

APPLICATION OF THE LOGISTIC MODEL
TO ANALYZING CATEGORICAL DATA

by

James E. Grizzle

Department of Biostatistics

School of Public Health

University of North Carolina

This research was supported by Cancer Chemo-
therapy National Service Center of the National
Institutes of Health under Contract No.
SA-43-PH-1932.

Technical Report No. II
Contract No. SA-43-PH-1932
Institute of Statistics Mimeograph Series No. 251
April 1960

ACKNOWLEDGMENTS

I am indebted to the members of my committee, H. L. Lucas, R. J. Monroe, D. B. Duncan, J. E. Legates and E. R. Barrick for their assistance in preparation of this dissertation. I am especially grateful to Dr. Duncan who initiated this work and to Dr. Lucas for supervision and helpful suggestions during its preparation.

To the Department of Biostatistics of the University of North Carolina and National Institutes of Health Cancer Chemotherapy Service Center I am thankful for financial support provided by Contract No. SA-43-PH-1932. I wish to express my appreciation to B. G. Greenberg for his encouragement and for the opportunity of working with him.

For her understanding and constant encouragement I am particularly grateful to my wife, Barbara.

TABLE OF CONTENTS

	Page
LIST OF TABLES	vi
LIST OF FIGURES	vi
THE PROBLEM AND REVIEW OF THE LITERATURE	1
The Problem	1
Review of the Literature	2
Analogues to Correlation	2
Analogues to Regression	3
Summarizing Remarks	8
BASIC THEORY	11
Conditions on Probability Density and Hypotheses	11
Tests of the Fit of the Model	12
Tests on the Parameters in the Model	15
Remarks	16
THEORY FOR THE LOGISTIC MODEL	17
The Model	17
Conventional Methods of Testing Hypotheses	19
A New Approach to Testing Hypotheses for the Logistic Model	20
Choice of Root to Complete the Estimation	23
Remarks	25
COMPUTATIONS FOR THE NEW METHOD AND EXAMPLES	27
Choice of Computational Procedure	27
Computations for the One d.f. Test	27
Computation for the Two d.f. Test and General Case	29
Important Special Case Which does not Require Matrix Inversion	30
Shorter Methods for Computing X^2 and Fitting the Model	34
NON-CENTRALITY PARAMETERS FOR OBTAINING THE ASYMPTOTIC POWER OF TESTS OF HYPOTHESES FOR THE LOGISTIC MODEL	46
The Utility of Non-Centrality Parameters	46
Non-Centrality Parameters for Tests of the Fit of the Model	46
Non-Centrality Parameters for Power of Tests of Parameters in the Model	50

TABLE OF CONTENTS (continued)

	Page
STUDIES ON THE SMALL SAMPLE PROPERTIES OF THE TESTS . . .	55
Choice of Sampling Populations	55
Studies on the Exact Distribution of χ^2_1	56
Methods	56
Results	57
Random Sampling Investigations	61
Methods	61
Results	66
Alternative Approach for Small Samples	70
LIST OF REFERENCES	74

LIST OF TABLES

	Page
1.0. The Incidence of Leukoplakia among Cottonmill Employees Who Use Unburnt Tobacco, Smoke Only, or do Neither.	10
4.0. Linear Restraints for Testings Hypotheses for Data in Table 1.0	38
4.1. Data on Number of Mothers with Previous Infant Losses	42
6.0. P Configurations Used in Sampling	56
6.1. Distribution of X_1^2 for $n = 1$ in Each Cell	58
6.2. Distribution of X_1^2 for $n = 2$ in Each Cell	59
6.3. Observed and Expected Cell Probabilities	62
6.4. Parameters Tested, Sample Sizes and Number of Times Each Experiment Was Repeated	62
6.5. Probability of a Type I Error and Power Observed for an .05 Test	65
6.6. Probability of a Type I Error and Power Observed for an .01 Test	65
6.7. Mean and Variance of X_1^2 when H_0 Is True	66

LIST OF FIGURES

6.0. Observed and Expected Power	69
6.1. 95% Confidence Region for α and β	71

CHAPTER I

THE PROBLEM AND REVIEW OF THE LITERATURE

1.0 The Problem

In some research areas an experiment may consist of taking samples from several binomial distributions, each with parameters P_i and n_i , where P_i is the probability of a success in the i -th distribution and n_i is the sample size. The experimental situation envisioned is analogous to the conventional experimental designs for which data are analyzed by the analysis of variance, except that here the data in each cell are obtained by sampling from a binomial distribution. A more general case would be the same type of experiment with a multinomial distribution in each cell. However, this will not be considered here.

Usually the experimenter will want to test hypotheses about the P_i themselves or about parameters in a mathematical model. For example, in the bioassay of a toxic substance a commonly assumed model is

$$\text{probit (or logit) } P_i = \alpha + \beta x_i, \quad i = 1, \dots, r,$$

where α and β are unknown parameters to be estimated, and x_i is the log dose. The investigator may want to test

$$H_0: \beta = \beta_0$$

or if two lines are being tested for parallelism, the hypothesis can be written

$$H_0: \beta_1 = \beta_2$$

where β_1 and β_2 are the slopes of the respective lines.

In experiments more complex than the example just given, the treatments may be applied at several levels in several combinations as in, for example, a factorial experiment.

Historically, data of this type have been analyzed by testing hypotheses of independence between row and column ways of classification. In many respects these tests are analogous to those tested in correlation models. In recent years the theory underlying this approach has been greatly extended and methods of analysis that are analogous to regression have been developed. It is the objective of this thesis to develop further the regression approach for a particular model. Before proceeding to do this, the literature relevant to both approaches will be reviewed briefly.

1.1 Review of the Literature

1.1.a Analogues to Correlation. Lancaster and Irwin (1949) showed how to make a separation of the total chi square into single degrees of freedom by dividing an $r \times s$ table into $(r-1)(s-1)$ 2×2 tables. These component chi squares do not add to the total as does the orthogonal partitioning of Lancaster (1951).

In Lancaster's (1951) orthogonal partitioning, the total chi square is divided into several components, each component being supposed to test some hypothesis. After examining Lancaster's (1951) test of interaction in detail, Mitra (1955) and Corsten (1957) concluded that it requires main effects to be nonexistent for its validity.

Earlier, Bartlett (1935), following a suggestion from Fisher, had given an unusual test of interaction that uses the root of a polynomial

equation in the test statistic. Mitra and Corsten concluded that this test does not suffer from the defect of Lancaster's. Norton (1945) extended Bartlett's test and outlined a computational method for obtaining the proper roots of the equations involved. Roy and Kastenbaum (1955), (1956) extended Bartlett's and Norton's results and supplied a more logical basis for their derivations.

1.1.b Analogues to Regression. The regression approaches differ among themselves in two important respects, the model and the method of estimation. Usually it is assumed that treatment effects are linearly related to the response when the response is measured on an appropriate scale. Choosing the correct scale is a critical issue in this approach.

The common methods of estimation are maximum likelihood (ML) and best asymptotic normal (BAN), (see Neyman [1949] and Bhapkar [1959]). Both methods have the same asymptotic properties, but in general BAN estimates are easier to obtain. The small-sample properties of estimates and the resulting test statistics obtained by the two methods have not been extensively investigated. Berkson (1955) conducted a sampling investigation comparing the two methods of estimation for the parameters in the model $\log \frac{P_i}{Q_i} = \alpha + \beta x_i$, and concluded that BAN estimates have a smaller mean square error than do ML estimates.

The first approach made to testing linear hypotheses was to make a variance stabilizing transformation. Perhaps the oldest and most widely used transformation is the arcsin or the arccos. The transformation $y = \arccos(1-2p)$, where p is the observed proportion,

was used by Fisher (1922) in analyzing data from genetics experiments. Bartlett, Yates, and Bliss used the equivalent transformation $y = \arcsin\sqrt{p}$ in analyzing a multifactor experiment with data arising by sampling from several binomial distributions. This transformation not only stabilizes the variance for adequate size n , but also, for a large class of data, provides a scale of measurement on which the treatment effects are linearly related to the transformed variable, except at extreme values of P . The problem of testing hypotheses is thus considerably simplified by permitting straightforward application of the analysis of variance and related techniques.

Eisenhart (1947) gives an excellent examination of the arcsin transformation. He concludes,

"...that while in large samples the variance of a transformed proportion is independent of the true proportion for most practical purposes, the variance of the transformed value, $y = 2 \arcsin\sqrt{p}$, is still proportional to $1/n$ where n is the number of observations upon which the observed proportion is based. Consequently, if p_1, p_2, p_3, \dots are proportions based on different numbers of observations n_1, n_2, n_3, \dots , and if the n 's differ widely, then the variance of the corresponding angular values may be so unequal, because of variation in the n 's that the advantages of the transformation are lost almost entirely...."

Eisenhart's investigation shows that for n as large as 10 the variance of y still depends markedly on P . Therefore, the arcsin transformation cannot be used validly in an analysis of variance unless the sample sizes are large and equal for each treatment combination.

In his book, Probit Analysis, Finney (1952) gives a method of analyzing factorial experiments using the probit transformation. The main reason for using the probit transformation for data of the type

being discussed is to provide a scale on which the treatment effects are additive. A long history of use in bioassay indicates that the probit transformation is usually successful in accomplishing this objective. The methods given by Finney present no new theoretical problems, but the computations for obtaining maximum likelihood estimates are lengthy.

Dyke and Patterson (1952) assumed that the treatment effects were linearly related to $\log_e \frac{P}{1-P}$, or logit P as this transformation is called in bioassay. They obtained maximum likelihood estimates of contrasts and their variances an iterative technique that is very similar to a weighted multiple regression analysis. A numerical example of the estimation of main effects, some interactions and their variances in a 2^4 factorial arrangement is given in their paper. Both the logistic and probit models lead to test statistics that are asymptotically distributed as chi square when the null hypothesis is true.

Reiersøl (1954) used the theory of BAN estimation to test hypotheses when the treatment effects are assumed to be linearly related on the P_i or untransformed scale. The hypotheses for which Reiersøl developed tests are parallel to those tested in the analysis of variance, and the test statistics are asymptotically distributed as chi square when the null hypothesis is true. Mitra (1955) extended Reiersøl's work to the multinomial distribution. Bhapkar (1959) extended this work still further to a variety of experimental situations, including the case where the responses have a natural ranking. Corsten (1957) applied a model equivalent to Reiersøl's to a certain problem and obtained maximum likelihood estimates of the cell probabilities

subject to the null hypothesis being true.

Mitra (1955) assumed that the investigator is interested in studying variation on the ratio $\frac{P}{1-P}$ and developed an analysis for 2^n factorial experiments. For convenience he expressed hypotheses as functions of $\log \frac{P}{1-P}$ and then obtained maximum likelihood estimates of the P_i subject to the null hypothesis being true. His tests all have the same form as Bartlett's test for interaction. Previously Dyke and Patterson had observed that Bartlett's test could be derived by assuming that treatment effects are additive on the logit scale, but they did not demonstrate the procedure.

Mitra and Reiersøl introduced a new way of obtaining the test statistic. To test $H_0: \theta_1 - \theta_2 = c_0$, where θ_1 and θ_2 are parameters in the model, they construct functions $f_1(P)$ and $f_2(P)$ such that if H_0 is true, $f_1(P) - f_2(P) = c_0$. Estimates of the P_i are then obtained subject to conditions implied by these functions. In this way hypotheses are tested about parameters without estimating them. In some cases this procedure is more flexible than conventional methods.

In practice, the methods reviewed in this section and the previous one have encountered difficulties not implicit in their theoretical development. Often in the correlation approach, little or no attention has been paid to distinguishing between ways of classification that are random and those that are controlled by the experimenter in constructing the hypotheses to test, and little or no attention has been given to alternatives to the null hypothesis. In the regression approach some researchers have not stipulated the model underlying the proposed test statistics.

As a case in point, Corsten examined Bartlett's test as applied to the data in Bartlett's paper. He concludes that, since one of the marginal totals is obviously fixed by the design of the experiment, the test statistic is incorrect. As a result he proposes a new test. Both of the tests are correct, but for different models. The test Bartlett gives is proper when treatment effects are additive on the logit scale regardless of whether the grand total or one set of marginals is fixed and the alternative one proposed by Corsten is correct when the treatment effects are additive on the P or untransformed scale.

Roy and his co-workers, Kastenbaum (1955), (1956), Mitra (1955), Diamond (1958) and Bhapkar (1959), proceeding along the lines of Cramér (1945), obtained general results for large samples that encompass both the approaches given above. Their approach is characterized by (1) abandoning conditional probabilities in deriving the tests, (2) careful differentiation between variates, i.e., success and failure, and factors, i.e., ways of classification that are controlled by the experimenter, and (3) specification of the model. They point out that the distinction made in (2) may not alter the test statistics in some cases, but it must be made in studying the power of the tests and in setting up meaningful hypotheses to test. To emphasize this essential concept, they divide their results into two general classes, those for single multinomial samples and those for product multinomial samples. A single multinomial sample is one divided into r categories where the probability of falling into any category is P_{ij} with $\sum_{ij} P_{ij} = 1$. A product multinomial sample is one in which r samples are classified into s categories, and the probability of falling into a category is

$$P_{ij} \text{ with } \sum_{j=1}^s P_{ij} = 1 .$$

Aitchison and Silvey (1958), and Silvey (1959), using much the same approach as Roy and his students, obtained many of the same results for more general probability density functions.

1.2 Summarizing Remarks

It is doubtful that a model exists which is correct for all binomial experiments with multiway classifications. However, it should be possible to ascertain models that can be used for broad classes of experiments. A large volume of bioassay data has shown that both logit P and probit P are effective in transforming the observed proportions to a scale on which the treatment effects are additive for biological data. There has been considerable discussion about the relative merits of the probit and logit transformations, but we do not wish to enter into this controversy here. Suffice it to say that no important practical differences in the two models have been demonstrated, see Finney (1952).

In the simpler binomial experiments, it is found almost invariably that the relationship between a stimulus and a response is described by a sigmoid curve. For this reason the validity of tests made by assuming a simple linear model, especially when the stimulus is applied at several levels, seems questionable. However, if we measure the response on a scale which linearizes the relationship between the stimulus and treatment effects, the problem will be simplified considerably. The arcsin, logit and probit transformations have been used to accomplish this.

The arcsin transformation will not be considered because it may not provide a linear relationship over the entire range of P , and, if it is desired to obtain maximum likelihood estimates of parameters, it does not appreciably shorten the computations. This transformation does provide, however, an adequate approximate analysis for many purposes when the sample sizes are equal or, with suitable weighting, when the sample sizes are unequal.

To present the problem more clearly, the data in Table 1.0, for which I am indebted to Drs. Peacock, Brawley and Greenberg (1959), are given as an example of a type frequently encountered in some research areas, for example medicine, where conventional methods do not test the most meaningful hypotheses. We will propose a new approach which combines the Dyke-Patterson (i.e., logistic) model with the methods of obtaining estimates of the P_i subject to restraints suggested by the work of Mitra and Reiersøl. Within this framework some new methods will be developed for testing hypotheses, and the non-centrality parameters for these tests will be obtained. Using these results the asymptotic power of these tests against specified alternatives can be computed if values can be assumed for the true parameter points.

As a by-product, we will obtain shorter, more flexible methods of testing hypotheses for the logistic model and also a shorter method of fitting the model to data. Lastly, the small sample properties of the tests will be examined with respect to power and agreement of the distribution of the test statistic with the chi square distribution.

Table 1.0. The Incidence of Leukoplakia among Cottonmill Employees Who Use Unburnt Tobacco, Smoke Only, or Do Neither

Erwin Mills (Durham Branch) 1958

	<u>White Males</u>			<u>White Females</u>		
	<u>Affected</u>	<u>Unaffected</u>	<u>Total</u>	<u>Affected</u>	<u>Unaffected</u>	<u>Total</u>
Age 25-34						
Unburnt Tobacco	4	16	20	4	5	9
Smokers Only	4	89	93	1	54	55
Neither	1	33	34	1	40	41
Age 35-44						
Unburnt Tobacco	13	28	41	11	34	45
Smokers Only	17	78	95	5	55	60
Neither	5	25	30	7	108	115
Age 45-54						
Unburnt Tobacco	26	47	73	21	35	56
Smokers Only	23	66	89	8	32	40
Neither	4	27	31	3	90	93
Age 55-64						
Unburnt Tobacco	26	47	73	21	35	56
Smokers Only	9	40	49	3	1	4
Neither	6	21	27	5	31	36

CHAPTER II

BASIC THEORY

In this chapter some of the basic results of Roy and his students will be presented. The notation and style will follow closely that of Diamond (1958). All of the theorems Diamond proves require that both the probability density and the mathematical model satisfy certain analytic conditions. We will now give these conditions and follow with a statement of the theorems.

2.0 Conditions on Probability Density and Hypotheses

Suppose we are given rs functions $P_{ij}(\theta_1, \dots, \theta_t)$ where $i=1, \dots, r$ and $j=1, \dots, s$, of $t \leq rs-r$ unknown parameters $\theta_1, \dots, \theta_t$ such that, for all points of a non degenerate open interval $(\bar{\theta})$ in the t -dimensional space of the θ_k for $k=1, \dots, t$ the $P_{ij}(\underline{\theta})$, where $\underline{\theta}' = (\theta_1, \dots, \theta_t)$, satisfy the following conditions:

- (a) $\sum_{j=1}^s P_{ij}(\underline{\theta}) = 1$ for $i = 1, \dots, r$.
- (b) $0 < c^2 \leq P_{ij}(\underline{\theta}) < 1$ for all i and j .
- (c) Every $P_{ij}(\underline{\theta})$ has continuous first and second order derivatives with respect to the θ_k .
- (d) The matrix $\left(\frac{P_{ij}(\underline{\theta})}{\partial \theta_k} \right)$ is of rank t .

Note that the P_{ij} are to be arranged lexicographically (i.e., $P_{11}, P_{12}, \dots, P_{1s}; P_{21}, P_{22}, \dots, P_{2s}; \dots; P_{r1}, P_{r2}, \dots, P_{rs}$) and the

matrix $\left(\frac{\partial^2 P_{ij}(\underline{\theta})}{\partial \theta_k^2} \right)$ has rs rows and t columns.

Suppose we are given $u < t$ functions of the θ_k $f_1(\underline{\theta}), \dots, f_u(\underline{\theta})$, such that for every point of $(\bar{\Gamma})$ the following conditions are satisfied:

(e) Every $f_m(\underline{\theta})$ has continuous first and second order derivatives with respect to θ_k .

(f) The matrix $\left(\frac{\partial f_m(\underline{\theta})}{\partial \theta_k} \right)$ where $m = 1, \dots, u$ and $k = 1, \dots, t$ is of rank u .

And let $\underline{\theta}_0$, where $\underline{\theta}'_0 = (\theta_1^0, \dots, \theta_t^0)$, be a point of $(\bar{\Gamma})$.

Assuming these analytic requirements are fulfilled, Diamond proves the following.

2.1 Tests of the Fit of the Model

Consider the hypothesis

$$H_0: P_{ij} = P_{ij}(\underline{\theta}) \quad \text{for } i = 1, \dots, r \text{ and } j = 1, \dots, s$$

subject to

$$f_m(\underline{\theta}) = 0, \quad m = 1, \dots, u < t,$$

and the alternative

$$H_n: P_{ijn} = P_{ij}(\underline{\theta}) + \frac{d_{ij}}{\sqrt{n}}$$

subject to

$$f_m(\underline{\theta}) = 0,$$

where

$$\sum_{j=1}^s d_{ij} = 0 \text{ but not all } d_{ij} = 0 ,$$

and where

$$P_{ijn} \neq P_{ij}(\underline{\theta}) \quad \text{for any } \underline{\theta} \text{ in } \underline{\Omega} .$$

Then the equations

$$(2.1.1) \quad \sum_{i=1}^r \sum_{j=1}^s \frac{n_{ij} - n_{i.} P_{ij}(\underline{\theta})}{P_{ij}(\underline{\theta})} \left(\frac{\partial P_{ij}(\underline{\theta})}{\partial \theta_k} \right) + \sum_{m=1}^u \lambda_m \frac{\partial f_m(\underline{\theta})}{\partial \theta_k} = 0 ,$$

$$k = 1, \dots, t ,$$

$$f_m(\underline{\theta}) = 0 ,$$

$$m = 1, \dots, u ,$$

for minimizing chi square in the modified sense, which for this case is the same as maximum likelihood (see Cramér [1945]), subject to H_0 , have exactly one system of solutions

$$\hat{\underline{\theta}}' = (\hat{\theta}_1, \dots, \hat{\theta}_t) , \quad \underline{\lambda} = (\lambda_1, \dots, \lambda_u)$$

such that $\hat{\underline{\theta}} \rightarrow \underline{\theta}_0$ in probability as $n \rightarrow \infty$ subject to the ratios $\frac{n_{i.}}{n}$ being held fixed and the statistic

$$(2.1.2) \quad X_*^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{(n_{ij} - n_{i.} P_{ij}(\hat{\underline{\theta}}))^2}{n_{i.} P_{ij}(\hat{\underline{\theta}})}$$

is in the limit, distributed under H_0 as a central chi square variate with $rs-r-t+u$ d.f. and under H_1 as a non-central chi square variate with a non-centrality parameter

$$(2.1.3) \quad \Delta^* = \underline{d}' \left[\underline{I} - B_*(B_*' B_*)^{-1} B_*' \right] \underline{d}$$

where

$$\underline{d}_{rsxl} = \left(\frac{d_{ij}}{\sqrt{P_{ij}^0}} \sqrt{\frac{n_{i\cdot}}{n}} \right), \quad i = 1, \dots, r \text{ and } j = 1, \dots, s,$$

$$(2.1.4) \quad \underline{B}_{rsxt-u}^* = \left[\frac{1}{\sqrt{P_{ij}^0}} \sqrt{\frac{n_{i\cdot}}{n}} \left(\frac{\partial P_{ij}^*(\underline{\theta})}{\partial \theta_{k''}} \right)_0 \right]$$

with $P_{ij}^*(\underline{\theta})$ being $P_{ij}^0(\underline{\theta})$ expressed in terms of the t-u independent θ_k ; $\left(\frac{\partial P_{ij}^*(\underline{\theta})}{\partial \theta_{k''}} \right)_0$ and $P_{ij}^0(\underline{\theta})$ indicate that the derivative and $P_{ij}(\underline{\theta})$ are evaluated at θ_0 .

Alternatively B_* may be expressed as

$$(2.1.5) \quad B_* = B_2 - B_1 F_1^{-1} F_2$$

with

$$(2.1.6) \quad \underline{B}_1 = \left[\frac{1}{\sqrt{P_{ij}^0}} \sqrt{\frac{n_{i\cdot}}{n}} \left(\frac{\partial P_{ij}(\underline{\theta})}{\partial \theta_{k'}} \right)_0 \right]$$

where $k' = 1, \dots, u$

denotes the θ_k made dependent under H_0 ,

$$(2.1.7) \quad \underline{B}_2 = \left[\frac{1}{\sqrt{P_{ij}^0}} \sqrt{\frac{n_{i\cdot}}{n}} \left(\frac{\partial P_{ij}(\underline{\theta})}{\partial \theta_{k''}} \right)_0 \right]$$

where $k'' = u+1, \dots, t$

denotes any θ_k remaining independent under H_0 ,

$$(2.1.8) \quad \underline{F}_1 = \left[\left(\frac{\partial f_m(\underline{\theta})}{\partial \theta_{k'}} \right)_0 \right]$$

and

$$F_{2, \text{uxt-u}} = \left[\left(\frac{\partial f_m(\underline{\theta})}{\partial \theta_{k''}} \right)_0 \right]$$

An important special case occurs where there are no restrictions on the parameters of the model, i.e., $u = 0$. Then $B_* = B_2$, if we set $u = 0$ in B_2 so that $k'' = 1, \dots, t$, and the degrees of freedom are $rs-s-t$.

2.2 Tests on the Parameters in the Model

Suppose we are given $P_{ij} = P_{ij}(\underline{\theta})$, $i = 1, \dots, r$ and $j = 1, \dots, s$. Consider the hypothesis

$$H_0: f_m(\underline{\theta}) = 0, \quad m = 1, \dots, u \leq t,$$

and the alternative

$$H_n: f_m(\underline{\theta}) = \frac{c_m}{\sqrt{n}}$$

where not all the $c_m = 0$. Let X^2 denote the statistic obtained by fitting the model $P_{ij}(\underline{\theta})$ without restraints, and $\hat{\underline{\theta}}$ and $\hat{\underline{\lambda}}$ be the unique consistent solution of the equations given by (2.1.1), and X_*^2 be the corresponding test statistic defined by (2.1.2), then

$$X_1^2 = X_*^2 - X^2$$

is, in the limit, independent of X^2 in probability and is distributed under H_0 as a central chi square variate with u d. f. and under H_n as a non-central chi square variate with u d. f. and a non-centrality parameter,

$$(2.2.1) \quad \Delta_1 = \underline{c}' F_1^{-1} B_1' \left[I - B_* (B_*' B_*)^{-1} B_*' \right] B_1 F_1^{-1} \underline{c},$$

where

$$\underline{c}' = (c_1, \dots, c_u)$$

and B_* , B_1 , B_2 and F_1 are defined by (2.1.5), (2.1.6), (2.1.7) and (2.1.8) .

2.3 Remarks on 2.1 and 2.2

Because of their generality, the results given in sections 2.1 and 2.2 constitute a powerful set of tools for the solution of many problems in the statistical analysis of categorical data. Both the model and the hypotheses are completely unspecified except for certain analytic conditions that must be satisfied. Thus the experimenter or statistician, as the case may be, has wide latitude in selecting a realistic model and in formulating meaningful hypotheses to test.

The results presented in section 2.1 give the theory necessary to test the goodness of fit of a proposed model and also give the non-centrality parameter of the test. To test hypotheses about parameters in a model and to obtain the non-centrality parameters of the tests, section 2.2 should be applied.

CHAPTER III

THEORY FOR THE LOGISTIC MODEL

3.0 The Model

"The product multinomial sample", as Roy, Mitra and Diamond have referred to it, will be assumed to be the probability sampling model. However, as noted in Chapter I, we shall be concerned only with the product binomial sample. Hence, the number of events "R" denoted by a_i and "not R" denoted by b_i , will be the variates and $a_i + b_i = n_i$ is fixed for each cell by the design of the experiment. Diamond's notation can be simplified since we are concerned only with the binomial distribution. In this special case $s = 2$, therefore $P_{i2} = 1 - P_{i1} = Q_i$, and no double subscript is needed. Hereafter all summations will be from $i = 1, \dots, r$ unless otherwise noted.

The relationship assumed to hold between treatments and P_i is

$$(3.0.1) \quad \ell_i = \log_e \frac{P_i}{1-P_i} = \sum_{k=1}^t x_{ik} \theta_k, \quad i = 1, \dots, r,$$

or written more compactly,

$$\begin{matrix} L \\ r \times 1 \end{matrix} = \begin{matrix} X \\ r \times t \end{matrix} \begin{matrix} \theta \\ t \times 1 \end{matrix}$$

where X is a matrix of real numbers of rank t determined by the design of the experiment.

Equation (3.0.1) implies

$$(3.0.2) \quad P_i(\underline{\theta}) = \frac{1}{1 + \exp(-\sum_{k=1}^t x_{ik} \theta_k)}$$

which is the most convenient form for applying the results given in Chapter II.

First we must verify that the logistic model given by (3.0.1) satisfies the analytic requirements set forth in section 2.0.

- (a) $P_i(\underline{\theta}) + Q_i(\underline{\theta}) = 1$ by definition ,
- (b) $0 < c^2 \leq P_i(\underline{\theta}) < 1$ for all real values of x_{ik} and θ_k by inspection of (3.0.2) ,

(c)
$$\frac{\partial P_i(\underline{\theta})}{\partial \theta_k} = P_i(\underline{\theta})Q_i(\underline{\theta})x_{ik} ,$$

$$\frac{\partial^2 P_i(\underline{\theta})}{\partial \theta_k \partial \theta_{k'}} = P_i(\underline{\theta})Q_i(\underline{\theta})[1 - 2P_i(\underline{\theta})]x_{ik}x_{k'} , \quad \text{for } k \text{ and } k' = 1, \dots, t ,$$
 and these are continuous.

- (d) The matrix $\frac{\partial P_i(\underline{\theta})}{\partial \theta_k} = [P_i(\underline{\theta})Q_i(\underline{\theta})x_{ik}]$ can be written

$\begin{matrix} D & X \\ r \times r & r \times t \end{matrix}$, where D is a diagonal matrix of rank r with $P_i(\underline{\theta})Q_i(\underline{\theta})$ on the diagonal and zeros elsewhere, and X is non-singular and of rank t by hypothesis. Since $t \leq r$, DX is of rank t .

Suppose we wish to test linear hypotheses,

- (e) $H_0: \begin{matrix} F & \underline{\theta} \\ u \times t & t \times 1 \end{matrix} = \underline{0}$ where F is of rank $u < t$,

about parameters of the logit model using the results in sections 2.1 or 2.2. Taking matrix derivatives,

$$\frac{\partial(F\underline{\theta})}{\partial \underline{\theta}} = F ,$$

and

$$\frac{\partial^2(F\underline{\theta})}{\partial^2 \underline{\theta}} = 0 ;$$

these are constants and hence continuous.

(f) The rank of $\frac{\partial F\theta}{\partial \theta} = F$ is $u < t$ by assumption.

Therefore, we conclude that sections 2.1 and 2.2 can be applied to the estimation of parameters and testing linear hypotheses about parameters of the logit model.

3.1 Conventional Methods of Testing Hypotheses

Cramér's modified chi square minimum method of estimation is identical to maximum likelihood (ML) estimation for the case being considered. The following method of obtaining ML estimates of the parameters of the logistic model is similar to the method given by Dyke and Patterson. This method, though relatively laborious can be used to obtain the proper test statistic for the tests given in sections 2.1 and 2.2. Some alternative methods, which in many cases will be shorter, for testing the same hypotheses will be presented later.

To obtain estimates of the parameters in the model given by (3.0.1) we observe that the likelihood for a product binomial sample is written

$$(3.1.1) \quad \phi = \prod_{i=1}^r \frac{n_i!}{a_i! b_i!} P_i^{a_i}(\theta) Q_i^{b_i}(\theta)$$

Then

$$\log \phi = \sum \left[\log n_i! - \log a_i! - \log b_i! + a_i \log P_i(\theta) + b_i \log Q_i(\theta) \right]$$

and the maximizing equations are

$$(3.1.2) \quad \frac{\partial \log \phi}{\partial \theta_k} = \sum_i \left[-a_i - n_i P_i(\theta) \right] x_{ik} = 0, \quad k = 1, \dots, t.$$

These equations cannot be solved explicitly for $\hat{\underline{\theta}}$, the estimate of $\underline{\theta}$. They are usually solved by the Newton-Raphson method. That is, (3.1.2) is expanded in a Taylor series about a trial value $\hat{\underline{\theta}}$, non-linear terms are neglected, and the resulting equations are solved by iteration. To test the hypothesis

$$H_0: P_i = P_i(\underline{\theta}), \quad i = 1, \dots, r,$$

we use the fact that

$$(3.1.3) \quad X^2 = \sum_{i=1}^r \frac{(a_i - n_i P_i(\hat{\underline{\theta}}))^2}{n_i P_i(\hat{\underline{\theta}}) Q_i(\hat{\underline{\theta}})},$$

where $P_i(\hat{\underline{\theta}})$ is $P_i(\underline{\theta})$ evaluated at $\hat{\underline{\theta}}$, is asymptotically distributed as a central chi square variate with $r-t$ degrees of freedom if H_0 is true. Given $P_i = P_i(\underline{\theta})$, this method can be used to test hypotheses about any sub-set of the θ 's. The details are much the same as in normal regression theory except that the regression equation must be fitted by iteration. Berkson (1957) has published a table which is helpful in the computations.

3.2 A New Approach to Testing Hypotheses for the Logistic Model

The methods derived here will apply to designs for which linear functions of $\log \frac{P_i}{Q_i} = \lambda_i$ can be easily constructed such that

$$(3.2.1) \quad \sum_{i=1}^r f_i \lambda_i = \theta_k$$

where the f_i are real numbers dictated by the model and the design of the experiment. To test

$$H_0: \theta_k = h_k$$

or

$$H_0: P_i = P_i(\theta) ,$$

subject to

$$\sum f_i^* \lambda_i = h_k ,$$

we obtain maximum likelihood estimates of P_i ($P_i(\theta)$ will henceforth be written P_i for convenience.), subject to the restraint

$$(3.2.2) \quad \sum f_i^* \lambda_i = h_k .$$

The log likelihood with one restraint is

$$(3.2.3) \quad \psi = \log \phi - \lambda \left[\sum f_i^* \lambda_i - h_k \right] ,$$

where λ is a Lagrangian multiplier. The maximizing equations are

$$\frac{\partial \psi}{\partial P_i} = a_i - n_i P_i - f_i^* \lambda = 0$$

giving

$$(3.2.4) \quad \hat{P}_i = \frac{a_i - f_i^* \lambda}{n_i} , \quad i = 1, \dots, r ,$$

which still have the Lagrangian multiplier, λ , that must be eliminated. The restraint (3.2.2) implies

$$\sum f_i^* \log \frac{\hat{P}_i}{Q_i} = h_k ,$$

or alternatively,

$$\Pi \left(\frac{\hat{P}_i}{Q_i} \right)^{f_{ik}^*} = e^{h_k} = g_k .$$

Substituting for \hat{P}_i , we have

$$(3.2.5) \quad \sum f_i^* \left[\log (a_i - f_i^* \lambda) - \log (b_i + f_i^* \lambda) \right] = h_k ,$$

or

$$\Pi \left(\frac{a_i - f_i^* \lambda}{b_i + f_i^* \lambda} \right)^{f_i^*} = \varepsilon_k .$$

Either of these equations may be solved for λ .

The extension to the general case is easily seen from the 2 d. f. test. To test

$$H_0: \theta_1 = h_1$$

$$\theta_2 = h_2$$

or

$$H_0: P_i = P_i(\theta) ,$$

subject to

$$\Sigma f_{i1}^* \ell_i = h_1 \quad \text{and} \quad \Sigma f_{i2}^* \ell_i = h_2 ,$$

the two restraints are

$$(3.2.6) \quad \Sigma f_{i1}^* \ell_i = h_1$$

$$\Sigma f_{i2}^* \ell_i = h_2 .$$

The log likelihood with two restraints is

$$\psi = \log \phi - \lambda_1 (\Sigma f_{i1}^* \ell_i - h_1) - \lambda_2 (\Sigma f_{i2}^* \ell_i - h_2) ,$$

and

$$\frac{\partial \psi}{\partial P_i} = a_i - n_i P_i - f_{i1}^* \lambda_1 - f_{i2}^* \lambda_2 = 0$$

which gives

$$(3.2.7) \quad \hat{P}_i = \frac{a_i - f_{i1}^* \lambda_1 - f_{i2}^* \lambda_2}{n_i} .$$

To eliminate λ_1 and λ_2 from (3.2.7) either

$$(3.2.8) \quad \Sigma f_{i1}^* \sqrt{\log(a_i - f_{i1}^* \lambda_1 - f_{i2}^* \lambda_2) - \log(b_i + f_{i1}^* \lambda_1 + f_{i2}^* \lambda_2)} = h_1,$$

$$\Sigma f_{i2}^* \sqrt{\log(a_i - f_{i1}^* \lambda_1 - f_{i2}^* \lambda_2) - \log(b_i + f_{i1}^* \lambda_1 + f_{i2}^* \lambda_2)} = h_2,$$

or

$$(3.2.9) \quad \Pi \left(\frac{a_i - f_{i1}^* \lambda_1 - f_{i2}^* \lambda_2}{b_i + f_{i1}^* \lambda_1 + f_{i2}^* \lambda_2} \right)^{f_{i1}^*} = \xi_1,$$

$$\Pi \left(\frac{a_i - f_{i1}^* \lambda_1 - f_{i2}^* \lambda_2}{b_i + f_{i1}^* \lambda_1 + f_{i2}^* \lambda_2} \right)^{f_{i2}^*} = \xi_2$$

must be solved simultaneously.

3.3 Choice of Root to Complete the Estimation

The question of which root to use now arises. Mitra (1955) proved that for a 2x2 factorial experiment, which generates a third degree polynomial, the smallest real root is the proper one to use. Lancaster (1951) said that Bartlett's equation had only one real root in the asymptotic case. However, Mitra constructs a counter example to disprove this. For the data used in Bartlett's paper (1935), the equation has only one real root. For larger experiments which give rise to higher degree polynomials, Mitra recommended taking the root that minimized X^2 as being a safe procedure.

In general the root which maximizes the likelihood in the numerical sense should be used. This may be difficult to determine for large experiments which generate high degree polynomials. It seems reasonable to restrict ourselves to solutions of (3.2.5) which yield positive estimates of P_i and Q_i when substituted into (3.2.4). The following considerations show that (3.2.5) has at most one root satisfying this

criterion. Using the log form of (3.2.5) we have

$$U = \sum f_i^* \sqrt{\log(a_i - f_i^* \lambda) - \log(b_i + f_i^* \lambda)} - h = 0 ,$$

and

$$\frac{dU}{d\lambda} = - \sum \frac{f_i^{*2}}{n_i \hat{P}_i \hat{Q}_i} .$$

If \hat{P}_i and \hat{Q}_i are to be positive for $i = 1, \dots, r$, then, for any acceptable solution, $\frac{dU}{d\lambda}$ must be negative, i.e., U is monotone decreasing. Therefore, (3.2.5) can have at most one root satisfying the criterion we have set up for acceptability. Let us call this root λ^* , then

$$(3.3.1) \quad X_*^2 = \lambda^{*2} S ,$$

where

$$S = \sum f_i^{*2} \left(\frac{1}{a_i - f_i^* \lambda^*} + \frac{1}{b_i + f_i^* \lambda^*} \right) ,$$

and $X_1^2 = X_*^2 - X^2$ are distributed as central chi squares with 1 d. f. if H_0 is true. X_1^2 and X_*^2 should be used depending on whether the hypotheses discussed in section 2.1 or 2.2 are being tested.

By an argument similar to the one given for the 1 d. f. case, it can be shown there is not more than one solution of (3.2.8) such that \hat{P}_i and \hat{Q}_i are positive.

$$U_1 = \sum_{i1} f_{i1}^* \sqrt{\log(a_{i1} - f_{i1}^* \lambda_1 - f_{i2}^* \lambda_2) - \log(b_{i1} + f_{i1}^* \lambda_1 + f_{i2}^* \lambda_2)} - h_1 = 0 .$$

$$U_2 = \sum_{i2} f_{i2}^* \sqrt{\log(a_{i2} - f_{i1}^* \lambda_1 - f_{i2}^* \lambda_2) - \log(b_{i2} + f_{i1}^* \lambda_1 + f_{i2}^* \lambda_2)} - h_2 = 0 .$$

The first partial derivatives always have the same sign in the

region of interest. Therefore, they can both equal zero simultaneously at most, once.

Then

$$(3.3.2) \quad X_*^2 = \Sigma (f_{i1}^* \lambda_1^* + f_{i2}^* \lambda_2^*)^2 \left(\frac{1}{a_i - f_{i1}^* \lambda_1^* - f_{i2}^* \lambda_2^*} + \frac{1}{b_i + f_{i1}^* \lambda_1^* + f_{i2}^* \lambda_2^*} \right)$$

and $X_1^2 = X_*^2 - X^2$ have central chi square distributions with 2 d. f. if H_0 is true. X_*^2 or X_1^2 is used as the test statistic depending on which section of Chapter II is being applied.

3.4 Remarks

The preceding results can be applied to the solution of other problems encountered in the analysis of data. The problem of combining 2x2 tables which is often encountered in the literature can be approached as the analysis of a single experiment rather than a series of small experiments. The model assumed for the analysis should include parameters which measure the differences in experimental material from one table to another. If the logistic model can be assumed, the methods presented here are appropriate. In other cases the results of Reiersøl (1954) or Bhapker's (1959) extension of them may be appropriate.

Occasionally data are encountered from a single multinomial sample in which it is desired to study how the occurrence of events A and B, where B is in some sense the complement of A, are influenced by the other classifications. If we let $\lambda = \log_e \frac{P_A}{P_B}$, where P_A and P_B are the probabilities of the occurrence of events A and B respectively, then the techniques given in this chapter and Chapter IV may be used, even though they were not developed with this kind of a sample in

mind. However, the non-centrality parameters given in Chapter V are no longer appropriate.

The methods of testing hypotheses and the computational techniques in the next chapter may be useful in bioassay when two or more compounds and mixtures of these compounds are administered.

The new methods presented in this chapter have the added advantage of ease of interpretation as compared to the conventional methods discussed in section 3.1. The effects of assuming the null hypothesis is true are conveniently displayed on the observed scale even though a different scale is used in formulating the hypothesis.

CHAPTER IV

COMPUTATIONS FOR THE NEW METHOD AND EXAMPLES

4.0 Choice of Computational Procedure

In the examination of (3.2.5) and (3.2.8) given in section 3.3, it was shown that not more than one solution of these equations satisfying the criterion $0 < \hat{P}_i < 1$ exists. Hence, we do not need to be concerned with the method of computation chosen converging to the wrong roots, as long as the solution gives $0 < \hat{P}_i < 1$. Of course, in practice, speed of convergence is important.

Scarborough (1958) examines the Newton-Raphson method of computing the roots of equations. In the case at hand, it is not possible to examine the method, generally, for speed of convergence because it varies according to the observations. However, experience has shown that the Newton-Raphson method works in practice and is adaptable to either desk or high speed computers. In fact, in the well known method of fitting the probit or the logistic model, the Newton-Raphson method is used to compute the solution of the ML equations, as in example 4.3.

4.1 Computations for the One d. f. Test

To proceed to obtain the solution of (3.2.5) by the Newton-Raphson method we expand it in a Taylor series about an initial value of λ_{i0}^* , λ_{i0} say, neglect non-linear terms, and solve for $(\lambda_i - \lambda_{i0})$. λ_{i0} may be chosen anywhere in the interval which will make the P's positive, or it may be taken to be zero if $a_i \neq 0$ and $b_i \neq 0$. In practice, the closer the initial guess is to λ^* , the faster the convergence.

Before expanding (3.2.5), let us write it as

$$\sum_i f_i^* \log(a_i - f_i^* \lambda) = \sum_i f_i^* \log(b_i + f_i^* \lambda) + h$$

Expanding this form about the initial value we obtain

$$(4.1.1) \quad \sum_i f_i^* \log(a_i - \lambda_0 f_i^*) - \delta_\lambda \sum_i f_i^{*2} \left[\frac{1}{a_i - \lambda_0 f_i^*} \right] =$$

$$h + \sum_i f_i^* \log(b_i + \lambda_0 f_i^*) + \delta_\lambda \sum_i f_i^{*2} \left[\frac{1}{b_i + \lambda_0 f_i^*} \right],$$

and

$$\delta_\lambda = \frac{\sum_i f_i^* \log(a_i - \lambda_0 f_i^*) - h - \sum_i f_i^* \log(b_i + \lambda_0 f_i^*)}{S}$$

where

$$\delta_\lambda = (\lambda - \lambda_0), \quad S = \sum \frac{f_i^{*2}}{n_i \hat{P}_i \hat{Q}_i}$$

and \hat{P} is \hat{P} evaluated at λ_0 .

Applying the same technique to the alternative form of (3.2.5), we obtain

$$\delta_\lambda = \frac{R_1 - R_2 g}{R_1 S_1 + R_2 S_2 g}$$

where

$$R_1 = \Pi (a_i - \lambda_0 f_i^*)^{f_i^*},$$

$$R_2 = \Pi (b_i + \lambda_0 f_i^*)^{f_i^*},$$

$$S_1 = \sum \frac{f_i^{*2}}{(a_i - \lambda_0 f_i^*)}$$

and

$$S_2 = \sum \frac{f_i^{*2}}{(b_i + \lambda_0 f_i^*)}$$

$\lambda_0 + \delta_\lambda$ is the first approximation to λ^* . $\lambda_0 + \delta_\lambda$ is then

substituted into (3.2.5) and the equation is solved again to obtain a second correction, δ_λ . This process is continued until δ_λ is as small as desired.

4.2 Computation for the Two d. f. Test and General Case

For the 2 d. f. test the algebra is more straightforward if we do not take the terms of (3.2.8) involving b_i to the right hand side. A solution can be obtained from (3.2.5) if desired. Expanding (3.2.8) about the trial points λ_{10} and λ_{20} we obtain

$$(4.2.1) \quad \begin{aligned} \underline{F}_1^* \underline{L} - \delta_\lambda \underline{F}_1^* \underline{D} \underline{F}_1^{*'} - \delta_\lambda \underline{F}_2^* \underline{D} \underline{F}_2^{*'} &= \underline{h}_1 \\ \underline{F}_2^* \underline{L} - \delta_\lambda \underline{F}_1^* \underline{D} \underline{F}_1^{*'} - \delta_\lambda \underline{F}_2^* \underline{D} \underline{F}_2^{*'} &= \underline{h}_2 \end{aligned}$$

where \underline{L} is a vector of elements $(l_i = \log \frac{\hat{P}}{\hat{A} \hat{Q}})$,

$$\underline{F}_1^* = (f_{11}^*, f_{21}^*, \dots, f_{r1}^*) \quad , \quad \underline{F}_2^* = (f_{12}^*, f_{22}^*, \dots, f_{r2}^*) \quad ,$$

\underline{D} is a diagonal matrix of elements

$$1/(n_i \hat{P}_i \hat{Q}_i) \quad ,$$

\hat{P}_i is \hat{P}_i evaluated at λ_0 , and $\delta_{\lambda i} = (\lambda_i - \lambda_{i0})$. A more compact way of writing (4.2.1) is

$$(4.2.2) \quad \underline{F}^* \underline{L} - \underline{\delta}_\lambda \underline{F}^* \underline{D} \underline{F}^{*'} = \underline{h} \quad ,$$

where

$$\underline{F}^{*'} = \left[\underline{F}_1^{*'} : \underline{F}_2^{*'} \right] \quad ,$$

giving

$$\underline{\delta}_\lambda = (\underline{F}^* \underline{D} \underline{F}^{*'})^{-1} (\underline{F}^* \underline{L} - \underline{h}) \quad .$$

The solution to the general case is identical with (4.2.2) except \hat{P} is now defined as

$$(4.2.3) \quad \hat{P}_i = (a_i - \sum_{j=1}^u f_{ij}^* \lambda_{j0}) / n_i$$

and the elements of $\underline{\delta}_\lambda$, \dot{D} and \dot{L} are redefined accordingly. To further bring out the analogy of the form of the solution with multiple regression, it might be noted that for most practical problems

$$\underline{h} = \underline{0} \quad ;$$

therefore

$$(4.2.4) \quad \underline{\delta}_\lambda = (F^* \dot{D} F^*)^{-1} F^* \dot{L}$$

As in the other cases we have considered, the first approximation to λ^* is $\underline{\lambda} = \underline{\lambda}_0 + \underline{\delta}_\lambda$. This $\underline{\lambda}$ now becomes $\underline{\lambda}_0$ and a second correction $\underline{\delta}_\lambda$ is obtained. The process is continued until the elements of $\underline{\delta}_\lambda$ are sufficiently small.

4.3. Important Special Case which does not Require Matrix Inversion

There is one type of test having more than 1 d. f. that occurs frequently in practice and deserves special mention because matrices do not have to be inverted in computing the roots. It is analogous to the test of homogeneity of a set of treatment effects in the analysis of variance. The notation will be clearer if we let i become a double subscript ij where $i = 1, \dots, u$ and $j = 1, \dots, v$. For the method to be useful there must be linear functions of the f_{ij} such that

$$(4.3.1) \quad \begin{aligned} \sum_i f_{i1}^* f_{i1} &= \theta_1 \quad , \\ &\cdot \\ &\cdot \\ &\cdot \end{aligned}$$

(4.3.1) cont.

$$\begin{aligned} \sum_i f_{ij}^* \ell_{ij} &= \theta_j \quad , \\ &\vdots \\ &\vdots \\ &\vdots \\ \sum_i f_{iv}^* \ell_{iv} &= \theta_v \quad . \end{aligned}$$

To test the hypotheses given that $P_{ij} = P_{ij}(\theta)$

$$H_0: \theta_1 = \theta_2 = \dots = \theta_v$$

or

$$H_0: P_{ij} = P_{ij}(\theta) \quad \text{for } i = 1, \dots, u \quad , \quad j = 1, \dots, v \quad ,$$

subject to $\theta_1 = \theta_2 = \dots = \theta_v$, we must obtain maximum likelihood estimates of the P_{ij} subject to

$$(4.3.2) \quad \sum_i f_{i1}^* \ell_{i1} = \sum_i f_{i2}^* \ell_{i2} = \dots = \sum_i f_{iv}^* \ell_{iv} \quad ,$$

or

$$\sum_i f_{i1}^* \ell_{i1} - \sum_i f_{iv}^* \ell_{iv} = 0 \quad ,$$

$$\vdots$$

$$\sum_i f_{ij}^* \ell_{ij} - \sum_i f_{iv}^* \ell_{iv} = 0 \quad ,$$

$$\vdots$$

$$\sum_i f_{i,v-1}^* \ell_{i,v-1} - \sum_i f_{iv}^* \ell_{iv} = 0 \quad .$$

The log likelihood with $v-1$ restraints is

$$\psi = \log \phi - \lambda_1 \left(\sum_i f_{i1}^* \ell_{i1} - \sum_i f_{iv}^* \ell_{iv} \right) - \dots - \lambda_{v-1} \left(\sum_i f_{i,v-1}^* \ell_{i,v-1} - \sum_i f_{iv}^* \ell_{iv} \right)$$

and

$$\frac{\partial \psi}{\partial P_{ij}} = a_{ij} - n_{ij} P_{ij} - \lambda_j f_{ij}^* = 0 \quad \text{for } i=1, \dots, u \quad \text{and } j=1, \dots, v-1 \quad ,$$

and

$$\frac{\partial \psi}{\partial P_{iv}} = a_{iv} - n_{iv} P_{iv} + f_{iv}^* \sum_{j=1}^{v-1} \lambda_j .$$

Therefore

$$\hat{P}_{ij} = \frac{a_{ij} - f_{ij}^* \lambda_j}{n_{ij}} \quad i = 1, \dots, v-1 ,$$

and

$$\hat{P}_{iv} = \frac{a_{iv} + f_{iv}^* \sum_{j=1}^{v-1} \lambda_j}{n_{iv}} = \frac{a_{iv} - f_{iv}^* \lambda_v}{n_{iv}}$$

where

$$\lambda_v = -\sum_{j=1}^{v-1} \lambda_j .$$

Substituting the estimates of P_{ij} into (4.3.2) we have

$$(4.3.3) \quad \sum_i f_{i1}^* [\log(a_{i1} - \lambda_1 f_{i1}^*) - \log(b_{i1} + \lambda_1 f_{i1}^*)] = \dots =$$

$$\sum_i f_{iv}^* [\log(a_{iv} - \lambda_v f_{iv}^*) - \log(b_{iv} + \lambda_v f_{iv}^*)] .$$

Note that now $\sum_{j=1}^v \lambda_j = 0$.

Expanding about the points $\lambda_{10}, \lambda_{20}, \dots, \lambda_{v0}$, where $\sum_{j=1}^v \lambda_{j0} = 0$, we have

$$(4.3.4) \quad T_1 - \delta_{\lambda_1} S_1 = \dots = T_v - \delta_{\lambda_v} S_v$$

where

$$\delta_{\lambda_j} = \lambda_j - \lambda_{j0} ,$$

$$T_j = \sum_i f_{ij}^* [\log(a_{ij} - \lambda_{j0} f_{ij}^*) - \log(b_{ij} + \lambda_{j0} f_{ij}^*)] .$$

and

$$S_j = \sum_i \frac{f_{ij}^{*2}}{n_{ij} \hat{p}_{ij} \hat{q}_{ij}}$$

Let us choose the j-th and k-th members of (4.3.4) and solve for δ_{λ_k} .

Then

$$\delta_{\lambda_k} = \frac{T_k}{S_k} - \frac{T_j}{S_k} + \delta_{\lambda_j} \frac{S_j}{S_k} .$$

Since $\sum_{j=1}^v \lambda_j = 0$ and any solution must satisfy $\sum_{j=1}^v \lambda_j^* = 0$ then

$\sum_{k=1}^v \delta_{\lambda_k} = 0$. Therefore,

$$0 = \sum_{k=1}^v \delta_{\lambda_k} = \sum_{k=1}^v \frac{T_k}{S_k} - T_j \sum_{k=1}^v \frac{1}{S_k} + \delta_{\lambda_j} S_j \sum_{k=1}^v \frac{1}{S_k}$$

or

$$0 = H - \frac{T_j}{S} + \delta_{\lambda_j} \frac{S_j}{S}$$

and

$$(4.3.5) \quad \delta_{\lambda_j} = (T_j - HS) \frac{1}{S_j} ,$$

where

$$H = \sum_{k=1}^v \frac{T_k}{S_k} \quad \text{and} \quad \frac{1}{S} = \sum_{k=1}^v \frac{1}{S_k} .$$

Note that the subscript k has the same range as j and, in fact, may be replaced by j. As in previous cases $\lambda_j = \lambda_{j0} + \delta_{\lambda_j}$ is only the first approximation to λ_j^* .

If desired, the antilog can be taken of each term of (4.3.2) and a solution derived for this form. The solution is

$$(4.3.6) \quad \delta_{\lambda_j} = (R_j - \frac{1}{Sh}) \frac{1}{S_j R_j}$$

where

$$R_j = \prod_{i=1}^u \left(\frac{a_{ij} - \lambda_{j0} f_{ij}^*}{b_{ij} + \lambda_{j0} f_{ij}^*} \right)^{f_{ij}^*},$$

S_j and $\frac{1}{S}$ are as in (4.3.4) and (4.3.5), and

$$h = \sum_{k=1}^v \frac{1}{R_k S_k}.$$

For tests of hypotheses using sections 2.2 or 2.3, $X_*^2 = \sum_{j=1}^v \lambda_j^2 S_j$ and $X_1^2 = X_*^2 - X^2$ are distributed asymptotically as central chi square variates with $v-1$ d.f. if H_0 is true.

Often tests of interaction can be put into this form as will be illustrated by examples 4.1 and 4.3. Norton's extension of Bartlett's test can be derived as a special case of 4.3.2 and his computing formulas can be obtained from (4.3.6).

4.4 Shorter Methods for Computing X^2 and Fitting the Model

The method of setting up hypotheses and the computing formulas previously given can also be utilized in fitting the logistic model to data and in computing the X^2 of section 2.2. Obtaining X^2 and fitting the model using these methods are not as obvious as testing hypotheses about parameters in the model, but they are easily obtained because of analogies with the analysis of variance.

Recall from section 2.2 that X^2 is computed by fitting the model with no restraints on the parameters. Then X^2 is given by

$$X^2 = \sum \frac{(a_i - n_i P_i(\hat{\theta}))^2}{n_i P_i(\hat{\theta}) Q_i(\hat{\theta})}.$$

In the analysis of variance, the error often can be thought of as being composed of block \times treatment interactions or in some instances as high order treatment interactions. A more convenient way of computing X^2 is to set up linear functions of f_i that would compose the error in an analysis of variance, set these functions equal to zero and obtain ML estimates of the P_i subject to these restraints. Then

$$X^2 = \sum \frac{(a_i - n_i \hat{P}_i)^2}{n_i \hat{P}_i \hat{Q}_i} .$$

The advantage of this method over conventional methods lies in the fact that $t-u$ is usually less than u . Thus the order of the matrices that must be inverted has been reduced. Furthermore, many times the P 's can be estimated without inverting matrices, under the hypothesis of no interaction, by applying section 4.3. It should be stressed that these \hat{P} 's have been obtained under the hypothesis of no interaction among effects included in the model. Although never explicitly stated, this is really what we are doing in the conventional fitting methods. Hence, the estimates of the P_i used in computing X^2 can be used to obtain the provisional logits and weights needed to fit the model and no further iteration will be required. This technique is illustrated and convenient computing formulas are given in example 4.3.

In addition to providing greater flexibility in testing hypotheses for the logistic model and giving computing formulas which can effect some saving in computation, several bits of apparently unconnected statistical theory have been shown to be related closely; namely, Bartlett's test of interaction and Norton's extension of it, Mitra's

approach to testing hypotheses and the work of many people on the logistic model.

Example 4.1 Illustration of testing hypotheses in a complex experiment without inverting matrices

An example of a type of data that the preceding techniques might be applied to is given in Table 1.0. We will apply the techniques developed in this chapter to test some of the hypotheses that would be tested by an analysis of variance.

The model assumed is

$$\log \frac{P_{ijk}}{Q_{ijk}} = \ell_{ijk} = \mu + \alpha_i + \tau_j + \theta_k + (\alpha\tau)_{ij} + (\alpha\theta)_{ik} + (\tau\theta)_{jk} + (\alpha\tau\theta)_{ijk}$$

where $\alpha_i = \text{effect of } i\text{-th age,}$ $\sum_{i=1}^4 \alpha_i = 0$,

$\tau_j = \text{effect of } j\text{-th type of tobacco use,}$ $\sum_{j=1}^3 \tau_j = 0$,

$\theta_k = \text{effect of sex,}$ $\sum_{k=1}^2 \theta_j = 0$.

For this type of data we have no knowledge of what parameters are necessary to include in the model to fit the data. Hence the complete model is assumed and, section 2.1 is applied to get estimates of the P_{ijk} subject to certain restraints, the restraints being chosen so that the desired hypothesis is tested.

As in the analysis of normally distributed data, the tests of main effects are not very helpful in interpreting the data when interactions are significant. For this reason it is desirable to start with the

tests of high order interactions and be guided by the results of these tests in setting up the most meaningful hypotheses to test about main effects. In the analysis that follows the highest order interaction is significant. When this is the case it is difficult to construct meaningful tests of lower order interactions and main effects.

It might be argued that the highest order interaction is really error, as it is often considered to be in normal analysis of an unreplicated factorial experiment. Then one proceeds using F tests. However, the X^2 's are not independent. It may be that whenever the sample size is large enough for the test statistic to follow chi square when the null hypothesis is true, independence has been attained. At present this is not known.

For the sake of example, we will proceed to set up the linear functions to test the hypotheses and show how to compute some of the test statistics even though they will not be particularly helpful in interpreting these particular data. We will use the notation,

$$\sum_k \ell_{ijk} = \ell_{ij.} \quad , \quad \sum_j \ell_{ijk} = \ell_{i.k} \quad , \quad \sum_{ij} \ell_{ijk} = \ell_{..k} \quad , \quad \text{etc..}$$

ML estimates of the P_{ijk} are obtained subject to the restraints shown in Table 4.0.

For all the tests having more than 1 d.f. the roots were computed by the methods in section 4.3. The roots for the 1 d.f. tests were computed by (4.1.1) or its alternative form. The roots and X_*^2 's are shown in the table on page 39.

Table 4.0. Linear Restraints for Testing Hypotheses for Data in Table 1.0.

	Hypotheses	Linear Restraints		
Main Effects	Homogeneity of Age Effects	$\lambda_{1..} = \lambda_{2..} = \lambda_{3..} = \lambda_{4..}$ $\lambda_{1..} - \lambda_{4..} = 0$ $\lambda_{2..} - \lambda_{4..} = 0$ $\lambda_{3..} - \lambda_{4..} = 0$	or	
	Homogeneity of Sex Effects	$\lambda_{..1} - \lambda_{..2} = 0$		
	Unburnt Tobacco Users vs. Smokers and Non-Users (A)	$2\lambda_{.1.} - \lambda_{.2.} - \lambda_{.3.} = 0$		
	Smokers vs. Non-Users (B)	$\lambda_{.2.} - \lambda_{.3.} = 0$		
First Order Interactions	Age x Sex	$\lambda_{1.1} - \lambda_{1.2} = \lambda_{2.1} - \lambda_{2.2} = \lambda_{3.1} - \lambda_{3.2}$ $= \lambda_{4.1} - \lambda_{4.2}$ $\lambda_{1.1} - \lambda_{1.2} - \lambda_{4.1} + \lambda_{4.2} = 0$ $\lambda_{2.1} - \lambda_{2.2} - \lambda_{4.1} + \lambda_{4.2} = 0$ $\lambda_{3.1} - \lambda_{3.2} - \lambda_{4.1} + \lambda_{4.2} = 0$	or	
	Age x (A)	$2\lambda_{11.} - \lambda_{12.} - \lambda_{13.} = 2\lambda_{21.} - \lambda_{22.}$ $- \lambda_{23.} = 2\lambda_{31.} - \lambda_{32.} - \lambda_{33.} = 2\lambda_{41.}$ $- \lambda_{42.} - \lambda_{43.}$		
	Age x (B)	$\lambda_{12.} - \lambda_{13.} = \lambda_{22.} - \lambda_{23.} = \lambda_{32.} - \lambda_{33.}$ $= \lambda_{42.} - \lambda_{43.}$		
	Sex x (A)	$2\lambda_{.11} - \lambda_{.21} - \lambda_{.31}$ $= 2\lambda_{.12} - \lambda_{.22} - \lambda_{.32}$		
	Sex x (B)	$\lambda_{.21} - \lambda_{.31} = \lambda_{.22} - \lambda_{.32}$		
	Second Order Interactions	Age x Sex x (A)	$2\lambda_{111} - 2\lambda_{112} - \lambda_{121} - \lambda_{131} + \lambda_{122}$ $+ \lambda_{132} = 2\lambda_{211} - 2\lambda_{212} - \lambda_{221} - \lambda_{231}$ $+ \lambda_{222} + \lambda_{232} = 2\lambda_{311} - 2\lambda_{312} - \lambda_{321}$ $- \lambda_{331} + \lambda_{322} + \lambda_{332} = 2\lambda_{411} - 2\lambda_{412}$ $- \lambda_{421} - \lambda_{431} + \lambda_{422} + \lambda_{432}$	
		Age x Sex x (B)	$\lambda_{121} - \lambda_{122} - \lambda_{131} + \lambda_{132} = \lambda_{221} - \lambda_{222}$ $- \lambda_{231} + \lambda_{232} = \lambda_{321} - \lambda_{322} - \lambda_{331}$ $+ \lambda_{332} = \lambda_{421} - \lambda_{422} - \lambda_{431} + \lambda_{432}$	

Test	λ^*	$\chi^2_{X^*}$
Homogeneity of Age Effects	$\lambda_1 = -2.522, \lambda_2 = -.283, \lambda_3 = 1.294, \lambda_4 = 1.511$	27.70
Homogeneity of Sex Effects	$\lambda = .306$.80
(A)	$\lambda = .994$	109.45
(B)	$\lambda = .844$	7.09
Age x Sex	$\lambda_1 = -.261, \lambda_2 = 1.451, \lambda_3 = .158, \lambda_4 = -1.348$	5.67
Age x (A)	$\lambda_1 = 1.070, \lambda_2 = -.025, \lambda_3 = .771, \lambda_4 = -1.816$	17.32
Age x (B)	$\lambda_1 = -.421, \lambda_2 = -1.743, \lambda_3 = 1.579, \lambda_4 = .585$	5.43
Sex x (A)	$\lambda = -.015$	1.06
Sex x (B)	$\lambda = -.533$	2.38
Age x Sex x (A)	$\lambda_1 = -.412, \lambda_2 = -.301, \lambda_3 = -.867, \lambda_4 = 1.580$	8.45
Age x Sex x (B)	$\lambda_1 = .493, \lambda_2 = 1.075, \lambda_3 = -.235, \lambda_4 = -1.333$	4.53

To illustrate the computations we will show one iteration of the Age x Sex x (B) interaction. The equations may be rewritten in any of several different forms. We will use the form,

$$\left(\frac{4-\lambda_1}{89+\lambda_1}\right)\left(\frac{1-\lambda_1}{40+\lambda_1}\right)\left(\frac{33-\lambda_1}{1+\lambda_1}\right)\left(\frac{54-\lambda_1}{1+\lambda_1}\right) = \left(\frac{17-\lambda_2}{78+\lambda_2}\right)\left(\frac{7-\lambda_2}{108+\lambda_2}\right)\left(\frac{25-\lambda_2}{5+\lambda_2}\right)\left(\frac{55-\lambda_2}{5+\lambda_2}\right) =$$

$$\left(\frac{23-\lambda_3}{66+\lambda_3}\right)\left(\frac{3-\lambda_3}{90+\lambda_3}\right)\left(\frac{23-\lambda_3}{4+\lambda_3}\right)\left(\frac{32-\lambda_3}{8+\lambda_3}\right) = \left(\frac{9-\lambda_4}{40+\lambda_4}\right)\left(\frac{5-\lambda_4}{31+\lambda_4}\right)\left(\frac{21-\lambda_4}{6+\lambda_4}\right)\left(\frac{1-\lambda_4}{3+\lambda_4}\right)$$

Choose $\lambda_{10} = \lambda_{20} = \lambda_{30} = \lambda_{40} = 0$ to start. Then,

$$\begin{array}{lll}
 R_1 = 2.002247 & S_1 = 3.335058 & \frac{1}{R_1 S_1} = .149754 \\
 R_2 = .776947 & S_2 = .681942 & \frac{1}{R_2 S_2} = 1.887387 \\
 R_3 = .313636 & S_3 = .846361 & \frac{1}{R_3 S_3} = 3.767198 \\
 R_4 = .042339 & S_4 = 1.915988 & \frac{1}{R_4 S_4} = 12.326220 \\
 h = 18.130559 & \frac{1}{Sh} = .191373 &
 \end{array}$$

$$\lambda_1' = (2.002247 - .191373) \cdot .149754 = .27 \quad \text{and by similar calculation } \lambda_2' = 1.10, \lambda_3' = .46, \lambda_4' = -1.83 .$$

These values of the λ 's are substituted into the equations and they are solved for a second approximation. For this set of equations the solutions are accurate for most purposes after four iterations. If, by sufficient insight, we had chosen $\lambda_{10} = .5$, $\lambda_{20} = 1.0$, $\lambda_{30} = -.25$, and $\lambda_{40} = -1.25$, convergence to the roots would have come in fewer iterations. It might further be noted that these computations are easily programed for high speed computers.

Example 4.2 Illustration of the use of a linear restraint in fitting a simple model

This example is given because it illustrates fitting models by setting up restraints that are considered to belong to error and estimating the P's subject to these restraints being zero.

Hodges (1958) gives a new method of fitting the logistic model in a simple linear regression case by ML. To illustrate the procedure, he gives the following example:

x_i	n_i	a_i
0	10	3
1	10	8
2	10	6

His estimates of the \hat{P} 's from the logistic model are

$$\hat{P}_1 = .4145, \quad \hat{P}_2 = .5710, \quad \text{and} \quad \hat{P}_3 = .7145 .$$

To obtain estimates of the \hat{P} 's assuming that they fall on the line $l_i = \alpha + \beta x_i$, we obtain estimates subject to the restraint $l_1 - 2l_2 + l_3 = 0$. The equation

$$\left(\frac{3 - \lambda}{7 + \lambda} \right) \left(\frac{6 - \lambda}{4 + \lambda} \right) = \left(\frac{8 + 2\lambda}{2 - 2\lambda} \right)^2$$

must be solved for λ^* . In this case $\lambda^* = -1.145$ which gives estimates of $\hat{P}_1 = .414$, $\hat{P}_2 = .571$, $\hat{P}_3 = .714$ (agreeing with Hodge's estimates to three places) and the X^2 for the test of lack of fit of the line is 3.33. It is believed that with some practice this could become a very fast method of fitting 3 point assays when the x 's are equally spaced. It would not be advantageous for assays with more than three points because each cycle of iteration would require solving two or more equations depending on the number of points.

Example 4.3 Illustration of the use of the \hat{P} 's obtained under the hypotheses of no interaction in fitting models

Cochran (1954) gives the following data which he obtained from Dr. Martha Rogers.

Table 4.1. Data on Number of Mothers with Previous Infant Losses

Birth Order		No. of Mothers with		Total
		Losses	None	
2	Problems	20	82	102
	Controls	10	54	64
3-4	Problems	26	41	67
	Controls	16	30	46
5+	Problems	27	22	49
	Controls	14	23	37

It is desired to compare the mothers of children in Baltimore schools who have been referred by their teachers as presenting behavioral problems to mothers of a comparable group of control children. For each mother, it is recorded whether she had suffered any infant losses previous to the child in the study.

Suppose it is desired to fit the model

$$l_{ij} = \mu + \alpha_i + \beta_j$$

where α_i is the effect of problem or control depending on whether $i = 1$ or 2 , and β_j is the effect of the j -th birth order, $j = 1, 2, 3$.

Fitting the model will use 4 degrees of freedom, leaving 2 for error. These two may be thought of as (Problem vs. Control) x (Birth order) interaction. If there is no interaction

$$l_{11} - l_{12} = l_{21} - l_{22} = l_{31} - l_{32} \quad .$$

To estimate the μ 's subject to these restrictions we must solve

$$\left(\frac{20-\lambda_1}{82+\lambda_1}\right)\left(\frac{54-\lambda_1}{10+\lambda_1}\right) = \left(\frac{26-\lambda_2}{41+\lambda_2}\right)\left(\frac{30-\lambda_2}{16+\lambda_2}\right) = \left(\frac{27-\lambda_3}{22+\lambda_3}\right)\left(\frac{23-\lambda_3}{14+\lambda_3}\right)$$

for λ_1^* , λ_2^* and λ_3^* where $\lambda_1^* + \lambda_2^* + \lambda_3^* = 0$. The solution is $\lambda_1^* = -.5057$, $\lambda_2^* = -1.2126$ and $\lambda_3^* = 1.7183$ and the estimates of P_{ij} under the restriction of no interaction are $\hat{P}_{11} = .2010$, $\hat{P}_{12} = .1483$, $\hat{P}_{21} = .4062$, $\hat{P}_{22} = .3215$, $\hat{P}_{31} = .5160$ and $\hat{P}_{32} = .4248$. X^2 with 2 d.f. is .85. The logits corresponding to these \hat{P} 's are -1.38 , -1.75 , $-.38$, $-.75$, $.06$ and $-.31$.

Earlier in this chapter it was suggested that estimates of the \hat{P} 's subject to restraints which ordinarily are called error could be used to get ML estimates of the parameters in a model without iteration. Let us derive a form convenient for this purpose.

(3.1.2) may be written

$$\begin{matrix} X' & \delta_p & = & \underline{0} \\ \text{txr} & \text{rxl} & & \end{matrix}$$

where δ_p is a rxl vector of elements

$$n_i \sqrt{p_i - P_i(\underline{\theta})}, \quad i = 1, \dots, r$$

Expanding this equation about a trial point $\underline{\theta}$, we obtain

$$\begin{matrix} X' & \delta_p & = & X' & D_1 & X & \delta_\theta & = & \underline{0} \\ \text{txr} & \text{rxl} & & \text{txr} & \text{rxr} & \text{rxt} & \text{txl} & & \text{txl} \end{matrix}$$

where δ_p has elements $\sqrt{n_i(p_i - P_i(\underline{\theta}))}$, D_1 is a matrix with $n_i P_i(\underline{\theta}) Q_i(\underline{\theta})$ on the diagonal and zeros elsewhere; δ_θ is the txl vector $(\underline{\theta} - \underline{\theta})$; and $P_i(\underline{\theta})$ is $P_i(\underline{\theta})$ evaluated at $\underline{\theta}$. Add $(X'D_1X)\underline{\theta}$ to the left and $X'D_1\underline{\theta}^*$, where $\underline{\theta}^*$ is the vector of provisional logits, to the right hand side of the above equation. Thus

we obtain

$$(X'D_1X)\underline{\theta} = X'\underline{\delta}_p + X'D_1\underline{\ell}^*$$

Now let

$$\underline{L}_* = \underline{\ell}^* + D_1^{-1}\underline{\delta}_p$$

or

$$\underline{\ell}^* = \underline{L}_* - D_1^{-1}\underline{\delta}_p$$

then

$$\begin{aligned} (X'D_1X)\underline{\theta} &= X'\underline{\delta}_p + X'D_1(\underline{L}_* - D_1^{-1}\underline{\delta}_p) \\ &= X'D_1\underline{L}_* \end{aligned}$$

and

$$\underline{\theta} = (X'D_1X)^{-1}X'D_1\underline{L}_*$$

Let us use the \hat{P} 's obtained under the no interaction hypothesis for the elements of D_1 and $\underline{\delta}_p$, reparameterize the model into a multiple regression form so that X is of rank 4, and solve for the new parameters. Let $\alpha = \alpha_1 - \alpha_2$ and $\tau_1 = \beta_1 - \beta_2$ and $\tau_2 = \beta_1 - \beta_3$. Then

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

and the equations

$$\begin{bmatrix} 71.93772 & -17.62000 & -1.72980 & 3.18720 \\ -17.62000 & 71.93772 & -2.17140 & -5.10068 \\ -1.72980 & -2.17140 & 50.65988 & 24.46504 \\ 3.18720 & -5.10068 & 24.46504 & 45.74288 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha \\ \tau_1 \\ \tau_2 \end{bmatrix} = \begin{bmatrix} -52.4868 \\ 3.5318 \\ -23.0773 \\ -34.6770 \end{bmatrix}$$

have the solution

$$\hat{\mu} = -.752, \quad \hat{\alpha} = -.185, \quad \hat{\tau}_1 = -.187, \quad \hat{\tau}_2 = -.627.$$

The predicted logits are

$$l_{11} = -1.38, \quad l_{12} = -1.75, \quad l_{21} = -.38, \quad l_{22} = -.75$$

$$l_{31} = .06 \quad \text{and} \quad l_{32} = -.31,$$

which are exactly those used to obtain \hat{D}_1 and $\hat{\delta}_p$ for this solution. Any further cycles of iteration would give the same estimates unless a more accurate table of logits was used. If it is desired to fit models to data, this technique may be used to shorten appreciably the computations in many cases.

CHAPTER V

NON-CENTRALITY PARAMETERS FOR OBTAINING THE ASYMPTOTIC POWER
OF TESTS OF HYPOTHESES FOR THE LOGISTIC MODEL

5.0 The Utility of Non-Centrality Parameters

In Chapter II the non-centrality parameters derived by Diamond (1958) were given as general forms with $P_{ij}(\underline{\theta})$ being unspecified except for certain analytic conditions. Since the asymptotic power is a function of the non-centrality parameters and the d.f. only, the non-centrality parameter provides the reference point needed for empirical investigations of the small sample power of tests. Without such a reference point, the way in which various parameters, viz. sample size, cell probabilities, hypotheses and alternatives, affect the power of a test is difficult if not impossible to determine and empirical studies become sterile exercises in computation. Also the non-centrality parameters may have other important uses such as, for example, designing experiments which maximize the power of the tests. This particular aspect may be of interest in bioassay.

In this chapter we will obtain the non-centrality parameters for tests on the logistic model from Diamond's general forms. Studies of the small sample power of the tests will be deferred until the next chapter.

5.1 Non-Centrality Parameters for Tests of the Fit of the Model

To test the hypothesis

$$H_0: P_i = P_i(\underline{\theta}) \quad ,$$

subject to

$$f_m(\underline{\theta}) = 0, \quad m = 1, \dots, u < t,$$

against the alternative

$$H_n: P_{in} = P_i(\underline{\theta}) + \frac{d_i}{\sqrt{n}}$$

subject to

$$f_m(\underline{\theta}) = 0 \quad \text{and} \quad d_i + d_i' = 0$$

where

$$\frac{d_i'}{\sqrt{n}} = Q_{in} - Q_i(\underline{\theta}),$$

and

$$P_{in} \neq P_i(\underline{\theta}) \quad \text{for any } \underline{\theta} \text{ in } \underline{\Omega},$$

we note that

$$\Delta^* = \underline{d}' \underline{I} - B_*(B_* B_*') B_*' \underline{d}$$

where \underline{d} and B_* are defined by (2.1.3) and (2.1.4). Recall that

$$B_* = B_2 - B_1 F_1^{-1} F_2$$

where B_1 , B_2 , F_1 and F_2 are the general forms defined by (2.1.6), (2.1.7) and (2.1.8). Let

$$f_m(\underline{\theta}) = 0, \quad m = 1, \dots, u,$$

be the linear functions of the θ_k ,

$$(5.1.1) \quad \begin{aligned} f_{11}\theta_1 + f_{12}\theta_2 + \dots + f_{1t}\theta_t &= 0, \\ &\vdots \\ f_{u1}\theta_1 + f_{u2}\theta_2 + \dots + f_{ut}\theta_t &= 0, \end{aligned}$$

where the matrix of coefficients, F , is of rank u . We may write

(5.1.1) more compactly as

$$(5.1.2) \quad \underset{uxt}{F} \underset{txl}{\theta} = \underset{-}{0} .$$

u of the θ_k may be arbitrarily chosen as the dependent θ 's as long as the matrix of coefficients is of rank u . This corresponds to the partitions

$$\underset{uxt}{F} = \left[\underset{u}{F_1} : \underset{t-u}{F_2} \right] u$$

and

$$\underset{txl}{\theta} = \begin{bmatrix} \theta_1 \\ \dots \\ \theta_2 \end{bmatrix} \begin{matrix} t \\ \\ t-u \end{matrix} .$$

Therefore,

$$\underset{-}{F\theta} = \left[\underset{-}{F_1} : \underset{-}{F_2} \right] \begin{bmatrix} \theta_1 \\ \dots \\ \theta_2 \end{bmatrix} = F_1 \theta_1 + F_2 \theta_2 ,$$

giving

$$(5.1.3) \quad \frac{\partial \underset{-}{F\theta}}{\partial \theta_1} = F_1 \quad \text{and} \quad \frac{\partial \underset{-}{F\theta}}{\partial \theta_2} = F_2 .$$

On substituting the logit model for $P_i(\theta)$ and taking partial derivatives, we obtain the matrices

$$(5.1.4) \quad B_1 = \begin{bmatrix} \sqrt{\frac{n_1}{n}} P_1^O Q_1^O x_{11} & \dots & \sqrt{\frac{n_1}{n}} P_1^O Q_1^O x_{1u} \\ \sqrt{\frac{n_1}{n}} Q_1^O P_1^O x_{11} & \dots & \sqrt{\frac{n_1}{n}} Q_1^O P_1^O x_{1u} \\ \vdots & & \vdots \\ \sqrt{\frac{n_r}{n}} P_r^O Q_r^O x_{r1} & \dots & \sqrt{\frac{n_r}{n}} P_r^O Q_r^O x_{ru} \\ \sqrt{\frac{n_r}{n}} Q_r^O P_r^O x_{r1} & \dots & \sqrt{\frac{n_r}{n}} Q_r^O P_r^O x_{ru} \end{bmatrix} ,$$

$$(5.1.5) \quad B_2 = \begin{bmatrix} \sqrt{\frac{n_1}{n}} P_1^O Q_1^O x_{1u+1} & \dots & \sqrt{\frac{n_1}{n}} P_1^O Q_1^O x_{1t} \\ -\sqrt{\frac{n_1}{n}} Q_1^O P_1^O x_{1u+1} & \dots & -\sqrt{\frac{n_1}{n}} Q_1^O P_1^O x_{1t} \\ \vdots & & \vdots \\ \sqrt{\frac{n_r}{n}} P_r^O Q_r^O x_{ru+1} & \dots & \sqrt{\frac{n_r}{n}} P_r^O Q_r^O x_{rt} \\ -\sqrt{\frac{n_r}{n}} Q_r^O P_r^O x_{ru+1} & \dots & -\sqrt{\frac{n_r}{n}} Q_r^O P_r^O x_{rt} \end{bmatrix}$$

When multiplied together

$$B_i' B_j = X_i' D_w X_j ,$$

$$\underline{d}' B_i = \underline{d}^* X_i$$

for

$$i \text{ and } j = 1, 2, * ,$$

or no subscript, and

$$\underline{d}' \underline{d} = \underline{d}^* D_w^{-1} \underline{d}^* ,$$

where

$$\underline{d}^* = \left[\frac{n_1}{n} d_1, \dots, \frac{n_r}{n} d_r \right] ,$$

D_w is an rxr diagonal matrix with elements $\frac{n_i}{n} P_i^O Q_i^O$ on the diagonal,

$$(5.1.6) \quad X_* = X_2 - X_1 F_1^{-1} F_2$$

where

$$(5.1.7) \quad X_1 = \begin{bmatrix} X_{1j} \end{bmatrix} \quad \begin{matrix} i = 1, \dots, r , \\ j = 1, \dots, u , \end{matrix}$$

and

$$X_2 = \begin{bmatrix} X_{1j} \end{bmatrix} \quad \begin{matrix} i = 1, \dots, r , \\ j = u+1, \dots, t . \end{matrix}$$

$$\Delta_1 = \underline{c}' F_1^{-1} B_1' \left[I - B_* (B_*' B_*)^{-1} B_*' \right] B_1 F_1^{-1} \underline{c}$$

as given by (2.2.1). On substituting for F_1^{-1} , B_1 , B_* and their products we find

$$(5.2.2) \quad \Delta_1 = \underline{c}' F_1^{-1} X_1' \left[D_w - D_w X_* (X_*' D_w X_*)^{-1} X_*' D_w \right] X_1 F_1^{-1} \underline{c}$$

A hypothesis that occurs frequently is

$$H_0: \theta_k = h_k,$$

against the alternative

$$H_n: \theta_k = h_k + \frac{c_k}{\sqrt{n}}, \quad k = 1, \dots, u.$$

The non-centrality parameter is given by

$$\Delta_1 = \underline{c}' B_1 \left[I - B_2 (B_2' B_2)^{-1} B_2' \right] B_1 \underline{c},$$

where $\underline{c} = [c_1, \dots, c_u]$, B_1 is as in (2.1.6) and B_2 is as in (2.1.7). This is a special case of (5.2.2) in which $F_1 = I$ and $F_2 = 0$. The non-centrality parameter for the logistic model then becomes

$$(5.2.3) \quad \Delta_1 = \underline{c}' X_1' \left[D_w - D_w X_2 (X_2' D_w X_2)^{-1} X_2' D_w \right] X_1 \underline{c}.$$

Example 5.1

To obtain the non-centrality parameters for the test in bioassay

$$H_0: \beta = \beta_0$$

against the alternative

$$\beta = \beta_0 + \frac{c}{\sqrt{n}},$$

we set

$$X_1' = (x_1, x_2, \dots, x_r) ,$$

and

$$X_2' = (1, 1, \dots, 1)$$

in (5.2.3) . Multiplying the matrices together and simplifying, we obtain

$$\Delta_1 = c^2 \left[\begin{array}{c} r \\ \sum_{i=1}^r w_i x_i^2 - \frac{\left(\sum_{i=1}^r w_i x_i \right)^2}{\sum_{i=1}^r w_i} \end{array} \right] ,$$

$$= c^2 S_{w x}^2$$

where

$$w_i = \frac{n_i}{n} P_{iQ_i}^{O O} .$$

Let us now define w_i' as $\frac{n_i}{n} P_{iQ_i}^{O O}$, then

$$\Delta_1 = (\beta - \beta_0)^2 S_{w' x}^2 .$$

Since the power is a monotone increasing function of Δ_1 , the above result shows that if an experiment is planned to minimize the variance of b , the estimate of β , the power is maximized.

Example 5.2

To obtain the non-centrality parameter for the test of parallelism of two lines, i.e.,

$$H_0: \beta_1 = \beta_2$$

against the alternative

$$H_n: \beta_1 - \beta_2 = \frac{c}{\sqrt{n}} ,$$

in (5.4.2) we let

$$X_1 = \begin{bmatrix} x_{11} \\ x_{12} \\ \vdots \\ \vdots \\ x_{1k} \\ 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix}, \text{ and } X_2 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ x_{21} & 0 & 1 \\ x_{22} & 0 & 1 \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ x_{2k} & 0 & 1 \end{bmatrix}$$

$$F_1 = 1, \text{ and } F_2 = \begin{bmatrix} -1 & 0 & 0 \end{bmatrix}.$$

Therefore

$$X_* = \begin{bmatrix} x_{11} & 1 & 0 \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ x_{1k} & 1 & 0 \\ x_{21} & 0 & 1 \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ x_{2k} & 0 & 1 \end{bmatrix}$$

and multiplying the matrices together and simplifying we obtain

$$\Delta_1 = c^2 \left[\frac{Sw_1 x_1^2 Sw_2 x_2^2}{Sw_1 x_1^2 + Sw_2 x_2^2} \right]$$

where

$$Sw_1 x_1^2 = \sum_{j=1}^k w_{1j} x_{1j}^2 - \frac{\left(\sum_{j=1}^k w_{1j} x_{1j} \right)^2}{\sum_{j=1}^k w_{1j}},$$

$$Sw_2^2 x_2^2 = \sum_{j=1}^{\ell} w_{2j} x_{2j}^2 - \frac{\left(\sum_{j=1}^{\ell} w_{2j} x_{2j} \right)^2}{\sum_{j=1}^{\ell} w_{2j}},$$

and

$$w_{ij} = \frac{n_{ij}}{n} P_{ij}^0 Q_{ij}^0.$$

Redefining w'_{ij} as $n_{ij} P_{ij}^0 Q_{ij}^0$, we have

$$\Delta_1 = \frac{(\beta_1 - \beta_2)^2}{\frac{1}{Sw_1^2 x_1^2} + \frac{1}{Sw_2^2 x_2^2}}.$$

Hence we see that the power of the test is maximized when the sum of the variances of b_1 and b_2 is minimized.

CHAPTER VI

STUDIES ON THE SMALL SAMPLE PROPERTIES OF THE TESTS

6.0 Choice of Sampling Populations

In order to find the sample size at which the distribution of the test statistics is adequately approximated by the chi square distribution and to check on the agreement between the observed power and the power predicted from non-central chi square distribution using the non-centrality parameters given in Chapter V, some sampling experiments were conducted for a 2x2 factorial experiment.

The four cell probabilities were obtained by assuming that, in the logistic model,

$$\log_e \frac{P_{ij}}{Q_{ij}} = \mu + \alpha_i + \beta_j + I_{ij}$$

where

$$i \text{ (and } j) = 1, 2 \quad ,$$

$$\alpha = \alpha_1 - \alpha_2 = -.3 \quad ,$$

$$\beta = \beta_1 - \beta_2 = .6 \quad ,$$

and

$$I_{ij} = 0$$

in all the configurations sampled. The individual configurations were obtained by letting $\mu = 0$, $\mu = -1$ and $\mu = -2$ in the A, B, and C configurations respectively. Configuration C was computed, for example, by

$$\log_e \frac{P_{11}}{Q_{11}} = -2 + (-.3) + .6 = -1.7 \quad ,$$

$$\log_e \frac{P_{12}}{Q_{12}} = -2 - (-.3) + .6 = -1.1 \quad ,$$

$$\log_e \frac{P_{21}}{Q_{21}} = -2 + (-.3) - .6 = -2.9 \quad ,$$

$$\log_e \frac{P_{22}}{Q_{22}} = -2 - (-.3) - .6 = -2.3 \quad .$$

The P's corresponding to these logits were obtained from Berkson's tables (1957). Using this technique, the P configurations shown in Table 6.0 were obtained.

Table 6.0. P Configurations Used in Sampling

Configurations	P_{11}	P_{12}	P_{21}	P_{22}
A	.57	.71	.29	.43
B	.33	.47	.13	.21
C	.15	.25	.052	.091

These cell probabilities were used in investigations of the exact sampling distributions for $n = 1$ and $n = 2$ in each cell and in an empirical investigation of significance level and power for larger sample sizes.

6.1 Studies on the Exact Distribution of X_1^2

6.1.a Methods. The distribution of the test statistic was found in two cases by enumeration of all possible outcomes for $n = 1$ and $n = 2$ in each cell. There are 16 and 81 outcomes respectively. The

data were analyzed by assuming they had been generated by configurations A, B, and C. Only the null cases were investigated. That is, the hypotheses tested were $\alpha_0 = -.3$, $\beta_0 = .6$, and $I_0 = 0$.

To test these hypotheses, ML estimates of the P_{ij} were obtained subject to the following restraints:

$$(6.1.1) \quad H_{\alpha_0}: \ell_{11} - \ell_{12} + \ell_{21} - \ell_{22} = 4(\alpha_0) \quad ,$$

$$H_{\beta_0}: \ell_{11} + \ell_{12} - \ell_{21} - \ell_{22} = 4(\beta_0) \quad ,$$

$$H_{I_0}: \ell_{11} - \ell_{12} - \ell_{21} + \ell_{22} = 4(I_0) \quad .$$

The root of a third degree polynomial which is required to complete the estimation was computed by methods given in section 4.1.

Since we are assuming existence of both main effects and interaction, there is no X^2 for lack of fit and the test statistic is

$$X_1^2 = X_*^2 = \lambda^* S \quad .$$

From these computations the exact central distribution of X_1^2 for each of the P configurations in Table 6.0 can be found.

6.1.b Results. The distribution for $n = 1$ is shown in Table 6.1.

Table 6.1. Distribution of X_1^2 for $n = 1$ in Each Cell

Test	X_1^2	Configuration								
		A			B			C		
		$f(X_1^2)$	Mean	Vari- ance	$f(X_1^2)$	Mean	Vari- ance	$f(X_1^2)$	Mean	Vari- ance
α	0.00	.88086	.42	1.43	.92304	.27	.96	.97627	.08	.31
	2.89	.09157			.05914			.01823		
	5.52	.02756			.01718			.00550		
β	0.00	.81807	.47	1.39	.88248	.30	.95	.96375	.09	.31
	2.15	.16679			.10773			.03323		
	7.43	.01513			.00978			.00302		
I	0.00	.89953	.40	1.44	.93508	.26	.97	.97997	.08	.31
	4.00	.10046			.06491			.02003		

The frequency distributions for the case $n = 2$ are given in Table 6.2. The frequencies for A, B, and C models followed the same general trend, except that they became more skewed, with the frequency of $X_1^2 = 0$ increasing and the frequency of the larger values decreasing correspondingly, as the P's were shifted more toward zero.

It will be observed in Tables 6.1 and 6.2 that the I test does not have as many possible values of X_1^2 as the other tests. There is a symmetry present in this test that is not in the others. The hypothesis tested was $I = 0$ and not $I = I_0$, $I_0 \neq 0$, as in the other tests. This gives rise to more configurations having the same X_1^2 , in fact twice as many, as the other tests. If we had tested $\alpha = 0$ and $\beta = 0$, they too would have had fewer possible values of X_1^2 , but the results would not have been comparable with the random sampling studies.

Table 6.2. Distribution of X_1^2 for $n = 2$ in Each Cell

X^2	Configuration								
	A			B			C		
	$f(X_1^2)$	Mean	Variance	$f(X_1^2)$	Mean	Variance	$f(X_1^2)$	Mean	Variance
α Test									
0.00	.49502	.94	2.33	.67996	.61	1.67	.90299	.19	.56
0.18	.04075			.01653			.00149		
0.40	.13050			.08165			.01867		
1.61	.14214			.09869			.04403		
1.64	.07066			.04573			.00992		
3.70	.06026			.03662			.00881		
4.29	.04168			.03082			.01240		
5.91	.00869			.00331			.00033		
8.28	.00958			.00637			.00133		
10.82	.00075			.00032			.00003		
	<u>1.00003</u>			<u>1.00000</u>			<u>1.00000</u>		
β Test									
0.00	.37140	1.02	2.41	.58067	.66	1.74	.86601	.20	.59
0.15	.15378			.09783			.02203		
0.73	.04075			.01653			.00149		
1.05	.21120			.16549			.07701		
2.49	.12476			.08121			.01854		
2.93	.04738			.02955			.00656		
4.40	.02682			.01136			.00105		
6.27	.02006			.01504			.00681		
13.44	.00364			.00223			.00049		
14.52	.00024			.00009			.00001		
	<u>1.00003</u>			<u>1.00000</u>			<u>1.00000</u>		
I Test									
0.00	.58165	.92	2.23	.72547	.60	1.59	.91703	.19	.54
0.89	.20116			.12738			.02859		
2.67	.16184			.11318			.04705		
5.34	.05028			.03190			.00715		
8.00	.00510			.00207			.00018		
	<u>1.00003</u>			<u>1.00000</u>			<u>1.00000</u>		

The exact distributions of X_1^2 for these cases show several things that would be difficult to prove conclusively with sampling experiments unless an enormous amount of sampling was done. The distribution is not the same for all the tests in a single configuration. For the case $n = 1$ the means of X_1^2 for the various tests are to be related to the size of the treatment effect and the variances are all almost equal. In the case $n = 2$ the effect of the size of the parameter being tested is even more pronounced with both the means and variances of X_1^2 varying directly as the size of the treatment effect. As would be expected, the moments of X_1^2 depart more from the moments of chi square as the assumed P's are shifted more toward zero. Also the distribution of X_1^2 when $n = 2$ is not unimodal.

From this investigation it is apparent that the distribution of X_1^2 is not closely approximated by chi square in experiments with small numbers in the cells. This may or may not be true for experiments with more cells. Considering the assumptions made in deriving the asymptotic distribution of X_1^2 , these results are not surprising. However, one always hopes that the large sample theory may hold reasonably well for small samples or that some modification can be made which will improve the approximation. We have not been successful in discovering such a modification.

Since increasing n in each cell rapidly increases the amount of enumeration and computation, no further work was done toward obtaining exact distributions. The remainder of the work was done by random sampling investigations.

6.2 Random Sampling Investigations

6.2.a Methods. The original configurations of P's were retained. Configurations A_1 , A_2 , and C_1 and C_2 will be referred to from time to time. The subscripts refer to different investigations on the A and C configurations. The difference between A_1 and A_2 , and C_1 and C_2 is shown in Table 6.4.

In the sampling studies that follow, random binomial variates were generated by a program obtained from the Royal McBee Company for use with their computer, the LGP-30. The program generates variates by squaring a random number, a , multiplying a^2 by c where c is some large number, and, if there is any overflow, rounding off the most significant part of a^2c and then adding the result to a^2 . The result of these operations now becomes the number operated upon and the whole process is repeated once more. Several digits are extracted from the middle of this generated number. These are presumed to be a variate from a uniform distribution with limits $0 \leq x \leq R$, where R is the maximum size of the extracted number. The uniform variate is transformed to a binomial variate by comparing the generated number, x , with a number r , where r is chosen so that $\frac{r}{R}$ is the cell probability. If $0 \leq x < r$ a failure is recorded or if $r \leq x \leq R$ a success is recorded. By the use of this binomial variate generator, the predetermined size sample was taken for each of the 4 cells of a 2×2 factorial experiment.

The expected cell probabilities and those observed from sampling are given in Table 6.3.

Table 6.3. Observed and Expected Cell Probabilities

Configuration	P_{11}		P_{12}		P_{21}		P_{22}	
	Exp.	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.	Obs.
A_1	.57	.53	.71	.71	.29	.32	.43	.43
A_2	.57	.57	.71	.68	.29	.30	.43	.43
B	.33	.32	.47	.46	.13	.12	.21	.21
C_1	.15	.15	.25	.26	.052	.042	.091	.077
C_2	.15	.14	.25	.24	.052	.043	.091	.074

The hypotheses tested, sample sizes in each cell and the number of times each experiment was repeated are given in Table 6.4.

Table 6.4. Parameters Tested, Sample Sizes and Number of Times Each Experiment Was Repeated

Configuration	μ	α Test		β Test		I Test		Sample Size in Each Cell	No. Times Exp. Rep't'd.
		α_0	α_1	β_0	β_1	I_0	I_1		
A_1	0	-.3	.1	.6	.5	0	.3	10	50
A_2	0	-.3	0	.6	.5	0	.9/.8	20	50/50
B	-1	-.3	0	.6	.5	0	.8	30	100
C_1	-2	-.3	0	.6	.5	0	.8	45	100
C_2	-2	-.3	.5	.6	0	0	-.3	45	100

To check on the significance level of the test, the parameter values shown under α_0 , β_0 and I_0 in Table 6.4 which were used in

generating the samples were tested by obtaining ML estimates of the P_{ij} subject to the restraints given by (6.1.1), computing X_1^2 and noting the percent of the time X_1^2 exceeded the tabular values of chi square. The power of the tests was estimated by obtaining estimates of P_{ij} subject to

$$\begin{aligned} H_{\alpha_1}: \quad & l_{11} - l_{12} + l_{21} - l_{22} = 4\alpha_1 \quad , \\ (6.2.1) \quad H_{\beta_1}: \quad & l_{11} + l_{12} - l_{21} - l_{22} = 4\beta_1 \quad , \\ H_{I_1}: \quad & l_{11} - l_{12} - l_{21} + l_{22} = 4I_1 \quad . \end{aligned}$$

These estimates were used in computing X_1^2 and the power was estimated by noting the percent of the time X_1^2 exceeded the tabular value of chi square.

Configuration A_1 was performed as a pilot trial to observe the random variable generator and the computations required to perform the test. If these appeared to work properly an idea could be obtained about the sample sizes and alternative parameters to use in a more extensive sampling program. It appeared from the results of A_1 that the sample size of 10 was much too small and the alternative parameters α_1 , β_1 , and I_1 , were too close to the parameters used in generating the sample. The sample size was increased to 20 and the alternatives were changed as shown in A_2 . However, a check of A_2 after the fiftieth experiment revealed that the observed power was higher than anticipated against the alternative I_1 . I_1 was, therefore, decreased from .9 to .8 for the remaining 50 experiments. Because the P 's were shifted more toward zero, the sample size was increased to

30 for the B configuration.

Before decreasing the cell probabilities still further to obtain the C_1 configurations it was decided to try to relate the agreement with chi square to some parameter which is a function of both the sample size and the cell probabilities. The parameter chosen was

$$Z = \sum_{ij} \frac{1}{n_{ij} P_{ij} Q_{ij}} \cdot$$

This was chosen partly from intuition and partly because functions of $\frac{1}{nPQ}$ occur in the moments of the goodness of fit test, see Irwin (1950). For A_1 , $Z = 1.79$; for A_2 , $Z = .89$, and for B, $Z = .78$. As a rule of thumb for determining sample size it was decided to try $Z \leq 1$. For the C_1 and C_2 cases, 45 observations per cell is the number for which $Z = 1$. Applying this rule to A_2 and B would have given sample sizes of 18 and 23 in each cell.

The change of parameters from C_1 to C_2 was to observe the effect of switching the alternatives on the power of the test and to get a point on the long middle section of the power curve that had not previously been examined.

The observed probability of a Type I error for .05 and .01 tests, and the power of these tests against the alternatives shown were obtained by observing the percent of the time X_1^2 exceeded 3.84 and 6.64, the .05 and .01 tabular values of chi square respectively.

The non-centrality parameters were computed by letting

$$X_1^i = (1, 1, -1, -1)$$

and

Table 6.5. Probability of a Type I Error and Power Observed for an .05 Test

	α Test			β Test			I Test			
	Obs. Type I Error	Δ_1	Expect- ed Power	Obs. Type I Error	Δ_1	Expect- ed Power	Obs. Type I Error	Δ_1	Expect- ed Power	
A ₁	4%	.37	9.5%	0	.09	5.5%	0	.80	15%	24%
A ₂	1%	1.61	25.0%	10%	.18	7.2%	10%	14.46	98%	94%
B	9%	1.85	29.0%	4%	.21	7.5%	7%	13.18	93%	88%
C ₁	6%	1.43	23.0%	5%	.16	7.0%	9%	10.17	89%	76%
C ₂	4%	10.17	90.0%	8%	5.72	67.0%	69%	1.43	23%	24%

Table 6.6. Probability of a Type I Error and Power Observed for an .01 Test

	α Test			β Test			I Test		
	Obs. Type I Error	Δ_1	Expect- ed Power	Obs. Type I Error	Δ_1	Expect- ed Power	Obs. Type I Error	Δ_1	Expect- ed Power
A ₁	0	2.5%	0	0	1.2%	0	0	4.7%	6%
A ₂	0	9.5%	2%	4%	2.0%	2%	0	89.0%	86%
B	0	11.0%	16%	2%	1.5%	2%	0	79.0%	82%
C ₁	0	8.5%	15%	0	1.7%	0	0	85.0%	82%
C ₂	2%	73.0%	79%	0	43.0%	55%	0	73.0%	76%

$$X_2 = \begin{bmatrix} 1, & 1, & 1, & 1 \\ 1, & -1, & 1, & -1 \\ 1, & -1, & -1, & 1 \end{bmatrix}$$

for the test of β , for example, and applying (5.2.3). The expected power was obtained by referring the non-centrality parameters, Δ_1 , to Fix's tables (1949).

6.2.b Results. The non-centrality parameters, observed probability of a Type I error and the power for an .05 test and an .01 chi square test are shown in Tables 6.5 and 6.6. (See page 65). The sample means and variances of X_1^2 are given in Table 6.7. The discussion of the results will be confined to the .05 test since the .01 test results seem to be similar to those for the .05 test.

Table 6.7. Mean and Variance of X_1^2 when H_0 Is True

Configuration	α		β		I	
	Mean	Variance	Mean	Variance	Mean	Variance
A_1	.79	1.02	.86	.78	.93	.98
A_2	.57	.61	1.30	3.27	.95	1.27
B	1.18	2.10	.89	2.10	.85	1.18
C_1	1.40	1.84	1.26	1.56	1.15	1.39
C_2	1.29	1.99	.98	1.73	1.16	1.36

Even though the experiment was repeated only 50 times, it is clear from the results of A_1 that a sample size of 10 in each cell is not adequate for the large sample theory to hold for this configuration of P's. Instead of getting an .05 test, the test is probably about an .01 test.

This is understandable when we look at the first two moments. The mean is not so far from the expected value of 1 as to be outside the range of sampling variation, but the second moment is about the same size as the first, rather than twice as large, which would explain the very conservative test obtained. That is, the mass of the distribution of X_1^2 for this sample size and P configuration is more closely concentrated about the mean than chi square. The observed power would seem to be at least as high as the expected power obtained from the computed non-centrality parameters.

There is considerable variation in the performances of the tests of the A_2 configuration when they are looked at individually. Instead of the expected 5% Type I errors for the three tests, they were 1%, 10% and 2% for the α , β , and I tests respectively. However, the average Type I error rate for the experiment is 4%. For the experiment as a whole, X_1^2 for a sample size of 20 follows chi square adequately, but when the tests are considered individually the number of Type I errors may deviate considerably from the expected 5%.

The performance of the test in the B configuration was very satisfactory. The null hypothesis was rejected 4% of the time for the β and I test, and 9% of the time for the α test. This would make the Type I error for the experiment 5.7% which is certainly adequate agreement.

The number of Type I errors for C_1 and C_2 fall in a range acceptable by most standards. From these results of C_1 and C_2 the rule of thumb for a one d.f. test, $\sum_{ij} \frac{1}{n_{ij} P_{ij} Q_{ij}} \leq 1$, does not seem too

conservative. Further work needs to be done for larger experiments and tests having more than 1 d.f. In the meantime it appears that the rules given by Cochran (1954) would not lead one astray even though they were not formulated with these tests in mind.

When the sample size is large enough for distribution of the test statistic to be adequately approximated by the chi square distribution when the null hypothesis is true, the non-centrality parameters predict the power of the experiment with adequate accuracy. The agreement between predicted and observed power is shown graphically in Figure 6.0. It is generally good except that the test for interaction has lower power than predicted. However, even for this test the disagreement is not enough to negate the usefulness of the non-centrality parameters and, according to these data, they should predict the power to within 7% for a 5% test and to within 3% for a 1% test, on the average.

The number of Type I errors, power of the test and the moments are consistent in the sense that the tests with large variances tended to have more Type I errors and those with small variances had fewer. The number of Type I errors for both the .05 and the .01 test seems to be related to the absolute size of the parameter being tested. This is in accord with the results of the exact studies described in section 6.1 where it was observed that the size of the treatment effects seemed to influence the mean and variance of the tests.

The agreement between the distribution of X_1^2 observed in sampling and chi square is not as close as one would expect intuitively by looking at the .05 points and the first two moments of the exact

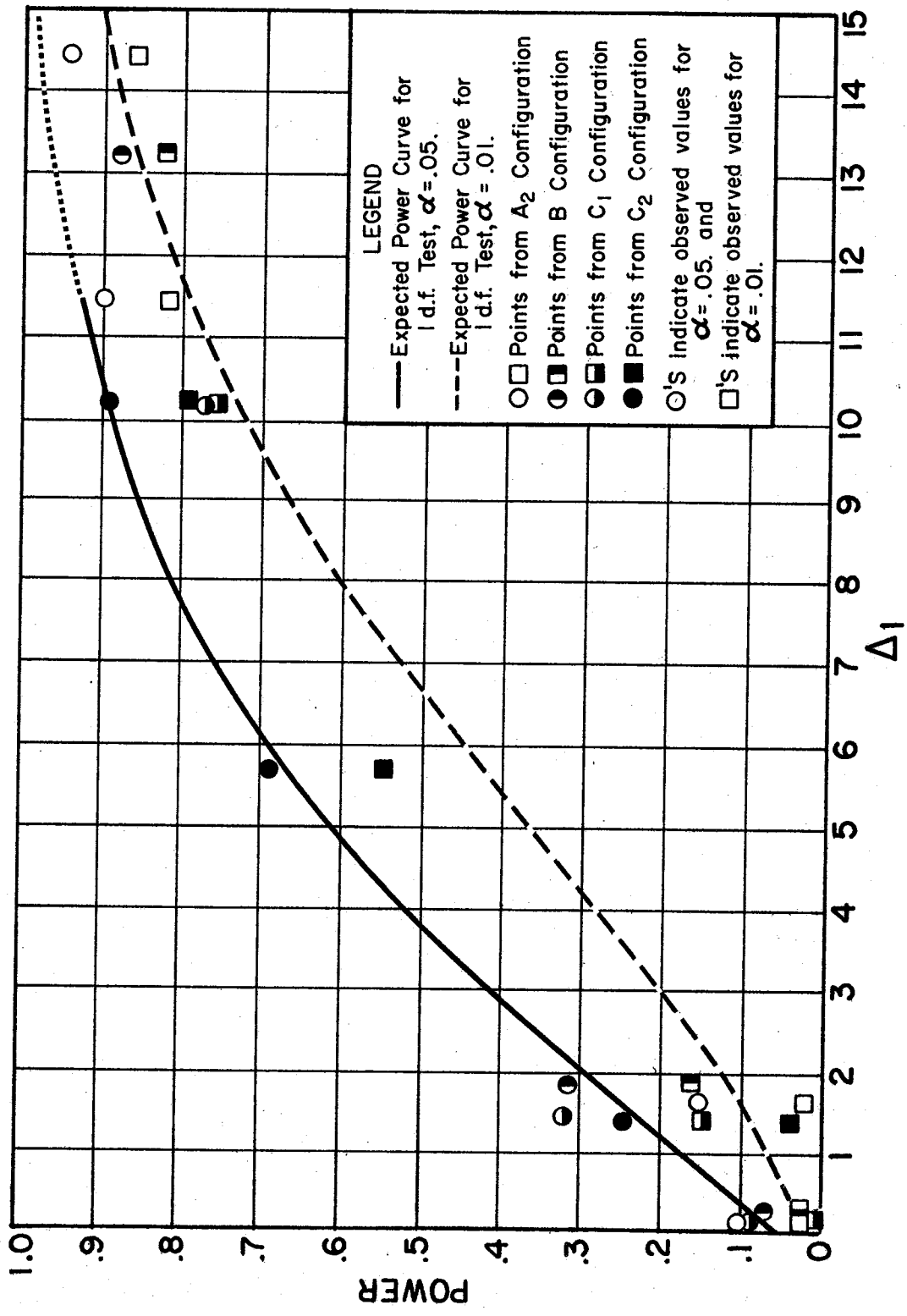


Figure 6.0. OBSERVED AND EXPECTED POWER

distributions for $n = 1$ and 2 . The reason for this lack of agreement, if it is other than sampling variation, is not apparent at this time.

6.3 Alternative Approach for Small Samples

As an alternative to a conventional test of significance for small samples the following proposal is offered which would be feasible for high speed computers. Let us assume the simplest case, viz., 2×2 factorial with no interaction in the model. We will find a joint confidence region for the two main effects. Assume hypothetical values for the α and β effects; using Berkson's tables we can find the cell probabilities which correspond to the assumed α and β . The probability of observing this sample with the cell probabilities implied by the α and β chosen can be computed. In this way we find a region inside which the probability of observing the sample obtained is greater than .05.

The process is sped up if we imagine α is assigned the horizontal axis and β the vertical axis as illustrated in Figure 6.1. Then let $\beta = 0$ and vary α until the .05 point is found; then let α and β both be positive and equal, and find the .05 point. Continue to select pairs of points until a region is inscribed in the α , β plane. If the inscribed region cuts the α axis we reject $H_0: \alpha = 0$, and if it cuts the β axis we reject $H_0: \beta = 0$.

For an illustration, let us consider a simple example with a sample size of 2 and with 1 success in each cell. We then find the joint 95% confidence region for α and β .

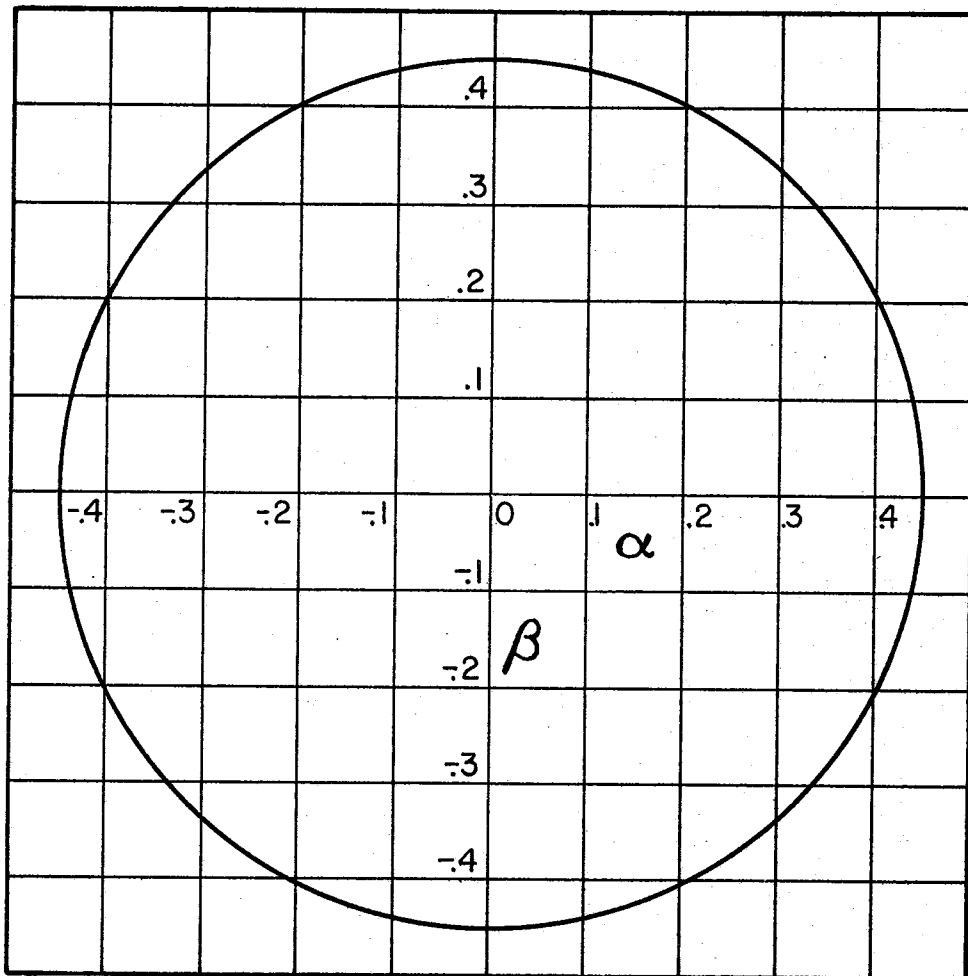


Figure 6.1. 95% CONFIDENCE REGION FOR α AND β

<u>α</u>	<u>β</u>	<u>Probability of Observing</u>
.00	.45	.05
.18	.41	.05
.32	.32	.05
.41	.18	.05
.45	.00	.05
-.18	.41	.05
-.32	.32	.05
-.41	.18	.05
-.45	.00	.05

The second half which is symmetric with the above computations inscribes a circular region in the α , β plane. In general the confidence region for two effects will be an ellipse. This one is circular because the treatment effects are equal. Extensions of this idea could be programmed for high speed computers.

It is obvious from the preceding discussion in this chapter that the small sample problem still remains troublesome. At the present time all that can be done is to make recommendations of a sample size at which the large sample theory holds. This we have done for the special case of a 2x2 factorial arrangement with equal numbers of observations in each cell. There still remains the problem of determining an adequate sample size for larger experiments. This has yet to be covered adequately for the more conventional chi square tests, not to mention the tests newly proposed here and elsewhere.

An enormous amount of computing time could be expended in investigating the effect of varying the parameters which affect the closeness of the chi square approximation. To do this thoroughly one would need to vary the n_{ij} and P_{ij} configurations, and the degrees of freedom. Also various combinations of n_{ij} and P_{ij} need to be examined.

If the problem is approached empirically, it seems that the best approach is to develop some critical parameter such that when this parameter is of the desired magnitude, the chi square approximation is adequate regardless of n_{ij} or P_{ij} configurations or of their interrelations. This we have attempted to do for the case described.

Perhaps with the increasing utilization of high speed computing equipment, small sample techniques such as that proposed above or variations of it, will result in more acceptable tests for small samples. This remains to be seen.

LIST OF REFERENCES

- Aitchison, J., and S. D. Silvey. 1958. Maximum likelihood estimation of parameters subject to restraints. *Ann. Math. Stat.* 29: 813-828.
- Bartlett, M. S. 1935. Contingency table interactions. Supplement to the *Jour. of the Royal Stat. Soc.* 11: 248-252.
- Berkson, J. 1955. Maximum likelihood and minimum chi square estimates of the logistic function. *Jour. of the Amer. Stat. Assoc.* 50: 130-162.
- Berkson, J. 1957. Tables for the maximum likelihood estimates of the logistic function. *Biometrics.* 13: 28-34.
- Bhapkar, V. P. 1959. Contribution to the statistical analysis of experiments with one or more responses (not necessarily normal). North Carolina Institute of Statistics Mimeograph Series No. 229.
- Cochran, W. G. 1940. The analysis of variance when experimental errors follow the poisson or binomial laws. *Ann. of Math. Stat.* 11: 335-347.
- Cochran, W. G. 1943. Analysis of variance for percentages based on unequal numbers. *Jour. of Amer. Stat. Assoc.* 38: 287-301.
- Cochran, W. G. 1952. The chi square test of goodness of fit. *Ann. of Math. Stat.* 23: 315-345.
- Cochran, W. G. 1954. Some methods for strengthening the common chi square tests. *Biometrics.* 10: 417-451.
- Corsten, L. C. A. 1957. Partitions of experimental vectors connected with multinomial distributions. *Biometrics.* 13: 451-484.
- Cramér, H. 1945. *Mathematical methods of statistics.* Princeton University Press, Princeton.
- Diamond, E. L. 1958. Asymptotic power and independence of certain classes of tests on categorical data. North Carolina Institute of Statistics Mimeograph Series No. 196.
- Dyke, G. V., and H. D. Patterson. 1952. Analysis of factorial arrangements when the data are proportions. *Biometrics.* 8: 1-12.
- Eisenhart, C., M. W. Hastay, and W. A. Wallis. 1947. *Techniques of statistical analysis.* McGraw-Hill Book Co. Inc., New York.

LIST OF REFERENCES (continued)

- Finney, D. J. 1952. Statistical methods in biological assay. Hafner Publishing Company, New York.
- Finney, D. J. 1952. Probit analysis. Cambridge University Press, Cambridge.
- Fisher, R. A. 1922. On the dominance ratio. Proceedings of the Royal Society of Edinburgh. 42: 321-341.
- Fisher, R. A. 1954. The analysis of variance with various binomial transformations. Biometrics. 10: 130-139.
- Fix, E. 1949. Tables of noncentral chi square. University of California Publications in Statistics 1: 15-19.
- Hodges, J. L. 1958. Fitting the logistic by maximum likelihood. Biometrics. 14: 453-561.
- Haldane, J. B. S. 1939. The mean and variance of chi square when used as a test of homogeneity when expectations are small. Biometrika. 31: 346-355.
- Irwin, J. O. 1949. A note on the subdivision of chi square into components. Biometrika. 36: 130-134.
- Irwin, J. O. 1950. Biological analysis with special reference to biological standards. Jour. of Hygiene. 48: 215-238.
- Kastenbaum, M. A. and D. E. Lamphiear. 1959. Calculation of chi square to test the no three factor interaction hypothesis. Biometrics. 15: 107-115.
- Landcaster, H. O. 1949. The derivation and partition of chi square in certain discrete distributions. Biometrika. 36: 117-219.
- Landcaster, H. O. 1951. Complex contingency tables treated by the partition of chi square. Jour. of the Royal Stat. Soc. Series B. 13: 242-249.
- Mitra, S. K. 1955. Contributions to the statistical analysis of categorical data. North Carolina Institute of Statistics Mimeograph Series No. 142.
- Neyman, J. 1949. Contributions to the theory of chi square tests. Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability. 239-273.

LIST OF REFERENCES (continued)

- Norton, H. W. 1945. Calculation of chi squared for complex contingency tables. Jour. of Amer. Stat. Assoc. 40: 251-258.
- Peacock, E. E. Jr., B. Brawley and B. G. Greenberg. 1959. Incidence of leukoplakia among cottonmill employees. Unpublished, Departments of Surgery and Biostatistics, Chapel Hill.
- Reiersøl, Olav. 1954. Tests of linear hypotheses concerning binomial experiments. Skandinavisk Aktuaritidskrift. 37: 38-59.
- Roy, S. N., and M. A. Kastenbaum. 1955. A generalization of analysis of variance and multivariate analysis to data based on frequencies in qualitative categories or class intervals. North Carolina Institute of Statistics Mimeograph Series No. 131.
- Roy, S. N., and M. A. Kastenbaum. 1956. On the hypothesis of no "interaction" in a multiway contingency table. Ann. of Math. Stat. 27: 749-757.
- Scarborough, James B. 1958. Numerical mathematical analysis. The Johns Hopkins Press, Baltimore.
- Silvey, S. D. 1959. The lagrangian multiplier test. Ann. of Math. Stat. 30: 389-407.
- Winsor, C. P. 1948. Factorial analysis of a multiple dichotomy. Human Biology. 20: 195-204.