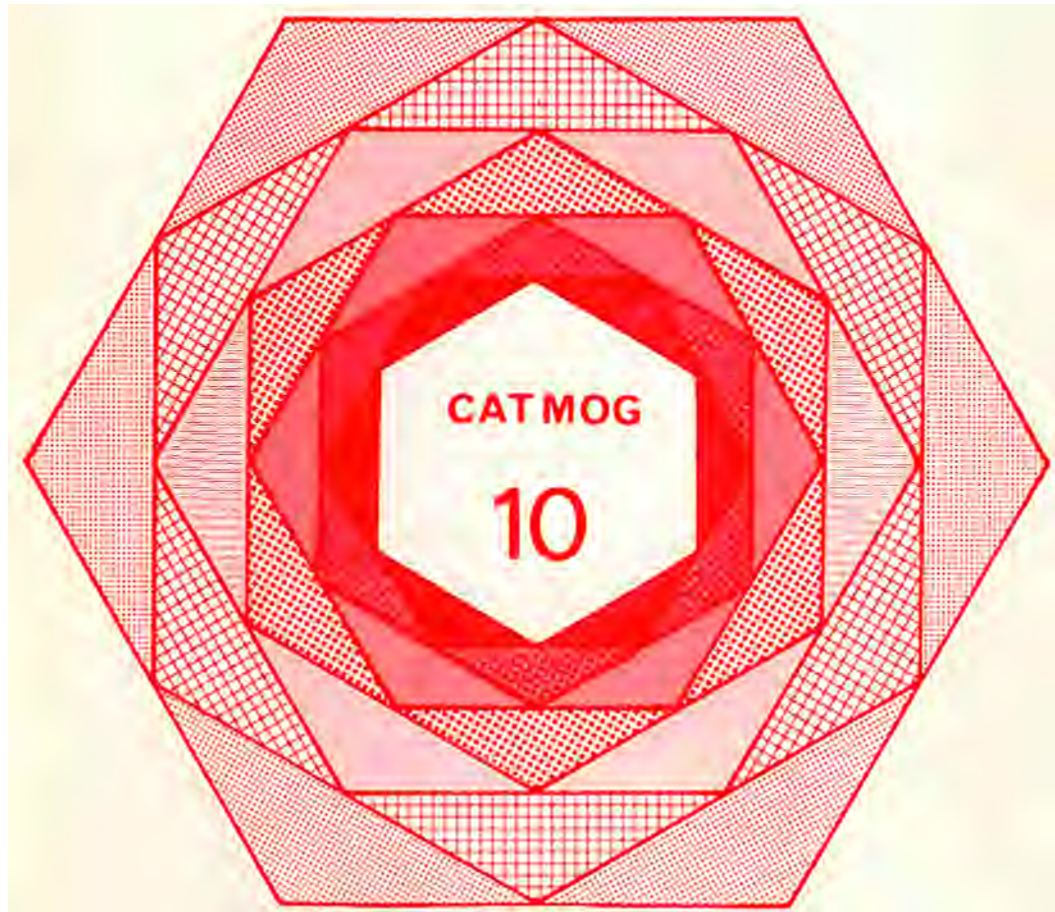
INTRODUCTION TO THE USE OF LOGIT MODELS IN GEOGRAPHY

Neil Wrigley



ISBN 0 902246 62 3

ISSN 0306-6142

CATMOG

(Concepts and Techniques in Modern Geography)

CATMOG has been created to fill a teaching need in the field of quantitative methods in undergraduate geography courses. These texts are admirable guides for the teachers, yet cheap enough for student purchase as the basis of classwork. Each book is written by an author currently working with the technique or concept he describes.

1. Introduction to Markov chain analysis - L. Collins
2. Distance decay in spatial interactions - P.J. Taylor
3. Understanding canonical correlation analysis - D. Clark
4. Some theoretical and applied aspects of spatial interaction shopping models - S. Openshaw
5. An introduction to trend surface analysis - D. Unwin
6. Classification in geography - R.J. Johnston
7. An introduction to factor analytical techniques - J.B. Goddard & A. Kirby
8. Principal components analysis - S. Daultrey
9. Causal inferences from dichotomous variables - N. Davidson
10. Introduction to the use of logit models in geography - N. Wrigley
11. Linear programming: elementary geographical applications of the transportation problem - A. Hay
12. An introduction to quadrat analysis - R.W. Thomas
13. An introduction to time-geography - N.J. Thrift

Other titles are in preparation, and any suggestions as to new titles should be sent to the Editor, Dr P.J. Taylor, Dept of Geography, University of Newcastle upon Tyne. Prospective authors are invited to submit an outline of their proposed text for consideration.

This series, Concepts and Techniques in Modern Geography is produced by the Study Group in Quantitative Methods, of the Institute of British Geographers.
For details of membership of the Study Group, write to the Institute of British Geographers, 1 Kensington Gore, London, S.W.7.

The series is published by Geo Abstracts Ltd., University of East Anglia, Norwich, NR4 7TJ, to whom all other enquiries should be addressed.

CONCEPTS AND TECHNIQUES IN MODERN GEOGRAPHY No. 10

AN INTRODUCTION TO THE USE OF LOGIT MODELS IN. GEOGRAPHY

by

Neil Wrigley

(University of Bristol)

CONTENTS

Acknowledgements

The author is grateful to K. Jones for permission to reproduce the data in Table 1, and also to A. Burn and the Cartographic Unit, University of Southampton.

I INTRODUCTION

(i) Purpose and pre-requisites

Regression models are probably the most widely used statistical models in geographical research. A regression model can be generally defined as a statistical model that embodies assumptions about a dependence relationship between one variable, the response or dependent variable, and one or more other variables, the explanatory or independent variables. Whilst most geographers are familiar with normal-theory regression models in which all variables are continuously distributed, many are less familiar with models in which either the response variable or one or more of the explanatory variables is a categorized variable. The aim of this monograph is to present an introduction to methods suitable for the situation in which the response variable in such a relationship is a categorized variable. The case in which the independent or explanatory variables are categorized or are a mixture of continuously distributed and categorized is discussed in the context of geographical applications by Silk (1976). The monograph assumes that the reader is familiar with normal-theory regression models. At certain points in the text some knowledge of matrix algebra will prove useful but the general reader who has no knowledge of matrix algebra will be able to proceed by simply missing out such sections.

(ii) Levels of measurement

Several geographical text books, for example Harvey (1969, p.308) and King (1969, p.10), consider the question of levels or scales of measurement. Four scales of measurement are generally recognized: the nominal scale, the ordinal scale, the interval scale and the ratio scale.

The nominal scale is the lowest level of measurement and nominal scaling amounts to categorizing objects or events and numbering them. For example, if individuals are classified by sex into two categories, male and female, these categories can be numbered 0 and 1.

At a slightly higher level than the nominal is the ordinal scale. In addition to categorizing objects or events, this scale orders the categories. In other words the objects or events in one category are not just recognized as being different from the objects or events in the other categories but are also recognized as standing in some kind of relation to them. For example, they may be larger, smaller, more preferred, less preferred. Such relationships allow a ranking or ordering of the categories in ascending or descending order, and each category is then assigned a number corresponding to its position in the ranking. For example, individuals may be classified according to socio-economic status and ranked into low, average or high. Numbers can then be assigned, for instance 1 = low, 2 = average, 3 = high, which reflect the ranking. The intervals (distances) which separate categories on the scale are unknown and may not be equal, consequently we cannot perform the standard arithmetic operations on this scale.

Above the ordinal scale are two scales at which measurement considerably stronger than ordinality can be achieved. These are the interval and ratio scales. Both scales have all the characteristics of an ordinal scale and in addition the intervals (distances) between any two numbers on the scale are of known size, thus allowing standard arithmetic operations. The interval scale differs from the ratio scale in having no true zero point as its origin. This can be seen in the case of two well known interval scales, the temperature scales centigrade and Fahrenheit. The zero point in measuring temperature is arbitrary and different for the two scales, however both scales contain the same amount and same kind of information and they are linearly related. Ratio scales on the other hand have true zero points as their natural origins and thus in addition to the properties of the interval scale, ratio scales have the property that the ratio of any two scale points is independent of the unit of measurement. For example, if we used the two ratio scales pounds and grams to measure the weights of two different objects, we would find that the ratio of the two pound weights was indentical to the ratio of the two gram weights.

Although many of the variables used in geographical research, such as height, distance and temperature, can be measured at high levels, at the ratio or interval scales, of equal importance are those which can only be measured at low levels, at the nominal or ordinal scales. Variables measured at the low levels can be termed categorizations of an unordered (nominal scale) or ordered (ordinal scale) nature. Categorizations are particularly important in human geography as a considerable amount of the information used by human geographers is collected by social survey methods, and the typical response to a social survey question is a categorized response. Throughout the monograph the examples used will be taken from human geography, however many of the variables used in physical geography are also categorized variables and consequently the methods to be discussed are generally applicable.

II THE CLASSICAL LINEAR REGRESSION MODEL. A BRIEF REVIEW.

In Section III of the monograph we will see that categorized response variables cannot be analysed satisfactorily using classical normal linear regression models. It is useful at this point therefore to state the assumptions of such models. (In addition see Poole and O'Farrell, 1971). Also, although the monograph assumes that the reader is familiar with such models, this section will serve as a brief summary of their form and will introduce the particular notation used in the rest of the monograph.

For simplicity we will assume that we have a linear regression model with only a single explanatory variable. The model can be written for each of a series of i localities (i=1,...N) as

$$Y_i = \alpha + \beta X_i + \varepsilon_i \qquad (1)$$

Y_i is the response variable, X_i the explanatory variable, ε_i the <u>error term</u> or <u>stochastic disturbance</u> and α and β are the unknown <u>regression parameters</u>. The regression model describes a stochastic relationship between X_i and Y_i and this implies that for every value X_i there is a whole probability distribution of possible values of Y_i. As a result the value of Y_i can never be forecast exactly. This results from the fact that the stochastic disturbance ε_i is random and thus imparts randomness to Y_i. In the linear regression model (1) therefore, the probability distribution of Y_i is determined by the value of X_i and by the distribution of ε_i. Because of this relationship a full specification of a regression model must state not only the form of the regression equation but also a set of assumptions concerning the probability distribution of ε_i. A particularly important specification is that which is termed the <u>classical normal linear regression model</u> and many geographical applications of regression techniques explicitly or implicitly assume such a specification.

The full specification of the classical normal linear regression model consists of a set of seven basic assumptions.

(A1) The linearity assumption. The regression equation should be <u>linear in the unknown parameters.</u> This implies that not only do simple straight line relationships of the form (1) meet the linearity condition but also curved line relationships such as

$$Y_i = \alpha + \beta X_i^2 + \varepsilon_i \qquad \text{or} \qquad Y_i = \alpha + \beta \frac{1}{X_i} + \varepsilon_i \qquad (2)$$

which being nonlinear in the variables but linear in the parameters are termed <u>intrinsically linear</u>. The intrinsic linearity of such models can be seen by simply redefining the explanatory variables. In the case of the models in (2), redefining $X_i^2 = V_i$ and $1/X_i = Z_i$, the models then take the same simple linear form as (1). Models which do not meet the linearity condition are those which are nonlinear in the parameters, for example

$$Y_i = \alpha X_i^\beta + \varepsilon_i \qquad (3)$$

(A2) The assumption that the values of the explanatory variables in the regression model can be measured without error.

(A3) The assumption that the values of the explanatory variable(s) are fixed or nonstochastic. (This assumption can easily be relaxed and the explanatory variable(s) allowed to be random. In this case the model simply requires a <u>conditional</u> formulation in which we concentrate on the conditional distribution of Y_i given observed values of the random explanatory variable(s).)

(A4) The zero error mean assumption. This is written $E(\varepsilon_i) = 0$. ($E(\varepsilon_i)$ is known as the expectation or expected value of ε_i, it is simply the mean of the probability distribution of possible values of ε_i.)

(A5) The constant error variance or homoscedasticity assumption. This is written $E(\varepsilon_i^2) = Var(\varepsilon_i) = \sigma^2$.

(A6) The independent error terms assumption. This is written $E(\varepsilon_i \varepsilon_j) = 0$, for $i \neq j$. In the case where i and j refer to geographical localities this assumption is known as the spatially independent error terms or no spatial autocorrelation assumption.

(A7) The assumption that ε_i is normally distributed.

Having stated the seven basic assumptions of the classical normal linear regression model we can now see that by taking expectations of both sides of (1) it follows, because of assumptions (A3) and (A4) that

$$E(Y_i) = E(\alpha + \beta X_i + \varepsilon_i) = \alpha + \beta X_i \qquad (4)$$

(By assumption (A3) X_i is a fixed value, thus $E(X_i) = X_i$.) This relationship (4) which gives the mean value of Y_i for the value of X_i at each of the i localities defines what is known as the population regression line. It is represented in Figure 1 by the dashed line. The regression parameters, a the intercept and β the slope, which define this line are however unknown and we must estimate their values from a sample of observations. When α and β are estimated from a sample we write them as $\hat{\alpha}$ and $\hat{\beta}$, and on the basis of their values we can determine the sample regression line, represented in Figure 1 by the unbroken line which is defined by the equation

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i \qquad (5)$$

across the i localities, where $\hat{Y}_i$ is termed the predicted or fitted value of Y_i. Most of the observed values of Y_i will not lie exactly on the sample regression line so the values of Y_i and $\hat{Y}_i$ will differ. This difference is called a residual and is designated e_i.

$$Y_i - \hat{Y}_i = Y_i - (\hat{\alpha} + \hat{\beta} X_i) = e_i \qquad (6)$$

In general, because $\hat{\alpha}$ and $\hat{\beta}$ are likely to differ from the true values of α and β, the residual e_i is different from the stochastic disturbance or error term ε_i, for ε_i is given by the relationship

$$Y_i - E(Y_i) = Y_i - (\alpha + \beta X_i) = \varepsilon_i \qquad (7)$$

ε_i is a population term and cannot be observed. The value of the residual e_i can thus be regarded as a sample estimate of ε_i.

The method of estimating the parameters of a linear regression model most familiar to geographers is the method of least squares. (See for example King 1969, p.121). Under the assumptions of the classical normal linear regression model outlined above, the ordinary least squares (O.L.S.) estimators $\hat{\alpha}$ and $\hat{\beta}$ can be shown to be what are termed the best linear unbiased estimators (BLUE's). Best linear unbiased estimators have a number of properties which intuitively we would like a "good" estimator to possess. (See the discussion by Kmenta 1971 p. 154-93, Huang 1970 p.26-32, Wonnacott and Wonnacott 1970 p. 40-47). They are unbiased; in other words each estimator has a sampling distribution with a mean equal to the parameter to be estimated. Each estimator also has a variance which is smaller than that of any other linear unbiased estimator (best linear unbiasedness).

Having discussed the assumptions of the classical normal linear regression model and the properties of the O.L.S. estimators under these assumptions, we are now in a position to consider the implications of having a categorized response variable. Instead of having a continuously distributed response variable Y_i which can take a whole range of possible values, we will now move on to consider the situation in which we have a categorized response variable Y_i which can take only a limited number of possible values.

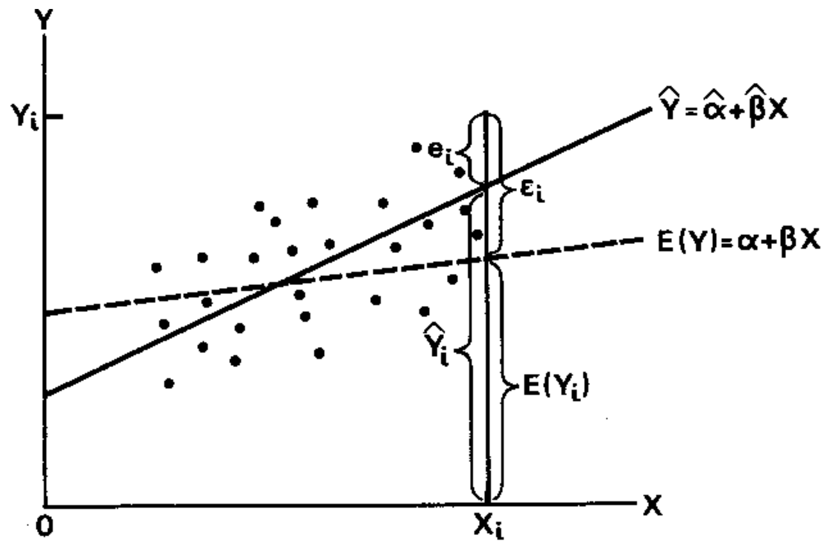


Fig. 1 Population and sample regression lines.

III TWO POSSIBLE RESPONSE CATEGORIES. THE DICHOTOMOUS CASE.

(i) Introduction

The simplest categorized response variable is a random variable with only two possible outcomes. In this section we will examine this simplest case. To make the discussion more readily comprehensible a specific example will be taken. In this example the categorized response variable measures the presence or absence of chronic bronchitis in adult males. We wish to attempt an explanation of the spatial incidence of chronic bronchitis in a particular city using a regression model with two explanatory variables, the amount of air pollution as measured by the smoke levels in the localities where each of the adult males live, and their cigarette consumption.

(ii) Problems associated with the classical linear regression model when the response variable is categorized

If we code the two possible outcomes of the categorized response variable 1 and 0, 1 representing presence of chronic bronchitis, 0 absence, and try to use such a response variable in a classical linear regression model with two explanatory variables X_{i1} and X_{i2}, of the form

$$Y_i = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i \qquad (8)$$

we will encounter a number of problems.

The first problem we will encounter concerns the violation of the constant error variance assumption (A5). This follows from the fact that the error term

$$\varepsilon_i = Y_i - (\alpha + \beta_1 X_{i1} + \beta_2 X_{i2}) \qquad (9)$$

can in this case only have one of two possible values:

$$\varepsilon_i = \begin{cases} 1 - (\alpha + \beta_1 X_{i1} + \beta_2 X_{i2}) & \text{if } Y_i = 1 \\ -(\alpha + \beta_1 X_{i1} + \beta_2 X_{i2}) & \text{if } Y_i = 0 \end{cases} \qquad (10)$$

These two possible values of ε_i must occur with probabilities P_i and $1-P_i$ respectively because of the binomial distribution of the response variable. Thus the classical assumption (A4) $(E(\varepsilon_i) = 0)$, implies

$$P_i[1 - (\alpha + \beta_1 X_{i1} + \beta_2 X_{i2})] + (1-P_i)[-(\alpha + \beta_1 X_{i1} + \beta_2 X_{i2})] = 0 \qquad (11)$$

and thus

$$P_i = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} \qquad (12)$$

$$1-P_i = 1 - (\alpha + \beta_1 X_{i1} + \beta_2 X_{i2}) \qquad (13)$$

The error variance can therefore be written as

$$\begin{aligned} E(\varepsilon_i^2) &= P_i[1-(\alpha+\beta_1 X_{i1}+\beta_2 X_{i2})]^2 + (1-P_i)[-(\alpha+\beta_1 X_{i1}+\beta_2 X_{i2})]^2 \\ &= (\alpha + \beta_1 X_{i1} + \beta_2 X_{i2})[1 - (\alpha + \beta_1 X_{i1} + \beta_2 X_{i2})] \end{aligned} \qquad (14)$$

Clearly the error variance is not a constant but depends upon the values of the explanatory variables. The classical assumption of constant error variance or homoscedasticity is therefore violated. Consequently if ordinary least squares estimation of the unknown parameters is used, then the O.L.S. estimators $\hat{\alpha}$, $\hat{\beta}_1$ and $\hat{\beta}_2$ of the parameters, although unbiased are not 'best' in the sense of being the minimum variance estimators of all linear unbiased estimators of α, β_1 and β_2.

The second problem we will encounter concerns the predicted values which are generated using the regression model (8). Since Y_i can only assume two different values, 1 and 0, $E(Y_i)$ the expected value of Y_i is a simple weighted average of the two possible values of Y_i with weights given by the respective probabilities of occurrence of the possible values. That is to say

$$E(Y_i) = [1 \times P_i] + [0 \times (1-P_i)] = P_i \qquad (15)$$

where P_i is the probability that Y_i=1 or in other words that the individual at locality i is a chronic bronchitic. By taking expectations of both sides

of (8) and using (15) we then have

$$P_i = E(Y_i) = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} \qquad (16)$$

In other words, in the case of our example it is useful and reasonable to interpret the expectation of Y_i given X_{i1} and X_{i2} as the probability of being a chronic bronchitic given X_{i1} and X_{i2}. The problem with this interpretation however concerns the predicted values

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} \qquad (17)$$

generated from model (8). In view of the probability interpretation of $E(Y_i)$ these predicted values are interpreted as predicted probabilities, i.e. $\hat{Y}_i = \hat{P}_i$. However, whereas probability is defined to lie between 0 and 1, the predictions (17) generated using (8) are unbounded and may take values from $-\infty$ to $+\infty$. Consequently the predictions may lie outside the meaningful range of probability and thus be inconsistent with the probability interpretation advanced.

(iii) Alternative solutions

As a means of analysing our categorized response variable the classical normal linear regression model has thus been shown to be seriously deficient. We will now consider alternative models which overcome these problems.

In order to allow the probability interpretation discussed above it is necessary that the condition

$$0 \leq \hat{P}_i \leq 1 \qquad (18)$$

is satisfied by the predictions of the regression model. In the case of equation (17), the simplest way to satisfy this condition is to impose the following arbitrary definition of $\hat{P}_i$

$$\hat{P}_i = \begin{cases} 0 & \text{if } \hat{\alpha} + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} < 0 \\ \hat{Y}_i & \text{if } 0 \leq \hat{\alpha} + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} \leq 1 \\ 1 & \text{if } 1 < \hat{\alpha} + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} \end{cases} \qquad (19)$$

and to use straightforward ordinary least squares estimation of the regression parameters. This solution is often referred to as the linear probability model. Unfortunately although it appears to be a simple solution to the pretícted probabilities problem, the model has a number of serious limitations which are discussed by Domencich and McFadden (1975, p. 102-108, 119). An important problem occurs for example if a number of our sample localities have associated values of the explanatory variables X_{i1} and X_{i2} drawn from the ranges of the explanatory variables where the probabilities take the extreme values. In this case the linear probability model will produce parameter estimates which are substantially biased below their true values. Moreover the linear probability model does not satisfactorily extend to the important multiple response category case which we will go on to discuss in Section IV. In the multiple response category case Domencich and McFadden (p. 119) conclude that the linear probability model 'does not yield a practical estimator with satisfactory statistical properties'.

As a result of the limitations of the linear probability model we must seek a more general method of satisfying condition (18). Perhaps the simplest model in which condition (18) is automatically satisfied is

$$P_i = \frac{e^{\alpha+\beta_1 X_{i1}+\beta_2 X_{i2}}}{1+e^{\alpha+\beta_1 X_{i1}+\beta_2 X_{i2}}} \quad (20)$$

$$1-P_i = \frac{1}{1+e^{\alpha+\beta_1 X_{i1}+\beta_2 X_{i2}}} \quad (21)$$

Although (20) and (21) are nonlinear models it is a simple matter to rewrite them as

$$\frac{P_i}{1-P_i} = e^{\alpha+\beta_1 X_{i1}+\beta_2 X_{i2}} \quad (22)$$

Then remembering that by the definition of a logarithm if $y=e^x$, $x=\log_e y$ (where $\log_e$, log to the base e, denotes the natural or Naperian logarithm) we have

$$\log_e \frac{P_i}{1-P_i} = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} \quad (23)$$

We have thus achieved a linear model in which the left hand side is a transformation of P_i. This transformation has the property of increasing from $-\infty$ to $+\infty$ as P_i increases from 0 to 1. It is called the logit transformation. Defining

$$L_i = \log_e \frac{P_i}{1-P_i} \quad (24)$$

and substituting into (23) we have what is termed the linear logit model for the individual at locality i.

The predictions of the logit model

$$\hat{L}_i = \hat{\alpha} + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} \quad (25)$$

can take any values in the range $-\infty$ to $+\infty$ but the predicted probabilities which can be found by substituting $\hat{\alpha}$, $\hat{\beta}_1$ and $\hat{\beta}_2$ of (25) into (20) are confined to the range $0 \leq \hat{P}_i \leq 1$ thus satisfying condition (18). The linear logit model or its nonlinear alternative (20) which we call the nonlinear logit model thus overcome the second of the problems we identified in Section III (ii) and represent feasible alternative regression models.

Another model in which condition (18) is also automatically satisfied is

$$P_i = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{t_i-5} e^{-\frac{1}{2}u^2} du \quad (26)$$

($\int$ is the integral sign). Just as the logit transformation was derived from (20) so in the case of (26) another transformation called the probit transformation can be derived. Unfortunately the probit transformation has a rather more complex definition. The probit of P_i in (26) is defined (see Finney, 1971) as t_i, where t_i equals the normal equivalent deviate (NED) such that a proportion P_i of the standard normal distribution falls to the left of NED, plus five to avoid negative values. That is

$$\text{Probit}(P_i) = t_i = \text{NED} + 5 \quad (27)$$

Another way to express this definition is to say that the probit of P_i is the abscissa which corresponds to a probability P_i in a normal distribution with mean 5 and variance 1. Although this definition is rather complex, in practice probits for specified values of P_i are available directly from tables, see Fisher and Yates, 1963. (Similar tables giving logits for specified values of P_i are also available, see Berkson, 1953). Using the definition of the probit, a probit regression model for the individual at locality i in the case of our example can then be written as

$$\text{Probit}(P_i) = t_i = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} \quad (28)$$

Although there are other models which automatically satisfy condition (18) and from which other transformations with similar properties to the logit and probit can be derived, the logit and probit are by far the most widely used. Unless the probability P_i is very close to 1 or 0 both transformations are virtually equivalent. As a result the choice between logit and probit models as alternative regression models more suited to the categorized response variable situation is essentially a matter of computational convenience. Logit models particularly those suited to the analysis of multiple response categories are without doubt computationally more convenient than the the equivalent probit models and consequently the majority of recently published analyses of categorized response variables in related disciplines adopt logit models in preference to probit models. In the rest of the monograph we will likewise adopt logit models as the alternative regression models we have been seeking.

(iv) Estimating the parameters of a linear logit model

Having thus selected logit models as regression models which are more suited to the categorized response variable situation, how do we then estimate the parameters of these models? The parameters of the nonlinear logit model can be directly estimated using the method of maximum likelihood. However, as this method of estimation is less familiar to most geographers than the method of least squares, discussion of it and of the nonlinear logit model will be deferred until the final section of the monograph. Instead we will concentrate upon the linear logit model and a least squares method of estimation.

We begin our estimation of the parameters of the linear logit model by first replacing the unobservable probability values on the left hand side of (23) with observable relative frequencies. In order to derive these relative frequencies it is necessary to group our sample of observations into K sets, with as near as possible a constant probability of being a chronic bronchitic within each set. These sets are determined on the basis of characteristic values of the explanatory variables. An example of this procedure is to be found in the next sub-section (III,v). For each of these sets (j=1,...K) we now replace the probability value P_j in the left hand side of the linear logit model (23) by the observed relative frequency f_j of the incidence of chronic bronchitis in set j. The model (23) can thus be rewritten as

$$\bar{L}_j = \log_e \frac{f_j}{1-f_j} = \alpha + \beta_1 X_{j1} + \beta_2 X_{j2} + (\bar{L}_j - L_j) \qquad (29)$$

The term $(\bar{L}_j - L_j)$ on the right-hand side of (29) is introduced to take account of the fact that the left-hand side now contains only an estimate of the true logit, an estimate based upon relative frequencies. Grouping of individual observations into sets is necessary because the transformed variable $\log_e P_i/1-P_i$ on the left hand side of (23) has by definition the value infinity whenever $P_i=1$ or 0. If we simply replace P_i by the observed value at locality i, in other words by 1 or 0, the transformation will thus always by definition be infinity unless grouping is undertaken.

Although the linear logit model satisfies condition (18) allowing the probability interpretation, the first of our original problems has not been solved. The error variances are not constant across the j sets of (29). Theil (1970 , p. 137) has shown that these error variances take the asymptotic form

$$\frac{1}{n_j P_j (1-P_j)} \qquad (30)$$

(where asymptotic means as the sample size gets larger and larger and approaches infinity'). This implies that the error variance may vary from set to set. Knowledge of the nature of the heteroscedasticity (the unequal nature of the error variances) can however be usefully exploited to solve our problem.

We can replace the ordinary least squares estimators of α, β_1 and β_2 by another set of least squares estimators known as weighted least squares - estimators, which effectively exploit the knowledge we have of the nature of the heteroscedasticity. In this way we are able to derive estimators which have all the properties we normally value. These weighted least squares estimators which will be discussed below use as weights terms of the form

$$n_j f_j (1-f_j) \qquad (31)$$

Weights of this form imply that as the number of individuals n_j in a set increases, more weight is allocated to that set in the estimation procedure. Given n_j on the other hand, as f_j approaches 0 or 1 less weight is allocated

for in these situations the logit $\bar{L}_j$ (see equation 29) takes large negative or positive values and is thus highly sensitive to small changes in f_j. Little weight is accordingly allocated to these highly unstable observations. In the case where f_j is either 0 or 1 the weight (31) is thus zero for in this case the logit becomes infinitely large and cannot be handled. The system of weights thus effectively excludes a set j in which the observed relative frequency is 0 or 1. Berkson (1953, 1955) has suggested however that such an exclusion represents an unwarranted waste of information. He advocates that we use instead replacement working values of the following form.

$$1/2n_j \quad \text{instead of 0 when } f_j = 0.$$

$$1 - 1/2n_j \quad \text{instead of 1 when } f_j = 1.$$

In ordinary least squares estimation of the parameters of a linear regression model (8) with two explanatory variables, a set of three simultaneous equations known as the normal equations are derived. These normal equations are shown below, they contain three unknown terms $\hat{\alpha}$, $\hat{\beta}_1$ and $\hat{\beta}_2$.

$$\begin{aligned}
\hat{\alpha}N + \hat{\beta}_1 \sum_i X_{i1} + \hat{\beta}_2 \sum_i X_{i2} &= \sum_i Y_i \\
\hat{\alpha}\sum_i X_{i1} + \hat{\beta}_1 \sum_i X_{i1}^2 + \hat{\beta}_2 \sum_i X_{i1}X_{i2} &= \sum_i Y_i X_{i1} \qquad (32) \\
\hat{\alpha}\sum_i X_{i2} + \hat{\beta}_1 \sum_i X_{i1}X_{i2} + \hat{\beta}_2 \sum_i X_{i2}^2 &= \sum_i Y_i X_{i2}
\end{aligned}$$

Given a set of data, solving the three equations for the three unknown terms produces the required estimates of the regression parameters.
Weighted least squares estimation of the model (29) also produces a set of three normal equations with three unknown terms similar to those of (32). However, these normal equations now contain weights of the type discussed above. Writing for simplicity the weights as

$$n_j f_j (1-f_j) = w_j \qquad (33)$$

then the modified normal equations are shown below:

$$\begin{aligned}
\hat{\alpha}\sum_j w_j + \hat{\beta}_1 \sum_j w_j X_{j1} + \hat{\beta}_2 \sum_j w_j X_{j2} &= \sum_j w_j \bar{L}_j \\
\hat{\alpha}\sum_j w_j X_{j1} + \hat{\beta}_1 \sum_j w_j X_{j1}^2 + \hat{\beta}_2 \sum_j w_j X_{j1} X_{j2} &= \sum_j w_j \bar{L}_j X_{j1} \qquad (34) \\
\hat{\alpha}\sum_j w_j X_{j2} + \hat{\beta}_1 \sum_j w_j X_{j1} X_{j2} + \hat{\beta}_2 \sum_j w_j X_{j2}^2 &= \sum_j w_j \bar{L}_j X_{j2}
\end{aligned}$$

Solving for the three unknown terms produces the required weighted least squares estimators.

Respondent	Chronic Bronchitis	Cigarette Consumption	Smoke Level
1	N	5.15	67.1
2	Y	0.00	66.9
3	N	2.50	66.7
4	N	1.75	65.8
5	N	6.75	64.4
6	N	0.00	64.4
7	Y	0.00	65.1
8	V	9.50	66.2
9	N	0.00	65.9
10	N	0.75	67.1
11	N	5.25	67.9
12	Y	8.00	68.1
13	Y	5.15	67.0
14	Y	30.00	66.3
15	N	0.00	65.7
16	N	0.00	65.2
17	N	5.25	64.2
18	N	10.05	64.6
19	N	0.00	63.5
20	V	3.40	63.0
21	N	0.00	62.7
?2	N	0.55	62.7
23	Y	9.50	62.1
24	Y	12.50	63.7
25	N	0.00	63.1
26	N	3.40	63.0
27	N	2.20	62.7
28	N	6.70	63.1
29	N	1.10	62.4
30	N	1.80	64.4
31	N	0.00	64.2
32	Y	3.60	64.2
33	N	1.60	63.0
34	N	6.20	62.2
35	N	14.75	62.3
36	N	0.35	63.7
37	Y	13.75	63.8
38	N	0.00	63.1
39	V	7.50	62.7
40	N	1.00	62.9
41	N	0.00	62.5
42	Y	14.80	61.7
43	Y	3.50	61.6
44	N	0.00	61.6
45	N	0.00	61.4
46	N	0.25	61.4
47	N	1.55	62.0
48	V	0.00	61.8
49	N	0.00	60.9
50	N	5.90	60.8
51	N	16.45	60.6
52	N	2.65	62.9
53	Y	12.50	62.6
54	N	0.00	62.1
55	N	14.55	61.7
56	Y	11.00	61.0
57	Y	6.75	62.7
58	N	0.00	62.7
59	r	0.00	61.7
60	N	1.75	60.9
61	N	2.40	60.6
62	N	10.05	60.4
63	Y	12.75	61.7
64	N	0.00	61.9
65	N	5.00	61.3
66	N	0.60	60.7
67	N	0.00	60.8
68	N	0.85	60.5
69	N	0.90	59.7
70	N	0.00	59.5
71	Y	8.75	59.6
72	N	0.80	59.1
73	Y	6.60	59.4
74	N	1.00	58.5
75	N	0.00	60.0
76	Y	8.15	59.8
77	N	0.00	59.7
78	Y	5.00	59.4
79	N	2.55	59.2
80	N	1.20	58.6
81	N	0.00	60.8
82	Y	11.25	60.4
83	N	0.00	60.2
84	N	2.00	60.0
85	N	1.90	59.4
86	N	0.45	59.8
87	Y	0.00	59.7
88	N	0.00	59.0
89	Y	6.90	59.0
90	N	2.35	58.6
91	N	3.95	59.7
92	N	0.60	59.6
93	Y	15.00	59.4
94	N	0.00	59.4
95	N	0.95	59.4
96	N	0.00	59.3
97	N	1.40	54.2
98	N	0.50	54.0
99	N	0.60	53.8
100	N	0.00	53.7
101	N	2.45	53.7
102	N	1.75	53.1
103	N	0.00	54.4
104	N	3.10	54.2
105	N	10.05	53.9
106	N	0.55	53.2
107	N	0.85	53.2
108	N	1.10	54.9
109	N	0.00	54.5
110	N	0.00	54.5
111	N	1.45	54.2
112	N	2.05	54.2
113	Y	10.50	54.0
114	N	0.50	55.8
115	Y	9.20	55.5
116	N	0.55	55.6
117	N	0.00	55.5
118	N	0.96	54.9
119	N	1.00	54.6
120	N	0.00	56.9
121	N	5.25	56.4
122	V	0.00	55.9
123	N	9.00	55.8
124	N	1.60	55.6
125	V	10.90	57.6
126	N	0.00	57.7
127	N	0.00	57.6
128	N	2.25	57.8
129	N	2.65	57.8
130	N	0.55	58.4
131	N	0.00	58.2
132	Y	4.50	58.0
133	N	15.00	58.1
134	N	0.00	57.9
135	N	0.00	57.3
136	N	4.20	58.3
137	N	0.55	58.1
138	Y	10.00	57.9
139	N	0.00	57.6
140	N	7.10	57.3
141	N	3.20	57.1
142	Y	0.00	58.9
143	r	6.80	58.6
144	N	0.00	58.7
145	N	0.00	57.5
146	N	2.35	57.2
147	N	24.90	58.0
148	N	2.65	57.9
149	Y	3.70	57.2
150	N	17.10	57.3
151	N	0.00	57.5
152	N	0.95	57.2
153	N	10.05	53.1
154	N	1.15	53.0
155	V	18.25	53.0
156	N	10.00	52.9
157	N	0.75	52.6
158	N	0.00	53.1
159	N	4.20	53.0
160	N	0.80	52.9
161	N	0.55	52.7
162	N	0.95	52.6
163	N	0.00	52.1
164	N	3.10	54.1
165	N	0.80	53.7
166	N	1.55	53.1
167	N	0.40	53.3
168	N	6.20	53.0
169	N	0.60	53.0
170	N	0.40	53.9
171	Y	7.50	53.7
172	N	7.15	53.4
173	N	0.25	53.2
174	N	3.60	53.4
175	N	0.95	53.2
176	N	2.80	54.9
177	V	20.25	54.9
178	N	0.95	54.6
179	N	4.25	54.1
180	N	4.15	54.2
181	N	10.00	57.4
182	N	3.40	57.3
183	N	0.00	57.3
184	N	3.60	56.7
185	N	0.90	56.5
186	N	0.00	56.8
187	N	0.00	56.6
188	r	6.40	56.5
189	N	0.95	56.3
190	N	1.06	56.3
191	N	13.30	56.2
192	N	1.10	56.6
193	N	17.20	55.9
194	N	1.65	56.0
195	V	5.00	55.8
196	N	2.10	55.7
197	N	0.60	57.0
198	V	8.25	56.7
199	N	0.90	56.4
200	N	0.00	56.5
201	V	12.30	55.2
202	N	1.15	56.9
203	N	2.20	56.7
204	N	3.60	56.0
205	V	10.00	55.5
206	N	0.60	55.3
207	N	9.50	56.5
208	N	0.70	56.3
209	V	9.00	56.1
210	N	0.00	55.9
211	N	0.50	55.5
212	N	0.90	55.4

Table 1

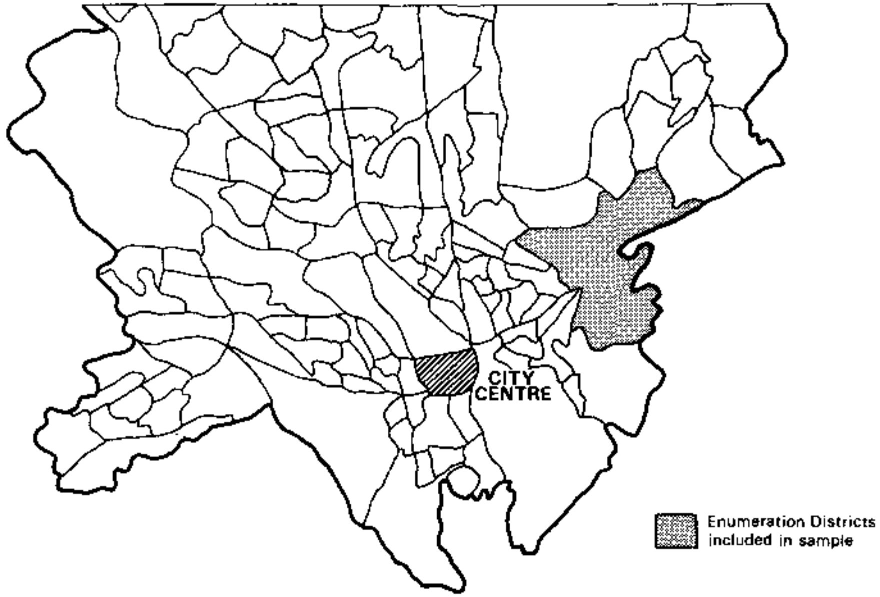


Fig. 2 Cardiff Enumeration Districts and Sample Area

(v) An example. Chronic bronchitis in an area of Cardiff

Having discussed how to estimate the parameters of a linear logit model suitable for our categorized response variable we will now work through the estimation procedure using a real set of data. The data reproduced in Table 1 give the presence or absence of chronic bronchitis in a sample of 212 adult males in the small area of Cardiff depicted in Figure 2. Also given are smoke level values ($\mu g/m^3$) for the locality where each respondent lives and the respondent's cigarette consumption. These figures are reproduced from a much larger study of chronic bronchitis in Cardiff conducted in 1974 by Jones*. The smoke level figures at each sample site were interpolated from 13 air pollution monitoring stations in the city. The presence or absence of chronic bronchitis in the sample was determined using a special form of questionnaire devised for use in the investigation of chronic bronchitis by the Medical Research Council (1962). In tests, detection of chronic bronchitis using this form of questionnaire have been found to be almost completely consistent with clinical diagnosis (McNab et al.1966). In the same questionnaire details of the respondent's smoking history were also collected and information on a set of other variables not reported here. The cigarette consumption values recorded in Table 1 were determined from the average number of cigarettes smoked per day and the number of years a respondent had smoked, expressed in hundreds of cigarettes. Consumption of pipe tobacco was converted into cigarette equivalents. The survey was

*Jones, K. (1975) ' A geographical contribution to the aetiology of chronic bronchitis', Unpublished B.Sc. dissertation, University of Southampton

Smoke Level \ Cigarette Consumption	≦0-0.75	0.76-3.75	3.76-6.75	6.76-10.74	≧10.75
≧61.0	$j=1$ $n_j=23$ $cb=4$ $X_{j1}=0.0826$ $X_{j2}=63.4826$	$j=2$ $n_j=13$ $cb=3$ $X_{j1}=2.3115$ $X_{j2}=63.4308$	$j=3$ $n_j=9$ $cb=2$ $X_{j1}=5.8000$ $X_{j2}=64.4333$	$j=4$ $n_j=5$ $cb=4$ $X_{j1}=8.9100$ $X_{j2}=64.7400$	$j=5$ $n_j=9$ $cb=7$ $X_{j1}=15.1778$ $X_{j2}=62.8222$
56.0 - 60.9	$j=6$ $n_j=33$ $cb=2$ $X_{j1}=0.1227$ $X_{j2}=58.5545$	$j=7$ $n_j=30$ $cb=1$ $X_{j1}=1.8970$ $X_{j2}=57.9467$	$j=8$ $n_j=8$ $cb=4$ $X_{j1}=5.2250$ $X_{j2}=58.5625$	$j=9$ $n_j=11$ $cb=7$ $X_{j1}=8.5909$ $X_{j2}=58.1182$	$j=10$ $n_j=8$ $cb=3$ $X_{j1}=15.4875$ $X_{j2}=58.4500$
≦55.9	$j=11$ $n_j=22$ $cb=1$ $X_{j1}=0.3068$ $X_{j2}=54.1591$	$j=12$ $n_j=23$ $cb=0$ $X_{j1}=1.6243$ $X_{j2}=54.0609$	$j=13$ $n_j=5$ $cb=1$ $X_{j1}=4.7600$ $X_{j2}=54.0200$	$j=14$ $n_j=9$ $cb=4$ $X_{j1}=9.2722$ $X_{j2}=54.2000$	$j=15$ $n_j=4$ $cb=3$ $X_{j1}=17.0000$ $X_{j2}=54.7500$

Table 2

SETS	w_j	$\bar{L}_j$
j = 1	3.3042	-1.5583
2	2.3079	-1.2036
3	1.5554	-1.2528
4	0.8000	1.3863
5	1.5554	1.2529
6	1.8786	-2.7411
7	0.9657	-3.3697
8	2.0000	0.0000
9	2.5453	0.5598
10	1.8750	-0.5108
11	0.9554	-3.0428
12	0.4883	-3.8077
13	0.8000	-1.3863
14	2.2222	-0.2233
15	0.7500	1.0986

SUMS	
$\sum_j w_j$	= 24.0034
$\sum_j w_j \bar{L}_j$	= -20.3017
$\sum_j w_j \bar{L}_j X_{j1}$	= 12.1292
$\sum_j w_j \bar{L}_j X_{j2}$	= -1197.3087
$\sum_j w_j X_{j1}$	= 147.0314
$\sum_j w_j X_{j2}$	= 1429.3462
$\sum_j w_j X_{j1}^2$	= 1609.5028
$\sum_j w_j X_{j2}^2$	= 85437.008
$\sum_j w_j X_{j1} X_{j2}$	= 8640.3761

Table 3

restricted to men over fifteen years of age who had lived at their location at the time of the survey for at least ten years. In the larger survey from which this data set is taken a 94.5% response rate was achieved.

As stated in the introduction to this section we now wish to attempt an explanation of the determinants of chronic bronchitis using a linear logit model of the form (29), and a weighted least squares estimation procedure. Our first task is to group our sample of respondents into sets on the basis of characteristic values of the two explanatory variables. Table 1 reveals that the range of smoke level values is from 52.1 to 68.1. The cigarette consumption values range from 0.00 to 30.00. By dividing the smoke level observations into 3 groups and the cigarette consumption observations into 5, we can next derive 3x5 = 15 sets. The total number of respondents in each set, the number with chronic bronchitis (cb) and the average values of the two explanatory variables cigarette consumption (X_{j1}) and smoke level (X_{j2}) in each set j are shown in Table 2. Table 2 is then used to produce the information required to solve the normal equations (34). This information is given in Table 3. Substituting the values given in Table 3 into the normal equations (34) we get

$$\hat{\alpha}24.0034 + \hat{\beta}_1 147.0314 + \hat{\beta}_2 1429.3462 = -20.3017 \quad (35a)$$

$$\hat{\alpha}147.0314 + \hat{\beta}_1 1609.5028 + \hat{\beta}_2 8640.3761 = 12.1292 \quad (35b)$$

$$\hat{\alpha}1429.3462 + \hat{\beta}_1 8640.3761 + \hat{\beta}_2 85437.008 = -1197.3087 \quad (35c)$$

The values of the unknown terms $\hat{\alpha}$, $\hat{\beta}_1$, and $\hat{\beta}_2$ which satisfy these three simultaneous equations can be found by following the detailed example given by Unwin (1975, p. 10) in an earlier CATMOG. They are found to be

$$\hat{\alpha} = -8.74264 \qquad \hat{\beta}_1 = 0.21054 \qquad \hat{\beta}_2 = 0.11096$$

As would be expected these parameter estimates suggest that both explanatory variables, cigarette consumption and smoke level are positively related to the incidence of chronic bronchitis in this area of Cardiff.

Having derived the parameter estimates we can now determine what our model predicts the probability of chronic bronchitis to be in each of the sets, using the following equation:

$$\hat{P}_j = \frac{e^{\hat{L}_j}}{1+e^{\hat{L}_j}} \quad (36)$$

where

$$\hat{L}_j = \hat{\alpha} + \hat{\beta}_1 X_{j1} + \hat{\beta}_2 X_{j2} \quad (37)$$

We can also test 'the validity of our particular model using a test statistic based upon the weighted discrepancies between the observed logit $\bar{L}_j$ based upon relative frequencies, and $\hat{L}_j$ the predicted logit. The form of this test statistic is

$$\sum_j (\bar{L}_j - \hat{L}_j)^2 w_j \quad (38)$$

SETS	$\bar{L}_j$	$\hat{L}_j$	$(\bar{L}_j-\hat{L}_j)^2 w_j$
j = 1	-1.5583	-1.6814	0.0501
2	-1.2036	-1.2179	0.0005
3	-1.2528	-0.3722	1.2061
4	1.3863	0.3166	0.9154
5	1.2529	1.4234	0.0452
6	-2.7411	-2.2198	0.5105
7	-3.3697	-1.9137	2.0473
8	0.0000	-1.1447	2.6205
9	0.5598	-0.4853	2.7801
10	-0.5108	1.0035	4.2996
11	-3.0428	-2.6687	0.1337
12	-3.8077	-2.4022	0.9645
13	-1.3863	-1.7466	0.1038
14	-0.2233	-0.7766	0.6804
15	1.0986	0.9114	0.0263
d.f. = 12		$\sum_j (\bar{L}_j-\hat{L}_j)^2 w_j = 16.3842$	

Table 4

and it is asymptotically distributed as chi-squared (χ^2) with degrees of freedom equal to the number of sets minus the number of parameters estimated. Table 4 gives the observed and predicted logits, the weighted discrepancies, the value of the test statistic (38) and the degrees of freedom for the example.

If the computed value of the test statistic (38) is greater than the value of χ^2 with appropriate degrees of freedom obtained from statistical tables we reject the null hypothesis of no significant difference between the logits predicted from our model and the observed logits, and we suggest that our model is not adequately representing the observed variation in the incidence of chronic bronchitis.

In the case of our example we have 12 degrees of freedom. At the conventional 0.05 and 0.01 levels of significance the tabulated χ^2 values are 21.0 and 26.2 respectively. The computed value of the test statistic for the example is 16.38, thus the null hypothesis cannot be rejected and we can suggest that there is an adequate agreement between the predicted logits from our model and the observed logits.

(vi) A matrix formulation

Although the X^2 value in the above example does not reveal a significant discrepancy between the predicted logits from our model and the observed logits at the conventional 0.05 level, it is in fact rather high. This suggests that our model could be improved, perhaps by the introduction of more explanatory variables. Certainly it would be useful for example to introduce the occupation of the respondent into the list of explanatory variables. However, when more than two explanatory variables are included in a model, the normal equations (34) become cumbersome to write out and to solve by hand. At this point we normally turn to matrix notation and matrix algebra to help us.

In matrix notation the normal equations (34) can be simplified to the equivalent form

$$(X'U^{-1}X)\hat{\beta} = X'U^{-1}L \qquad (39)$$

where, remembering that there are K sets (j=1,...K)

$$X'U^{-1}X = \begin{bmatrix} \sum_j w_j & \sum_j w_j X_{j1} & \sum_j w_j X_{j2} \\ \sum_j w_j X_{j1} & \sum_j w_j X_{j1}^2 & \sum_j w_j X_{j1} X_{j2} \\ \sum_j w_j X_{j2} & \sum_j w_j X_{j1} X_{j2} & \sum_j w_j X_{j2}^2 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 1 & \dots & 1 \\ X_{11} & X_{21} & \dots & X_{K1} \\ X_{12} & X_{22} & \dots & X_{K2} \end{bmatrix} \begin{bmatrix} w_1 & 0 & \dots & 0 \\ 0 & w_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & w_K \end{bmatrix} \begin{bmatrix} 1 & X_{11} & X_{12} \\ 1 & X_{21} & X_{22} \\ \vdots & \vdots & \vdots \\ 1 & X_{K1} & X_{K2} \end{bmatrix}$$

$$X'U^{-1}L = \begin{bmatrix} \sum_j w_j L_j \\ \sum_j w_j L_j X_{j1} \\ \sum_j w_j L_j X_{j2} \end{bmatrix} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ X_{11} & X_{21} & \dots & X_{K1} \\ X_{12} & X_{22} & \dots & X_{K2} \end{bmatrix} \begin{bmatrix} w_1 & 0 & \dots & 0 \\ 0 & w_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & & w_K \end{bmatrix} \begin{bmatrix} L_1 \\ L_2 \\ \vdots \\ L_K \end{bmatrix}$$

and $$\hat{\beta} = \begin{bmatrix} \hat{\alpha} \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}$$

By the rules of matrix algebra the solution of (39) is then

$$\hat{\beta} = (X'U^{-1}X)^{-1}X'U^{-1}L \qquad (40)$$

The reader will now be able to see that it is a simple matter to include another explanatory variable. All that is involved is the addition of an extra column to the matrix of observations X and thus X', the transpose, will have an extra row. In other words, with one extra explanatory variable X and X' become

$$X = \begin{bmatrix} 1 & X_{11} & X_{12} & X_{13} \\ 1 & X_{21} & X_{22} & X_{23} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{K1} & X_{K2} & X_{K3} \end{bmatrix} \qquad X' = \begin{bmatrix} 1 & 1 & \dots & 1 \\ X_{11} & X_{21} & \dots & X_{K1} \\ X_{12} & X_{22} & \dots & X_{K2} \\ X_{13} & X_{23} & \dots & X_{K3} \end{bmatrix}$$

The inclusion of an additional explanatory variable will also involve re-grouping the observations into K new sets, the sets now being determined on the basis of characteristic values of three explanatory variables rather than two as in the above example.

At this point more advanced readers may wish to consider how the inferential tests we can conduct on our logit models can be expressed in matrix algebra. General readers are advised to move straight on to Section IV.

In matrix notation the test statistic (38) becomes

$$(L - X\hat{\beta})'\, U^{-1}(L - X\hat{\beta}) \qquad (41)$$

It also becomes a simple matter to compute the standard errors of the parameter estimates. These standard errors are given by the square roots of the diagonal elements of (42),

$$(X'U^{-1}X)^{-1} \qquad (42)$$

If we wish to test the validity of adding an extra explanatory variable or set of explanatory variables to the model of our example, we can either simply inspect the parameter/standard error ratio(s) or use a general test statistic of the form (Grizzle et al 1969)

$$\hat{\beta}'C'\left[C(X'U^{-1}X)C'\right]^{-1}C\hat{\beta} \qquad (43)$$

to test the null hypothesis H_0: $C\beta = 0$; in other words to test that the additional parameter(s) is(are) zero. C is a suitably specified matrix. In our case if we wish to test the validity of adding only a single extra explanatory variable to the previous model, C is a one row matrix, in other words a vector, of the form.

$$C = \left[0,0,0,1\right]$$

Therefore,

$$\underline{C\beta} = \begin{bmatrix}0,0,0,1\end{bmatrix} \begin{bmatrix}\alpha \\ \beta_1 \\ \beta_2 \\ \beta_3\end{bmatrix}$$

The test statistic (43) has under the null hypothesis an asymptotic chi-squared distribution with degrees of freedom equal to the number of rows of C.

IV MULTIPLE RESPONSE CATEGORIES. THE POLYCHOTOMOUS CASE

(i) Introduction

The case we examined in the previous section, the case in which the categorized response variable has only two possible outcomes, is only the simplest categorized response variable problem. There are far more cases in which the geographer is faced with categorized response variables with more than two possible outcomes. We call these polychotomous rather than dichotomous variables. In this section it will be shown how the linear logit model can be extended to cover such cases. To aid the discussion it is useful to think in terms of a particular example. Let us presume that a survey of opinions about a proposed road improvement scheme has been conducted in an area surrounding the site of the proposed improvement. In response to a questionnaire the sample of individuals surveyed have indicated whether they are in favour', 'against' or 'undecided' about the scheme. We now wish to consider how people's opinions about the scheme are affected by their proximity to its location, where proximity of individual i is measured by the straight line distance X_i from the location of the proposed scheme to the home of individual i.

(ii) The extended linear logit model

We can begin to extend our previous dichotomous linear logit model by first arbitrarily scoring the three possible responses 2 (in favour), 1 (against) and 0 (undecided). Taking any pair of these possible responses, say 'in favour' and 'against', scored 2 and 1 respectively, we can then write a linear logit relationship for the pair (compare it with 23) as

$$\log_e \frac{P_{2/i}}{P_{1/i}} = \alpha_{21} + \beta_{21}X_i \qquad (44)$$

where $P_{2/i}$ is the probability that individual i will respond 'in favour', α_{21} and β_{21} denote that the parameters refer to that pair of responses and we have only the single explanatory variable X_i, distance. Taking another pair of categories, for example 'against' (1) and 'undecided' (0), we can then add the linear logit relationship for this pair to that for the first pair and we get

$$\log_e \frac{P_{2/i}}{P_{1/i}} + \log_e \frac{P_{1/i}}{P_{0/i}} = \alpha_{21} + \alpha_{10} + X_i(\beta_{21} + \beta_{10}) \qquad (45)$$

or

$$\log_e\left(\frac{P_{2/i}}{P_{1/i}} \times \frac{P_{1/i}}{P_{0/i}}\right) = \alpha_{21} + \alpha_{10} + X_i(\beta_{21} + \beta_{10}) \qquad (46)$$

The left hand side of (46) can however be simplified to give the linear logit relationship

$$\log_e \frac{P_{2/i}}{P_{0/i}} = \alpha_{21} + \alpha_{10} + X_i(\beta_{21} + \beta_{10}) \qquad (47)$$

Following the form of (44) this linear logit relationship can also be written as

$$\log_e \frac{P_{2/i}}{P_{0/i}} = \alpha_{20} + \beta_{20}X_i \qquad (48)$$

Thus if we compare the last two equations (48) and (47) we can write

$$\beta_{20} = \beta_{21} + \beta_{10} \quad \text{and} \quad \alpha_{20} = \alpha_{21} + \alpha_{10} \qquad (49)$$

$$\beta_{21} = \beta_{20} - \beta_{10} \quad \text{and} \quad \alpha_{21} = \alpha_{20} - \alpha_{10} \qquad (50)$$

Theil (1970, p. 119) notes that the right-hand sides of (50) can then be rewritten as

$$\beta_{21} = \beta_2 - \beta_1 \quad \text{and} \quad \alpha_{21} = \alpha_2 - \alpha_1 \qquad (51)$$

and this enables us to rewrite (44) as

$$L_{21/i} = \log_e \frac{P_{2/i}}{P_{1/i}} = (\alpha_2 - \alpha_1) + X_i(\beta_2 - \beta_1) \qquad (52)$$

Equation (52) forms the basis of the extended linear logit model, and it is sufficient to consider (52) for all responses different from the arbitrarily chosen denominator response (1, 'against' in our case). In addition the right hand side of (52) shows that only the differences of the α's and β's matter and therefore we can impose the constraints

$$\alpha_1 = \beta_1 = 0 \qquad (47)$$

on (52) without loss.

Instead of a single linear logit equation for each individual as in (23) we now have in the extended case of our example (remembering that the constraints (53) have been imposed) two linear logit equations of the form (52).

$$\tilde{L}_{21/i} = \log_e(P_{2/i}/P_{1/i}) = \alpha_2 + \beta_2 X_i \qquad (54)$$

$$L_{01/i} = \log_e(P_{0/i}/P_{1/i}) = \alpha_0 + \beta_0 X_i$$

(iii) Estimating the parameters of the extended linear logit model

Estimation of the parameters of the extended linear logit model (54) proceeds in exactly the same manner as in the estimation of the parameters of the simple linear logit model (23). That is to say individuals are grouped into sets and for each set j (j=1,...K), (54) is rewritten as

$$\tilde{L}_{21/j} = \log_e(f_{2/j}/f_{1/j}) = \alpha_2 + \beta_2 X_j + (\tilde{L}_{21/j} - L_{21/j}) \qquad (55)$$

$$\tilde{L}_{01/j} = \log_e(f_{0/j}/f_{1/j}) = \alpha_0 + \beta_0 X_j + (\tilde{L}_{01/j} - L_{01/j})$$

The parameters of (55) are then estimated using the weighted least squares method discussed above which in this general case is known as <u>generalised least squares.</u> However, because there are now two equations for each set j, the estimation method becomes slightly more complex. The error terms $(\tilde{L}_{21/j} - L_{21/j})$ and $(\tilde{L}_{01/j} - L_{01/j})$ on the right hand side of (55) are correlated for each set j. They are correlated because the probabilities of the three responses must sum to one in each set. As a result the weight now has the form

$$\underline{W}_j = n_j \begin{bmatrix} f_{2/j}(1-f_{2/j}) & -f_{2/j}f_{0/j} \\ -f_{0/j}f_{2/j} & f_{0/j}(1-f_{0/j}) \end{bmatrix} \qquad (53)$$

In other words instead of being a single element or <u>scalar,</u> it is now a 2 x 2 matrix.

It should also be noted that in this extended case Grizzle et al. (1969) have suggested that replacement working values like those discussed in III (iv) above should be employed when the observed relative frequency is 0 or 1. When there are r possible response categories (r=3 in our case) they suggest using

$$1/rn_j \text{ instead of } 0, \text{ and } 1 - 1/rn_j \text{ instead of } 1.$$

Rather than write out in conventional algebra the normal equations for our extended linear logit model, it is less cumbersome to use the equivalent matrix form as we did in III (vi). As we saw in III (vi) in matrix notation the normal equations have the form.

$$(\underline{X}'\underline{U}^{-1}\underline{X})\underline{\hat{\beta}} = \underline{X}'\underline{U}^{-1}\underline{\tilde{L}} \qquad (56)$$

Remembering that in terms of our example we have only one explanatory variable, proximity, and K sets (j=1,...K), then the elements of (57) are

$$\underline{X}'\underline{U}^{-1}\underline{X} = \begin{bmatrix} 1 & 0 & 1 & 0 & \dots & 1 & 0 \\ X_1 & 0 & X_2 & 0 & \dots & X_K & 0 \\ 0 & 1 & 0 & 1 & \dots & 0 & 1 \\ 0 & X_1 & 0 & X_2 & \dots & 0 & X_K \end{bmatrix} \begin{bmatrix} \underline{W}_1 & & 0 & 0 & \dots & 0 & 0 \\ & & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \underline{W}_2 & & \dots & 0 & 0 \\ 0 & 0 & & & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & \underline{W}_K & \\ 0 & 0 & 0 & 0 & \dots & & \end{bmatrix} \begin{bmatrix} 1 & X_1 & 0 & 0 \\ 0 & 0 & 1 & X_1 \\ 1 & X_2 & 0 & 0 \\ 0 & 0 & 1 & X_2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_K & 0 & 0 \\ 0 & 0 & 1 & X_K \end{bmatrix}$$

$$\underline{X}'\underline{U}^{-1}\underline{\tilde{L}} = \begin{bmatrix} 1 & 0 & 1 & 0 & \dots & 1 & 0 \\ X_1 & 0 & X_2 & 0 & \dots & X_K & 0 \\ 0 & 1 & 0 & 1 & \dots & 0 & 1 \\ 0 & X_1 & 0 & X_2 & \dots & 0 & X_K \end{bmatrix} \begin{bmatrix} \underline{W}_1 & & 0 & 0 & \dots & 0 & 0 \\ & & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \underline{W}_2 & & \dots & 0 & 0 \\ 0 & 0 & & & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & \underline{W}_K & \\ 0 & 0 & 0 & 0 & \dots & & \end{bmatrix} \begin{bmatrix} \tilde{L}_{21/1} \\ \tilde{L}_{01/1} \\ \tilde{L}_{21/2} \\ \tilde{L}_{01/2} \\ \vdots \\ \tilde{L}_{21/K} \\ \tilde{L}_{01/K} \end{bmatrix}$$

$$\underline{\hat{\beta}} = \begin{bmatrix} \hat{\alpha}_2 \\ \hat{\beta}_2 \\ \hat{\alpha}_0 \\ \hat{\beta}_0 \end{bmatrix}$$

By the rules of matrix algebra the solution of (57) is then

$$\underline{\hat{\beta}} = (\underline{X}'\underline{U}^{-1}\underline{X})^{-1}\underline{X}'\underline{U}^{-1}\underline{\tilde{L}} \qquad (58)$$

as in III (vi). The significance tests and standard error formulas given in III (vi) also apply directly.

(iv) More than one explanatory variable or more than three response categories

The example discussed above is the simplest possible case of an extended linear logit model. If more than one explanatory variable is to be included in the above model all that is involved is the addition of an extra two columns in the matrix of observations X and thus an extra two rows in X' for every additional explanatory variable. With one extra explanatory variable X therefore becomes

$$\begin{bmatrix} 1 & X_{11} & X_{12} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & X_{11} & X_{12} \\ 1 & X_{21} & X_{22} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & X_{21} & X_{22} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{K1} & X_{K2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & X_{K1} & X_{K2} \end{bmatrix}$$

If the categorized response variable has more than three possible outcomes, an extra linear logit equation is added to (54) and thus to (55) for every additional possible outcome. The $\underline{X}$, $\underline{U}^{-1}$, $\underline{\hat{L}}$ and $\underline{\hat{\beta}}$ matrices and vectors are thus increased in size. These matrices are set out for the case of four possible outcomes in Wrigley (1975, p. 190-91).

(v) Predicted probabilities from the extended linear logit model

In the case of our original example, three response categories and one explanatory variable, our model predicts the probability of each response in each set to be

$$\hat{P}_{2/j} = \frac{e^{\hat{L}_{2/j}}}{1+e^{\hat{L}_{2/j}}+e^{\hat{L}_{0/j}}} \qquad \hat{P}_{0/j} = \frac{e^{\hat{L}_{0/j}}}{1+e^{\hat{L}_{2/j}}+e^{\hat{L}_{0/j}}} \qquad \hat{P}_{1/j} = \frac{1}{1+e^{\hat{L}_{2/j}}+e^{\hat{L}_{0/j}}} \tag{59}$$

where for example

$$\hat{L}_{2/j} = \hat{\alpha}_2 + \hat{\beta}_2 X_j \tag{60}$$

These predicted probabilities sum to one for each set j. Unlike the predicted probabilities in the dichotomous case (36) however, the predicted probabilities in the polychotomous case are not necessarily a monotonically increasing or decreasing function of the values of the explanatory variable(s) In other words, unlike the classical linear regression model or the dichotomous linear logit model, as the value(s) of the explanatory variable(s) rises or falls consistently, so the predicted probability of a particular response may first increase and then decrease. This feature is shown in Figure 3 which is an example of a four response category case taken from a study of consumer shopping behaviour by the author. As accessibility between households and retail outlets, measured by mean spatial separation, becomes worse (as mean spatial separation moves from negative to positive), so the probability that households (in which the housewife works) will choose a particular form of shopping behaviour known as a 3-shop-type diversification level, first increases and then decreases

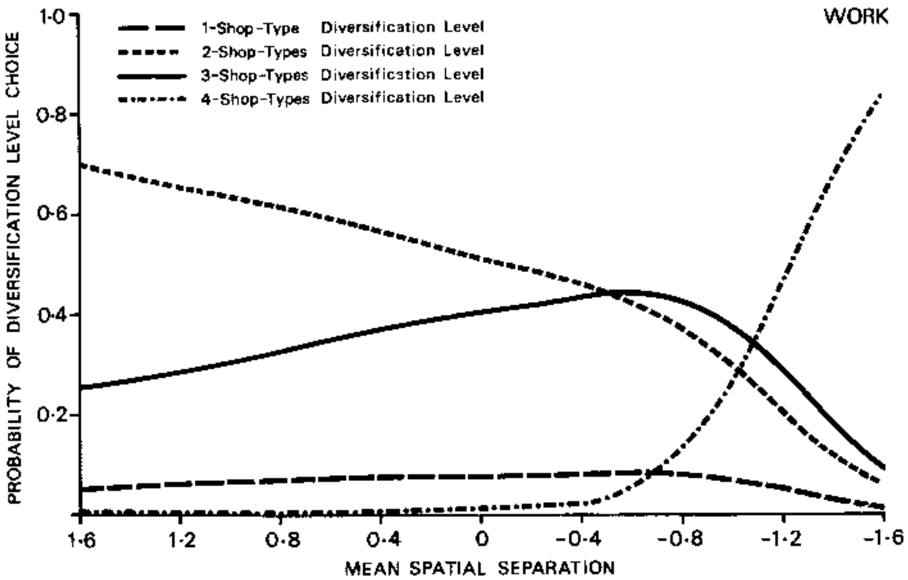


Fig. 3 An illustration of the pattern of predicted probabilities in the polychotomous case

V EXTENSIONS AND PROBLEM AREAS

(i) Maximum likelihood estimation and nonlinear logit models

In the previous sections of this monograph we have concentrated upon the linear logit model and a least squares method of estimation. As the reader who has attempted to work through the example in III(v) will already have found, the great drawback of this method of estimating the parameters of a linear logit model is the fact that grouping observations into sets is a laborious procedure. When a whole series of models are to be estimated using the same basic set of data, re-grouping must be undertaken for every postulated model, and unless a highly flexible computer routine is available the work involved is a major deterrent. It would be extremely useful therefore to have a method of estimation which does not require the grouping of observations into sets, which works instead with the individual observations. In fact we have such an estimation method. It is known as the method of maximum likelihood, and by using this method we can estimate the parameters of a nonlinear logit model of the form (20) without first having to convert it to a linear model (23).

The method of maximum likelihood was developed by R.A. Fisher (see Edwards 1972). If they come from a discrete distribution the likelihood of a set of observations is their joint probability of occurrence. In the case of the chronic bronchitis example of Section III the likelihood of the set of observations in Table 1 is

$$\Lambda = \prod_{i=1}^{N_1} P_i \prod_{i=N_1+1}^{N} (1-P_i) \qquad (61)$$

(where Π means multiply the elements together, e.g. $\Pi P_i = P_1 \times P_2 \times \ldots P_N$)
From (20) and (21) we thus have

$$\Lambda = \prod_{i=1}^{N_1} \frac{e^{\alpha+\beta_1 X_{i1}+\beta_2 X_{i2}}}{1+e^{\alpha+\beta_1 X_{i1}+\beta_2 X_{i2}}} \prod_{i=N_1+1}^{N} \frac{1}{1+e^{\alpha+\beta_1 X_{i1}+\beta_2 X_{i2}}} \qquad (62)$$

In other words whenever $Y_i = 1$ (the individual at locality i is a chronic bronchitic) the likelihood contains a term (20). Whenever $Y_i = 0$, the likelihood contains a term (21). The likelihood (62) states that there are N_1 chronic bronchitics and $N-N_1$ non chronic bronchitics in the sample. In the case of the data set of Table 1, $N = 212$, $N_1 = 46$.

As specified the likelihood depends upon a set of unknown parameters. These parameters are then estimated by taking as estimates those values of the parameters which maximize (62). Rather than maximize the likelihood itself however, it is usual to maximize instead the logarithm of the likelihood. In the case of (62) this implies maximizing

$$\log_e\Lambda = \sum_{i=1}^{N_1} (\alpha + \beta_1 X_{i1} + \beta_2 X_{i2}) - \sum_{i=1}^{N} \log_e(1 + e^{\alpha + \beta_1 X_{i1} + \beta_2 X_{i2}}) \qquad (63)$$

The so called first order conditions for a maximum occur when the first partial derivatives are set to zero.

$$\frac{\partial \log_e\Lambda}{\partial\alpha} = \frac{\partial \log_e\Lambda}{\partial\beta_1} = \frac{\partial \log_e\Lambda}{\partial\beta_2} = 0 \qquad (64)$$

The extension of the maximum likelihood estimation method to the multiple response category case is set out, using a four possible response category example, in Wrigley (1975, p. 192-3). Details of significance tests which are applicable when maximum likelihood estimation is used are given in Wrigley (1976).

Determination of the maximum of the log likelihood (63), or its extension in any particular multiple response category case, requires a numerical optimization computer routine. A set of such routines have been written by the author, and details of their availability will be supplied on request.

(ii) Probability surface mapping

One of the oldest and simplest techniques used by the geographer in spatial analysis is trend surface mapping. Trend surface mapping attempts to decompose a spatial series into two components, a trend or regional component and a residual or local component. (See Bassett 1972, Chorley and Haggett 1965, Unwin 1975). A trend surface model is basically a linear regression model in which the explanatory variables are the geographical co-ordinates U_i and V_i of each site or locality i, and the response variable or variable to be mapped Z_i measures the strength of any spatially distributed variable.

Although trend surface mapping has been widely used in a number of disciplines its range of applicability has been limited by the apparent universal acceptance that the variable to be mapped must be measured at a high level, at least interval scale. Categorizations have previously been viewed as unmappable by the trend surface method. Logit models provide us however with a solution and in a recent paper (Wrigley 1976) the author has demonstrated that surface mapping of such variables can be achieved using logit models. The resulting maps can be called probability surface maps, and they are illustrated in the paper using the empirical example of the perception of aircraft noise disturbance around Manchester (Ringway) Airport.

(iii) Residuals and testing for spatial autocorrelation

An integral part of surface mapping is the ability to map the residuals or local component values, and the ability to assess whether such a map displays any discernible spatial organisation. This ability to map residual values and assess such maps for the presence or absence of spatial organisation or spatial autocorrelation is also important in the context of the standard logit models discussed in Sections III and IV.

In the case of the dichotomous linear logit model estimated by weighted least squares, as discussed in Section III, a residual can be defined for each pair of values X_{j1} and X_{j2} used in the estimation procedure as

$$f_j - \hat{P}_j = e_j \qquad (65)$$

where $\hat{P}_j$ is given by equation (36). However, such a residual refers to a set j, the set having been defined on the basis of the explanatory variables. Unless, therefore, the list of explanatory variables were to include geographical co-ordinates, a set will be a non-areal entity and thus it will be impossible to attribute to the residual the locational information which is necessary in any mapping or test of spatial autocorrelation.

In the context of direct maximum likelihood estimation of the nonlinear logit model, individual observations are used in the estimation and thus a residual can be defined in the dichotomous case as

$$Y_i - \hat{P}_i = e_i \qquad (66)$$

Cox (1970, p.96) has suggested standardising such residuals to have a mean of zero and a unit variance as follows,

$$\frac{Y_i - \hat{P}_i}{\sqrt{\hat{P}_i(1-\hat{P}_i)}} = e_i \qquad (67)$$

In contrast to (65), (67) has the required locational information, but as Y_i can only take the value 0 or 1, it will typically have a highly non-normal distribution. In particular residuals close to the value zero will not occur except for extreme values of P_i. In these locations where there is a high or low probability of first category selection, or in terms of the example of Section III, of being a chronic bronchitic, the residuals have a very skew distribution. Thus in locations where the probability of being a chronic bronchitic is high, the residuals are either small and positive or large and negative. In locations where the probability of being a chronic bronchitic is low, the residuals are either small and negative or large and positive.

In any attempts to test for the presence of spatial autocorrelation amongst the residuals (67), we are hampered not only by extreme non-normality of the residuals but also by the fact that the statistic geographers commonly use (Cliff and Ord 1972) is not applicable to residuals derived from a non-linear model. Some form of generalisation of the Cliff-Ord spatial auto-correlation statistic for regression residuals is required. In the absence of such a statistic it is suggested that we attempt purely visual or ad hoc statistical assessments of the presence of spatial autocorrelation amongst our map of residuals, or employ in an informal manner Cliff and Ord's (1969) earlier spatial autocorrelation statistic for spatially distributed variables using the version of the test based upon the assumption of randomisation rather than that based upon the assumption of normality. These suggestions are however only tentative and much work remains to be done in this area.

In the multiple response category case the definition of a residual (67) must be generalised to the form

$$\frac{Y_{r/i} - \hat{P}_{r/i}}{\sqrt{\hat{P}_{r/i}(1-\hat{P}_{r/i})}} = e_{r/i} \qquad (68)$$

where $\hat{P}_{r/i}$ is the predicted probability that the individual at locality i will choose the rth response category, and $Y_{r/i}$ = 1 if category r occurs and 0 otherwise. In the multiple response category case there are thus as many residual maps as there are response categories.

(iv) Ordered categorizations

Throughout the monograph no account has been taken of any ordering which might exist amongst the response categories. All categorized response variables have been treated as unordered. A potential extension to the logit models outlined is to attempt to take account in the specification of the models and in the estimation procedures of any ordering which might exist amongst the response categories. Cox (1970, p.104) presents a nonlinear model which takes account of the ordering of the responses in a particular three category situation.

(v) Conclusion

It is the contention of this monograph that logit models are an extremely valuable aid to the geographer, enabling him to deal with the categorized response variables which he often wishes to study. Logit models are becoming widely used in a range of related disciplines, and as applications multiply it will become increasingly important for the geographer to understand their characteristics and potential.

BIBLIOGRAPHY

A. BASIC SOURCES ON THE LOGIT MODEL

Ashton, W.D. (1972) *The logit transformation: with special reference to its uses in bioassay.* (London, Griffin).

Berkson, J. (1944) Application of the logistic function to bio-assay. *Journal American Statistical Association,* 39, 357-65.

Berkson, J. (1953) A statistically precise and relatively simple method of estimating the bio-assay with quantal response based on the logistic function. *Journal American Statistical Association,* 48, 565-99.

Berkson, J. (1955) Maximum likelihood and minimum x^2 estimates of the logistic function. *Journal American Statistical Association,* 50, 130-62.

Cox, D.R. (1970) *The analysis of binary data.* (London, Methuen).

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

Gurland, J., Lee, I. and Dolan, P.A. (1960) Polychotomous quantal response in biological assay. *Biometrics,* 16, 382-98.

Lindsey, J.K. (1973) *Inferences from sociological survey data: a unified approach.* (Amsterdam, Elservier).

Mantel, N. (1966) Models for complex contingency tables and polychotomous dosage response curves. *Biometrics,* 22, 83-95.

McFadden, D. (1974) Conditional logit analysis of qualitative choice behaviour in: *Frontiers in econometrics,* ed P. Zarembka, (New York, Academic Press).

Theil, H. (1969) A multinomial extension of the linear logit model. *International Economic Review,* 10, 251-59.

Theil, H. (1970) On the estimation of relationships involving qualitative variables. *American Journal of Sociology,* 76, 103-54.

Wrigley, N. (1975) Analysing multiple alternative dependent variables. *Geographical Analysis,* 7, 187-95.

B. SELECTED RECENT APPLICATIONS

Cragg, J.G. and Uhler, R.S. (1970) The demand for automobiles. *Canadian Journal of Economics,* 3, 386-406.

de Donnea, F.X. (1971) *The determinants of transport mode choice in Dutch cities.* (Rotterdam, U.P.).

Domencich, T.A. and McFadden, D. (1975) *Urban travel demand. A behavioural analysis.* (Amsterdam, North-Holland).

Mantel, N. and Brown, C. (1973) A logistic re-analysis of Ashford and Sowden's data on respiratory symptoms in British coal miners. *Biometrics,* 29, 649-65.

Richards, M.G. and Ben-Akiva, M.E. (1975) *A disaggregate travel demand model.* (Farnborough, Saxon House).

Schmidt, P. and Strauss, R. (1975) The prediction of occupation using multiple logit models. *International Economic Review,* 16, 471-86.

Watson, P.L. (1974) *The value of time. Behavioural models of modal choice.* (Farnborough, D.C. Heath).

Wrigley, N. (1976) Probability surface mapping: a new approach to trend surface mapping. *Transactions Institute of British Geographers.*

C. OTHER REFERENCES

Bassett, K. (1972) Numerical methods for map analysis. in: *Progress in geography,* 4, eds C. Board, et al. (London, Arnold).

Chorley, R.J. and Haggett, P. (1965) Trend surface mapping in geographical research. *Transactions Institute of British Geographers,* 37, 47-67.

Cliff, A.D. and Ord, J.K. (1969) The problem of spatial autocorrelation. in: *London papers in regional science,* 1, *Studies in regional science,* ed A.J. Scott, (London, Pion).

Cliff, A.D. and Ord, J.K. (1972) Testing for spatial autocorrelation among regression residuals. *Geographical Analysis,* 4, 267-84.

Edwards, A.W.F. (1972) *Likelihood.* (Cambridge, U.P.).

Finney, D.J. (1971) *Probit analysis.* (3rd Edition) (Cambridge, U.P.).

Fisher, R.A. and Yates, F. (1963) *Statistical tables for biological, agricultural and medical research. (6th* Edition) (Edinburgh, Oliver and Boyd).

Harvey, D. (1969) *Explanation in geography.* (London, Arnold).

Huang, D.S. (1970) *Regression and econometric methods.* (New York, Wiley).

King, L.J. (1969) *Statistical analysis in geography.* (Englewood Cliffs, N.J., Prentice-Hall).

Kmenta, J. (1971) *Elements of econometrics.* (New York, Macmillan).

McNab, G.R., Slater, E.J. and Stewart, C.J. (1966) Response to a questionnaire on chronic bronchitis symptoms in East Anglia. *British Journal of Preventive and Social Medicine,* 20, 181-88.

Medical Research Council (1962) *Instructions for the use of the short questionnaire on respiratory symptoms.* (London).

Poole, M.A. and O'Farrell, P.N. (1971) The assumptions of the linear regression model. *Transactions Institute of British Geographers,* 52, 145-58.

Silk, J. (1976) A comparison of regression lines using dummy variable analysis. *Geographical Papers,* 44, *University of Reading, Department of Geography.*

Unwin, D.J. (1975) *An introduction to trend surface analysis.* Concepts and techniques in modern geography, 5. (Norwich, Geo Abstracts Ltd).

Wonnacott, R.J. and Wonnacott, T.H. (1970) *Econometrics.* (New York, Wiley).