi-square test statistic with d=1 degree of freedom. If $\pi_{11} = \pi_{61}$, then $2\beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5 = 0$ and C is defined to be the (1 x 6)vector [0 2 1 1 1 1].

The Bonferroni inequality (see, for example, Neter and Wasserman (1974, p. 147)) can be used to provide an overall level of confidence on the entire set of (s)(s-1)/2 pairwise comparisons. For the Walker example, if a family confidence coefficient of .90 is desired, then the hypothesis of equal pairwise probabilities will be rejected for any comparison with a p-value less then (.10/15)=.0066. The pairwise comparisons are summarized in Table 3. It is concluded that treatment 4 will yield a higher response rate (lower non-response

rate) than treatments 1, 5, and 6. In addition, treatments 2 and 3 yield a higher response rate than treatment 6.

TABLE 3
PAIRWISE COMPARISONS FOR WALKER INCENTIVE DATA

Comparison	Degrees of Freedom	Chi-Square	P-Value
1 vs 2	1	1.82	.1779
1 vs 3	1	.91	.3405
1 vs 4	1	11.80	.0006*
1 vs 5	1	.55	.4595
1 vs 6	1	2.97	.0847
2 vs 3	1	.16	.6928
2 vs 4	1	4.28	.0386
2 vs 5	1	4.59	.0321
2 vs 6	1	10.03	.0015*
3 vs 4	1	6.10	.0135
3 vs 5	1	3.00	.0830
3 vs 6	1	7.55	.0060*
4 vs 5	1	18.74	.0001*
4 vs 6	1	29.51	.0001*
5 vs 6	1	1.03	.3110

* = Significant at a .10 family confidence level ($p < .0066$).

Incentive	4	2	3	1	5	6
Non-response Rate	.4583	.6064	.6344	.7011	.7500	.8105

Underscore represents treatments that are not significantly different.

The GSK methodology can also be used to fit the weighted least squares logit model described by Green et al. (1977). In order to fit the logit model, $\underline{F}(\pi)$ is defined to contain the non-linear functions $f_m(\pi) = \ln [\pi_{m1} / \pi_{m2}]$ for m=1,...,6. The advantage of the logit function over the linear function used previously is that the predicted cell probabilities are constrained to the range of zero to one. For the Walker data, the chi-square test statistic used to test the incentive factor in the logit model is 30.90 with 5 degrees of freedom.

As mentioned by Flath and Leonard (1979), there are instances where the maximum likelihood logit model is preferred to the weighted least squares model. In general, the weighted least squares model requires many observations per cell. Kleinbaum and Kupper (1978, p. 459) recommend cell sizes of at least 10 and preferably 25 observations or more. For experiments such as the Walker study in which these requirements are met, there should be little practical difference between the two models.

Extensions to More Than One Factor and More Than Two Response Categories

Experimental designs with more than one factor can be handled in the GSK methodology by suitable modifications of the design matrix. As an example, consider the data reported by Green et al. (1977) shown in Table 4(a). This data consists of three factors, income (I), education (E), and mobility (M), where each factor has two levels. The response variable is dichotomous and represents whether or not a resident adopted a particular telecommunications service. A proposed model might include the three first-order interaction terms (IE, IM, and EM) and the second-order interaction term (IEM).

TABLE 4
RESULTS OF GREEN EXPERIMENT

(a) Frequency Count*

Treatment	Number of Adopters	Number of Non-Adopters
I_1 E_1 M_1	153	2007
I_1 E_1 M_2	226	911
I_1 E_2 M_1	61	825
I_1 E_2 M_2	233	858
I_2 E_1 M_1	147	1216
I_2 E_1 M_2	139	408
I_2 E_2 M_1	287	1638
I_2 E_2 M_2	382	1033

I_1 = Income level $\leq$ \$12,500
I_2 = Income level > \$12,500
E_1 = High School Level Education or Below
E_2 = Some College Education or Above
M_1 = Non-Mobile
M_2 = Mobile

*The cell frequencies have been estimated from the percentages reported by Green.

(b) Estimated Parameters and "ANOVA" Table

$$\hat{\underline{\beta}} = \begin{bmatrix} 1.7253 \\ .2437 \\ .0641 \\ .5330 \\ -.0492 \\ .0872 \\ .0210 \\ -.0513 \end{bmatrix}$$

Source	Degrees of Freedom	Chi-Square	P-Value
I	1	62.77	.0001
E	1	4.35	.0371
M	1	300.16	.0001
IE	1	2.55	.1101
IM	1	8.04	.0046
EM	1	.47	.4944
IEM	1	2.78	.0956

If indicator variables are defined as

$$X_k = \begin{cases} 1 \text{ if first level of factor k, and} \\ -1 \text{ if second level of factor k} \end{cases} \quad (5)$$

for k=1, 2, 3, then

$$X = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & 1 & -1 & -1 & -1 \\ 1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 \end{bmatrix} \quad (6)$$

is the appropriate design matrix.[2] The logit function is fit by defining $f_m(\pi) = \ln[\pi_{m1} / \pi_{m2}]$ for m=1,...,8.

[2]This particular coding for the indicator variables is used in the SAS procedure FUNCAT and is also recommended by Neter and Wasserman (1974, p. 633) for a model that contains interaction terms.

The results shown in Table 4(b) indicate that the IE, EM and IEM interactions are not significant and can be dropped from the model. If these three terms are deleted, the other four terms (I, E, M, and IM) all have p-values less than .0017 and the goodness of fit test statistic has a value of 6.17 with 3 degrees of freedom (p = .1039). It can therefore be assumed that the reduced model provides a good fit. Multiple comparisons may now be conducted as in the single-factor design described previously. Since the IM interaction is significant, one may wish to collapse over E and do pairwise comparisons of the probabilities in the resulting four cells.

To illustrate how the GSK methodology can be extended to the case of a polytomous response variable consider again the Walker data but assume now that r = 3. For example, some of the questionnaires returned in the survey were incomplete and of poor quality. The response variable could therefore be defined to have three categories, no response, incomplete response, and complete response. For sake of brevity, assume now that there are only s = 3 populations. The three populations are homogeneous if $\pi_{11} = \pi_{21} = \pi_{31}$, and $\pi_{12} = \pi_{22} = \pi_{32}$. Thus, for each of the three populations, define two response functions $f_{m1}(\pi) = \pi_{m1}$ and $f_{m2}(\pi) = \pi_{m2}$ for m = 1, 2, 3, and

$$[\underline{F}(\pi)]' = [\pi_{11}\ \pi_{21}\ \pi_{31}\ \pi_{12}\ \pi_{22}\ \pi_{32}]. \qquad (7)$$

When more than one function is constructed for each population, the design matrix must be slightly modified. For the present problem, the appropriate form is

$$X = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & -1 & -1 \end{bmatrix} \qquad (8)$$

with $\underline{\beta}' = [\beta_{01}\ \beta_{11}\ \beta_{21}\ \beta_{02}\ \beta_{12}\ \beta_{22}]$. In terms of the parameters contained in $\underline{\beta}$, $\beta_{11} = \beta_{21} = \beta_{12} = \beta_{22} = 0$ implies $\pi_{11} = \pi_{21} = \pi_{31} = \beta_{01}$ and $\pi_{12} = \pi_{22} = \pi_{32} = \beta_{02}$ and thus homogeneity of the three populations. This hypothesis is tested by defining

$$C = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \qquad (9)$$

and computing the chi-square test statistic with d = 4 degrees of freedom.

Summary

This paper has illustrated the Grizzle, Starmer, and Koch (1969) methodology for analyzing contingency tables in which a response variable and one or more experimental factors can be logically defined. The procedure is based on weighted least squares and should therefore be useful for researchers with a basic understanding of regression analysis. Programs written to perform the required computations include the FUNCAT procedure of the Statistical Analysis System (SAS, 1979), the program GENCAT described in Landis, Stanish, Freeman, and Koch (1976), and the program BASCAT, described in Sawyer (1979). Sawyer (1981) has also developed the program DIFCAT which performs the GSK methodology without forming $\underline{F}(\pi)$ in terms of log, linear, or exponential operators.

References

David Flath and E.W. Leonard, "A Comparison of Two Logit Models in the Analysis of Qualitative Marketing Data," Journal of Marketing Research, XVI (November, 1979), 533-538.

Paul E. Green, Frank J. Carmone, and David P. Wachspress, "On the Analysis of Qualitative Data in Marketing Research," Journal of Marketing Research, XIV (February, 1977), 52-59.

James E. Grizzle, C. Frank Starmer, and Gary G. Koch, "Analysis of Categorical Data by Linear Models," Biometrics, 25 (September, 1969), 489-504.

Jane T. Helwig and Káthryn A. Council, editors, SAS User's Guide, 1979 Edition (Raleigh, North Carolina: SAS Institute Inc., 1979).

David G. Kleinbaum and Lawrence L. Kupper, Applied Regression Analysis and Other Multivariable Methods (North Scituate, Massachusetts: Duxbury Press, 1978).

J.R. Landis, W.M. Stanish, J.L. Freeman, and G.G. Koch, "A Computer Program for the Generalized Chi-Square Analysis of Categorical Data Using Weighted Least Squares (GENCAT)," Computer Programs in Biomedicine, 6(4), (1976), 196-231.

John Neter and William Wasserman, Applied Linear Models (Homewood, Illinois: Richard D. Irwin, Inc., 1974).

John W. Sawyer, Jr., "BASCAT: A BASIC Program for Grizzle, Starmer, and Koch Analysis Which is Suitable for Mini-Computers," 1979 Proceedings of the Statistical Computing Section, American Statistical Association, (1979), 400-401.

_____, "A Computer Program for Weighted Least Squares Analysis of Categorical Data Without Log, Linear, and Exponential Operators." Paper presented at 1981 Annual Meetings of the American Statistical Association, Statistical Computing Section, Detroit, Michigan.

Bruce J. Walker, "The Effects of Incentives on Response Rate with a Personally-Distributed, Mail-Returned Questionnaire," Working Paper, Arizona State University, 1981.