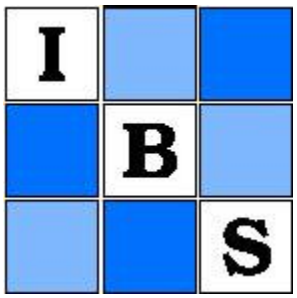

BIOMETRICS 29, 205–220
March 1973

THE ANALYSIS OF RESIDUALS IN CROSS-CLASSIFIED TABLES[1]

SHELBY J. HABERMAN

Department of Statistics, University of Chicago, Chicago, Illinois 60637, U.S.A.

SUMMARY

Techniques are proposed for analysis of residuals associated with log-linear models for frequency tables. Results are applied to two-way tables and logit models.

1. INTRODUCTION

In regression analysis and analysis of variance, examination of residuals forms an important part of good statistical practice (see Draper and Smith [1966]). More general uses of residuals have also been explored by Cox and Snell [1968]. In this paper, methods of residual analysis developed for log-linear models for cross-classified tables by Haberman [1972b] are used to examine models for quantal response and for complete and incomplete $r \times c$ contingency tables.

Section 2 contains a brief summary of results of Haberman [1972b] which are used in this paper. A more complete summary is provided in the appendix. In section 3, residual analysis is used in the hypothesis of row and column independence in a complete $r \times c$ table. In section 4, methods are developed appropriate for analysis of quasi-independence models in incomplete $r \times c$ tables, and in section 5, methods for use in logit analysis are considered.

2. BASIC PROPERTIES OF RESIDUALS IN LOG-LINEAR MODELS

Following the notation of Haberman [1970; 1972a, b, c], we consider a frequency table $\mathbf{n} = \{n_i : i \in I\}$ indexed by a set I with $q > 0$ elements. This table is assumed to be an I-tuple with mean $\mathbf{m} = \{m_i\}$ such that $m_i > 0$ for $i \in I$. The table satisfies a log-linear model if the log-mean vector $\boldsymbol{\mu} = \{\log m_i\}$ is assumed to be in a linear subspace $\mathfrak{m}$ and the vector $\mathbf{n}$ of observations results from any one of the sampling procedures described in the Appendix.

Under any of these sampling schemes, we define a vector $\mathbf{e} = \{(n_i - \hat{m}_i)/\sqrt{\hat{m}_i}\}$ of standardized residuals, where $\hat{\mathbf{m}}$ is the maximum likelihood estimate of $\mathbf{m}$ if $\boldsymbol{\mu} \in \mathfrak{m}$ and if $\mathbf{n}$ is obtained by the Poisson or multinomial

[1] This research was carried out in the Department of Statistics, University of Chicago, under partial support by Research Grant No. NSF GS-31967x from the Division of the Social Sciences of the National Science Foundation.

sampling procedures described in the Appendix. As noted in the Appendix, the vector $\mathbf{e}$ has the property that for any value of $\boldsymbol{\mu}$ in $\mathfrak{m}$, the asymptotic variance v_i of element e_i is less than or equal to 1 for any $i \in I$.

The definition of $\mathbf{e}$ is to some extent arbitrary, for vectors such as $\{2[\sqrt{n_i} - \sqrt{\hat{m}_i}]\}$ and $\{\sqrt{n_i} + \sqrt{(n_i + 1)} - \sqrt{(4\hat{m}_i + 1)}\}$ are asymptotically equivalent to $\mathbf{e}$. One virtue of the definition of $\mathbf{e}$ employed in this paper is that the Pearson chi-square statistic X^2 satisfies

$$\begin{aligned} X^2 &= \sum (n_i - \hat{m}_i)^2/\hat{m}_i \\ &= \sum e_i^2 . \end{aligned} \tag{1}$$

Some useful analysis is possible if v_i does not vary much for $i \in I$, particularly if each asymptotic variance v_i is close to 1. At the cost of some extra computational labor, one may perform a more precise analysis by use of the vector $\mathbf{d}$ of adjusted residuals $d_i = e_i/\sqrt{\hat{v}_i}$, where $\hat{v}_i$ is the maximum likelihood estimate of v_i. Each coordinate d_i is approximately distributed has a standard normal deviate, even if v_i differs substantially from 1. In some cases, a compromise may be achieved between computational labor and variable variances of standardized residuals by considering an alternative adjustment $e^* = \{e_i/\sqrt{\hat{u}_i}\}$, where $\hat{u}_i$ is the maximum likelihood (ML) estimate of the variance u_i of the ith standardized residual corresponding to the model $\boldsymbol{\mu} \in \mathfrak{n}$, where $\mathfrak{n}$ is a linear subspace of $\mathfrak{m}$. In some cases, a suitable choice of $\mathfrak{n}$ may result in an easily computed $\hat{\mathbf{u}}$. If $\boldsymbol{\mu} \in \mathfrak{m}$, each coordinate e_i^* is approximately normally distributed with mean 0 and variance v_i/u_i between v_i and 1. Such a compromise is examined in section 5.

Given these standardized and adjusted residuals, one may apply various graphical and analytical techniques to detect deviations from the proposed model. These techniques, which are discussed by Anscombe and Tukey [1963] and Draper and Smith [1966], depend on the structure of the data. To illustrate their use, we consider the analysis of two-way tables and simple logit models.

3. COMPLETE TWO-WAY TABLES

The classical problem in the analysis of an $r \times c$ table is the test for independence of the variables representing the rows and columns of the table. If the table arises from a single multinomial sample of size N with cell probabilities $\{p_{ij}\}$, then the independence hypothesis is the hypothesis that

$$p_{ij} = p_{i+}p_{+j}, \quad 1 \le i \le r, \quad 1 \le j \le c, \tag{2}$$

where

$$p_{i+} = \sum_{j=1}^{c} p_{ij}, \qquad 1 \le i \le r,$$

and

$$p_{+j} = \sum_{i=1}^{r} p_{ij}, \qquad 1 \le j \le c.$$

As shown by Bishop and Fienberg [1969], the hypothesis that (2) holds is equivalent if $p_{ij} > 0$, $1 \leq i \leq r$, $1 \leq j \leq c$, to the hypothesis that for some α, $\{\beta_i\}$, and $\{\gamma_j\}$, the vector $\boldsymbol{\mu} = \{\log (Np_{ij})\}$ of log-means satisfies

$$\mu_{ij} = \alpha + \beta_i + \gamma_j, \quad 1 \leq i \leq r, \quad 1 \leq j \leq c, \tag{3}$$

where

$$\sum \beta_i = \sum \gamma_j = 0.$$

Under Poisson or multinomial sampling (see Appendix), the ML estimate $\hat{\mathbf{m}}$ is $\{n_{i+}n_{+j}/n_{++}\}$, provided $n_{i+} > 0$, $1 \leq i \leq r$, and $n_{+j} > 0$, $1 \leq j \leq c$. Thus the standardized residual has coordinates

$$e_{ij} = (n_{ij} - n_{i+}n_{+j}/n_{++})/\sqrt{n_{i+}n_{+j}/n_{++}}\,. \tag{4}$$

As shown in Haberman [1972a], the asymptotic variance is

$$v_{ij} = (1 - m_{i+}/m_{++})(1 - m_{+j}/m_{++}), \quad 1 \leq i \leq r, \quad 1 \leq j \leq c, \tag{5}$$

and its ML estimate is

$$\hat{v}_{ij} = (1 - n_{i+}/n_{++})(1 - n_{+j}/n_{++}), \quad 1 \leq i \leq r, \quad 1 \leq j \leq c. \tag{6}$$

Thus the vector $\mathbf{d}$ of adjusted residuals has coordinates $e_{ij}/\sqrt{\hat{v}_{ij}}$, where e_{ij} satisfies (4) and $\hat{v}_{ij}$ satisfies (6). Since $\mathbf{d}$ is easily computed in this example, it appears reasonable to base all analysis on adjusted residuals.

To illustrate use of adjusted residuals in data analysis, consider the data presented in Table 1, which is taken from Davies [1961]. The observa-

TABLE 1

PISTON-RING FAILURES IN FOUR COMPRESSORS

Comp. No.	Leg			Total
	North	Center	South	
1	17	17	12	46
2	11	9	13	33
3	11	8	19	38
4	14	7	28	49
Total	53	41	72	166

tions in this 4 × 3 table represent the number of piston ring failures in each leg of four compressors at an Imperial Chemical Industries plant. If $p(j \mid i)$ is the probability of failure in leg j given failure in compressor i, then one hypothesis of interest is that $p(j \mid i) = p(j \mid i')$ for $i \neq i'$ and $1 \leq j \leq 3$. This hypothesis is equivalent to the hypothesis that the log-mean μ satisfies (3). Since the Pearson chi-square for this table is 11.7 and there are 6 degrees of freedom (D.F.), there is some relatively weak evidence that the probabilities of failure vary for different compressors. Davies [1961] looks at special two-way tables constructed from Table 1 to find whether further evidence exists that the behavior of the compressors varies. A more effective investigation can be conducted by use of the adjusted residuals, which are given in Table 2.

TABLE 2

ADJUSTED RESIDUALS FOR PISTON-RING FAILURES

Comp. No.	Leg		
	North	Center	South
1	.86	2.27	-2.78
2	.19	.38	-.52
3	-.45	-.59	.94
4	-.60	-2.01	2.32

Inspection of Table 2 shows that 4 adjusted residuals—d_{12} , d_{13} , d_{42} , and d_{43}—all have magnitudes greater than 2. The largest residual in magnitude is d_{13} , which is -2.78. The probability that a standard normal deviate exceeds 2.78 in absolute value is only 0.0054. Since d_{13} has been selected out of 12 residuals, one cannot conclusively state that the South leg of compressor 1 does not conform to the independence model, but there are obvious reasons for suspicion.

A graph of the adjusted residuals on normal probability paper shows further evidence that the model may not be satisfactory. Following Tukey [1962], we assign the kth smallest adjusted residual d_{ij} to the probability coordinate $(3k - 1)/(3rc + 1) = (3k - 1)/37$. The resulting graph is shown

in Figure 1. Ideally, the residuals should lie near the straight line in the figure, which represents percentiles for the standard normal distribution. The residuals d_{13} , d_{42} , d_{12} , and d_{43} all deviate sharply from this line. Thus

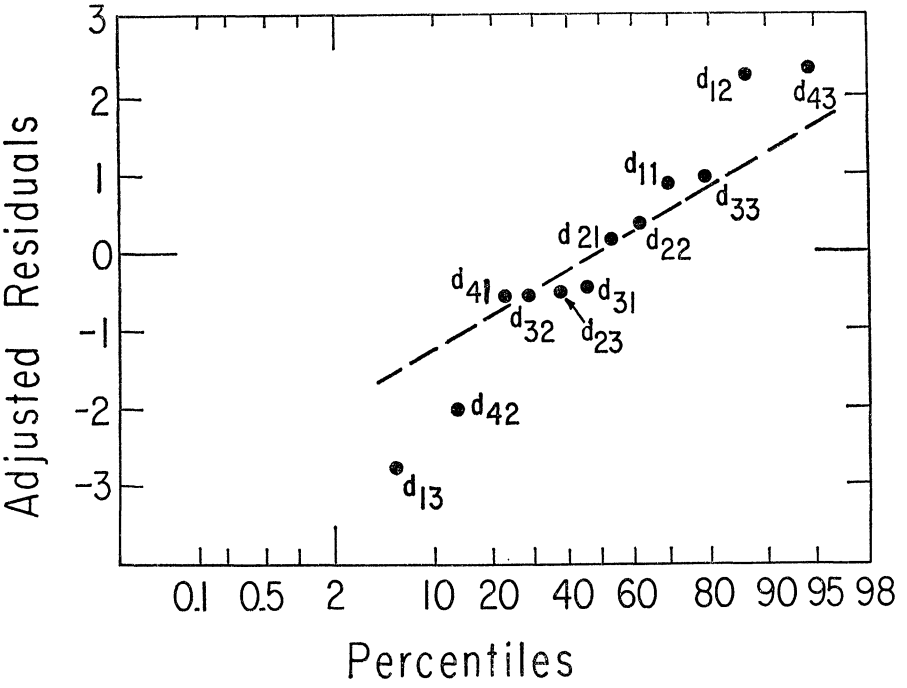

FIGURE 1

FULL NORMAL PLOT OF ADJUSTED RESIDUALS FOR PISTON-RING FAILURES

there is ample reason to suspect deviations from the model in compressors 1 and 4. These deviations may represent lack of independence or they may indicate that the number of piston ring failures is more variable than should be the case if they have a Poisson distribution.

In this example, use of adjusted residuals rather than standardized residuals has a substantial effect on the analysis. The largest standardized residual in magnitude is e_{13} , which is -1.78. Thus the size of the standardized residuals provides no suggestion of inadequacies in the model. Inspection of (5) suggests that use of adjusted residuals rather than standardized residuals is likely to make a substantial difference in the analysis unless r and c are large enough so that all expected row and column totals m_{i+} and m_{+j} are small fractions of the expected total number of observations m_{++} .

4. INCOMPLETE TWO-WAY TABLES

In some $r \times c$ tables, cell counts n_{ij} are only examined for indices (i, j) in some proper subset I of $\bar{r} \times \bar{c}$. To facilitate analysis of such tables, Bishop

and Fienberg [1969], Fienberg [1970; 1972], Goodman [1968], Mantel [1970], and Wagner [1970] have considered quasi-independence models for which, in analogy to (3), one has

$$\mu_{ij} = \alpha + \beta_i + \gamma_j , \quad (i, j) \in I, \tag{7}$$

where

$$\sum \beta_i = \sum \gamma_j = 0.$$

The set of vectors $\mathfrak{y}$ which satisfy (7) is a linear manifold with dimension depending on I. The conditions for multinomial sampling are satisfied if the entire table is a single multinomial sample, if each row $\{n_{ij} : j \in K_i\}$, $i = 1, \cdots, r$, is an independent multinomial sample, or if each column $\{n_{ij} : i \in L_j\}$, $j = 1, \cdots, c$, is an independent multinomial sample, where $K_i = \{j: (i, j) \in I\}$ and $L_j = \{i: (i, j) \in I\}$. The conditions of this paper are also satisfied if all cell counts are independent Poisson random variates or if sampling is conditional on the row and column totals.

Depending on I, the ML estimate $\hat{\mathfrak{m}}$ for Poisson or multinomial sampling may have an expression in closed form or it may be computed by iterative proportional fitting (see Goodman [1968], Bishop and Fienberg [1969], Fienberg [1970], and Haberman [1972b]). Computation of $\hat{\theta}_i$ is more difficult than in the independence model for complete tables, but formulas from Haberman [1972b] may be used to compute $\hat{\theta}_i$ if $\hat{\mathfrak{m}}$ has a closed-form expression. If $\hat{\mathfrak{m}}$ does not have a closed-form expression, then $\hat{\theta}_{ij} = 1 - \gamma_{ij}^{(i,j)}$, where $\gamma^{(i,j)}$ is the solution of the normal equations

$$\sum_{j' \in K_i} \gamma_{ij'}^{(i,j)} \sqrt{\hat{m}_{ij'}} = \sqrt{\hat{m}_{ij}}\, \delta_{ii'} , \qquad i' = 1, \cdots, r, \tag{8}$$

$$\sum_{i' \in L_j} \gamma_{i'j}^{(i,j)} \sqrt{\hat{m}_{i'j}} = \sqrt{\hat{m}_{ij}}\, \delta_{jj'} , \qquad j' = 1, \cdots, c. \tag{9}$$

To illustrate use of residual analysis in an incomplete table, consider Table 3, which was originally analyzed by Bishop and Fienberg [1969] by use of a quasi-independence model. In this table, stroke patients are rated at entrance and discharge for severity of condition on a scale from A (least severe) to E (most severe). Since a patient is not discharged unless his condition is no worse than at entrance, only cell counts n_{ij} for which $i \geq j$ can be positive. Consequently, Bishop and Fienberg propose a quasi-independence model for the data with $I = \{(i, j): 1 \leq j \leq i \leq 5\}$. Given this model, Bishop and Fienberg obtain the ML estimate $\hat{\mathfrak{m}}$ shown in Table 4. The corresponding Pearson χ^2 of 8.37 on 6 D.F. is not large, but the sample is sufficiently small so that an examination of residuals is useful in order to ensure that no systematic deviations from the model are present. To accomplish this objective, Tables 3 and 4 may be employed to compute the standardized residuals shown in Table 5. These residuals are not large and exhibit no obvious patterns. It should be noted that the standardized residuals e_{11} and e_{55} are always 0 in triangular tables of this type.

In this example, the more precise analysis possible by use of adjusted

TABLE 3

INITIAL AND FINAL RATINGS ON DISABILITY OF 121 STROKE PATIENTS

Initial State	Final State				
	A	B	C	D	E
A	5	—	—	—	—
B	4	5	—	—	—
C	6	4	4	—	—
D	9	10	4	1	—
E	11	23	12	15	8

TABLE 4

ESTIMATED MEANS FOR INITIAL AND FINAL RATINGS ON DISABILITY OF 121 STROKE PATIENTS

Initial State	Final State				
	A	B	C	D	E
A	5.00	—	—	—	—
B	3.75	5.25	—	—	—
C	4.43	6.20	3.37	—	—
D	6.16	8.63	4.69	4.52	—
E	15.66	21.92	11.94	11.48	8.00

TABLE 5

STANDARDIZED RESIDUALS FOR INITIAL AND FINAL RATINGS ON DISABILITY OF 121 STROKE PATIENTS

Initial State	Final State				
	A	B	C	D	E
A	0.00	—	—	—	—
B	0.13	-0.11	—	—	—
C	0.75	-0.88	0.34	—	—
D	1.14	0.47	-0.32	-1.66	—
E	-1.18	0.23	0.02	1.04	0.00

residuals is possible but rather more tedious than in the complete case. By Haberman [1972b], the adjustment factor $\hat{v}_{ij}$ is

$$\hat{v}_{ij} = 1 - \hat{m}_{ij}\left[1/n_{+j} + 1/n_{i+} - \sum_{k=i}^{j}(1/A_k) + \sum_{k=i+1}^{j}(1/B_k)\right], \qquad (10)$$

where

$$n_{+j} = \sum_{i=j}^{5} n_{ij}\,, \qquad 1 \le j \le 5, \qquad (11)$$

$$n_{i+} = \sum_{j=1}^{i} n_{ij}\,, \qquad 1 \le i \le 5, \qquad (12)$$

$$A_k = \sum_{j=1}^{k}\sum_{i=k}^{5} n_{ij}\,, \qquad 1 \le k \le 5, \qquad (13)$$

$$B_k = \sum_{j=1}^{k-1}\sum_{i=k}^{5} n_{ij}\,, \qquad 2 \le k \le 5. \qquad (14)$$

Using these formulas, one obtains the adjusted residuals given in Table 6.

TABLE 6

ADJUSTED RESIDUALS FOR INITIAL AND FINAL RATINGS ON DISABILITY OF 121 STROKE PATIENTS

Initial State	Final State				
	A	B	C	D	E
A	——	——	——	——	——
B	0.18	-0.18	——	——	——
C	0.98	-1.29	0.43	——	——
D	1.50	0.66	-0.41	-2.17	——
E	-2.03	0.44	0.03	2.17	——

Adjusted residuals d_{11} and d_{55} are omitted since $\hat{v}_{11}$ and $\hat{v}_{55}$ are both 0. The equations $d_{21} = -d_{22}$ and $d_{44} = -d_{54}$ should also be noted.

The adjusted residuals differ considerably from the standardized residuals; however, given the number of residuals present, none of the adjusted residuals is excessively large. As shown in Figure 2, the full normal plot of the adjusted residuals tends to confirm this conclusion, although interpretation is complicated by the high correlations between pairs of adjusted residuals.

5. LOGIT ANALYSIS

In the logit model for quantal response, $N_i > 0$ subjects receive a log dosage t_i of some drug or poison, where $1 \leq j \leq r$. Two responses to the stimulus are possible, with n_{ik} subjects that receive log dosage t_i having response k, where $1 \leq j \leq r$ and $1 \leq k \leq 2$. If a logit model is employed, the probability of response 1 given log dosage t_i is

$$p(1 \mid j) = 1/[1 + \exp\{-(\alpha + \beta t_i)\}], \quad 1 \leq j \leq r, \tag{15}$$

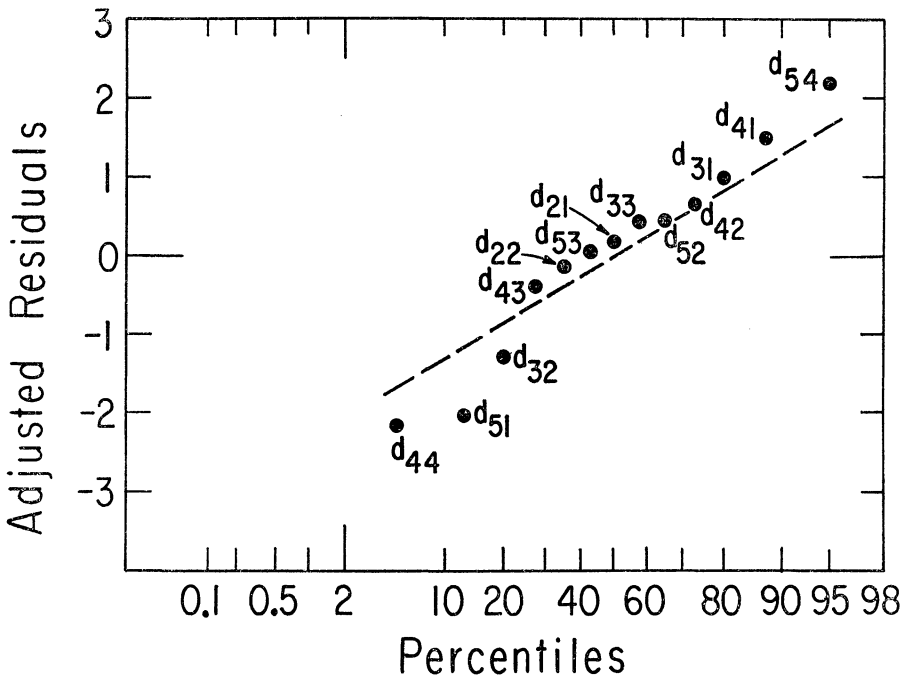

FIGURE 2

FULL NORMAL PLOT OF ADJUSTED RESIDUALS FOR INITIAL AND FINAL DISABILITY RATINGS

for some α and β. Equivalently, one may say that

$$\begin{aligned} \mu_{j1} - \mu_{j2} &= \log (m_{1j}/m_{2j}) \\ &= \log [p(1 \mid j)/p(2 \mid j)] \\ &= \alpha + \beta t_j , \qquad 1 \leq j \leq r. \end{aligned} \tag{16}$$

The set $\mathfrak{m}$ of $\mathfrak{u}$ such that

$$\mu_{j1} - \mu_{j2} = \alpha + \beta t_j , \quad 1 \leq j \leq r,$$

for some α and β is a linear subspace. Multinomial sampling is present with $s = r$, $I = \bar{r} \times \bar{2}$, $I_k = \{k\} \times \bar{2}$ for $1 \leq k \leq r$, and $\mathbf{v}^{(k)} = \{\delta_{jk'} : (j, k') \in \bar{r} \times \bar{2}\} \in \mathfrak{m}$ (see Haberman [1972a, b] and Appendix).

If for some j and j', $t_j \neq t_{j'}$, the likelihood estimate $\hat{\mathfrak{m}}$ may be determined by the Newton-Raphson algorithm described in Finney [1952]. The estimated asymptotic variance of $e_{jk} = (n_{jk} - \hat{m}_{jk})/\sqrt{(\hat{m}_{jk})}$ is shown in Haberman [1972a] to satisfy

$$v_{jk} = (m_{jk'}/N_j)\left[1 - \left(w_j \Big/ \sum_{j'=1}^{r} w_{j'}\right) - \left\{w_j(t_j - \bar{t})^2 \Big/ \sum_{j'=1}^{r} w_{j'}(t_{j'} - \bar{t})^2\right\}\right], \tag{17}$$

where k' is 2 if k is 1, k' is 1 if k is 2,

$$w_j = \hat{m}_{j1}\hat{m}_{j2}/N_j , \tag{18}$$

and

$$\bar{t}_i = (\sum w_j t_j)/\sum w_j \ . \tag{19}$$

Thus the adjusted residual may be written as

$$d_{ik} = \frac{(n_{ik} - \hat{m}_{ik})}{\left[\sqrt{(w_j)\left\{1 - \left(w_j \Big/ \sum_{j'=1}^{r} w_{j'}\right) - \left\{w_j(t_j - \bar{t})^2 \Big/ \sum_{j'=1}^{r} w_{j'}(t_{j'} - \bar{t})^2\right\}\right\}}\right]} \ . \tag{20}$$

If one wishes to work with a simpler expression, one may define $\mathfrak{n}$ as the set of vectors $\mathfrak{y}$ with $\mu_{ik} = c_i$ for some $\mathbf{c} = \{c_i\}$. In this case $\hat{u}_i = w_i$ and

$$e^*_{ik} = (n_{ik} - \hat{m}_{ik})/\sqrt{w_i} \ . \tag{21}$$

This adjustment is attractive since w_i is the estimated variance of n_{ik} under multinomial sampling. Since $n_{i+} = \hat{m}_{i+} = N_i$, $d_{i1} + d_{i2} = 0$ if $1 \leq j \leq r$. Thus it is sufficient to confine attention to the residuals d_{i1} or e^*_{i1} , $1 \leq j \leq r$.

To illustrate use of residual analysis with logit models, consider the data in Table 7 on toxicity of ethylene oxide on the grain beetle *Calandra*

TABLE 7

TOXICITY OF ETHYLENE OXIDE

Log Dosage	Number of Subjects	Number Killed	Expected Number Killed	Adjusted Residuals for Number Killed	
				d_{j1}	e^*_{j1}
0.394	30	23	27.2	-2.93	-2.72
0.391	30	30	27.1	2.00	1.78
0.362	31	29	26.6	1.36	1.23
0.322	30	22	23.2	-0.59	-0.44
0.314	26	23	19.6	1.68	1.66
0.260	27	7	15.8	-3.68	-3.31
0.225	31	12	14.2	-0.88	-0.72
0.199	30	17	11.0	2.60	2.27
0.167	31	10	8.3	0.83	.69
0.033	24	0	1.2	-1.27	-.97

granaria. These data have been analyzed by Finney [1952] by use of probits. Logit analysis results in expected values very similar to those found by Finney. Since the Pearson χ^2 statistic is 33.5 and there are 8 D.F., the model clearly fails to fit the data. The adjusted residuals d_{j1} for log dosages 0.394,

0.260, and 0.199 all exceed 2 in magnitude, and the adjusted residual for log dosage 0.260 is -3.68. Thus the failure of the model is not confined to one or two dosages. Use of the adjusted residuals e_{j1}^{*} , $1 \leq j \leq 10$, does not change this conclusion, for in this example, the two sets of adjusted residuals are rather similar.

To obtain more insight into the failure of the model, two graphs are appropriate. The first (Figure 3) is a full normal plot of the adjusted residuals

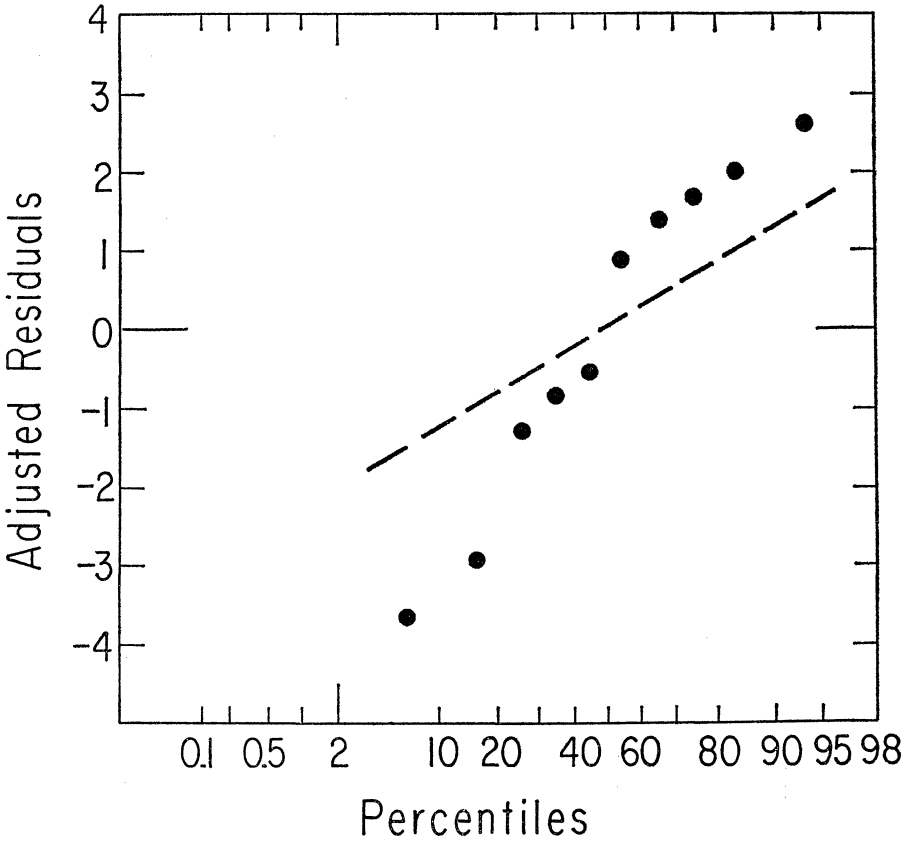

FIGURE 3

FULL NORMAL PLOT OF ADJUSTED RESIDUALS FOR TOXICITY STUDY

d_{j1} , $1 \leq j \leq 10$, constructed in the same manner as the full normal plot for the piston-ring data. The second graph (Figure 4) plots adjusted residuals against log dosages. Figure 3 suggests both that the adjusted residuals are substantially more variable than should be the case if the model were adequate and that this variability is not confined to only a few residuals. Figure 4 indicates the variability of the adjusted residuals is not a result of systematic deviations of the observed probabilities of survival from those predicted by (15), for the adjusted residuals display no readily discerned pattern.

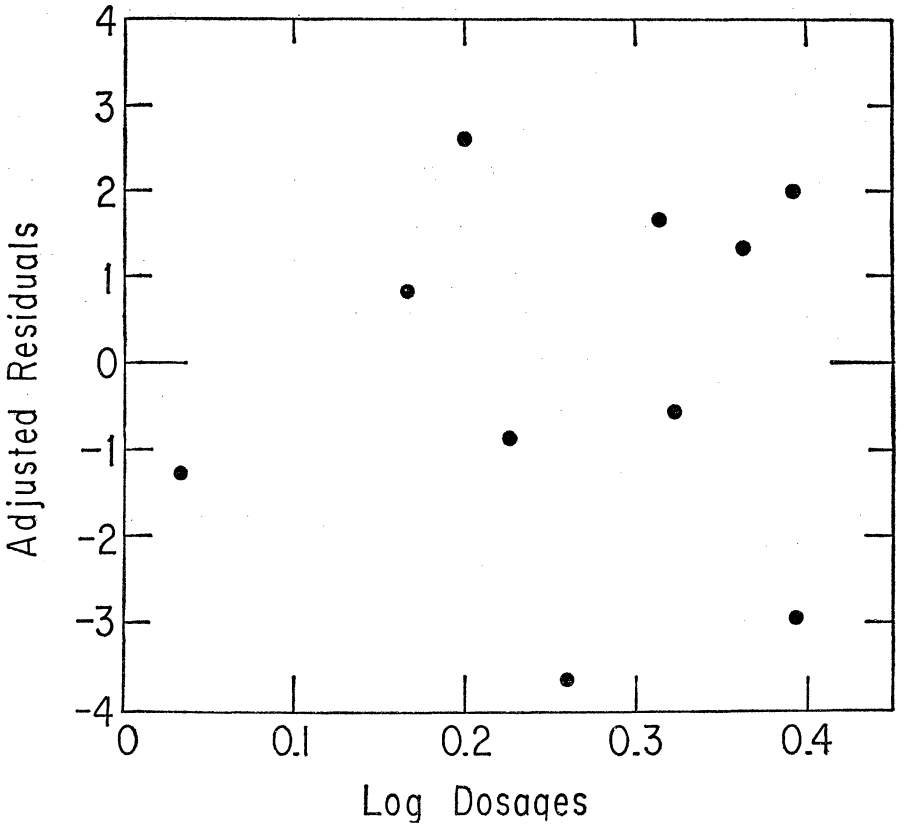

FIGURE 4

PLOT OF ADJUSTED RESIDUALS AGAINST LOG DOSAGE FOR TOXICITY STUDY

Given these graphs, one may reasonably conclude that variations in experimental conditions are present which are not revealed by the data. This conclusion is strongly supported by the contrast in results obtained by use of the nearly identical log dosages 0.394 and 0.391. Further support for the conclusion can be found in Busvine's [1938] account of difficulties encountered in conducting the experiments from which the data in Table 7 were derived.

L'ANALYSE DES RESIDUS DANS DES TABLES CROISEES

RESUME

On propose des techniques pour l'analyse des résidus associée aux modèles log-linéaires dans des tables de fréquences. Les résultats sont appliqués à des tables à deux entrées et aux modèles "logit".

REFERENCES

Anscombe, F. J. and Tukey, J. W. [1963]. The examination and analysis of residuals. *Technometrics 5*, 141–60.

Bishop, Y. M. M. and Fienberg, S. E. [1969]. Incomplete two-dimensional contingency tables. *Biometrics 25*, 119–28.
Busvine, J. R. [1938]. The toxicity of ethylene oxide to *Calandra oryzae*, *C. granaria*, *Tribolium castaneum*, and *cimex lectularis*. *Ann. Appl. Biol. 25*, 605–32.
Cox, D. R. and Snell, E. J. [1968]. A general definition of residuals. *J. Roy. Statist. Soc. B30*, 248–65.
Davies, O. L. [1961]. *Statistical Methods in Research and Production*. Oliver and Boyd, London.
Draper, N. R. and Smith, H. [1966]. *Applied Regression Analysis*. Wiley, New York.
Fienberg, S. E. [1970]. Quasi-independence and maximum likelihood estimation in incomplete contingency tables. *J. Amer. Statist. Ass. 65*, 1610–6.
Fienberg, S. E. [1972]. The analysis of incomplete multi-way contingency tables. *Biometrics 28*, 177–202.
Finney, D. J. [1952]. *Probit Analysis*. Cambridge University Press, Cambridge.
Goodman, L. A. [1968]. The analysis of cross-classified data: Independence, quasi-independence, and interactions in contingency tables with or without missing entries. *J. Amer. Statist. Ass. 63*, 1091–131.
Haberman, S. J. [1970]. The general log-linear model. Ph.D. dissertation, University of Chicago.
Haberman, S. J. [1972a]. The general log-linear model: Sufficient statistics and likelihood equations. Submitted to *Ann. Statist.*
Haberman, S. J. [1972b]. *The Analysis of Frequency Data*. Submitted to IMS Monograph Series.
Haberman, S. J. [1972c]. Conditional Poisson sampling: Asymptotic properties and applications. Submitted to *Ann. Statist.*
Mantel, N. [1970]. Incomplete contingency tables. *Biometrics 26*, 291–304.
Tukey, J. W. [1962]. The future of data analysis. *Ann. Math. Statist. 33*, 1–67.
Wagner, S. S. [1970]. Maximum likelihood estimate for contingency tables with zero diagonal. *J. Amer. Statist. Ass. 65*, 1362–83.

APPENDIX

SAMPLING PROPERTIES OF STANDARDIZED RESIDUALS

In this paper, the vector $\mathbf{n}$ is assumed to arise from one of the following three sampling procedures:

(A) *Poisson sampling.* The coordinates n_i , $i \in I$, are independent Poisson random variables such that $E(n_i) = m_i$.

(B) *Multinomial sampling.* The set I is partitioned into $s \geq 1$ disjoint, nonempty sets I_k , $1 \leq k \leq s$, such that the vectors $\{n_i : i \in I_k\}$ are independent multinomial vectors with means $\{m_i : i \in I_k\}$. It is assumed that each vector $\mathbf{v}^{(k)} \in \mathfrak{m}$, where for $1 \leq k \leq s$,

$$v_i^{(k)} = \begin{cases} 1 & \text{if } i \in I_k \\ 0 & \text{if } i \in I - I_k \end{cases}$$

(C) *Conditional Poisson sampling.* As in (A), the coordinates n_i , $i \in I$, are independent Poisson random variables such that $E(n_i) = m_i$. However, sampling is conditional on the restraints

$$(\mathfrak{n}, \mathbf{v}^{(k)}) = \sum n_i v_i^{(k)}$$
$$= (\mathfrak{m}, \mathbf{v}^{(k)}),$$

where $1 \leq k \leq s$, $v_i^{(k)}$ is an integer for $i \in I$, and $\mathbf{v}^{(k)} \in \mathfrak{m}$.

For example, in section 2, the set of vectors $\mathfrak{y}$ which satisfy (3) is a linear subspace $\mathfrak{m}$ of dimension $r + c - 1$. The conditions for multinomial sampling are satisfied in this section. Since one has $I = \bar{r} \times \bar{c}$, where $\bar{r}$ is the set of integers from 1 to r and $\bar{c}$ is the set of integers from 1 to c, $s = 1$, $I_1 = I$, and $\mathbf{v}^{(1)} = \{1: (i, j) \in \bar{r} \times \bar{c}\} \in \mathfrak{m}$.

The conditions for multinomial sampling continue to apply if each row is an independent multinomial sample of size N_i with probabilities $\{p(j \mid i): 1 \leq j \leq c\}$ such that $p(j \mid i) > 0$ for $1 \leq i \leq r$, $1 \leq j \leq c$. In this case, the homogeneity hypothesis that $p(j \mid i) = p(j \mid i')$ if $i \neq i'$ and $1 \leq j \leq c$ is equivalent to (3), where $\mathfrak{y} = \{\log [N_i\, p(j \mid i)]\}$. One now has multinomial sampling with $s = r$, $I_k = \{k\} \times \bar{c}$, and $\mathbf{v}^{(k)} = \{\delta_{ik} : (i, j) \in \bar{r} \times \bar{c}\} \in \mathfrak{m}$ for $1 \leq k \leq r$, where δ_{ik} is the Kronecker delta. Similar remarks apply if each column is an independent multinomial sample.

Poisson sampling arises if the table is generated from a sample of size N, where N is Poisson and the probability that a given element of the sample belongs to cell (i, j) is $p_{ij} > 0$, where p_{ij} satisfies (2). One has conditional Poisson sampling if row and column totals are both fixed.

The basic motivation for the definition of $\mathbf{e}$ is that under Poisson sampling, $(n_i - m_i)/\sqrt{m_i}$ has an approximate $N(0, 1)$ distribution and $\hat{m}_i$ is an obvious estimate of m_i . Further motivation for this choice of standardized residual comes from the observations that the ML estimate of $\mathbf{m}$ is the same under conditions (A) and (B) and under any of these conditions, Haberman [1972b] has shown that $\mathbf{e}$ has an approximate $N(\mathbf{0}, I - P_{\mathfrak{m}^*})$ distribution, where I is the identity operator,

$$\mathfrak{m}^* = \{\mu^* \in R^I: \mu_i^* = \mu_i \sqrt{m_i},\ i \in I \quad \text{and} \quad \mathfrak{y} \in \mathfrak{m}\}, \tag{A1}$$

and $P_{\mathfrak{m}^*}$ is the orthogonal projection on $\mathfrak{m}^*$.

Since the asymptotic covariance of $\mathbf{e}$ is $I - P_{\mathfrak{m}^*}$, the asymptotic variance of a standardized residual e_i is

$$v_i = (\boldsymbol{\delta}^{(i)},\ [I - P_{\mathfrak{m}^*}]\boldsymbol{\delta}^{(i)}), \tag{A2}$$

where $\boldsymbol{\delta}^{(i)}$ is the coordinate vector defined by

$$\delta_j^{(i)} = \begin{cases} 1 & \text{if} \quad j = i \\ 0 & \text{if} \quad j \in I - \{i\} \end{cases}. \tag{A3}$$

Since $I - P_{\mathfrak{m}^*}$ is an orthogonal projection and $\boldsymbol{\delta}^{(i)}$ has length 1, it follows that $v_i \leq 1$.

If u_i is used to approximate v_i , then one notes that

$$u_i = (\boldsymbol{\delta}^{(i)},\ [I - P_{\mathfrak{n}^*}]\boldsymbol{\delta}^{(i)}), \tag{A4}$$

where

$$\mathfrak{n}^* = \{\mu^* \in R^I: \mu_i^* = \mu_i \sqrt{m_i},\ i \in I \quad \text{and} \quad \mathfrak{y} \in \mathfrak{n}\}. \tag{A5}$$

Since $\mathfrak{n}^* \subset \mathfrak{m}^*$, $u_i \leq v_i$.

Received May 1972, *Revised October* 1972

Key Words: Logit analysis; Contingency tables; Residuals; Incomplete tables.