THE USE OF CATMOD FOR REPEATED MEASUREMENT ANALYSIS OF CATEGORICAL DATA

William M. Stanish, SAS Institute Inc.
Gary G. Koch, University of North Carolina at Chapel Hill

1. INTRODUCTION

Many studies use data collection designs in which each subject has a response measured under more than one condition. A major advantage of using such a design is that comparisons among the conditions are not obscured by subject-to-subject variability. The price that one pays for this advantage is that the responses within a subject are correlated, and those correlations must be taken into account in the data analysis. For continuous data, this may add a dimension of complexity to the analysis since, for example, it has the potential for changing a univariate analysis to a multivariate one. For categorical data, on the other hand, repeated measures change only the complexion of the analysis since the methods were specifically designed to deal with correlated functions: namely, the response probabilities.

One difficulty posed by repeated measures analysis of categorical data is that repeated measurement designs can result in a great proliferation of cells in a multidimensional contingency table. Thus, it is important to extract functions of the data which summarize the critical information, and to have the computing ability to extract those functions easily.

This paper reviews an important methodology for analyzing categorical data from repeated measurement experiments: weighted least squares (WLS) methods for fitting generalized linear models to correlated functions of response probabilities. In addition, the paper introduces a new SAS® procedure, CATMOD, that facilitates the WLS analysis of categorical data from repeated measurement experiments. Examples are given in which CATMOD is used to analyze previously published data.

Another important technique for repeated measurement analysis of categorical data is based on the use of randomization model test statistics such as those corresponding to nonparametric rank tests. This technique is appropriate if interest lies solely in hypothesis testing, rather than on model building and parameter estimation. For example, this method was used by Koch et al. (1983) to analyze data from a two-period cross-over design of a medical experiment. It was also used by Johnson et al. (1981) to analyze longitudinal categorical data from a multi-center clinical trial. One of the most useful types of randomization test statistics is the class of generalized Cochran-Mantel-Haenszel statistics (Landis, Heyman, and Koch, 1978), and some of these may be computed by PROC TFREQ in SAS 82, and by PROC FREQ in the next release of the SAS System.

A third method of repeated measurement analysis is to use maximum likelihood estimation to fit log-linear models to the joint probabilities of the contingency table (Bishop, Feinberg, and Holland, 1975, Chapter 8). With this method, hypotheses of marginal homogeneity must be expressed in terms of models corresponding to symmetry and quasi symmetry. For such analyses, one disadvantage of this method is that it may require larger sample sizes than those required by WLS analysis, since the asymptotic theory must apply to the individual cells of the contingency table, rather than to the marginal probabilities.

2. METHODOLOGY

The WLS analysis of categorical data from repeated measures experiments involves three components: (1) computation of appropriate functions of the data, (2) fitting asymptotic regression models to the functions through WLS estimation, and (3) testing hypotheses about linear combinations of the model parameters. The first two areas are examined in some detail; the third involves the use of Wald Statistics, and is essentially the same as the hypothesis testing in analyses that are not repeated measures analyses (Grizzle, Starmer, and Koch, 1969). Also highlighted in this section are the features of the SAS procedure CATMOD that facilitate the repeated measurement analysis of categorical data.

2.1 Function Selection

For repeated measurement data, the most useful functions to analyze are often the first-order marginal probabilities of the dependent variables. These probabilities summarize the information on a dependent variable in much the same way that the mean summarizes the important information in a continuous data analysis. Alternatively, if the response variable is ordinal, then it may be reasonable to assign scores to the levels of the variable, and to compute the mean score as the basic function to be analyzed. Some analyses, however, involve questions that are more complicated than homogeneity of means or marginal probabilities, and these require that specialized functions of the data be computed and analyzed. Finally, other considerations may dictate that the information from one or more repeated measurement factors be summarized by fewer functions than the number of dependent variables, and this provides a way of partitioning the data analysis into smaller components.

Marginal Probabilities

Suppose each subject in an experiment is graded at three time points with respect to his ability to perform a certain task (response = 1 for failing grade, 2 for passing grade). Then there is one dependent variable for each time point (T1, T2, and T3), and these variables correspond to the three levels of a repeated measurement factor which may be called TIME. There are eight possible responses for each subject, resulting from the 2x2x2 cross-classification of T1, T2, and T3.

Rather than analyzing all eight probabilities corresponding to the responses, we focus on the first-order marginal probabilities p_{ij} = the probability of response j at time i. There are six of these functions, but only three are linearly independent since the marginal probabilities sum to 1 for each time point. The three functions are computed for each population of interest, and the subsequent analysis is based on the variation among these functions within populations, as well as across populations.

If the grading had been done on a three-point scale (response = 1 for below average, 2 for average, 3 for above average), there would be 27(=3x3x3) possible responses, and nine(=3+3+3) marginal probabilities, of which six are linearly independent. As before, the six functions would be computed for each population of interest, and the subsequent analysis would be based on the variation among these functions.

Suppose the experiment given above (with response = 1 for failing grade, 2 for passing grade) is expanded, so that two tasks are done at each time point: one standard task and one experimental task. Then there are six dependent variables D11, D12, D13, D21, D22, and D23, where the first digit in the variable name denotes the task (1 for standard, 2 for experimental), and the second denotes the time point. Now, there are 64 possible responses for each subject, resulting from the cross-classification of the six dependent variables.

Rather than analyzing all 64 probabilities corresponding to the responses, we focus attention on the first-order marginal probabilities of the six dependent variables. In other words, we analyze the marginal probabilities p_{ijk} = the probability of response k for task i at time j. There are 12 of these functions, of which six are linearly independent. The six functions would be computed for each population, and their variation could then be studied.

If the grading had been done on a three-point scale, then there would be 729(=3**6) possible responses, and 18(=3*6) marginal probabilities, but only 12 linearly independent marginals. Thus, the use of marginal probabilities can sometimes result in substantial data reduction, and thereby facilitate interpretation of results.

The Mean Score Function

As the number of levels of the response variable increases, it becomes more unwieldy to deal with marginal probabilities in repeated measurement analyses because there is an exponential increase in the total number of response functions, as well as in the sample size required to support the asymptotic distributions of the functions. If the response variable is ordinal, then it may be reasonable to assign scores to the levels of the variable, and to compute the mean score for each dependent variable. For example, if nonparametric rank scores are used to compute the means for each population of interest, then the resulting analysis of variation is a nonparametric analysis of variance (Semenya et al., 1983). Also, the mean score function is particularly useful when the response variable represents frequency of occurence, because the scores then lie on an interval scale.

Other Functions

If the response variable has two levels, then an alternative to analyzing the marginal probability (p) of one of the levels is to analyze the logit function (=log{p/(1-p)}). This results in a logistic regression analysis. Guthrie(1981) used both marginal probabilities and logits to analyze some dichotomous repeated measurement data.

Occasionally, the investigator asks questions which require the computation of specialized functions of the data. In repeated measures experiments, the question of agreement between measures often arises, and there are several functions that measure the extent of agreement or disagreement. For example, the comparison of a standard diagnostic procedure vs. a test procedure might require the computation of sensitivity and specificity (MacMillan, Becker, Koch, Stokes, and Vandiviere, 1981), and a study of observer agreement would suggest the computation of generalized kappa statistics (Landis and Koch, 1977).

Ideally, one would like to compute at least one function for each dependent variable in the analysis. Sometimes, however, there are so many dependent variables and/or populations that the total number of functions produced would be too unwieldy to handle. In that case, one common solution is to summarize one repeated measurement factor by one or more ordinal dependent variables, and to analyze the mean scores of the new dependent variables, rather than the original responses. If the new dependent variable is formed by summing or averaging across the levels of the summarized factor, then the resulting analysis yields all those effects which do not include that factor. If the new dependent variables are formed from contrasts among the levels of the summarized factor, then the resulting analysis yields interactions between that factor and the other effects.

For example, in a study of obstetrically related pain (Johnson, Amara, Edwards, and Koch, 1981), each patient was observed once per hour for eight hours, and at each hour, her pain status was recorded on a five-point scale: 0 = none, 1 = a little, 2 = some, 3 = a lot, 4 = terrible. If a mean score were computed for each dependent variable, there would be eight means, one for each hour. But since there are 32 populations, this would create a total of 256 functions to analyze. One solution is to dichotomize the response variable (1 = little or no pain, 0 = some or more pain), and to summarize the repeated measurement factor TIME by the sum of the eight dependent variables: the number of hours for which a patient had little or no pain. The analysis is then based on the mean of this dependent variable.

Another example of this strategy (Koch, Grizzle, Semenya, and Sen, 1978) was in an analysis of an experiment which had two repeated measurement factors. The study was designed to evaluate the effect of various treatments on mastitis in dairy cows. The basic response was dichotomous (1=infected, 0 otherwise). The two repeated measurement factors were INTERVENTION (pre-treatment, post-treatment) and QUARTER (one level for each of the four quarters of a cow's udder). One part of the analysis summarized the repeated measurement factor QUARTER by the summation variable: number of infected quarters. Another part of the analysis summarized the factor INTERVENTION by the contrast variable: pre-treatment infection status minus post-treatment infection status. The effect labeled QUARTER in the latter analysis is actually an INTERVENTION*QUARTER interaction with respect to the original response variable.

2.2 Model Fitting via WLS

Once the functions of the probabilities p have been computed, they become part of a function vector F(p), which gets modeled as

$$E_A F(p) = X \beta,$$

where β is the parameter vector, X is the design matrix that defines the meaning of the parameters, and E_A denotes asymptotic expectation. See Grizzle, Starmer, and Koch (1969) and Koch et al. (1977) for the theory behind the parameter estimation, goodness-of-fit tests, and statistics for testing hypotheses about the parameters.

For analyses that are not repeated measurement analyses, the primary interest lies in the presence and description of main effects and interactions for the independent variables (which define the populations). For repeated measurement analyses, there must be additional focus on (1) the presence and description of main effects and interactions among the repeated measurement factors, as well as (2) the presence and description of interactions between the independent variables and the repeated measurement factors.

For example, if TIME were the repeated measurement factor (as in Section 2.1 when the dependent variables were T1, T2, and T3), and if SEX were an independent variable that defined two populations (male and female), then the initial model might contain effects for SEX, TIME, and SEX*TIME. If TIME and TASK were the repeated measurement factors (as in Section 2.1 when the dependent variables were D11, D12, D13, D21, D22, and D23), and SEX were the independent variable, then the initial model might contain effects for SEX, TIME, TASK, SEX*TIME, SEX*TASK, TIME*TASK, and SEX*TIME*TASK. Subsequent reduced models would contain only the significant effects.

Sometimes, the significant effects in a model may be partitioned into two or more components, and further model reduction may be appropriate if one or more of the components have no more variability than what would be expected on the basis of random error. For example, the TIME effect (say, with two df) could be partitioned into two components: linear and quadratic. If the quadratic component were nonsignificant, then a reduced model would contain only the linear component of TIME. As another example, the two effects, TIME and SEX*TIME, could be partitioned into two components: a TIME effect for males, and a TIME effect for females. Similarly, the two effects, SEX and SEX*TIME, could be partitioned into three components: a SEX effect within each of the three time points.

2.3 CATMOD

A new SAS procedure, CATMOD (available in the next release of the SAS System), does WLS analysis on generalized functions of categorical data, as well as maximum likelihood analysis on log-linear models. CATMOD has several features which facilitate the analysis of repeated measurement experiments:

1. The statement, RESPONSE MARGINALS, induces the computation of response functions that are the first-order marginal probabilities of the dependent variables.

2. The statement, RESPONSE LOGITS, induces the computation of response functions that are the generalized logits of the first-order marginal probabilities of the dependent variables.

3. The statement, RESPONSE MEANS, induces the computation of response functions that are the mean scores of numeric dependent variables.

4. The MODEL statement may contain the keyword _RESPONSE_, which allows modeling of the repeated measurement factors.

5. The REPEATED statement may be used to define the repeated measurement factors and to indicate which repeated measurement effects will be included in the model.

6. The design matrix may be specified directly on the MODEL statement so that complete model flexibility is available.

7. The POPULATION statement may be used to define the populations in terms of the independent variables.

CATMOD is used to analyze the data from the examples in Section 3. Shown in Figures 1 through 12 are the SAS statements required to analyze the data, followed by brief samples of the output.

3. EXAMPLES

Example 1: One Repeated Measurement Variable, Dichotomous Response

These data were previously analyzed by Marascuilo and Serlin (1977) and by Guthrie(1981). For each of three groups, subjects were measured at four trials with respect to a dichotomous variable (the data are hypothetical, and it is not indicated whether the trials correspond to an ordinally scaled variable or not). Specifying RESPONSE MARGINALS indicates to CATMOD that the response functions to be computed and analyzed are the marginal probabilities of 'success' at each trial. This specification thus yields 12 response functions: four marginal probabilities for each of the three groups.

The first analysis of interest might be an analysis of variance to assess the main effects for the repeated measurement factor (trial) and the independent variable (group), as well as their interaction. Figure 1 shows the CATMOD input statements, as well as the response frequencies. The MODEL statement is specified in the usual way, except that the term _RESPONSE_ is used to denote the factor TRIAL. The four variables on the left-hand side of the MODEL statement correspond to the four trials. These four variables yield 16(=2x2x2x2) possible responses, but only 13 are actually observed. These 13 probabilities are then reduced to four marginal probabilities within each group. In fitting the model, CATMOD uses a full-rank design matrix that reflects a center-point parameterization. The corresponding analysis of variance table shows that all of the effects are significant, including the GROUP*TRIAL interaction (Q=18.71, df=6).

Given the significant interaction, one is interested in the nature of the interaction, and for this purpose, the MODEL statement in Figure 2 can be used to assess the effect of TRIAL in each group. The corresponding analysis of variance table shows that the trial effect is significant only in group 3 (Q=75.07, df=3), and so one might be led to a final model that contains effects only for GROUP and TRIAL(GROUP=3). The MODEL statement for this case is given in Figure 3, and the corresponding analysis of variance table shows that the goodness-of-fit test for that model is Q=5.09, df=6.

If TRIAL corresponds to an ordinally scaled factor, such as time, then one might be interested in fitting polynomial trial effects within each level of group after discovering significant GROUP*TRIAL interaction in a preliminary run. Figure 4 shows that CATMOD allows direct specification of a design matrix on the right-hand side of the MODEL statement, and that the user may also specify that certain sets of parameters be tested for equality to zero. The design matrix in Figure 4 specifies linear, quadratic, and cubic polynomial trial effects within each level of group (the coefficients presume equal intervals between adjacent levels of TRIAL), and the statements following the design matrix provide labels for the sets of parameters which are to be tested for equality to zero. The corresponding analysis of variance table shows the test statistics, and provides guidelines for model reduction, such as elimination of all the quadratic and cubic trial effects.

Based on the saturated polynomial model and on resultant model reduction, a reasonable final model is one that contains effects only for group and for linear trial effects within groups 2 and 3. The appropriate MODEL statement is given in Figure 5, and the corresponding analysis of variance table shows that the model fits well (Q=0.85, df=7). The parameter estimates (not shown with the other output) are (.58, .13, -.03, -.05, -.135). The corresponding table of predicted values gives the predicted marginal probabilities based on this final model, together with their standard errors. Also shown are the observed marginals and their standard errors, and the resulting residuals.

It should be noted that the polynomial modeling yields a linear trial effect in group 2, whereas the previous models suggest that the trial effect is nonsignificant in that group. This is due to the fact that the polynomial modeling exploits the ordinal nature of the trial variable. Since its power is directed specifically at polynomial-type association, it is able to identify an additional feature of the data that would otherwise go unrecognized. Nevertheless, the result should perhaps be regarded with some caution because of the aggressive modeling strategy that was used.

Example 2: Two Repeated Measurement Variables, Dichotomous Response

These data appear in MacMillan et al. (1981) and represent the results of a study for the comparative evaluation of a test procedure and a standard procedure. Both procedures are done on each subject, and the standard and test results are then evaluated at each of two time points as being positive or negative. Thus, there are four dependent variables:

STD1 = results of standard procedure at Time 1
TEST1 = results of test procedure at Time 1
STD2 = results of standard procedure at Time 2
TEST2 = results of test procedure at Time 2

and these represent a cross-classification of the two repeated measurement factors, time and treatment. For the analysis with CATMOD (see Figure 6), specification of RESPONSE MARGINALS yields the marginal probability of a positive (or negative) evaluation for each of the four dependent variables, and specification of REPEATED TIME 2 TRTMENT 2 indicates to CATMOD that the four functions to be analyzed represent a 2x2 factorial of the repeated measurement factors, time and treatment.

The first issue of interest might be an overall comparison of the marginal probabilities of a positive (or negative) evaluation. Figure 6 shows the appropriate input for a CATMOD analysis. The four dependent variables appear on the left-hand side of the MODEL statement, and the term _RESPONSE_ on the right-hand side indicates that the model is based on the relationship among the response variables. The definition of _RESPONSE_ on the REPEATED statement indicates to CATMOD that the model is to be a saturated model with respect to the repeated measurement factors, time and treatment. The analysis of variance table from the CATMOD output shows that the TIME*TRTMENT interaction is nonsignificant (p=.12), as well as the average TIME effect (p=.36).

The reduced model (Figure 7) thus excludes these effects by redefining _RESPONSE_ in the REPEATED statement to include an effect only for TRTMENT. The goodness-of-fit statistic for this model tests the joint effect of the TIME and TIME*TRTMENT effects, and shows that the model fits well (Q=3.51, df=2, p=.17). The table of predicted values gives the observed and predicted marginal probabilities and their standard errors.

A second type of analysis on these data, one which MacMillan et al. essentially pursued, is an analysis of the accuracy of the test procedure relative to that of the standard procedure. Specifically, we compute the sensitivity (the conditional probability of a positive assessment by the test procedure, given a positive assessment by the standard procedure) and the specificity (the conditional probability of a negative assessment by the test procedure, given a negative assessment by the standard procedure) for each of the two follow-up visits. The RESPONSE statement in Figure 8 yields the following four functions.

SENS1 = sensitivity at Time 1
SPEC1 = specificity at Time 1
SENS2 = sensitivity at Time 2
SPEC2 = specificity at Time 2

The REPEATED statement indicates to CATMOD that these functions correspond to a 2x2 factorial of the repeated measurement factors ACCURACY and TIME, and that the first model to be fit is a saturated model with respect to these factors. The resulting analysis of variance table shows that the interaction effect is nonsignificant (Q=1.0, df=1, p=.32).

The final model on the accuracy measures (Figure 9) is therefore a main-effects model, which is obtained by specifying _RESPONSE_ = TIME ACCURACY on the REPEATED statement. The parameter estimates (not shown) are (.89, -.01, -.07). The resulting table of predicted functions shows that the predicted sensitivity and specificity estimates are .82 and .96, respectively, with a .01 adjustment depending on whether the estimates correspond to Time 1 or Time 2.

Example 3: One Repeated Measurement Variable, Polychotomous Response

These data are from a study of the geographic distribution of multiple sclerosis (Westland and Kurland, 1953), and were analyzed in an observer agreement context by Landis and Koch (1977). In each of two geographic areas, a neurologist in the area reviewed the medical records of patients from that area, and classified the patients into one of four categories:

1. Certain multiple sclerosis;
2. Probable multiple sclerosis;
3. Possible multiple sclerosis (odds 50:50);
4. Doubtful, or definitely not multiple sclerosis.

Each neurologist then reviewed the medical records of the patients from the other geographic area, and classified those patients in the same manner.

For questions that relate to the extent of observer agreement with respect to specific diagnoses of individual patients, the answers may be obtained by the analysis of joint probabilities and the computation of kappa statistics. For this example, the questions of interest relate to the presence of observer effects, geographic effects, and their interaction, and these questions may be answered by the analysis of marginal probabilities.

For this purpose, the SAS data step and the CATMOD input statements shown in Figure 10 yield a preliminary analysis of the data. The two dependent variables on the left-hand side of the MODEL statement, OBS1 and OBS2, cause CATMOD to form one response for each observed combination of the levels of the variables. Only 14 responses are actually observed, out of a possible 16(=4x4). Inclusion of the term GROUP on the right-hand side of the MODEL statement causes CATMOD to form two populations, one for each geographic area.

The statement, RESPONSE MARGINALS, causes CATMOD to compute three linearly independent marginal probabilities for each dependent variable (i.e., each observer), a total of six functions per population. The REPEATED statement defines the repeated measurement factor, OBSERVER, as having two levels. Since there are six functions per population, but only two levels of the repeated measurement factor, CATMOD groups the response functions into sets of three(=6/2), and analyzes the model effects with respect to these sets. Thus, the effects, INTERCEPT, GROUP, OBSERVER, and GROUP*OBSERVER, each have three degrees of freedom, rather than the one degree of freedom that they would have had if there had been only one response function for each dependent variable.

That this analysis is an appropriate one may be verified by noting that the results are invariant with respect to the order in which the data are entered (provided, of course, that the data are entered in such a way that the same marginals are chosen for both observers). In other words, the results are the same, regardless of which linearly independent marginals are chosen. If some effect (e.g., GROUP, with two levels) had been defined with only one degree of freedom for these data, then the invariance property would not hold since the effect would depend specifically on which three linearly independent marginals were chosen for analysis.

The results, following Figure 10, show that all of the effects are statistically significant ($p<.05$). An alternative way to look at the total effect of OBSERVER is to consider the observer effect as being nested within groups. The absence of such an effect is known as marginal homogeneity, and its test statistic may be computed by using the MODEL statement given in Figure 11. The corresponding results show that the hypothesis of marginal homogeneity, (i.e., the OBSERVER(GROUP) effect) must be rejected ($Q=69.01$, $df=6$, $p<=.0001$).

Finally, the OBSERVER(GROUP) effect may be partitioned into two components: an observer effect within each of the two groups. Since there is a significant GROUP*OBSERVER effect, such an analysis would indicate which group had the strongest observer effect. The MODEL statement shown in Figure 12 produces such an analysis. The corresponding results show that most of the evidence for an observer effect is from Group 1. The table of observed and predicted values shows that the residuals are all zero, as they must be when the model is saturated.

SAS is the registered trademark of SAS Institute Inc., Cary, NC, USA.

4. REFERENCES

1. Bishop, Y.M.M., Feinberg, S.E., and Holland, P.W. (1975). Discrete Multivariate Analysis: Theory and Practice. Cambridge, Massachusetts: The MIT Press.

2. Grizzle, J.E., Starmer, C.F., and Koch, G.G. (1969). "Analysis of Categorical Data by Linear Models,", Biometrics 25, 489-503.

3. Guthrie, D. (1981). "Analysis of Dichotomous Variables in Repeated Measures Experiments," Psychological Bulletin 90, 189-195.

4. Johnson, C., Amara, I., Edwards, S., and Koch, G.G. (1981). "Some Computational Strategies for the Analysis of Longitudinal Categorical Data," 1981 Proceedings of the Statistical Computing Section of the American Statistical Association, 293-298.

5. Koch, G.K., Gitomer, S.L., Skalland, L., and Stokes, M. (1983). "Some Non-Parametric and Categorical Data Analyses for a Change-Over Design Study and Discussion of Apparent Carry-Over Effects," Statistics in Medicine 2, 397-412.

6. Koch, G.K., Grizzle, J.E., Semenya, K., and Sen, P.K. (1978). "Statistical Methods for Evaluation of Mastitis Treatment Data," Journal of Dairy Science 61,829-847.

7. Koch, G.K., Landis, J.R., Freeman, J.L., Freeman, D.H., Lehnen, R.G. (1977). "A General Methodology for the Analysis of Experiments with Repeated Measurement of Categorical Data," Biometrics 33,133-158.

8. Landis, J.R. and Koch, G.G. (1977). "The Measurement of Observer Agreement for Categorical Data," Biometrics 33, 159-174.

9. Landis, J.R., Heyman, E.R., and Koch, G.G. (1978). "Average Partial Association in Three-Way Tables: A Review and Discussion of Alternative Tests," International Statistical Review 46, 237-254.

10. MacMillan, J., Becker, C., Koch, G.G., Stokes, M., and Vandiviere, H.M. (1981). "An Application of Weighted Least Squares Methods to the Analysis of Measurement Process Components of Variability in an Observational Study," 1981 Section on Survey Research Methods, Proceedings of the American Statistical Association, 680-685.

11. Marascuilo, L.A. and Serlin, R. (1977). "Interactions for Dichotomous Variables in Repeated Measures Designs," Psychological Bulletin 84, 1002-1007.

12. Semenya, K.A., Koch, G.G., Stokes, M.E., and Forthofer, R.N. (1983). "Linear Models Methods for Some Rank Function Analyses of Ordinal Categorical Data," Communications in Statistics: Theory and Methods 12, 1277-1298.

13. Westland, K.B. and Kurland, L.T. (1953). "Studies on Multiple Sclerosis in Winnipeg, Manitoba and New Orleans, Louisiana," American Journal of Hygiene 57, 380-396.

Figure 1

```
DATA; INPUT A B C D GROUP WT @@; CARDS;
      /*--INSERT MARASCUILO & SERLIN DATA HERE--*/
PROC CATMOD;
   TITLE M & S DATA, MODEL 1, SATURATED MODEL;
   WEIGHT WT;
   REPEATED TRIAL;
   RESPONSE MARGINALS;
   MODEL A*B*C*D= GROUP _RESPONSE_ _RESPONSE_*GROUP/FREQ;
```

RESPONSE PROFILES

VAR	1	2	3	4	5	6	7	8	9	10	11	12	13
A	0	0	0	0	0	0	0	0	1	1	1	1	1
B	0	0	0	0	1	1	1	1	0	0	0	0	1
C	0	0	1	1	0	0	1	1	0	0	1	1	1
D	0	1	0	1	0	1	0	1	0	1	0	1	1

RESPONSE FREQUENCIES

GROUP	1	2	3	4	5	6	7	8	9	10	11	12	13
1	0	4	2	1	4	0	0	0	3	0	1	0	0
2	2	1	2	2	1	1	0	1	1	1	0	1	2
3	0	3	0	5	0	2	1	2	0	0	0	2	0

ANALYSIS OF VARIANCE TABLE

SOURCE	DF	CHI-SQUARE	PROB
INTERCEPT	1	354.88	0.0001
GROUP	2	24.79	0.0001
RESPONSE	3	21.45	0.0001
GROUP*_RESPONSE_	6	18.71	0.0047
RESIDUAL	0	0.00	1.0000

Figure 2

```
TITLE M & S DATA, MODEL 2, TRIAL NESTED WITHIN GROUP;
MODEL A*B*C*D=GROUP _RESPONSE_(GROUP=1)
              _RESPONSE_(GROUP=2)  _RESPONSE_(GROUP=3);
```

ANALYSIS OF VARIANCE TABLE

SOURCE	DF	CHI-SQUARE	PROB
INTERCEPT	1	354.88	0.0001
GROUP	2	24.79	0.0001
RESPONSE(GROUP=1)	3	0.18	0.9812
RESPONSE(GROUP=2)	3	4.92	0.1780
RESPONSE(GROUP=3)	3	75.07	0.0001
RESIDUAL	0	0.00	1.0000

Figure 3

```
TITLE M & S DATA, MODEL 3, TRIAL NESTED WITHIN GROUP 3;
MODEL A*B*C*D=GROUP _RESPONSE_(GROUP=3)/NOPROFILE ;
```

ANALYSIS OF VARIANCE TABLE

SOURCE	DF	CHI-SQUARE	PROB
INTERCEPT	1	386.94	0.0001
GROUP	2	25.42	0.0001
RESPONSE(GROUP=3)	3	75.07	0.0001
RESIDUAL	6	5.09	0.5319

Figure 4

```
TITLE M & S DATA, COMPLETE POLYNOMIAL EFFECTS;
POPULATION GROUP;
MODEL A*B*C*D=( 1   1  0   -3  1 -1    0  0  0    0  0  0 ,
                1   1  0   -1 -1  3    0  0  0    0  0  0 ,
                1   1  0    1 -1 -3    0  0  0    0  0  0 ,
                1   1  0    3  1  1    0  0  0    0  0  0 ,
                1   0  1    0  0  0   -3  1 -1    0  0  0 ,
                1   0  1    0  0  0   -1 -1  3    0  0  0 ,
                1   0  1    0  0  0    1 -1 -3    0  0  0 ,
                1   0  1    0  0  0    3  1  1    0  0  0 ,
                1  -1 -1    0  0  0    0  0  0   -3  1 -1 ,
                1  -1 -1    0  0  0    0  0  0   -1 -1  3 ,
                1  -1 -1    0  0  0    0  0  0    1 -1 -3 ,
                1  -1 -1    0  0  0    0  0  0    3  1  1 )
                  (1='INTERCEPT',          2 3='GROUP',
4='LINEAR, GROUP 1',   7='LINEAR, GROUP 2',   10='LINEAR, GROUP 3',
5='QUADRATIC,GROUP 1',8='QUADRATIC,GROUP 2',11='QUADRATIC,GROUP 3',
6='CUBIC, GROUP 1',    9='CUBIC, GROUP 2',    12='CUBIC, GROUP 3');
```

ANALYSIS OF VARIANCE TABLE

SOURCE	DF	CHI-SQUARE	PROB
INTERCEPT	1	354.88	0.0001
GROUP	2	24.79	0.0001
LINEAR, GROUP 1	1	0.10	0.7497
LINEAR, GROUP 2	1	4.33	0.0375
LINEAR, GROUP 3	1	73.30	0.0001
QUADRATIC,GROUP 1	1	0.08	0.7810
QUADRATIC,GROUP 2	1	0.09	0.7623
QUADRATIC,GROUP 3	1	0.09	0.7623
CUBIC, GROUP 1	1	0.01	0.9115
CUBIC, GROUP 2	1	0.41	0.5232
CUBIC, GROUP 3	1	0.12	0.7280
RESIDUAL	0	0.00	1.0000

Figure 5

```
TITLE M & S DATA, LINEAR EFFECTS WITHIN GROUPS 2 AND 3;
POPULATION GROUP;
MODEL A*B*C*D=( 1  1  0  0  0 ,
                1  1  0  0  0 ,
                1  1  0  0  0 ,
                1  1  0  0  0 ,
                1  0  1 -3  0 ,
                1  0  1 -1  0 ,
                1  0  1  1  0 ,
                1  0  1  3  0 ,
                1 -1 -1  0 -3 ,
                1 -1 -1  0 -1 ,
                1 -1 -1  0  1 ,
                1 -1 -1  0  3 ) (  1 = 'INTERCEPT',
                                 2 3 = 'GROUP',
                                   4 = 'LINEAR, GROUP 2'
                                   5 = 'LINEAR, GROUP 3')/PREDICT;
```

ANALYSIS OF VARIANCE TABLE

SOURCE	DF	CHI-SQUARE	PROB
INTERCEPT	1	385.55	0.0001
GROUP	2	32.77	0.0001
LINEAR, GROUP 2	1	4.43	0.0354
LINEAR, GROUP 3	1	74.89	0.0001
RESIDUAL	7	0.85	0.9969

PREDICTED VALUES FOR RESPONSE FUNCTIONS

		OBSERVED		PREDICTED		
SAMPLE	FUNCTION NUMBER	FUNCTION	STANDARD ERROR	FUNCTION	STANDARD ERROR	RESIDUAL
1	1	0.733333	0.11418	0.715668	0.0182593	0.0176649
	2	0.733333	0.11418	0.715668	0.0182593	0.0176649
	3	0.733333	0.11418	0.715668	0.0182593	0.0176649
	4	0.666667	0.121716	0.715668	0.0182593	-0.0490018
2	1	0.666667	0.121716	0.700358	0.0950097	-0.0336915
	2	0.666667	0.121716	0.601273	0.0776131	0.0653933
	3	0.466667	0.128812	0.502189	0.0863501	-0.035522
	4	0.4	0.126491	0.403104	0.115432	-.00310386
3	1	0.866667	0.0877707	0.889339	0.0698734	-0.0226721
	2	0.666667	0.121716	0.618758	0.0457522	0.0479082
	3	0.333333	0.121716	0.348178	0.0354903	-0.0148448
	4	0.0666667	0.0644061	0.0775978	0.0487972	-0.0109312

Figure 6

```
DATA A;
INPUT STD1 $ TEST1 $ STD2 $ TEST2 $ WT @@; CARDS;
NEG NEG NEG NEG 509    NEG NEG NEG POS  4    NEG NEG POS NEG  17
NEG NEG POS POS   3    NEG POS NEG NEG 13    NEG POS NEG POS   8
NEG POS POS POS   8    POS NEG NEG NEG 14    POS NEG NEG POS   1
POS NEG POS NEG  17    POS NEG POS POS  9    POS POS NEG NEG   7
POS POS NEG POS   4    POS POS POS NEG  9    POS POS POS POS 170
PROC CATMOD;
   TITLE DIAGNOSTIC DATA, MARGINAL SYMMETRY, SATURATED MODEL;
   WEIGHT WT;
   RESPONSE MARGINALS;
   MODEL STD1*TEST1*STD2*TEST2 = _RESPONSE_ ;
   REPEATED TIME 2 TRTMENT 2/_RESPONSE_= TIME TRTMENT TIME*TRTMENT;
```

ANALYSIS OF VARIANCE TABLE

SOURCE	DF	CHI-SQUARE	PROB
INTERCEPT	1	2385.34	0.0001
TIME	1	0.85	0.3570
TRTMENT	1	8.20	0.0042
TIME*TRTMENT	1	2.40	0.1215
RESIDUAL	0	0.00	1.0000

Figure 7

```
REPEATED TIME 2 TRTMENT 2 / _RESPONSE_ = TRTMENT;
TITLE DIAGNOSTIC DATA, MARGINAL SYMMETRY, TRTMENT ONLY
```

ANALYSIS OF VARIANCE TABLE

SOURCE	DF	CHI-SQUARE	PROB
INTERCEPT	1	2386.97	0.0001
TRTMENT	1	9.55	0.0020
RESIDUAL	2	3.51	0.1731

PREDICTED VALUES FOR RESPONSE FUNCTIONS

		OBSERVED		PREDICTED		
SAMPLE	FUNCTION NUMBER	FUNCTION	STANDARD ERROR	FUNCTION	STANDARD ERROR	RESIDUAL
1	1	0.708701	0.0161348	0.706778	0.0154721	0.00192315
	2	0.723834	0.015877	0.732476	0.0151364	-.00864249
	3	0.706179	0.0161757	0.706778	0.0154721	-.00059892
	4	0.738966	0.0155964	0.732476	0.0151364	0.00648992

```
RESPONSE EXP 1 -1 0  0 0  0 0  0 ,                              Figure 8
             0  0 1 -1 0  0 0  0 ,
             0  0 0  0 1 -1 0  0 ,
             0  0 0  0 0  0 1 -1 LOG 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1
                                     0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1
                                     1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0
                                     1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0
                                     0 0 0 1 0 0 0 1 0 0 0 1 0 0 0 1
                                     0 0 1 1 0 0 1 0 0 1 1 0 0 1 1 1
                                     1 0 0 0 1 0 0 1 0 0 0 1 0 0 0 0
                                     1 1 0 0 1 1 0 1 1 0 0 1 1 0 0 0
MODEL STD1*TEST1*STD2*TEST2 = _RESPONSE_ /PRED;
REPEATED TIME 2 ACCURACY 2/_RESPONSE_ =TIME ACCURACY TIME*ACCURACY;
TITLE SENSITIVITY AND SPECIFICITY ANALYSIS, SATURATED MODEL;
```

ANALYSIS OF VARIANCE TABLE

SOURCE	DF	CHI-SQUARE	PROB
INTERCEPT	1	6444.69	0.0001
TIME	1	0.23	0.6312
ACCURACY	1	38.60	0.0001
TIME*ACCURACY	1	1.00	0.3178
RESIDUAL	0	-0.00	1.0000


```
REPEATED TIME 2 ACCURACY 2 /                                    Figure 9
         _RESPONSE_ = TIME ACCURACY;
TITLE SENSITIVITY AND SPECIFICITY ANALYSIS, NO INTERACTION MODEL;
```

ANALYSIS OF VARIANCE TABLE

SOURCE	DF	CHI-SQUARE	PROB
INTERCEPT	1	6448.79	0.0001
TIME	1	4.10	0.0428
ACCURACY	1	38.81	0.0001
RESIDUAL	1	1.00	0.3178

PREDICTED VALUES FOR RESPONSE FUNCTIONS

SAMPLE	FUNCTION NUMBER	OBSERVED FUNCTION	OBSERVED STANDARD ERROR	PREDICTED FUNCTION	PREDICTED STANDARD ERROR	RESIDUAL
1	1	0.822511	0.0251392	0.809654	0.0215946	0.0128566
	2	0.948399	0.00933164	0.950058	0.00918267	-.00165895
	3	0.815451	0.0254142	0.828291	0.0219238	-0.0128402
	4	0.969643	0.00725007	0.968694	0.00718759	.000948751

```
DATA ONE; INPUT GROUP OBS1 OBS2  WT @@; CARDS;                  Figure 10
1 4 4 10    1 4 1  3    1 4 2  7    1 4 3  3
1 1 4  1    1 1 1 38    1 1 2  5    1 1 3  0
1 2 4  0    1 2 1 33    1 2 2 11    1 2 3  3
1 3 4  6    1 3 1 10    1 3 2 14    1 3 3  5
2 4 1  1    2 4 2  2    2 4 3  4    2 4 4 14
2 3 1  2    2 3 2 13    2 3 3  3    2 3 4  4
2 2 1  3    2 2 2 11    2 2 3  4    2 2 4  0
2 1 1  5    2 1 2  3    2 1 3  0    2 1 4  0
PROC CATMOD;
  WEIGHT WT;
  RESPONSE MARGINALS;
  MODEL OBS1*OBS2 = GROUP _RESPONSE_  GROUP*_RESPONSE_ /PREDICT;
  REPEATED OBSERVER 2;
  TITLE MULTIPLE SCLEROSIS DATA ANALYSIS;
```

ANALYSIS OF VARIANCE TABLE

SOURCE	DF	CHI-SQUARE	PROB
INTERCEPT	3	875.79	0.0001
GROUP	3	37.62	0.0001
RESPONSE	3	47.90	0.0001
GROUP*_RESPONSE_	3	14.09	0.0028
RESIDUAL	0	0.00	1.0000

NOTE: _RESPONSE_ = OBSERVER

```
  MODEL OBS1*OBS2 = GROUP _RESPONSE_(GROUP);                    Figure 11
  TITLE TEST MARGINAL HOMOGENEITY BY OBSERVER(GROUP)
```

ANALYSIS OF VARIANCE TABLE

SOURCE	DF	CHI-SQUARE	PROB
INTERCEPT	3	875.79	0.0001
GROUP	3	37.62	0.0001
RESPONSE(GROUP)	6	69.01	0.0001
RESIDUAL	0	-0.00	1.0000

NOTE: _RESPONSE_ = OBSERVER

```
MODEL OBS1*OBS2 = GROUP _RESPONSE_(GROUP=1)
                        _RESPONSE_(GROUP=2)/PREDICT;
TITLE OBSERVER AGREEMENT DATA, OBSERVER NESTED WITHIN EACH GROUP;
```

Figure 12

ANALYSIS OF VARIANCE TABLE

SOURCE	DF	CHI-SQUARE	PROB
INTERCEPT	3	875.79	0.0001
GROUP	3	37.62	0.0001
RESPONSE(GROUP=1)	3	58.47	0.0001
RESPONSE(GROUP=2)	3	10.54	0.0145
RESIDUAL	0	-0.00	1.0000

PREDICTED VALUES FOR RESPONSE FUNCTIONS

		OBSERVED		PREDICTED		
SAMPLE	FUNCTION NUMBER	FUNCTION	STANDARD ERROR	FUNCTION	STANDARD ERROR	RESIDUAL
1	1	0.295302	0.0373716	0.295302	0.0373716	0
	2	0.315436	0.0380688	0.315436	0.0380688	0
	3	0.234899	0.0347302	0.234899	0.0347302	0
	4	0.563758	0.0406272	0.563758	0.0406272	0
	5	0.248322	0.0353941	0.248322	0.0353941	0
	6	0.0738255	0.0214218	0.0738255	0.0214218	0
2	1	0.115942	0.0385422	0.115942	0.0385422	0
	2	0.26087	0.0528625	0.26087	0.0528625	0
	3	0.318841	0.0561031	0.318841	0.0561031	0
	4	0.15942	0.0440694	0.15942	0.0440694	0
	5	0.42029	0.0594231	0.42029	0.0594231	0
	6	0.15942	0.0440694	0.15942	0.0440694	0