This research was supported by the National Institutes of Health, Institute of General Medical Sciences, Grant No. GM-12868-06.

THE ANALYSIS OF CATEGORICAL DATA FROM MIXED MODELS

by

Gary G. Koch and Donald W. Reinfurt
Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 672

MARCH 1970

The Analysis of Categorical Data From Mixed Models

by

Gary G. Koch and Donald W. Reinfurt
Department of Biostatistics, University of North Carolina
Chapel Hill, North Carolina 27514, U.S.A.

1. Introduction

In considering two-way contingency tables, one often is concerned with two types of models: the "two response, no factor" situation in which neither margin is fixed and the "one response, one factor" situation in which one margin (say row totals) is fixed and the other is not. For the first case, the hypothesis of primary interest is one of independence or no association between the responses; while in the other case, the hypothesis of primary interest is one of homogeneity over the factor levels or equality of the distributions in the different rows.

This paper is concerned with the recognition of a third type of model which will be called a "mixed categorical data model of order 2" in the case of two-way tables. The experimental situation corresponding to such a model involves exposing each of n randomly chosen subjects from some homogeneous population to both levels of a binary factor (eg. control vs. treated) and classifying each of the two responses into one of r categories. The resulting data are then represented in the matrix of an $r \times r$ contingency table which will be assumed to follow a multinomial distribution. It is apparent that this is a categorical data version of the classical matched pairs design. As such, it has been studied in a somewhat different context recently by Miettinen [1968, 1969]

The Analysis of Categorical Data From Mixed Models

by

Gary G. Koch and Donald W. Reinfurt
Department of Biostatistics, University of North Carolina
Chapel Hill, North Carolina 27514, U.S.A.

This paper is concerned with the recognition of the underlying structure of contingency tables which are analogous to the well-known mixed model in analysis of variance. The experimental situation corresponding to such a model involves exposing each of n subjects to each of the d levels of a given factor and classifying each of the d responses into one of r categories. The resulting data are represented in an $r \times r \times \ldots \times r$ contingency table of d-dimensions. The hypothesis of principal interest is one of first order marginal symmetry or equality of the one-dimensional marginal distributions. Alternatively, if the r categories may be quantitatively scaled, then attention is directed at the hypothesis of equality of the mean scores over the d first order marginals. Test statistics for these hypotheses are developed in terms of minimum Neyman χ^2 or equivalently weighted least squares analysis of underlying linear models. As such, they have a strong resemblance to the Hotelling T^2 procedures used with continuous data in mixed models. Several numerical examples are given to illustrate the use of the various methods discussed.

and others (Worcester [1964], Billewicz [1964, 1965], Bennett [1967], and Chase [1968]). In addition, the methods of Mantel and Haenszel [1959] are also of interest for this experimental situation. Although each of the previously cited papers deal with appropriate statistical procedures for the analysis of mixed categorical data models of order 2, they do not indicate the way in which such contingency table data can be analyzed as a special case of a unified general model.

One such model which will be the basis of inference in the remainder of this paper is that of Grizzle, Starmer, and Koch [1969]. As a result, test statistics are derived through weighted least squares analysis of certain appropriately formulated linear models and correspond identically with the minimum modified χ_1^2-statistics due to Neyman [1949] or equivalently the generalized quadratic form criteria due to Wald [1943]. Alternative methods are also appropriate in certain contexts; eg. those of Lewis [1968] and Goodman [1970] which are based on maximum likelihood and that of Ku and Kullback [1968] based on minimum discrimination information.

To apply the methods of Grizzle, Starmer, and Koch (GSK), it is first necessary to recognize the underlying factor-response structure for the data (see Bhapkar and Koch [1968]). Since only the sample size n is fixed, this is a "two response, no factor" situation. However, the hypothesis of interest is one of homogeneity over the factor levels; i.e., there is no difference between the effects of control and treated. This hypothesis can be formulated specifically as an hypothesis of marginal symmetry in the two-way table.

The hypothesis of marginal symmetry in two-way contingency tables has been considered by a number of authors including Stuart [1955], Bhapkar [1966], and most recently Ireland, Ku, and Kullback [1969]. However, not a great deal of attention has been given to motivating the situations in which tests of marginal symmetry are of the most interest with respect to the interpretation of data from the biological, physical, or social sciences. Hence, one of the principal purposes of this paper is to demonstrate that tests of marginal symmetry are not only applicable to mixed categorical data models, but are the only tests directed at the question of primary interest associated with such experimental situations; namely homogeneity over factor levels. One should also note that the traditional hypothesis of independence which is usually applied to "two response, no factor" tables is of only casual interest here. This remark follows from the fact that responses from the same subject (or units in the same matched pair) to different factor levels are naturally expected to be associated since they have the same initial source in common.

Finally, it is important to recognize that if the r categories are describable in terms of some quantitative scale, then the question of principal interest is that of equality of mean scores over the two marginals. This hypothesis, which bears a definite resemblance to the relationship tested by the classical matched pairs t-test used with continuous data, is weaker than the hypothesis of marginal symmetry in the sense that it specifies a restriction on only one degree of freedom rather than (r-1) degrees of freedom. On the other hand, test statistics directed at this weaker hypothesis are more powerful than those directed at the hypothesis of marginal symmetry when the inequality of the two mean scores is the only alternative of real interest.

2. Mixed Categorical Data Models Of Order 2

First let us consider the following data (see Table 1) from case

TABLE 1

7477 Women Aged 30-39; Unaided Distance Vision

Right Eye	Left Eye: Highest Grade (1)	Second Grade (2)	Third Grade (3)	Lowest Grade (4)	Total
Highest Grade (1)	1520	266	124	66	1976
Second Grade (2)	234	1512	432	78	2256
Third Grade (3)	117	362	1772	205	2456
Lowest Grade (4)	36	82	179	492	789
Total	1907	2222	2507	841	7477

records of the eye-testing of 7477 employees in Royal Ordnance factories which have been used as an illustrative example for tests of marginal symmetry by Stuart [1955], Bhapkar [1966], Grizzle, Starmer, and Koch [1969], and Ireland, Ku, and Kullback [1969]. These data may be viewed as arising from a mixed model in which each of n = 7477 subjects is classified according to both levels of a factor which will be called eye position (namely right eye and left eye) with the corresponding cell frequencies being assumed to follow a multinomial distribution. As such, the hypothesis of principal interest is that the distributions of grades of vision are the same for both eyes; in other words marginal symmetry. The values of test statistics, each approximately distributed as χ^2 with D.F. = 3 if the hypothesis is true, which have been obtained by the authors cited above are as follows:

Stuart	$X^2 = 11.96$
Bhapkar	$X^2 = 11.97$
GSK	$X^2 = 11.98$
IKK	$X^2 = 12.00$

These values are essentially the same and indicate that the two marginal distributions are significantly different.

Let us next look at the problem of marginal symmetry in a more formal way. If $\pi_{jj'}$ denotes the probability that a subject in the population is classified according to the j-th grade category on the right eye and the j'-th grade category on the left eye, then the hypothesis of marginal symmetry may be written as

$$\pi_{j.} = \pi_{.j'} \qquad j = j' = 1, 2, 3, 4 \tag{2.1}$$

where $\pi_{j.} = \sum_{j'=1}^{4} \pi_{jj'}$ and $\pi_{.j'} = \sum_{j=1}^{4} \pi_{jj'}$. Following the GSK approach, one would first form the unrestricted maximum likelihood estimates $p_{jj'} = (n_{jj'}/n)$ of $\pi_{jj'}$ where $n_{jj'}$ is the number of subjects classified into the jj'-th cell. From these, the desired estimates of $\pi_{j.}$ and $\pi_{.j'}$ may be obtained by taking the appropriate marginal sums of $p_{jj'}$; namely $p_{j.} = \sum_{j'=1}^{4} p_{jj'}$ and $p_{.j'} = \sum_{j=1}^{4} p_{jj'}$ respectively. If the following linear model is fitted to the $p_{j.}$ and $p_{.j'}$

$$E\begin{bmatrix} p_{1.} \\ p_{2.} \\ p_{3.} \\ p_{.1} \\ p_{.2} \\ p_{.3} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} \tag{2.2}$$

by weighted least squares, then the residual sum of squares with D.F. = 3 is the GSK test statistic for (2.1) which is the same as that of Bhapkar [1966] (except for round-off error).

Alternatively, if scores a_1, a_2, a_3, a_4 are assigned to the response categories, then the hypothesis of equality of mean scores over the two marginals may be written as

$$\alpha_{RE} \equiv \sum_{j=1}^{r} a_j \pi_{j.} = \sum_{j'=1}^{r} a_{j'} \pi_{.j'} \equiv \alpha_{LE} \qquad (2.3)$$

One test statistic for this hypothesis is the residual sum of squares with D.F. = 1, which is obtained by forming $\hat{\alpha}_{RE} = \sum_{j=1}^{r} a_j p_{j.}$ and $\hat{\alpha}_{LE} = \sum_{j'=1}^{r} a_{j'} p_{.j'}$ and fitting the linear model

$$E \begin{bmatrix} \hat{\alpha}_{RE} \\ \hat{\alpha}_{LE} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \mu$$

by weighted least squares. This statistic may be more compactly written as

$$X^2 = n(\hat{\alpha}_{RE} - \hat{\alpha}_{LE})^2 / \{ \sum_{j=1}^{r} \sum_{j'=1}^{r} [(a_j - a_{j'}) - (\hat{\alpha}_{RE} - \hat{\alpha}_{LE})]^2 p_{jj'} \} \qquad (2.4)$$

For the data in Table 1, the scores $a_1 = 1$, $a_2 = 2$, $a_3 = 3$, $a_4 = 4$ have been assigned to the grade categories. The appropriate value of X^2 may be computed as follows:

$$\hat{\alpha}_{RE} = \frac{17012}{7477} \qquad \hat{\alpha}_{LE} = \frac{17236}{7477} \qquad (\hat{\alpha}_{RE} - \hat{\alpha}_{LE})^2 = \left(\frac{-224}{7477} \right)^2$$

$$\sum_{j=1}^{r} \sum_{j'=1}^{r} (a_j - a_{j'})^2 p_{jj'} = (1) \left(\frac{1678}{7477}\right) + 4 \left(\frac{401}{7477}\right) + 9 \left(\frac{102}{7477}\right) = \frac{4200}{7477}$$

$$X^2 = 7477 \left(\frac{-224}{7477}\right)^2 / \{[(4200)(7477) - (224)^2] / (7477)^2\}$$

$$= 11.97$$

The similarity of the test given in (2.4) to the matched pairs t-test may be seen clearly by writing down the frequency distribution of the difference ($d_{jj'} = a_j - a_{j'}$) between the right eye grade and the left eye grade. These are shown in Table 2. The value of the average

TABLE 2

Difference (d)	Frequency (f)
-3	66
-2	202
-1	903
0	5296
1	775
2	199
3	36

difference $\bar{d}$ is $(-224/7477) \simeq -.03$ and its estimated standard error $s_{\bar{d}}$ is .0087. Hence

$$t^2 = \bar{d}^2 / s^2_{\bar{d}} = (-3.46)^2 = 11.97 \qquad (2.5)$$

which is identical to the value obtained for the X^2 test previously described.

As a second example, let us consider the following data (see Table 3) from Miettinen [1969] which pertain to a study of the extent to

TABLE 3

Previous History of Induced Abortion in Propositi with Ectopic Pregnancy and Matched Controls

			Number of Controls with Induced Abortion					
			0	1	2	3	4	Total
Propositus with induced abortion	no	0	5	1	0	0	0	6
	yes	1	3	5	3	0	1	12
	total		8	6	3	0	1	18

which induced abortions increase the risk of ectopic implantation in subsequent pregnancies. To each of 18 propositi with ectopic pregnancy following at least one earlier pregnancy, four control subjects were matched with respect to order of pregnancy, age, and husband's education. The history of induced abortions terminating any of the preceding pregnancies was then recorded for both the propositi and the controls. These data may be viewed as arising from an asymmetric mixed model in which each of the n = 18 sets of matched cases correspond to generalized subjects and classification is with respect to each level of the factor propositus vs. control. The complicating feature of these data is that the responses to the different factor levels are measured on seemingly different scales; i.e., propositus on a two-value scale and control on a five value scale. This difficulty can be resolved by formulating the hypothesis of marginal symmetry in an appropriate fashion. In particular, let $\pi_{jj'}$ denote the probability that a matched set is classified according to the j-th level with respect to propositus and the j'-th level with respect to controls.

If scores $a_0 = 0$, $a_1 = 1$ are attached to the responses for propositus and scores $b_0 = 0$, $b_1 = 1/4$, $b_2 = 2/4$, $b_3 = 3/4$, and $b_4 = 4/4$ are attached to the responses for controls, then

$$\alpha_p \equiv \sum_{j=0}^{1} a_j \pi_{j\cdot} = \sum_{j=0}^{1} \sum_{j'=0}^{4} a_j \pi_{jj'}$$
$$\alpha_c \equiv \sum_{j'=0}^{4} b_{j'} \pi_{\cdot j'} = \sum_{j=0}^{1} \sum_{j'=0}^{4} b_{j'} \pi_{jj'} \qquad (2.6)$$

represent the probability of observing an induced abortion in a propositus and the expected relative frequency of observing an induced abortion in a set of four controls respectively. Hence, the hypothesis of interest is H_0: $\alpha_p = \alpha_c$. Essentially, the same test indicated in (2.4) is applicable here.

Following the GSK approach, one would first apply the A-matrix

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ .00 & .25 & .50 & .75 & 1.00 & .00 & .25 & .50 & .75 & 1.00 \end{bmatrix}$$

to the data in Table 3 arranged in a single (10×1) column vector with the transpose of row 1 on top and that of row 2 underneath. This would generate $\hat{\alpha}_p$ and $\hat{\alpha}_c$ to which the linear model

$$E \begin{bmatrix} \hat{\alpha}_p \\ \hat{\alpha}_c \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \mu \quad \text{where} \quad \begin{bmatrix} \hat{\alpha}_p \\ \hat{\alpha}_c \end{bmatrix} = \begin{bmatrix} 0.667 \\ 0.222 \end{bmatrix}$$

is fitted by weighted least squares. The test statistic (with D.F. = 1) again is the residual sum of squares and for these data is given by $X^2 = 20.85$. This result is consistent with that of Miettinen [1969] who observed essentially $X^2 = 16.00$ without continuity correction and $X^2 = 13.69$

with continuity correction. This is not unexpected since both the test procedure given here and that of Miettinen are in some sense asymptotically equivalent. The advantage of the one proposed here is that it is an easily produced special case of a very general program.

3. Mixed Categorical Data Models of Higher Orders

In general a "mixed categorical data model of order d" involves exposing each of n randomly chosen subjects from some homogeneous population to each level of a factor with d levels and classifying each response into one of r categories. The resulting data are then represented in an $r \times r \times \ldots \times r$ contingency table of d dimensions which will be assumed to follow a multinomial distribution. This represents a categorical version of the classical mixed model with n subjects and d treatments (see Eisenhart [1947] and Koch and Sen [1968]). The hypothesis of interest is equality of treatment effects or homogeneity over factor levels. For the case $r = 2$, Cochran [1950] has proposed a test (sometimes called the Q-test) by a permutation argument. An alternative test due to Bhapkar [1965] has been formulated in terms of d-th order marginal symmetry and can be produced from the weighted least squares approach. This procedure is illustrated in Grizzle, Starmer, and Koch [1969] for data indicating the responses of 46 subjects to each of the drugs A, B, and C as either being favorable or not (see Table 4). This example

TABLE 4

Tabulation of Responses to Drugs A, B, and C

Response to A		favorable			not favorable			total
Response to B		fav.	not fav.	total	fav.	not fav.	total	
Response to C	fav.	6	2	8	2	6	8	16
	not fav.	16	4	20	4	6	10	30
	total	22	6	28	6	12	18	46

is a mixed categorical data model of order 3 with drug type being the factor and having levels A, B, and C. The hypothesis of interest is equality of the probability of a favorable response for each of the three drugs; i.e., equality of the corresponding marginal distributions or marginal symmetry of the three one-way margins.

In the continuous case, the hypothesis of equality of treatment effects in mixed models is most appropriately tested by a Hotelling T^2-procedure (see Scheffé [1959]) unless certain symmetry relationships hold in which case an analysis of variance test applies. Both of these tests together with certain non-parametric counterparts are described in Koch and Sen [1968]. Hence, it is apparent that in the general categorical data setting of interest here, the test statistics applied to the hypothesis of marginal symmetry bear the same analogy to the Hotelling T^2-test as the test statistics used in the previous section bore to the classical matched pairs t-test. For the data in Table 4, the GSK approach led to a $X^2 = 6.58$ with D.F. = 2 indicating a significant difference ($\alpha = .05$) between drugs. The reader is referred to that paper for the details leading to this result and the Appendix for the pertinent theory.

A more complex example can be seen in terms of the following data (see Table 5) of Lessler [1962] which have been described in Bhapkar and Koch [1968]. In this case, there are two groups of male subjects in a study of the nature of sexual symbolism in objects: Group A members were not told the purpose of the experiment while Group C subjects were. Each subject was asked to classify as masculine or feminine (M or F) certain objects which had been previously characterized as being culturally

TABLE 5

Responses of Males to Culturally Masculine and Anatomically Feminine Objects with Weak Intensity

Resp. at 1/5 Sec.			M			F			Total
Resp. at 1/100 Sec.			M	F	Total	M	F	Total	
A	Resp. at 1/1000 Sec.	M	171	6	177	7	7	14	191
		F	18	12	30	7	56	63	93
	Total		189	18	207	14	63	77	284
C	Resp. at 1/1000 Sec.	M	184	10	194	7	20	27	221
		F	38	14	52	7	114	121	173
	Total		222	24	246	14	134	148	394

masculine and anatomically feminine but with weak intensity at three different exposure rates: 1/5 second, 1/100 second, and 1/1000 second. Hence, this is a "three response, one factor" situation with Group being the factor and the M or F classification at different exposure rates being the responses. However, since it is of interest to compare the responses at different exposure rates, it is apparent that the data from either one of the groups A or C by themselves is a mixed categorical data model of order 3. Thus, one may call the data for the combined groups a "split plot categorical data" model. Hence, the hypotheses of interest are equality of group (whole-plot) effects, equality of exposure rate (split-plot) effects, and no interaction between exposure rates and groups. All of these may be readily formulated by using the GSK approach. First let us define the following relative frequency vectors for the two groups

$$\underset{\sim}{p}_A' = (171, 18, 6, 12, 7, 7, 7, 56)(1/284)$$

$$\underset{\sim}{p}_C' = (184, 38, 10, 14, 7, 7, 20, 114)(1/394) \qquad (3.1)$$

$$\underset{\sim}{p}' = (\underset{\sim}{p}_A', \underset{\sim}{p}_C')$$

The overall vector $\underset{\sim}{p}$ is to be pre-multiplied by the following A-matrix to determine functions $\underset{\sim}{F} = \underset{\sim}{A}\,\underset{\sim}{p}$ where

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \end{bmatrix} \qquad (3.2)$$

The resulting statistics are estimates of the probability of masculine classification at the exposure rates 1/5 second, 1/100 second, 1/1000 second respectively in Groups A and C respectively. The next step in the analysis is to fit the following linear model by weighted least squares

$$E\{\underset{\sim}{F}\} = \begin{bmatrix} 1 & 1 & -1 & 0 & 0 & 0 \\ 1 & 0 & 2 & 0 & 0 & 0 \\ 1 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} \mu_A \\ \lambda_{A_1} \\ \lambda_{A_2} \\ \mu_C \\ \lambda_{C_1} \\ \lambda_{C_2} \end{bmatrix} \quad \text{where } \underset{\sim}{F} = \begin{bmatrix} 0.729 \\ 0.715 \\ 0.673 \\ 0.624 \\ 0.599 \\ 0.561 \end{bmatrix} \qquad (3.3)$$

and μ_A is a mean effect in Group A, λ_{A_1} is the linear effect of exposure in Group A, λ_{A_2} is the non-linear effect of exposure in Group A, and μ_C, λ_{C_1}, λ_{C_2} are similarly defined quantities in Group C. The hypothesis of no group × exposure rate interaction may be formulated as

$$H_0: \lambda_{A_1} = \lambda_{C_1}, \ \lambda_{A_2} = \lambda_{C_2}$$

which may be tested by using the following hypothesis matrix

$$C = \begin{bmatrix} 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 \end{bmatrix}$$

in accordance with the principles of weighted regression analysis. The resulting $X^2 = 0.20$ with D.F. = 2 indicates that the data are consistent with this hypothesis of no interaction. Hence, the model is revised as

$$E\{\underset{\sim}{F}\} = \begin{bmatrix} 1 & 1 & 1 & -1 \\ 1 & 1 & 0 & 2 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & 0 & 2 \\ 1 & -1 & -1 & -1 \end{bmatrix} \begin{bmatrix} \mu \\ \gamma \\ \lambda_1 \\ \lambda_2 \end{bmatrix} \qquad (3.4)$$

in which μ is an overall mean, γ is the differential group effect, λ_1 is the linear effect of exposure, and λ_2 is the non-linear effect of exposure. The hypothesis of no group main effects may be formulated as

$$H_0: \gamma = 0$$

and tested by using the hypothesis matrix

$$C = [0 \quad 1 \quad 0 \quad 0]$$

while the hypothesis of no exposure rate main effects may be formulated as

$$H_0: \lambda_1 = \lambda_2 = 0$$

and tested by using the hypothesis matrix

$$C = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} .$$

The resulting test statistics respectively are $X^2 = 11.57$ with D.F. = 1 and $X^2 = 14.55$ with D.F. = 2. Hence, it is apparent that there are both significant differences between the two groups and among the three exposure rates. In fact, the differences among the three exposure rates are linear as can be seen by testing separately for the linear and for the non-linear effect of exposure. Thus, we see that split-plot contingency tables can be analyzed by the linear model approach as readily as the more standard types of multi-response, multi-factor experiments treated in Grizzle, Starmer, and Koch [1969].

4. Some Other Related Problems

The mixed models considered in this paper may be viewed as belonging to Case III of Koch and Sen [1968] for which the Hotelling T^2-test is applicable. However, since it is stated here that a single multinomial distribution applies to the entire table for the underlying homogeneous population, it is apparent that the discussion given here not only requires additive subject effects but also no subject effects at all. On the other hand if subject effects are additive and are selected at random, then this same analysis is applicable in an approximate sense.

The categorical situation, analogous to Case IV of Koch and Sen in which the analysis of variance F-test is appropriate, occurs when the responses of the same subject over repeated trials to the different levels of the factor are independent. This means, in the case of a mixed model of order d, that the $r \times r \times \ldots \times r$ contingency table with n observations in it can be replaced by a $d \times r$ table with nd observations in which the d rows correspond to the d different treatments or factor levels and the r columns correspond to the levels of the response. The frequencies in each row are the appropriate first-order marginal distributions corresponding to each of the d dimensions of the $r \times r \times \ldots \times r$ table. Hence, the hypothesis of marginal symmetry in the complex table may be translated here into the traditional hypothesis of homogeneity in a "one factor, one response" two-way table. Thus, it may be tested by applying the classical χ^2-test to this two-way table. One important difference between this test procedure and the one in the complex table is that it is based on an effective sample size of nd rather than n. Thus, if Case IV applies, then this type of test is naturally more powerful than those given for Case III. Finally, as pointed out in Bhapkar [1968], if the response is quantitatively scaled, then the hypothesis of equality of mean scores in the two-way table may be tested instead of the usual homogeneity test. Here, this test would be a Case IV alternative to the type of test described in (2.4).

The more general Cases I and II of Koch and Sen allow for complex subject effects, and hence in the case of categorical data, involve product multinomial distributions. As a result, the methods given here are not directly applicable. However, such experimental situations can be handled to some extent by fitting complex linear models with both treatment effects and subject effects. The details of this type of analysis of product multinomial models are given in Grizzle, Starmer and Koch [1969].

5. Appendix of Theoretical Results

Let $\pi_{j_1 j_2 \cdots j_d}$ denote the probability of cell $(j_1, j_2, \ldots, j_d)$ where $j_\xi = 1, 2, \ldots, r$ with $\xi = 1, 2, \ldots, d$ in a d-dimensional table. Define $\phi_{\xi j}$ to be the marginal probability of level j for the ξ-th dimention. Hence

$$\phi_{\xi j} = \sum_{\substack{\nu=1 \\ \nu\neq\xi}}^{d} \sum_{\substack{j_\nu=1 \\ j_\xi=j}}^{r} \pi_{j_1 j_2 \cdots j_d} \qquad (5.1)$$

The hypothesis of d-th order marginal symmetry on the one-way margins may be written

$$\phi_{1j} = \phi_{2j} = \ldots = \phi_{dj} \qquad j = 1, 2, \ldots, r \qquad (5.2)$$

This hypothesis may be readily tested by letting $n_{j_1 j_2 \cdots j_d}$ denote the observed frequency of cell $(j_1, j_2, \ldots, j_d)$. An A-matrix is then produced which generates the estimates $p_{\xi j}$ of $\phi_{\xi j}$ obtained as

$p_{\xi j} = \frac{1}{n} \sum_{\substack{\nu=1 \\ \nu\neq\xi}}^{d} \sum_{\substack{j_\nu=1 \\ j_\xi=j}}^{r} n_{j_1 j_2 \ldots j_d}$ where n is the total sample size. For example, if d = 3 and r = 3, we have

$$A = \begin{bmatrix} 1\,1\,1 & 1\,1\,1 & 1\,1\,1 & 0\,0\,0 & 0\,0\,0 & 0\,0\,0 & 0\,0\,0 & 0\,0\,0 & 0\,0\,0 \\ 0\,0\,0 & 0\,0\,0 & 0\,0\,0 & 1\,1\,1 & 1\,1\,1 & 1\,1\,1 & 0\,0\,0 & 0\,0\,0 & 0\,0\,0 \\ 0\,0\,0 & 0\,0\,0 & 0\,0\,0 & 0\,0\,0 & 0\,0\,0 & 0\,0\,0 & 1\,1\,1 & 1\,1\,1 & 1\,1\,1 \\ 1\,1\,1 & 0\,0\,0 & 0\,0\,0 & 1\,1\,1 & 0\,0\,0 & 0\,0\,0 & 1\,1\,1 & 0\,0\,0 & 0\,0\,0 \\ 0\,0\,0 & 1\,1\,1 & 0\,0\,0 & 0\,0\,0 & 1\,1\,1 & 0\,0\,0 & 0\,0\,0 & 1\,1\,1 & 0\,0\,0 \\ 0\,0\,0 & 0\,0\,0 & 1\,1\,1 & 0\,0\,0 & 0\,0\,0 & 1\,1\,1 & 0\,0\,0 & 0\,0\,0 & 1\,1\,1 \\ 1\,0\,0 & 1\,0\,0 & 1\,0\,0 & 1\,0\,0 & 1\,0\,0 & 1\,0\,0 & 1\,0\,0 & 1\,0\,0 & 1\,0\,0 \\ 0\,1\,0 & 0\,1\,0 & 0\,1\,0 & 0\,1\,0 & 0\,1\,0 & 0\,1\,0 & 0\,1\,0 & 0\,1\,0 & 0\,1\,0 \\ 0\,0\,1 & 0\,0\,1 & 0\,0\,1 & 0\,0\,1 & 0\,0\,1 & 0\,0\,1 & 0\,0\,1 & 0\,0\,1 & 0\,0\,1 \end{bmatrix}$$

Since the $\phi_{\xi j}$ necessarily satisfy restrictions of the type $\sum_{j=1}^{r} \phi_{\xi j} = 1$, the A-matrix is modified to generate $p_{\xi j}$ for only the combinations $\xi = 1, 2, \ldots, d$, $j = 1, 2, \ldots, (r-1)$. In the example, this means rows 3, 6, 9 would be deleted from A. To the resulting estimates $p_{\xi j}$, the following linear model would be fitted by weighted least squares

$$E\{p_{\xi j}\} = \mu_j \qquad \begin{array}{l} \xi = 1, 2, \ldots, d \\ j = 1, 2, \ldots, (r-1) \end{array} \tag{5.3}$$

This can be written more fully as

$$E \begin{bmatrix} p_{11} \\ p_{12} \\ \cdots \\ p_{1,r-1} \\ p_{21} \\ p_{22} \\ \cdots \\ p_{2,r-1} \\ \cdots \\ \cdots \\ \cdots \\ p_{d1} \\ p_{d2} \\ \cdots \\ p_{d,r-1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ & \cdots & & \\ 0 & 0 & \ldots & 1 \\ 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ & \cdots & & \\ 0 & 0 & \ldots & 1 \\ & \cdots & & \\ & \cdots & & \\ & \cdots & & \\ 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ & \cdots & & \\ 0 & 0 & & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \cdots \\ \mu_{r-1} \end{bmatrix} \tag{5.4}$$

The residual sum of squares with D.F. = (d-1)(r-1) is the test statistic for the hypothesis of marginal symmetry in (5.2). When (5.2) holds, this statistic is approximately distributed as χ^2 with D.F. = (d-1)(r-1).

Alternatively, if scores $a_1, a_2, \ldots, a_r$ are assigned to the response categories, then the hypothesis of equality of mean scores over the d marginals may be written as

$$\alpha_1 = \alpha_2 = \ldots = \alpha_d \qquad (5.5)$$

where $\alpha_\xi = \sum_{j=1}^{r} a_j \phi_{\xi j}$. The hypothesis (5.5) may be tested by obtaining the estimates $\hat{\alpha}_\xi = \sum_{j=1}^{r} a_j p_{\xi j}$ where $\xi = 1, 2, \ldots, d$ and fitting the linear model

$$E \begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \cdots \\ \hat{\alpha}_d \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \cdots \\ 1 \end{bmatrix} \mu \qquad (5.6)$$

by weighted least squares; again the test statistic is the residual sum of squares with D.F. = (d-1).

The tests given here for (5.2) and (5.5) appear similar to the classical ones of homogeneity and equality of mean scores in "one response, one factor" tables. However, here the underlying table is d-dimensional and the $p_{\xi j}$ are correlated with each other while in the "one response, one factor" tables the underlying table is two-dimensional and the $p_{\xi j}$ are independent of $p_{\xi' j}$ for all $\xi' \neq \xi$. Hence, in the underlying estimates of variance and covariance, the two types of procedures are quite different.

Finally, it is obvious that other types of hypotheses of marginal symmetry (eg. equality of all two-way margins) may be easily tested by this approach. All that is required is choosing an appropriate A-matrix, and then fitting a linear model specific to the hypothesis of interest. Since such tests are usually not readily interpreted experimentally in the context of mixed models, they will not be discussed in any more detail here.

REFERENCES

1. Bennett, B. M. [1967]. Tests of hypotheses concerning matched samples. J. R. Statist. Soc. B 29, 468-74.

2. Bhapkar, V. P. [1965]. Categorical data analogs of some multivariate tests. S. N. Roy Memorial Volume, University of North Carolina Press, Chapel Hill, N.C.

3. Bhapkar, V. P. [1966]. A note on the equivalence of two test criteria for hypotheses in categorical data. J. Amer. Statist. Assoc. 61, 228-35.

4. Bhapkar, V. P. [1968]. On the analysis of contingency tables with a quantitative response. Biometrics 24, 329-38.

5. Bhapkar, V. P. and Koch, G. G. [1968]. On the hypotheses of 'no interaction' in contingency tables. Biometrics 24, 567-94.

6. Billewicz, W. Z. [1964]. Matched samples in medical investigations. Brit. J. Prev. Soc. Med. 18, 167-73.

7. Billewicz, W. Z. [1965]. The efficiency of matched samples: an empirical investigation. Biometrics 21, 623-44.

8. Chase, G. R. [1968]. On the efficiency of matched pairs in Bernoulli trials. Biometrika 55, 365-69.

9. Cochran, W. G. [1950]. The comparison of percentages in matched samples. Biometrika 37, 256-66.

10. Eisenhart, C. [1947]. The assumptions underlying the analysis of variance. Biometrics 3, 1-21.

11. Goodman, L. A. [1970]. The multivariate analysis of qualitative data: interactions among multiple classifications. J. Amer. Statist. Assoc. 65.

12. Grizzle, J. E., Starmer, C. F., and Koch, G. G. [1969]. Analysis of categorical data by linear models. Biometrics 25, 489-504.

13. Ireland, C. T., Ku, H. H., and Kullback, S. [1969]. Symmetry and marginal homogeneity of an $r \times r$ contingency table. J. Amer. Statist. Assoc. 64, 1323-41.

14. Koch, G. G. and Sen, P. K. [1968]. Some aspects of the statistical analysis of the 'mixed model'. Biometrics 24, 27-48.

15. Ku, H. H. and Kullback, S. [1968]. Interaction in multidimensional contingency tables: an information theoretic approach. J. of Res. of the Nat. Bureau of Stds. - Math. Sci, 72B, 159-99.

16. Lessler, K. J. [1962]. The anatomical and cultural dimensions of sexual symbols. Unpublished Ph.D. thesis, Michigan State University.

17. Lewis, J. A. [1968]. A program to fit constants to multiway tables of quantitative and quantal data. <u>Appl. Statist. 17</u>, 33-42.

18. Mantel, N. and Haenszel, W. [1959]. Statistical aspects of the analysis of data from retrospective studies of disease. <u>J. Nat. Cancer Inst. 22</u>, 719-48.

19. Miettinen, O. S. [1968]. The matched pairs design in the case of all-or-none responses. <u>Biometrics 24</u>, 339-52.

20. Miettinen, O. S. [1969]. Individual matching with multiple controls in the case of all-or-none responses. <u>Biometrics 25</u>, 339-55.

21. Neyman, J. [1949]. Contribution to the theory of the χ^2 test. pp. 239-73 in <u>Proc. Berkeley Symp. Math. Statist. Prob.</u> University of California Press, Berkeley and Los Angeles.

22. Scheffé, H. [1959]. <u>The Analysis of Variance</u>. John Wiley and Sons, Inc., New York.

23. Stuart, A. [1955]. A test for homogeneity of the marginal distributions in a two-way classification. <u>Biometrika 42</u>, 412-16.

24. Wald, A. [1943]. Tests of statistical hypotheses concerning several parameters when the number of observations is large. <u>Trans. Amer. Math. Soc. 54</u>, 426-82.

25. Worcester, J. [1964]. Matched samples in epidemiologic studies. <u>Biometrics 20</u>, 840-48.