MEMORANDUM
RM-5734-PR
NOVEMBER 1968

ESTIMATION TECHNIQUES FOR DEPENDENT LOGIT MODELS

A. A. Cook, Jr. and A. J. Gross

PREPARED FOR:

UNITED STATES AIR FORCE PROJECT RAND

The RAND Corporation
SANTA MONICA · CALIFORNIA

MEMORANDUM
RM-5734-PR
NOVEMBER 1968

ESTIMATION TECHNIQUES FOR DEPENDENT LOGIT MODELS

A. A. Cook, Jr. and A. J. Gross

This research is supported by the United States Air Force under Project RAND—Contract No. F44620-67-C-0045—monitored by the Directorate of Operational Requirements and Development Plans, Deputy Chief of Staff, Research and Development, Hq USAF. Views or conclusions contained in this study should not be interpreted as representing the official opinion or policy of the United States Air Force.

DISTRIBUTION STATEMENT
This document has been approved for public release and sale; its distribution is unlimited.

The RAND Corporation
1700 MAIN ST. • SANTA MONICA • CALIFORNIA • 90406

Published by The RAND Corporation

PREFACE

This Memorandum draws on three previous RAND studies, [9], [11], and [12], to illustrate the use of a probability prediction method involving the logit model. It is shown how the model is used to develop the prediction of failure trends within aircraft and the prediction of the reenlistment rate of first-term airmen. The authors plan subsequent studies in which they will use the logit model to study the interrelated maintenance characteristics of different aircraft subsystems and to analyze the reenlistment probabilities of Category IV airmen.

To follow the procedures described here, readers should be acquainted with regression analysis and contingency table methods.

SUMMARY

This Memorandum discusses estimation techniques of logit models whose response variables are not necessarily statistically independent. The statistical techniques developed and utilized here include the Generalized Least Squares approach of Zellner [14]; stepwise regression and its relevance in estimating logit model parameters; and the "contingency table" concept of fixed marginal probabilities which is used to develop a bivariate logit model that possesses certain statistical and probabilistic properties concerning the correlation coefficient and the summation of individual cell probabilities to unity.

The techniques are designed for application to two Air Force problems. The first problem uses concomitant information to estimate the probability that first-term airmen reenlist. The second uses information concerning flying missions to estimate the probabilities that interdependent aircraft subsystems fail during these missions.

Some applications to the problems are presented in this Memorandum; however, this effort is mainly directed toward development of the necessary techniques. An example illustrating the various techniques is given in Sec. V of the Memorandum.

In a forthcoming study, these logit models will be applied to two problems, analyzing them in detail. First, existing aircraft sortie data will be used to study the interrelated maintenance characteristics of the different aircraft subsystems. Second, the survivability (as defined in the text) and reenlistment probabilities of category IV airmen will be analyzed.

ACKNOWLEDGMENTS

We would like to thank our RAND colleagues, John McCall and Anders Sweetland, for their interest and assistance; we are particularly indebted to Ernest Scheuer, George Fishman and John Lu for many helpful comments.

CONTENTS

GLOSSARY

Endogenous variable	A variable determined within the framework of a specific model.
Exogenous variable	A variable determined outside of or independent of the model.
AFQT score	Air Force Qualifying Test score.
Opportunity cost	The difference in dollars between reenlistment and the best civilian alternative employment, if the best civilian employment pays more and the choice is reenlistment.

I. INTRODUCTION

This Memorandum presents a probability prediction model, the logit model, and its theoretical development tailored to be used on two problems explored in previous RAND studies. The first area is aircraft performance; the second, airmen survivability and retention. In a forthcoming study we will extensively apply the theory developed here.

McGlothlin and Donaldson [12] describe an Air Force Base supply officer who became well known because of his ability to predict future aircraft performance based on previous maintenance records. The authors report that the 5th Bomb Wing at Travis Air Force Base claims that its method of increasing the probability of sortie success depends on a particular aircraft selection procedure for each sortie. In the same vein, they discuss aircraft labeled "dogs," which are inherently less reliable in terms of sortie success than other aircraft of the same type, flying the same type of sorties. The method in [12] and in a subsequent study by Donaldson and Sweetland [9] employs statistical analysis to discover the failure trends within aircraft and to delineate differences in performance and maintenance among aircraft, and thus to identify the inherent dogs.

In a seemingly different area, McCall and Wallace [11] have studied the training and retention of Air Force electronic specialists. Approaching their analysis from an economic standpoint, they generate a supply function for electronic specialists. Their basic function relates the probability that first-term airmen will reenlist to the opportunity cost of doing so, where the cost is defined as the difference between potential civilian earnings and Air Force remuneration. A similar problem is that of predicting the first-term enlistment survivability of various airmen groups.[†]

Although aircraft performance and airmen survivability and retention present different problems in terms of subject matter, both can be analyzed by probability prediction models. In the first case, the probability of failure is considered an endogenous variable, depending upon one or more exogenous variables such as type of sortie, previous maintenance actions on the total aircraft and/or its subsystems, and length of maintenance service. In the second case, the probability of reenlisting is an endogenous variable, depending on the opportunity cost of reenlisting; the probability of survival is an endogenous variable depending on exogenous variables such as age, prior education, race, and initial AFQT score. These two problems use similar predictive models and statistical tests.

To treat such problems and to indicate answers to others falling within the same framework, we analyze a predictive model known as the logit model. In this model, the probability P that an event E will occur depends on a vector $\underline{X}' = (X_1, \ldots, X_n)$ of exogenous variables and is written

$$P(E) = \frac{1}{1 + e^{-\underline{\beta}'\underline{X}}}, \tag{1a}$$

where $\underline{\beta}' = (\beta_1, \ldots, \beta_n)$ is a vector of the regression coefficients to be estimated. This is known as the logistic function. The logit of P(E), then, is simply the following transformation:

[†]Survivability denotes the capability of an airman to complete his first term of reenlistment.

(1b) $$\text{logit}\,[P(E)] = \ln\left[\frac{P(E)}{1 - P(E)}\right] = \underline{\beta}'\underline{X}\ .$$

The occurrence (or nonoccurrence) of E is usually assumed to be independent of the occurrence (or nonoccurrence) of other events. It is well known, however, that aircraft subsystems do not fail independently of one another during sorties. To verify this statement we have observed the writeup histories of three aircraft subsystems for 200 sorties: radar, inertial navigation, and fire control. Table 1 shows the observed cell frequencies and the expected cell frequencies under the hypothesis of the mutual independence of these writeups.

Table 1

FREQUENCIES OF DIFFERENT EVENTS FOR THREE SUBSYSTEMS

Frequency	SSS	SSF	SFS	FSS	SFF	FSF	FFS	FFF	Total
Observed	86.0	46.0	10.0	18.0	9.0	7.0	8.0	16.0	200.0
Expected	72.3	46.2	19.8	23.5	12.7	15.0	6.4	4.1	200.0

NOTE: S stands for no writeup and F for writeup. Thus, the event SFS is no writeup of the radar, writeup of the inertial navigation, and no writeup for fire control.

The analysis is essentially that of a 2x2x2 contingency table in which there are two outcomes for each subsystem: S for no writeup and F for writeup. Let f_{ijk} be the observed number of sorties in the cell (i,j,k), where i, j, k = S, F. Similarly, let $\hat{f}_{ijk}$ be the expected number of sorties in the same cell. Under the assumption of mutual independence of the writeups, the expected fraction p_{ijk} (of the total number of sorties) in the cell (i,j,k) is equal to the product of the marginals; that is $p_{ijk} = p_{i..}p_{.j.}p_{..k}$, where $p_{i..} = \sum_{j,k} f_{ijk}/N$, $i = S, F$, and similarly for the other two subsystems. Then $\hat{f}_{ijk} = Np_{ijk}$.

We then compute the statistic

$$\chi^2 = \sum_{i,j,k} \frac{(f_{ijk} - \hat{f}_{ijk})^2}{\hat{f}_{ijk}} .$$

The associated chi-square value is 49.0, with a single degree of freedom. This value is significant beyond the 0.0001 level. Thus, the hypothesis of mutual independence of writeups for these three subsystems is emphatically rejected. This dependence of aircraft subsystems as well as the dependence of airmen subgroups[†] has led us to couch our specific problems in a dependent equation model.

Section II presents the basic theory of the logit model. First, we briefly review the literature on the use of these models, and then describe the model as it applies to our needs. This includes a theoretical explanation of the Generalized Least Squares approach that Zellner [14] uses to estimate "seemingly unrelated regression equations." A short discourse on stepwise regression and its relevance in estimating the parameters of logit models is also included. Section III describes the estimation procedures that are subsequently used to establish the posited relationships. In Sec. IV, we present a bivariate logit model, describe its mathematics, and give reasons why a generalized dependent logit model might be better than individual independent logit models. Section V illustrates a problem solved by the methods described here. And the Appendix contains derivations of some equations in the text.

[†]Airmen subgroups are any groups that can be distinguished on some predetermined basis: for example, Caucasian and non-Caucasian; high school graduates and high school dropouts. These groups are clearly neither independent nor homogeneous.

II. BASIC THEORY

Much has been written about the theory and application of the logistic function. One pioneer in the use of the logistic function is Joseph Berkson [2-6,7]. He has applied the logistic function to quantal data (death or survival) in bioassay testing of animals; that is, the testing of a chemical compound in animals to determine lethal doses for various percentages of the particular animal population. Other important papers concerning the logistic function have been written by Cox [8], Dyke and Patterson [10], and Walker and Duncan [13].

Suppose that each observation in a series of trials can take one of two forms, which in our particular applications are "reenlistment" or "failure to reenlist" for the airmen data, and "subsystem writeup" or "no subsystem writeup" for the aircraft maintenance data. Suppose that corresponding to each trial there are one or more independent variables upon which the outcome of a particular trial is suspected to depend. Since each observation (the dependent variable) is dichotomous, we denote the observations by 0 and 1 and thus obtain a sequence of 0's and 1's for the outcome of the trials.

To formulate the logistic function, let $Y_1, \ldots, Y_n$ be mutually independent, binomially distributed random variables and $\underline{X}_1', \ldots, \underline{X}_n'$ be a set of fixed s-vectors; that is, $\underline{X}_i' = (X_{i0}, X_{i1}, \ldots, X_{i,s-1})$. We assume $s < n$. Further, let p_i be the probability of success and K_i be the number of trials at the i^{th} stage, $i = 1, 2, \ldots, n$. Then the binomial distribution may be written:

$$b_i(Y_i, p_i, K_i) = \binom{K_i}{Y_i} p_i^{Y_i}(1 - p_i^{K_i - Y_i}), \tag{2}$$

where $Y_i = 0, 1, \ldots, K_i$, $i = 1, 2, \ldots, n$. It is suspected that p_i is related to the vector $\underline{X}_i$.

One might conjecture a linear relationship between p_i and $\underline{X}_i$, $i = 1, \ldots, n$. That is,

$$p_i = \underline{\beta}'\underline{X}_i \equiv \beta_0 + \beta_1 X_{i1} + \ldots + \beta_{s-1} X_{i,s-1} .$$

However, p_i must lie between zero and unity so that the procedure for estimating the parameters $\beta_0, \beta_1, \ldots, \beta_{s-1}$ would be constrained by this condition on p_i, $i = 1, \ldots, n$. There are functional relationships between p_i and $\underline{X}_i$, $i = 1, \ldots, n$ that are free of the above constraint but yet insure the restriction that p_i lie between zero and unity is not violated. One such relationship is the logit model given by

$$\text{logit}\ (p_i) = \ln \frac{p_i}{1 - p_i} = \underline{\beta}'\underline{X}_i , \tag{3}$$

where $\underline{\beta}' = (\beta_0, \ldots, \beta_{s-1})$ is the regression vector upon which we draw inferences. Rewriting (3), we see that

$$p_i = \frac{1}{1 + e^{-(\underline{\beta}'\underline{X}_i)}} . \tag{4}$$

Ordinarily, $X_{i0} = 1$ $(i = 1, \ldots, s - 1)$ so that β_0 is the intercept parameter. We shall abide by this convention.

We place emphasis on the logit model in this Memorandum for the following reasons:

1. The procedure for estimating $\underline{\beta}$ is free of constraints such as those that exist for the linear model.

2. Estimates of the logit model parameters are easily obtained and possess desirable statistical properties.

3. In many applications, including those discussed here, the model has empirical justification.

4. Much has been written that we can use concerning the logit model and its application.

There are two standard procedures for estimating the regression coefficient vector $(\beta_0, \beta_1, \ldots, \beta_{s-1})$--maximum likelihood and minimum chi-square. Berkson compares these methods in [6].

We solve the equations†

$$\sum_{i=1}^{n} K_i(Y_i/K_i - \hat{p}_i) = 0 \tag{5a}$$

and

$$\sum_{i=1}^{n} K_i X_{ij}(Y_i/K_i - \hat{p}_i) = 0 \tag{5b}$$

$j = 0, \ldots, s - 1$, for $\hat{\underline{\beta}}' = (\hat{\beta}_0, \hat{\beta}_1, \ldots \hat{\beta}_{s-1})$, where

$$\hat{p}_i = \frac{1}{1 + e^{-(\hat{\underline{\beta}}'\underline{X}_i)}} . \tag{6}$$

$\hat{\underline{\beta}}$ is then the maximum likelihood estimate of $\underline{\beta}$.

The solution of the equations††

$$\sum_{i=1}^{n} K_i \frac{[\tilde{p}_i(1 - Y_i/K_i) + (1 - \tilde{p}_i)Y_i/K_i]}{\tilde{p}_i(1 - \tilde{p}_i)} (Y_i/K_i - \tilde{p}_i) = 0 \tag{7a}$$

---

†Equations (5a) and (5b) are the normal equations whose solution yields the maximum likelihood estimates of $\underline{\beta}$. The derivation of these equations is sketched in the Appendix.

††Equations (7a) and (7b) are the normal equations whose solution yields the minimum χ^2 estimates of $\tilde{\underline{\beta}}$. The derivation of these equations is also sketched in the Appendix.

and

(7b) $$\sum_{i=1}^{n} K_i X_{ij} \frac{[\tilde{p}_i(1 - Y_i/K_i) + (1 - \tilde{p}_i)Y_i/K_i]}{\tilde{p}_i(1 - \tilde{p}_i)} (Y_i/K_i - \tilde{p}_i) = 0$$

$j = 1, \ldots, s - 1$, in terms of $\tilde{\underline{\beta}}$, where

(8) $$\tilde{p}_i = \frac{1}{(1 + e^{-(\underline{\beta}'\underline{X}_i)})},$$

yields the minimum chi-square estimate $\tilde{\underline{\beta}}$ of $\underline{\beta}$. The solution of each set of equations (5a, 5b) and (7a, 7b) usually requires an iterative approach. This is so because the coefficients are functions of $\hat{p}$ (or $\tilde{p}$), and $\hat{p}$ (or $\tilde{p}$) is not linear in the parameters.

Berkson [6] discusses the properties of $\tilde{\underline{\beta}}$ and $\hat{\underline{\beta}}$ and, to some extent, their relative merits. For example, the estimates $\tilde{\underline{\beta}}$ are best asymptotically normal (BAN) and hence they have the same asymptotic properties as $\hat{\underline{\beta}}$. That is, both $\tilde{\underline{\beta}}$ and $\hat{\underline{\beta}}$ are consistent, asymptotically efficient, and asymptotically normally distributed estimates of $\underline{\beta}$. In the comparative sample Berkson worked with he demonstrated that the mean square errors of the minimum chi-square estimates were smaller than the mean square errors of the corresponding maximum likelihood estimates. He also discusses the minimum "logit χ^2," which has the desirable property that the parameters $\underline{\beta}$ can be estimated simply and directly.

The "minimum logit χ^2 estimate" of $\underline{\beta}$ is the solution to the following quantity defined as the "logit χ^2":

(9) $$\text{logit } \chi^2 = \sum_{i=1}^{N} n_i p_i(1 - p_i)(\ell_i - \hat{\ell}_i)^2,$$

where $\ell_i = \ln p_i/(1 - p_i)$ is the observed logit, and $\hat{\ell}_i = \underline{\beta}'\underline{X}$ is the estimated logit. The normal equations for obtaining the minimum logit χ^2 estimate of $\underline{\beta}$ can easily be written:[†]

(10a) $$\sum_{i=1}^{N} n_i p_i(1 - p_i)(\ell_i - \hat{\ell}_i) = 0,$$

(10b) $$\sum_{i=1}^{N} n_i p_i(1 - p_i)X_{ij}(\ell_i - \hat{\ell}_i) = 0.$$

Unfortunately, the usual logit model as it is structured applies only to binomially distributed variables and thus cannot be applied directly to a multinomial situation. Because of the dependent nature of our data (in the sense that aircraft subsystem writeups are correlated) and because efficient linear regression procedures are available when the observed proportions $r_i = X_i/K_i$ are not merely zero or unity, we now discuss these applicable procedures. The substance of the following derivations can be found in Zellner [14] and Zellner and Lee [15].

Suppose that independent samples of size $n_1, n_2, \ldots, n_m$ are observed on items (such as airmen reenlistments, and aircraft subsystem writeups) that have a dichotomous response. Suppose that r_i is the proportion of "1's" for the i^{th} sample. We then assume that

(11) $$r_i = p_i + u_i,$$

where u_i has a binomial distribution with mean zero and variance $[p_i(1 - p_i)]/n_i$, and where $-p_i < u_i < 1 - p_i$.

[†]The derivation of these normal equations is also sketched in the Appendix.

From (11) we see that

(12) $$\frac{r_i}{1 - r_i} = \frac{p_i}{1 - p_i}\left(\frac{1 + u_i/p_i}{1 - u_i/(1 - p_i)}\right).$$

Thus, letting $Z_i = \ln \frac{r_i}{1 - r_i}$ and using (3) we have

(13) $$Z_i = \underline{\beta}'\underline{X}_i + \ln (1 + u_i/p_i) - \ln [1 - u_i/(1 - p_i)].$$

Expanding the last two terms of (13), we obtain

(14a) $$Z_i = \underline{\beta}'\underline{X}_i + \frac{u_i}{p_i(1 - p_i)} + Q_i,$$

where Q_i denotes the remainder. Q_i is of order n_i^{-1} in probability, and its omission will not affect the asymptotic properties of the estimates of Z_i.

Thus,

(14b) $$Z_i \doteq \underline{\beta}'\underline{X}_i + \epsilon_i,$$

where

$$\epsilon_i = u_i \Big/ p_i(1-p_i), \qquad i = 1, \ldots, m.$$

Hence, the least squares of estimate $\underline{\beta}^*$ of $\underline{\beta}$ is given by

(15) $$\underline{\beta}^* = \left(X' \Sigma^{-1} X\right)^{-1} X' \Sigma^{-1} \underline{Z},$$

where Σ is a diagonal matrix whose i^{th} element on the diagonal is $[n_i p_i(1 - p_i)]^{-1}$. The least-squares procedure to obtain $\underline{\beta}^*$ is an application of Aitken's generalized least squares [1]. The variance-covariance matrix of $\underline{\beta}^*$ is given by

$$\text{(16)} \qquad \text{Var } \underline{\beta}^* = \left(X' \Sigma^{-1} X\right)^{-1} + o(n_i^{-1}),$$

where $o(n_i^{-1})$ signifies a matrix whose elements are of smaller order than n_i^{-1} in probability.

As demonstrated in the Introduction, with respect to subsystem writeups, while there is no apparent intracorrelation among writeups within a subsystem, there may be a correlation among groups of subsystems with respect to writeups. Thus, we make some changes in the model (14) for the case of ℓ characteristics. We again follow Zellner and Lee's lead. Let

$$\text{(17)} \qquad \begin{bmatrix} \underline{Z}_1 \\ \underline{Z}_2 \\ \vdots \\ \underline{Z}_\ell \end{bmatrix} = \begin{bmatrix} \underline{J} & \underline{0} & \cdots & \underline{0} & \underline{R}_1 & \cdots & \underline{R}_s \\ \underline{0} & \underline{J} & \cdots & \underline{0} & \underline{R}_1 & \cdots & \underline{R}_s \\ \vdots & & & \vdots & & & \\ \underline{0} & \underline{0} & & \underline{J} & \underline{R}_1 & & \underline{R}_s \end{bmatrix} \begin{bmatrix} \beta_{01} \\ \vdots \\ \beta_{0\ell} \\ \beta_1 \\ \vdots \\ \beta_s \end{bmatrix} + \begin{bmatrix} \underline{V}_1 \\ \underline{V}_2 \\ \vdots \\ \underline{V}_\ell \end{bmatrix},$$

where $\underline{J}$ is an n-vector, each of whose components is unity:

$$\underline{R}'_m = (X_{m1}, \ldots, X_{mn}),^{\dagger} \quad m = 1, 2, \ldots, s,$$
$$V_{ij} = u_{ij}/p_{ij}(1 - p_{ij}).$$

For the remainder of Sec. II, the $\underline{X}_i$ vectors will denote the vector of trials for the i^{th} exogenous variable (the $\underline{R}_i$ vectors in the footnote).

We require that the $\underline{Z}_i$'s each have the same dimension--N x 1. Thus, writing (17) in compact notation we see that

$$\underline{Z} = X\,\underline{\beta} + V\,, \tag{18}$$

where

$$\underline{Z}' = (\underline{Z}'_1, \underline{Z}'_2, \ldots, \underline{Z}'_\ell),$$
$$\underline{\beta}' = (\beta_{01}, \ldots, \beta_{0\ell}, \ldots, \beta_1, \ldots, \beta_s),$$
$$\underline{V}' = (\underline{V}'_1, \ldots, \underline{V}'_\ell)$$

---

† It is important to note the relationship between the X vectors at the bottom of page 5 and the R vectors. The individual components of the vectors are such that XX' = R'R, where we define the matrices X and R as follows:

$$X = [\underline{X}_1 \ldots \underline{X}_n] = \begin{bmatrix} X_{11} & X_{21} & \cdots & X_{n1} \\ \vdots & & & \\ X_{1s} & X_{2s} & \cdots & X_{ns} \end{bmatrix}$$

and

$$R = [\underline{R}_1 \ldots \underline{R}_s] = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1s} \\ \vdots & & & \\ X_{n1} & X_{n2} & \cdots & X_{ns} \end{bmatrix},$$

where the X_{ij} represents the value of the j^{th} exogenous variable for the i^{th} trial.

and

(19) $$X = \begin{bmatrix} \underline{j} & \underline{0} & \cdots & \underline{0} & \underline{X}_1 & \cdots & \underline{X}_s \\ \underline{0} & \underline{j} & & \underline{0} & \underline{X}_1 & \cdots & \underline{X}_s \\ \cdot & \cdot & & \cdot & \cdot & & \cdot \\ \underline{0} & \underline{0} & \cdots & \underline{j} & \underline{X}_1 & \cdots & \underline{X}_s \end{bmatrix}^{\dagger}.$$

The model for determining the proportion of airmen reenlistments is the generalization of (14) that Zellner and Lee [15] discuss. It can also be written as (18) with the same definition for $\underline{Z}'$ and $\underline{V}'$, but with $\underline{\beta}' = (\underline{\beta}_1', \ldots, \underline{\beta}_\ell')$, and

(20) $$X = \begin{bmatrix} \underline{X}_1 & 0 & \cdots & 0 \\ 0 & \underline{X}_2 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & \underline{X}_\ell \end{bmatrix},$$

where $\underline{X}_i$ denotes the set of exogenous variables for the i^{th} distinct group of airmen.

The variance-covariance matrix for $\underline{V}$, which is independent of $X\underline{\beta}$, will now be developed in general. Again, following the development of Zellner and Lee [15], the variance-covariance matrix Σ is given by

(21) $$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1\ell} \\ \Sigma_{21} & \Sigma_{22} & \cdots & \Sigma_{2\ell} \\ \vdots & \vdots & & \vdots \\ \Sigma_{\ell 1} & \Sigma_{\ell 2} & \cdots & \Sigma_{\ell\ell} \end{bmatrix},$$

and

† $\underline{j}' = (11, \ldots, 1)$.

(22) $$\Sigma_{ii} = \begin{bmatrix} \sigma^i_{11} & 0 & \dots & 0 \\ 0 & \sigma^i_{22} & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & \sigma^i_{nn} \end{bmatrix}$$

and $\sigma^i_{jj} = [n_j p_{ij}(1 - p_{ij})]^{-1}$.

To evaluate Σ_{ij}, $(i \neq j)$, we need some additional theory. First of all, $E(V_{i\nu}V_{j\mu}) = 0$ unless $\nu = \mu$. To calculate $E(V_{i\nu}V_{j\nu})$ we sketch the approach in [15]. First of all we note from the definition of $V_{k\nu}$, $k = i, j$ and Eq. (11) that

(23a) $$\sigma^\nu_{ij} \equiv E(V_{i\nu}V_{j\nu}) = [p_{i\nu}p_{j\nu}(1-p_{i\nu})(1-p_{j\nu})]^{-1} E(r_{i\nu}-p_{i\nu})(r_{j\nu}-p_{j\nu}).$$

$r_{k\nu}$ $k = i, j$ can be written as the average of n_ν dichotomous variables so that

$$E(r_{i\nu}-p_{i\nu})(r_{j\nu}-p_{j\nu}) = \frac{1}{n_\nu^2} E\left(\sum_{t=1}^{n_\nu} (y_{i\nu}(t)-p_{i\nu}) \sum_{t=1}^{n_\nu} (y_{j\nu}(t)-p_{j\nu})\right),$$

where

$$y_{k\nu}(t) = \begin{cases} 1 \text{ with probability } p_{k\nu} \\ 0 \text{ with probability } 1-p_{k\nu} \end{cases} \quad k = i, j.$$

We now observe that $E(y_{i\nu}(t) - p_{i\nu})(y_{j\nu}(t') - p_{j\nu}) = 0$ unless $t = t'$, in which case

$$E(y_{i\nu}(t) - p_{i\nu})(y_{j\nu}(t) - p_{j\nu}) = p^{ij}_{11}(\nu) - p_{i\nu}p_{j\nu}.$$

Thus,

(23b) $$\sigma^\nu_{ij} = [n_\nu p_{i\nu}p_{j\nu}(1 - p_{i\nu})(1 - p_{j\nu})]^{-1} [p^{ij}_{11}(\nu) - p_{i\nu}p_{j\nu}].$$

$p_{11}^{ij}(\nu)$ is the probability that an individual in the ν^{th} group has value "1" in both characteristics i and j. Thus,

(24) $$\Sigma_{ij} = \begin{bmatrix} \sigma_{ij}^1 & 0 & \dots & 0 \\ 0 & \sigma_{ij}^2 & \dots & 0 \\ 0 & 0 & \dots & \sigma_{ij}^n \end{bmatrix}, \quad \text{for } i \neq j.$$

Since we want to test hypotheses about $\underline{\beta}$ and derive corresponding confidence intervals, we now discuss some general principles concerning testing of hypotheses and derivation of confidence regions for the estimated parameters. Suppose we wish to test the hypothesis that the s-dimensional parameter space $\underline{\beta}$, based on the estimate $\underline{\beta}^*$ given by (15), can be reduced to a $s' < s$ dimensional space. The null hypothesis H_0 can be written

(25) $$H_0: C\,\underline{\beta} = \underline{0},$$

where $C_{(s-s')xs}$ exhibits the restriction under H_0. To test H_0, Zellner [14] gives as a test statistic:

(26) $$F_{s-s',n-s} = \frac{(n-s)}{(s-s')} \frac{\underline{Z}'\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}C'[C(X'\Sigma^{-1}X)^{-1}C']C(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\underline{Z}}{\underline{Z}'\Sigma^{-1}\underline{Z}-\underline{Z}'\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\underline{Z}}.$$

That is, we reject H_0 at the α-level of significance if and only if

(27) $$F_{s-s',n-s} > F_{s-s',n-s,\alpha}.$$

$F_{s-s',n-s,\alpha}$ is the upper α percentile of the F distribution with s-s' and n-s' degrees of freedom.

To obtain a confidence region for $\underline{\beta}$, we note that since $\underline{\beta}^*$ is a best linear unbiased estimator of $\underline{\beta}$, then based on the normality assumption of $\underline{Z}$,[†]

$$(28) \qquad F_{s,n-s} = \left(\frac{n-s}{s}\right)\frac{(\underline{\beta}^* - \underline{\beta})'(\underline{\beta}^* - \underline{\beta})}{[\underline{Z}'\Sigma^{-1}\underline{Z}-\underline{Z}'\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\underline{Z}]} .$$

Thus, a $(1-\alpha)$ level of confidence region for $\underline{\beta}$ is given by

$$(29) \qquad \left(\frac{n-s}{s}\right)\frac{(\underline{\beta}^* - \underline{\beta})'(\underline{\beta}^* - \underline{\beta})}{\underline{Z}'\Sigma^{-1}\underline{Z}-\underline{Z}'\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\underline{Z}} \le F_{s,n-s;\alpha} ,$$

where $F_{s,n-s;\alpha}$ is similar to $F_{s-s',n-s;\alpha}$.

In applying the logit models considered here, we note that there is a relationship between the dependent variables and several exogenous or independent variables. For example, the failure (or success) of interdependent aircraft subsystems is related to exogenous variables such as type of mission (for example, escort mission, bombing mission), length of mission, maintenance actions prior to the mission, and time of previous overhaul. In these applications it is important to identify, if possible, those exogenous variables which "best explain" the behavior of the dependent variables in the sense that by eliminating those exogenous variables that are not "best," little is lost in the accuracy of predicting the outcome of the dependent variables. Stepwise regression affords a methodology for choosing a subset from the set of exogenous variables that are among the best with regard to "explaining" the

[†]The asymptotic normality of $\underline{Z}$ is a consequence of the asymptotic normality of $\underline{u}$. The rate of convergence of $\underline{u}$ to normality depends on $\underline{p}$. For p_i small or large, the rate of convergence of u_i is slower than for moderate values of p_i.

behavior of the dependent variables. Briefly, the stepwise regression "explains" the behavior of the dependent variable by maximizing the regression sum of squares sequentially until some prespecified criterion is achieved. That is, we examine all simple regressions†

$$(30)\qquad Z_i' = \beta_\nu X_{\nu i}' + \epsilon_i$$

$$X_{\nu i}' = (X_{\nu i} - \bar{X}_\nu);$$

$$Z_i' = (Z_i - \bar{Z});\ i = 1, 2, \ldots, n,\quad \nu = 1, 2, \ldots, s.$$

We then choose $\hat{\beta}_{\nu_1}'$ such that

$$(31)\qquad \text{regression s.s.}(\hat{\beta}_{\nu_1})^{\dagger\dagger} = \max_{\nu = 1, 2, \ldots, s}\left(\hat{\beta}_\nu \cdot \sum_{i=1}^{n} X_{\nu i}' X_i'\right)$$

$$= \max_{\nu = 1, 2, \ldots, s}\left(\sum_{i=1}^{n} X_{\nu i}' Z_i' \Big/ \sum_{i=1}^{n} X_{\nu i}'^{2}\right).$$

Next, we let $Z_i'' = Z_i' - \hat{\beta}_{\nu_1} X_{\nu_1}$, and examine the s-1 remaining regressions. Thus, we consider

$$(32)\qquad Z_i'' = \beta_{\nu'} X_{\nu' i}' + \epsilon_i,$$

$$X_{\nu' i}' = (X_{\nu' i} - \bar{X}_{\nu'});\ i = 1, 2, \ldots, n,\quad \nu' = 1, \ldots, s,\quad \nu' \neq \nu_1.$$

Now we choose $\hat{\beta}_{\nu_2}$ such that

$$(33)\qquad \text{regression s.s}\left(\hat{\beta}_{\nu_2}\right)^{\dagger\dagger} = \max_{\substack{\nu'=1,\ldots,s\\ \nu'\neq\nu_1}}\left(\left(\sum_{i=1}^{n} X_{\nu' i}' Z_i''\right)^2 \Big/ \sum_{i=1}^{n} X_{\nu' i}'^{2}\right).$$

---

†We assume $E(\epsilon_i \epsilon_j) = \delta_{ij}\sigma^2$, where δ_{ij} is the Kronecker delta, i.e., $\delta_{ii} = 1$, $\delta_{ij} = 0$, $i \neq j$.

††Regression s.s.$(\hat{\beta}_i)$, i = 1, 2 is the regression sum of squares due to $\hat{\beta}_i$.

This procedure is continued until one of the following occurs: (1) a predetermined amount of total variation is explained by the subset $(\beta_{\nu_1}, \ldots, \beta_{\nu_{s_1}})$; (2) a predetermined number of regression coefficients, say s_1, is chosen; or (3) some other practical cutoff point is reached.

As indicated in the footnote on page 15, this development assumes uncorrelated and homoscedastic observations; however, the modifications discussed earlier enable us to apply the stepwise procedure to our problem.

III. ESTIMATION PROCEDURES

In describing the techniques used to estimate the parameters of the logit model, we divide the following discussion into two major parts: (1) the estimation of single equation models, and (2) the estimation of dependent equation models.

A single equation model is one that estimates the parameters of a single posited relationship when this relationship is uncorrelated with any others. For example, the equation that relates the probability of reenlistment for airmen to exogenous factors falls into this category. Consider now

(34a) $$Z_i = \underline{\beta}'\underline{X}_i + \epsilon_i, \qquad i = 1, \ldots, n,$$

where

$$\underline{\beta}' = (\beta_0, \ldots, \beta_{s-1})$$
$$\underline{X}_i' = (X_{0i}, X_{1i}, \ldots, X_{s-1,i})$$
$$\epsilon_i = u_i(r_i(1 - r_i))^{-1}$$
$$E(\epsilon_i) = 0,$$

and for $i = j$, $E(\epsilon_i\epsilon_j) = [N_i r_i(1 - r_i)]^{-1}$

Using the procedure of ordinary least squares (OLS), we can obtain an estimator $\underline{b}^*$ of $\underline{\beta}$:

$$\underline{b}^* = (X'X)^{-1} X'\underline{Z},$$

where

$$X = (\underline{X}_1', \underline{X}_2', \ldots \underline{X}_n').$$

The estimator $\underline{b}^*$ is unbiased and is equal to the maximum likelihood estimator $\underline{\tilde{b}}$ (assuming, of course, that the ϵ_i are normally distributed). The disturbances, ϵ_i, exhibit heteroscedasticity, however, and hence

$\underline{b}^*$ is not a best linear unbiased estimator (BLUE). To get an asymptotically BLUE estimator, we use the following procedure. Since (34) is a single equation, it is easy to obtain $\underline{b}^*$. First we calculate the weights

(34b) $$w_i = [n_i r_i (1 - r_i)]^{\frac{1}{2}}, \quad i = 1, \ldots, n,$$

where r_i is the observed proportion of "1's" for the i^{th} sample as given in (11). A weighted regression is then performed using OLS on the following equation:

(35) $$w_i z_i = \underline{\beta}' w_i \underline{X}_i + \eta_i.$$

The OLS estimator $\underline{\beta}^*$ of $\underline{\beta}$ is now an asymptotically BLUE estimator and is equal to both the MLE $\hat{\underline{\beta}}$ and the minimum logit χ^2 estimator of (14). In addition, $\underline{\beta}^*$ is exactly the same as the $\underline{\beta}^*$ of (15).

However, the simple procedure does not suffice for dependent equation models, that is those that have two or more posited relationships which are interdependent. Two types of these models were explained in (17) and (20). Assuming that the entire model contains ℓ equations, we can write the system as

(36a) $$\begin{bmatrix} \underline{z}_1 \\ \\ \vdots \\ \underline{z}_\ell \end{bmatrix} = \begin{bmatrix} \underline{X}_1 & 0 & \ldots & 0 \\ 0 & \underline{X}_2 & \ldots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \ldots & \underline{X}_\ell \end{bmatrix} \begin{bmatrix} \underline{\beta}_1 \\ \underline{\beta}_2 \\ \vdots \\ \underline{\beta}_\ell \end{bmatrix} + \begin{bmatrix} \underline{u}_1 \\ \\ \vdots \\ \underline{u}_\ell \end{bmatrix},$$

or

(36b) $$\underline{Z} = X\underline{\beta} + \underline{u}\ .$$

As noted above, the $\underline{X}_\theta$ are not necessarily mutually exclusive (see (17)), nor is it essential that all the $\underline{\beta}_\theta$ be different. The solution of (36) can be written as

(37) $$\underline{\beta}^* = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\underline{Z}\ .$$

The method that Zellner has devised [14] is used to estimate the matrix Σ^{-1}. Essentially, it estimates each of the ℓ individual weighted regressions by OLS. Thus

(38) $$\underline{b}_\theta^* = (\underline{X}_\theta'\underline{X}_\theta)^{-1}\underline{X}_\theta'\underline{Z}_\theta, \qquad \theta = 1, \ldots, \ell.$$

The estimator $\underline{b}_\theta^*$ is used to obtain the MLE of the disturbance vector $\hat{\underline{u}}_\theta$ for equation θ:

(39) $$\hat{\underline{u}}_\theta = \underline{Z}_\theta - \underline{X}_\theta\underline{b}_\theta^*, \qquad \theta = 1, \ldots, \ell.$$

We can then compute the matrix

(40) $$\hat{\Sigma}_u^{-1} = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1\ell} \\ \vdots & & \vdots \\ \sigma_{\ell 1} & \cdots & \sigma_{\ell\ell} \end{bmatrix}^{-1}$$

such that $\hat{\Sigma}_u$ is of dimension $\ell \times \ell$ and has the general element $\hat{\sigma}_{\theta\emptyset}$ given by

$$\hat{\sigma}_{\theta\emptyset} = \frac{1}{n}\,(\hat{\underline{u}}_\theta'\hat{\underline{u}}_\emptyset) = \frac{1}{n}\sum_{i=1}^{n} \hat{u}_{\theta i}\hat{u}_{\emptyset i}\ ,$$

$$\theta, \emptyset = 1, \ldots, \ell.$$

The matrix $\hat{\Sigma}_u^{-1}$ is an estimate of Σ^{-1}. Replacing Σ^{-1} by $\hat{\Sigma}_u^{-1}$ in (37) yields estimates $\underline{\beta}^*$ for $\underline{\beta}$, where $\underline{\beta}^{*\prime} = (\underline{\beta}_1^{*\prime}, \ldots, \underline{\beta}_\theta^{*\prime}, \ldots, \underline{\beta}_\ell^{*\prime})$. These estimators $\underline{\beta}_\theta^*$ consider explicitly the intercorrelation between the ℓ equations.

The principal advantage of using the above method is that the estimators $\underline{\beta}_\theta^*$ are more efficient† than the $\underline{b}_\theta^*$. This is especially helpful to the decisionmaker when various constraints are imposed on the estimators as in (42) below.

However, there is a gain in efficiency using Zellner's method only if correlation exists between the dependent variables; that is, only if $E(\underline{Z}_\theta \underline{Z}_\emptyset) \neq 0$, for all θ, $\emptyset$ such that $\theta \neq \emptyset$.

If there is a good reason to believe that $E(\underline{Z}_\theta \underline{Z}_\emptyset) = 0$, then there is a considerably simpler method that may be used to estimate (17). Thus, let $\underline{y}' = (\underline{Z}_1', \ldots, \underline{Z}_\ell')$, $X = (\underline{X}_1, \ldots, \underline{X}_{s+\ell})$, and $\underline{\beta}' = (\beta_1, \ldots, \beta_{s+\ell})$. Then the following regression may be estimated by OLS:

$$\underline{y} = \sum_{\theta=\ell+1}^{s+\ell} \beta_\theta \underline{X}_\theta + \sum_{\theta=1}^{\ell} \beta_\theta \underline{X}_\theta + \underline{u} \,, \tag{41}$$

where $\underline{X}_\theta = 0, 1$ for all $\theta \in [1, \ell]$. Similarly, the parameters of (36) can be estimated by applying OLS separately to each equation of the model.

The system (17) we wish to estimate is merely the general system (36) with linear restriction on the estimates; viz., the "slope" coefficients are constrained to be the same among the equations, while

---

†That is,

$$\frac{\mathrm{Var}(\underline{\beta}_\theta^*)}{\mathrm{Var}(\underline{b}_\theta^*)} \leq 1, \qquad \text{for all } \theta.$$

the intercepts are allowed to vary. To handle linear restrictions in the present problem, we must minimize the following equation with respect to $\underline{\beta}$:

(42) $$(\underline{Z} - X\underline{\beta})'(\underline{Z} - X\underline{\beta}) - \lambda(R\underline{\beta} - \underline{r}),$$

where $\underline{Z}$, X, and $\underline{\beta}$ are defined in (18), λ is a row vector of Lagrange multipliers, and R and $\underline{r}$ are a matrix and a column vector, respectively, which together define the restrictions on the system. The estimator, $\underline{\beta}^{\Delta}$, will satisfy the relationship

$$R\underline{\beta}^{\Delta} = \underline{r} ,$$

where R is a matrix with a row for each restriction. For example, the restriction matrix for (17) would be

$$\begin{bmatrix} a_{11} & \cdots & a_{1\ell} & \cdots & a_{1T} \\ \vdots & & \vdots & & \vdots \\ a_{\ell 1} & \cdots & a_{\ell\ell} & \cdots & a_{\ell T} \end{bmatrix} \begin{bmatrix} \beta_{01} \\ \vdots \\ \beta_{0\ell} \\ \beta_1 \\ \vdots \\ \beta_s \end{bmatrix} ,$$

where

(i) $a_{ij} = 0$, for $i, j = 1 \ldots \ell$

(ii) $a_{ij} = 1$, for $j = \ell + i$; $i = 1, \ldots, \ell$; $j = 1, \ldots, T$ and $T = \ell \cdot S$

(iii) $a_{ij} = 1$, for $j = \ell + s + i$.

Minimizing (42), we obtain the following expression for β^{Δ}:

(43a) $$\underline{\beta}^{\Delta} = \hat{\underline{\beta}} + (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(\underline{r} - R\hat{\underline{\beta}}),$$

where $\hat{\underline{\beta}} = (X'X)^{-1}X'\underline{Z}$, the unconstrained least-squares estimator. The covariance matrix of the estimator $\underline{\beta}^{\Delta}$ is given by $V(\underline{\beta}^{\Delta})$:

$$V(\underline{\beta}^{\Delta}) = V(\hat{\underline{\beta}}) - V(\hat{\underline{\beta}})R'[R(X'X)^{-1}R']^{-1}RV(\hat{\underline{\beta}}). \tag{43b}$$

To account for the correlated error term between equations, we use the following estimator of $\underline{\beta}$:

$$\underline{b}^{\Delta} = \hat{\underline{b}} + (X'\Sigma_u^{-1}X)^{-1}R'[R(X'\Sigma_u^{-1}X)^{-1}R']^{-1}(\underline{r} - R\hat{\underline{b}}), \tag{44a}$$

and the covariance matrix is

$$V(\underline{b}^{\Delta}) = V(\hat{\underline{b}}) - V(\hat{\underline{b}})R'[R(X'\Sigma_u^{-1}X)^{-1}R']^{-1}RV(\hat{\underline{b}}). \tag{44b}$$

IV. BIVARIATE LOGIT MODELS

The development to this point has been oriented toward univariate logit models. That is, we have considered the probability that an aircraft system would either "fail" or "not-fail," that an airman would either "reenlist" or "not-reenlist," or that the subject under consideration would possess some characteristic on a "yes-no" basis. Then we analyzed the interdependent relationships of the various subjects. However, situations arise in which a subject can possess more than one characteristic, each on a "yes-no," or "pass-fail" basis.

For example, consider any weapon system with more than one component part (or in some circles, "black box"). In these instances it would be preferable to incorporate the dependence of the component parts directly into the probability functions. We have done this below for the simplest case--that of two variables--such that the marginal distribution of the probability of "success" for each variable is represented by a logistic function.

Consider the following 2x2 contingency table for two characteristics, say η_1 and η_2, of some subject on a "success-fail" basis such that success is represented by a one and failure by a zero, where

$$\sum_j p_{1j} = p_{1.}, \sum_i p_{i1} = p_{.1}, \sum_{i,j} p_{ij} = 1.$$

Table 2

CONTINGENCY TABLE FOR CHARACTERISTICS η_1 AND η_2

$\eta_2 \backslash \eta_1$	Success	Fail	Marginals
Success	p_{11}	p_{10}	$p_{1.}$
Fail	p_{01}	p_{00}	$p_{0.}$
Marginals	$p_{.1}$	$p_{.0}$	$p_{..}$

We now take into account the intercorrelation of η_1 and η_2. Let $p_{rs}(i)$, where r, s = 0,1 is the probability that on the i^{th} trial the characteristic η_1 takes on the value r, and η_2 the value s.

If η_1 and η_2 are independent characteristics we can then write

(45a) $$p_{1.}(i)p_{.1}(i) = \left[(1 + e^{\gamma_{1i}})(1 + e^{\gamma_{2i}})\right]^{-1} \equiv p_{11}(i),$$

where

$$\gamma_{1i} = -(\alpha_0 + \alpha_1 X_{i1} + \cdots + \alpha_{s-1} X_{i,s-1}),$$

and

$$\gamma_{2i} = -(\beta_0 + \beta_1 Z_{i1} + \cdots + \beta_{t-1}, Z_{i}, _{t-1}).$$

However, (45a) does not hold if characteristics η_1 and η_2 are dependent. Regardless of whether or not η_1 and η_2 are independent, the relationships of Table 2 must hold. Namely, $p_{11}(i) + p_{10}(i) = p_{1.}(i)$, $p_{11}(i) + p_{01}(i) = p_{.1}(i)$, $p_{10}(i) + p_{00}(i) = p_{.0}(i)$, and $p_{01}(i) + p_{00}(i) = p_{0.}(i)$. We shall require that these marginal probabilities be logit models so that when we deal with these characteristics marginally (separately) we can use the theory already developed.

Thus,

(46) (i) $$p_{1.}(i) = \left[1 + e^{\gamma_{1i}}\right]^{-1},$$

(ii) $$p_{.1}(i) = \left[1 + e^{\gamma_{2i}}\right]^{-1}.$$

A way to write $p_{11}(i)$ as a bivariate logit model to include the situation for which η_1 and η_2 are dependent while preserving the requirement that (46) holds, is to write

(45b) $$p_{11}(i) = \left[(1 + e^{\gamma_{1i}})(1 + e^{\gamma_{2i}}) + u(\gamma_{1i}, \gamma_{2i}, \Delta_{12})\right]^{-1},$$

so that $p_{10}(i)$, $p_{01}(i)$, and $p_{00}(i)$ are chosen in such a way that the relationships of Table 2 hold as well as Eq. (46). In order that (45') reduce to (45) when η_1 and η_2 are independent, we require $u_i \equiv u(\gamma_{1i}, \gamma_{2i}, \Delta_{12})$[†] to be so chosen that

(47) $$u(\gamma_{1i}, \gamma_{2i}, 0) = 0.$$

We now investigate further properties of the $p_{rs}(i)$. We require the following five conditions concerning $p_{rs}(i)$.

(i) The probability associated with any one cell is greater than zero and less than one. Hence, for $p_{11}(i)$ we have

$$0 < \left[\left(1 + e^{\gamma_{1i}}\right)\left(1 + e^{\gamma_{2i}}\right) + u_i\right]^{-1} < 1.$$

(ii) The probability, $p_{10}(i)$, can be written explicitly by using the equality $p_{11}(i) + p_{10}(i) = p_{1.}(i)$, which comes from Table 2. Hence,

(48) $$p_{10}(i) = \left(1 + e^{\gamma_{1i}}\right)^{-1} - \left[\left(1 + e^{\gamma_{1i}}\right)\left(1 + e^{\gamma_{2i}}\right) + u_i\right]^{-1}.$$

Since $p_{10}(i)$ has the same restriction as $p_{11}(i)$, we have

$$0 < \left(1 + e^{\gamma_{1i}}\right)^{-1} - \left[\left(1 + e^{\gamma_{1i}}\right)\left(1 + e^{\gamma_{2i}}\right) + u_i\right]^{-1} < 1.$$

[†]The manner in which the $u(\gamma_{1i}, \gamma_{2i}, \Delta_{12})$ is incorporated is by no means unique. The constant covariance term between the two characteristics seems appropriate for the present situation, however, and is in fact much simpler than some more general function such as $\Delta_{12i} = g(\gamma_{1i}, \gamma_{2i})$.

(iii) Similarly, since $p_{11}(i) + p_{01}(i) = p_{.1}(i)$,

$$p_{01}(i) = \left(1 + e^{\gamma_{2i}}\right)^{-1} - \left[\left(1 + e^{\gamma_{1i}}\right)\left(1 + e^{\gamma_{2i}}\right) + u_i\right]^{-1} . \tag{49}$$

Again, $0 < p_{01}(i) < 1$, so that

$$0 < \left(1 + e^{\gamma_{2i}}\right)^{-1} - \left[\left(1 + e^{\gamma_{1i}}\right)\left(1 + e^{\gamma_{2i}}\right) + u_i\right]^{-1} < 1.$$

(iv) Now, $p_{00}(i) = 1 - p_{10}(i) - p_{01}(i) - p_{11}(i)$, with the usual restriction $0 < p_{00}(i) < 1$. Thus,

$$p_{00}(i) = 1 - \left(1 + e^{\gamma_{1i}}\right)^{-1} - \left(1 + e^{\gamma_{2i}}\right)^{-1} + \left[\left(1 + e^{\gamma_{1i}}\right)\left(1 + e^{\gamma_{2i}}\right) + u_i\right]^{-1} \tag{50}$$

and

$$0 < 1 - \left(1 + e^{\gamma_{1i}}\right)^{-1} - \left(1 + e^{\gamma_{2i}}\right)^{-1} + \left[\left(1 + e^{\gamma_{1i}}\right)\left(1 + e^{\gamma_{2i}}\right) + u_i\right]^{-1} < 1.$$

(v) Let ρ_{12}^{i} be the correlation coefficient between the two characteristics η_1 and η_2. Then noting that

$$\rho_{12}^{i} = \frac{\mathrm{cov}(\eta_1, \eta_2 | \text{given } i)}{\sqrt{\mathrm{var}(\eta_1 | \text{given } i)\,\mathrm{var}(\eta_2 | \text{given } i)}} ,$$

we have

$$\rho_{12}^{i} = \frac{\left[\left(1+e^{\gamma_{1i}}\right)\left(1+e^{\gamma_{2i}}\right) + u_i\right]^{-1} - \left[\left(1+e^{\gamma_{1i}}\right)\left(1+e^{\gamma_{2i}}\right)\right]^{-1}}{\sqrt{e^{\gamma_{1i} + \gamma_{2i}} \left(1+e^{\gamma_{1i}}\right)^{-2} \left(1+e^{\gamma_{2i}}\right)^{-2}}} . \tag{51}$$

Since ρ_{12}^{i} is a correlation coefficient it must lie between -1 and 1. Thus,

$$-1 < \frac{\left[\left(1+e^{\gamma_{1i}}\right)\left(1+e^{\gamma_{2i}}\right) + u_i\right]^{-1} - \left[\left(1+e^{\gamma_{1i}}\right)\left(1+e^{\gamma_{2i}}\right)\right]^{-1}}{\sqrt{e^{\gamma_{1i}+\gamma_{2i}} \left(1+e^{\gamma_{1i}}\right)^{-2} \left(1+e^{\gamma_{2i}}\right)^{-2}}} < 1 .$$

We can consolidate conditions (i), (ii), and (iii) into the following theorem.

<u>Theorem 1</u>. Conditions (i), (ii), and (iii) hold if and only if

$$u_i > \max_{k=1,2}\left[-\left(e^{\gamma_{ki}} + e^{\gamma_{1i}+\gamma_{2i}}\right)\right]. \tag{52}$$

<u>Proof</u>. (a) Assume $u_i > \max_{k=1,2}\left[-\left(e^{\gamma_{ki}} + e^{\gamma_{1i}+\gamma_{2i}}\right)\right]$.
Then, $u_i > -\left(e^{\gamma_{1i}} + e^{\gamma_{2i}} + e^{\gamma_{1i}+\gamma_{2i}}\right)$, which implies $\left(1+e^{\gamma_{1i}}\right)\left(1+e^{\gamma_{2i}}\right) + u_i > 1$. Hence, condition (i) follows.

Furthermore,

$$\max_{k=1,2}\left(1+e^{\gamma_{ki}}\right) < \left(1+e^{\gamma_{1i}}\right)\left(1+e^{\gamma_{2i}}\right) + u_i,$$

and since $\left(1+e^{\gamma_{1i}}\right)\left(1+e^{\gamma_{2i}}\right) + u_i > 0$, we have

$$\max_{k=1,2}\left(1+e^{\gamma_{ki}}\right)^{-1} < 1 + \left[\left(1+e^{\gamma_{1i}}\right)\left(1+e^{\gamma_{2i}}\right) + u_i\right]^{1}.$$

Thus, conditions (ii) and (iii) follow.

(b) Assume conditions (i), (ii), and (iii) hold. Then from conditions (ii) and (iii) it follows that

$$u_i > \max_{k=1,2}\left[-\left(e^{\gamma_{ki}} + e^{\gamma_{1i}+\gamma_{2i}}\right)\right].$$

Q.E.D.

<u>Theorem 2</u>. It is necessary and sufficient that

$$\max_{k=1,2}\left[-\left(e^{\gamma_{ki}} + e^{\gamma_{1i}+\gamma_{2i}}\right)\right] < u_i < \frac{\left(1+e^{\gamma_{1i}}\right)\left(1+e^{\gamma_{2i}}\right)}{\left(1-e^{\gamma_{1i}+\gamma_{2i}}\right)}\, e^{(\gamma_{1i}+\gamma_{2i})} \tag{53}$$

in order that conditions (i)-(iv) hold, provided $\gamma_{1i} + \gamma_{2i} < 0$. Otherwise, (52) provides the necessary and sufficient condition.

Proof. (a) Assume $\gamma_{1i} + \gamma_{2i} < 0$, and that (53) holds. Then $u_i > \max_{k=1,2}\left[-\left(e^{\gamma_{ki}} + e^{\gamma_{1i}+\gamma_{2i}}\right)\right]$ implies conditions (i)-(iii) hold. Further, this implies

$$(54)\qquad \left[1+e^{\gamma_{1i}}\right]^{-1} + \left[1+e^{\gamma_{2i}}\right]^{-1} > \left[\left(1+e^{\gamma_{1i}}\right)\left(1+e^{\gamma_{2i}}\right) + u_i\right]^{-1},$$

since

$$\min_{k=1,2}\left[1+e^{\gamma_{ki}}\right]^{-1} > \left[\left(1+e^{\gamma_{1i}}\right)\left(1+e^{\gamma_{2i}}\right) + u_i\right]^{-1}.$$

Finally, $u_i < \dfrac{\left(1+e^{\gamma_{1i}}\right)\left(1+e^{\gamma_{2i}}\right)}{\left(1-e^{\gamma_{1i}+\gamma_{2i}}\right)}\, e^{(\gamma_{1i}+\gamma_{2i})}$ implies that

$$1 + \left[\left(1+e^{\gamma_{1i}}\right)\left(1+e^{\gamma_{2i}}\right) + u_i\right]^{-1} > \left(1+e^{\gamma_{1i}}\right)^{-1} + \left(1+e^{\gamma_{2i}}\right)^{-1},$$

for the case $\gamma_{1i} + \gamma_{2i} < 0$.

(b) Assume $\gamma_{1i} + \gamma_{2i} < 0$, and that conditions (i)-(iv) hold. Then a reversal of the argument in (a) establishes (53).

(c) If $\gamma_{1i} + \gamma_{2i} \geq 0$, it is then easy to show by an argument analogous to the proof of Theorem 1 that (52) provides a necessary and sufficient condition for conditions (i)-(iv) to hold.

Theorem 3. (a) Assume $\gamma_{1i} + \gamma_{2i} < 0$. Then condition (v) holds if and only if

$$(55)\qquad \left(1+e^{\gamma_{1i}}\right)\left(1+e^{\gamma_{2i}}\right)\left[\frac{e^{\left(\frac{\gamma_{1i}+\gamma_{2i}}{2}\right)}}{1-e^{\left(\frac{\gamma_{1i}+\gamma_{2i}}{2}\right)}}\right] \geq u_i$$

$$\geq -\frac{e^{\left(\frac{\gamma_{1i}+\gamma_{2i}}{2}\right)}\left(1+e^{\gamma_{1i}}\right)\left(1+e^{\gamma_{2i}}\right)}{\left[1+e^{\left(\frac{\gamma_{1i}+\gamma_{2i}}{2}\right)}\right]} .$$

(b) Assume $\gamma_{1i} + \gamma_{2i} \geq 0$. Then condition (v) holds if and only if

$$u_i \geq -\frac{e^{\left(\frac{\gamma_{1i}+\gamma_{2i}}{2}\right)}\left(1+e^{\gamma_{1i}}\right)\left(1+e^{\gamma_{2i}}\right)}{\left[1+e^{\left(\frac{\gamma_{1i}+\gamma_{2i}}{2}\right)}\right]}$$

Proof. Rewriting condition (v) we see that

$$-e^{\left(\frac{\gamma_{1i}+\gamma_{2i}}{2}\right)}\left[\left(1+e^{\gamma_{1i}}\right)\left(1+e^{\gamma_{2i}}\right)\right]^{-1} \leq \left[\left(1+e^{\gamma_{1i}}\right)\left(1+e^{\gamma_{2i}}\right)+u_i\right]^{-1}$$

$$-\left[\left(1+e^{\gamma_{1i}}\right)\left(1+e^{\gamma_{2i}}\right)\right]^{-1} \leq e^{\left(\frac{\gamma_{1i}+\gamma_{2i}}{2}\right)}\left[\left(1+e^{\gamma_{1i}}\right)\left(1+e^{\gamma_{2i}}\right)\right]^{-1} .$$

A manipulation of this inequality for the two cases, $\gamma_{1i} + \gamma_{2i} < 0$ and $\gamma_{1i} + \gamma_{2i} \geq 0$ establishes (54) and (55) respectively.

Theorem 4. Suppose $\xi < 0$. Then,

$$(56)\qquad \frac{e^{\xi/2}}{\left(1-e^{\xi/2}\right)} > \frac{e^{\xi}}{1-e^{\xi}} .$$

Proof. (56) follows because $e^{\xi/2} - e^{3/2\xi} > e^{\xi} - e^{3/2\xi}$.

Theorem 5. If $\gamma_{2i} > \gamma_{1i}$, then

$$(57a)\quad e^{\gamma_{1i}}\left(1+e^{\gamma_{2i}}\right) < \frac{e^{\left(\frac{\gamma_{1i}+\gamma_{2i}}{2}\right)}\left(1+e^{\gamma_{1i}}\right)\left(1+e^{\gamma_{2i}}\right)}{1+e^{\left(\frac{\gamma_{1i}+\gamma_{2i}}{2}\right)}} < e^{\gamma_{2i}}\left(1+e^{\gamma_{1i}}\right).$$

Otherwise, if $\gamma_{1i} \geq \gamma_{2i}$,

$$(57b)\quad e^{\gamma_{2i}}\left(1+e^{\gamma_{1i}}\right) \leq \frac{e^{\left(\frac{\gamma_{1i}+\gamma_{2i}}{2}\right)}\left(1+e^{\gamma_{1i}}\right)\left(1+e^{\gamma_{2i}}\right)}{1+e^{\left(\frac{\gamma_{1i}+\gamma_{2i}}{2}\right)}} \leq e^{\gamma_{1i}}\left(1+e^{\gamma_{2i}}\right).$$

Proof. If $\gamma_{2i} > \gamma_{1i}$, then

$$e^{\gamma_{1i}}\left[1+e^{\left(\frac{\gamma_{1i}+\gamma_{2i}}{2}\right)}\right] < e^{\left(\frac{\gamma_{1i}+\gamma_{2i}}{2}\right)}\left(1+e^{\gamma_{1i}}\right).$$

The other inequalities are proved in a similar manner.

Theorem 6. (a) Assume $\gamma_{1i} + \gamma_{2i} < 0$ and $\gamma_{2i} > \gamma_{1i}$. Then conditions (i)-(v) hold provided that

$$(58a)\quad -e^{\gamma_{1i}}\left(1+e^{\gamma_{2i}}\right) < u_i < \frac{\left(1+e^{\gamma_{1i}}\right)\left(1+e^{\gamma_{2i}}\right)}{\left(1-e^{\gamma_{1i}+\gamma_{2i}}\right)}\, e^{\left(\gamma_{1i}+\gamma_{2i}\right)}.$$

If $\gamma_{1i} \geq \gamma_{2i}$, then conditions (i)-(v) hold provided that

$$(58b)\quad -e^{\gamma_{2i}}\left(1+e^{\gamma_{1i}}\right) < u_i < \frac{\left(1+e^{\gamma_{1i}}\right)\left(1+e^{\gamma_{2i}}\right)}{\left(1-e^{\gamma_{1i}+\gamma_{2i}}\right)}\, e^{\left(\gamma_{1i}+\gamma_{2i}\right)}.$$

(b) Assume $\gamma_{1i} + \gamma_{2i} \geq 0$. If $\gamma_{2i} > \gamma_{1i}$, conditions (i)-(v) hold provided that

(59a) $$-e^{\gamma_{1i}}\left(1+e^{\gamma_{2i}}\right) < u_i .$$

If $\gamma_{1i} \geq \gamma_{2i}$, then conditions (i)-(v) hold provided that

(59b) $$-e^{\gamma_{2i}}\left(1+e^{\gamma_{1i}}\right) < u_i .$$

Proof. (a) Theorem 2 provides the upper inequality for u_i in order that conditions (i)-(iv) hold. Theorems 3 and 4 show that no adjustment is required when condition (v) is added. The lower inequality for u_i follows from Theorems 2, 3, and 5, depending on whether $\gamma_{2i} > \gamma_{1i}$ or $\gamma_{2i} \leq \gamma_{1i}$.

(b) If $\gamma_{1i} + \gamma_{2i} \geq 0$, there are no upper bound requirements on u_i with respect to conditions (i)-(v).

We now consider an example incorporating specific assumptions about the form of u. Let

$$u_i = e^{\gamma_{1i}+\gamma_{2i}}\,(e^{\Delta_{12}} - 1).$$

Then

(60) $$p_{11}(i) = \left(1 + e^{\gamma_{1i}} + e^{\gamma_{2i}} + e^{\gamma_{1i}+\gamma_{2i}+\Delta_{12}}\right) .$$

Clearly, $u_i > \max_{k=1,2}\left[-\left(e^{\gamma_{ki}} + e^{\gamma_{1i}+\gamma_{2i}}\right)\right]$, so that the lower inequalities in Theorem 6 are satisfied. We need only be concerned with the upper inequalities if $\gamma_{1i} + \gamma_{2i} < 0$. In this case, we must have

(61) $$e^{\Delta} - 1 < \frac{\left(1+e^{\gamma_{1i}}\right)\left(1+e^{\gamma_{2i}}\right)}{\left(1-e^{\gamma_{1i}+\gamma_{2i}}\right)}$$

satisfied for all γ_{1i} and γ_{2i} such that $\gamma_{1i} + \gamma_{2i} < 0$. Thus, $-\infty < \Delta_{12} < \ln 2$ assures that conditions (i)-(v) are met.

We now sketch the development of the maximum likelihood equation. Suppose we make observations on N individuals, each of which can be measured on a zero-one basis with regard to the two characteristics η_1 and η_2. Let $y_{jk}(i)$ be the random variable that takes the value 1 if the outcome on the i-th individual is (j,k) with respect to η_1 and η_2, respectively, $i = 1, \ldots, N$; $j, k = 0, 1$. Otherwise, $y_{jk}(i)$ takes the value zero.

(62) $$\text{pr}\{y_{jk}(i) = 1\} = p_{jk}(i).$$

Thus, the likelihood L of a sample of size N is equal to

(63) $$\prod_{j,k=0,1} \left\{ \prod_{i=1}^{N} [p_{jk}(i)]^{y_{jk}(i)} \right\},$$

where

(64) $$\sum_{j,k} p_{jk}(i) = \sum_{j,k} y_{jk}(i) = 1.$$

Letting L represent the likelihood and $\mathcal{L} = \ln_e L$, we have

(65) $$\mathcal{L} = \sum_{i=1}^{N} [y_{11}(i) \ln_e p_{11}(i) + y_{10}(i) \ln_e p_{10}(i) + y_{01}(i) \ln_e p_{01}(i) + y_{00}(i) \ln_e p_{00}(i)].$$

Now, Eq. (60) determines $p_{jk}(i)$ as a function of $\underline{\beta}$ and Δ_{12}. Thus, with some algebraic simplification

$$(66) \quad \mathcal{L} = - \sum_{i=1}^{N} \ln_e \left(1 + e^{\gamma_{1i}} + e^{\gamma_{2i}} + e^{\gamma_{1i}+\gamma_{2i}+\Delta_{12}}\right)$$

$$+ \sum_{i=1}^{N} y_{10}(i)\left[\gamma_{2i} + \ln_e \left(1 + e^{\gamma_{1i}+\Delta_{12}}\right) - \ln_e \left(1 + {}^{\gamma_{1i}}\right)\right]$$

$$+ \sum_{i=i}^{N} y_{01}(i)\left[\gamma_{1i} + \ln_e \left(1 + e^{\gamma_{2i}+\Delta_{12}}\right) - \ln_e \left(1 + e^{\gamma_{2i}}\right)\right]$$

$$+ \sum_{i=1}^{N} y_{00}(i)\Big[\left(\gamma_{1i} + \gamma_{2i}\right) + \ln_e \left(2 + e^{\gamma_{1i}} + e^{\gamma_{2i}} - e^{\Delta_{12}}\left(1 = e^{\gamma_{1i}+\gamma_{2i}}\right)\right.$$

$$\left. - \ln_e \left(1 + e^{\gamma_{1i}}\right) - \ln_e \left(1 + e^{\gamma_{2i}}\right)\Big].$$

Recalling that $\gamma_{1i} = [\alpha_0 + \underline{\alpha}'\underline{X}_i]$ and $\gamma_{2i} = [\beta_0 + \underline{\beta}'\underline{Z}_i]$, the equations whose solution yields the maximum likelihood estimators are given by

(67) $$\frac{\partial \mathcal{L}}{\partial \alpha_0} = 0 \quad ,$$

(68) $$\frac{\partial \mathcal{L}}{\partial \hat{\alpha}_i} = 0 \quad ,$$

$i = 1, \ldots, s-1$

(69) $$\frac{\partial \mathcal{L}}{\partial \hat{\beta}_0} = 0 \quad ,$$

(70) $$\frac{\partial \mathcal{L}}{\partial \hat{\beta}_j} = 0 \quad ,$$

and $j = 1, \ldots, t-1$

(71) $$\frac{\partial \mathcal{L}}{\partial \Delta_{12}} = 0 ,$$

where, for example, (67) is

(72) $$\sum_{i=1}^{N} y_{10}(i) + \sum_{i=1}^{N} y_{01}(i)\left\{\frac{e^{\gamma_{2i}+\Delta_{12}}}{1+e^{\gamma_{2i}+\Delta_{12}}} - \frac{e^{\gamma_{2i}}}{1+e^{\gamma_{2i}}}\right\}$$

$$+ \sum_{i=1}^{N} y_{00}(i)\left\{1 + \frac{e^{\gamma_{2i}}\left(1+e^{\gamma_{1i}+\Delta_{12}}\right)}{\left[2+e^{\gamma_{1i}}+e^{\gamma_{2i}}-e^{\Delta_{12}}\left(1-e^{\gamma_{1i}+\gamma_{2i}}\right)\right]} - \frac{e^{\gamma_{2i}}}{\left(1+e^{\gamma_{2i}}\right)}\right\}$$

$$= \sum_{i=1}^{N} \frac{e^{\gamma_{2i}}\left(1+e^{\gamma_{1i}+\Delta_{12}}\right)}{\left(1+e^{\gamma_{1i}}+e^{\gamma_{2i}}+e^{\gamma_{1i}+\gamma_{2i}+\Delta_{12}}\right)} ,$$

Thus, the fitted $p_{jk}(i)$ values denoted by $\hat{p}_{jk}(i)$ are given by

(73) $$\hat{p}_{11}(i) = \frac{1}{1+e^{\hat{\gamma}_{1i}}+e^{\hat{\gamma}_{2i}}+e^{\hat{\gamma}_{1i}+\hat{\gamma}_{2i}+\hat{\Delta}_{12}}} ,$$

(74) $$\hat{p}_{01}(i) = \frac{e^{\hat{\gamma}_{1i}}\left(1+e^{\hat{\gamma}_{2i}+\hat{\Delta}_{12}}\right)}{\left(1+e^{\hat{\gamma}_{1i}}+e^{\hat{\gamma}_{2i}}+e^{\hat{\gamma}_{1i}+\hat{\gamma}_{2i}+\hat{\Delta}_{12}}\right)\left(1+e^{\hat{\gamma}_{2i}}\right)} ,$$

(75) $$\hat{p}_{10}(i) = \frac{e^{\hat{\gamma}_{2i}}\left(1+e^{\hat{\gamma}_{1i}+\hat{\Delta}_{12}}\right)}{\left(1+e^{\hat{\gamma}_{1i}}+e^{\hat{\gamma}_{2i}}+e^{\hat{\gamma}_{1i}+\hat{\gamma}_{2i}+\Delta_{12}}\right)\left(1+e^{\hat{\gamma}_{1i}}\right)} ,$$

and

(76) $$\hat{p}_{00}(i) = \frac{e^{\hat{\gamma}_{1i}+\hat{\gamma}_{2i}}\left[2+e^{\hat{\gamma}_{1i}}+e^{\hat{\gamma}_{2i}}-e^{\hat{\Delta}_{12}}\left(1-e^{\hat{\gamma}_{1i}+\hat{\gamma}_{2i}}\right)\right]}{\left(1+e^{\hat{\gamma}_{1i}}\right)\left(1+e^{\hat{\gamma}_{2i}}\right)\left(1+e^{\hat{\gamma}_{1i}}+e^{\hat{\gamma}_{2i}}+e^{\hat{\gamma}_{1i}+\hat{\gamma}_{2i}+\hat{\Delta}_{12}}\right)} ,$$

$i = 1, 2, \ldots, N.$

In conclusion, we note that besides the predictive properties of $\hat{p}_{jk}(i)$, we can use the likelihood ratio technique to test various hypotheses concerning the parameters. For example, we can test the hypotheses:

(i) $\Delta_{12} = 0$,

(ii) $\beta_{01} = \beta_{02}$,

(iii) $\beta_{p1} = \beta_{p2} = \cdots = \beta_{pn} = 0$,

where $\beta_{p1}, \beta_{p2}, \ldots, \beta_{pn}$ is a subset from $\beta_1, \ldots, \beta_{s-1}$.

V. USE OF THE MODEL: AN EXAMPLE

Logit models provide excellent opportunities for specific types of decisionmaking. If a choice must be made between different courses of action, then the decisionmaker can consider the various probabilities of success as determined by the specified model. The parameters of this model would be estimated from past performance, using the techniques developed in the previous sections. Similarly, if the decision does not involve choosing between alternatives, but rather involves deciding whether to "choose one alternative" or not, the decisionmaker can consider the probability of success of this one course of action as determined by the posited logit model and then base his decision on some *a priori* probability level.

Frequently, the determinations of the probabilities of success of two different classes of alternatives are not independent. Hence, models that consider such possible dependence will be of more value to the decisionmaker. The dependent logit models developed above are such models.

This section presents an example that describes the determination of the logit model parameters under three different assumptions that will be explained. The data for the example are data on Category IV airmen who were admitted to the service in 1960 under a special program. These data were collected into a specific file known as the "Dual 25"[†] data file.

[†]To be admitted to this special program, a potential recruit had to achieve a score of 25 (out of 100) on at least two AQEs. However, a score of 40 (out of 100) on any one AQE would qualify him for "regular enlistment," and hence would preclude participation in the "Dual 25" program.

The basic hypothesis is that the probability of reenlistment after an initial 4-year term by Category IV airmen is related to their initial AFQT score, their race, and their age at enlistment by means of the logistic function. Thus the logit relation can be written

$$\text{(77)} \qquad \text{logit } (r) = \ln\left(\frac{r}{1-r}\right) = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \alpha_3 X_{3i},$$

where X_1 = the AFQT score, X_2 = a dummy variable for race (0 for Negro, 1 for Caucasian), and X_3 = an age scale variable.

If we assume that the reenlistment rate of all enlistees is determined by the independent variables X_1, X_2, and X_3 in the same way, then the parameters of (77) can be estimated straightforwardly by using "simple least squares" regression analysis. The results are

$$\text{(78)} \qquad \text{logit } (r) = \underset{(2.125)}{9.532} - \underset{(-1.907)}{0.326X_1} - \underset{(-2.244)}{1.696X_2} - \underset{(-1.526)}{0.215X_3},$$

$$R^2 = 0.1757,$$

$$(SE)^2 = 0.7509,$$

where the numbers in parentheses are the "t" values associated with the parameter estimates, and

	Mean	Variance
Logit (r)	-0.4711	0.8243
X_1	25.1310	0.6179
X_2	0.5548	0.0596
X_3	4.0095	1.7204

Now assume that the enlistees can be categorized into two groups: those with a high school education and those without. Separating the sample into its appropriate parts, we can estimate (77) for each group. Thus

(1) high school education completed:

$$\text{logit}\ (r) = \underset{(2.671)}{13.420} - \underset{(-2.926)}{0.594X_1} - \underset{(-1.385)}{1.630X_2} + \underset{(1.423)}{0.313X_3}, \tag{79}$$

$$R^2 = 0.3766,$$
$$(SE)^2 = 0.5242,$$

where the numbers in parentheses are the "t" values and

	Mean	Variance
X_1	24.971	0.6687
X_2	0.3762	0.0247
X_3	5.1143	0.7460
Logit (r)	-0.4238	0.8409

(2) high school education <u>not</u> completed:

$$\text{logit}\ (r) = \underset{(1.505)}{8.019} - \underset{(-1.068)}{0.216X_1} - \underset{(-1.309)}{1.124X_2} - \underset{(-2.625)}{0.762X_3}, \tag{80}$$

$$R^2 = 0.3569,$$
$$(SE)^2 = 0.4147,$$

where the "t" values are given in the parentheses and

	Mean	Variance
logit (r)	-0.4714	0.6449
X_1	25.2900	0.5161
X_2	0.7333	0.0308
X_3	2.9048	0.2538

Assume now, however, that the two groups do not function independently. We then estimate parameters of each equation, given this dependency assumption. The residual covariance matrix between Eqs. (79) and (80) is given by

$$\hat{\Sigma}_u = \begin{pmatrix} 0.5242 & 0.1463 \\ 0.1463 & 0.4147 \end{pmatrix}.$$

The inverse of $\hat{\Sigma}_u$ is then used as developed in Secs. II and III to obtain "better"† estimates for the parameters of (79) and (80). Thus we obtain the following estimate:

(1) high school education completed:

(81) $$\text{logit}(r) = \underset{(2.149)}{10.330} - \underset{(-2.497)}{0.487X_1} - \underset{(-1.195)}{1.338X_2} + \underset{(1.767)}{0.373X_3}$$

and

(2) high school education not completed:

(82) $$\text{logit}(r) = \underset{(1.074)}{5.446} - \underset{(-0.595)}{0.114X_1} - \underset{(-0.927)}{0.757X_2} - \underset{(-3.035)}{0.850X_3},$$

where the "t" values are given in parentheses.

Thus depending on the stringency of the assumptions the decision-maker is willing to endure, the best estimates for those assumptions may be obtained as above.

One other case may be illustrated by using this example. Assume that the group of high school graduates is different from, but not independent of, the group of non-high school graduates. Further, assume

---

†The estimates are "better" in the sense manifested in Secs. II and III, specifically Eq. (37).

that the difference is manifested merely by some constant amount, whereas the effect of each independent variable on the probability of success is the same in both groups. In this case we "constrain" the parameter values α_1, α_2, and α_3 of (77) to be the same for the estimates of (81) and (82). Thus we obtain

$$\text{(83)} \qquad \text{logit } (r) = \begin{Bmatrix} 10.050 \\ 10.810 \end{Bmatrix} - 0.383X_1 - 2.085X_2 - 0.025X_3,$$

where the intercept value 10.050 is associated with high school graduates and 10.810 is the intercept for non-high school graduates.

We have used this set of Category IV airmen data to illustrate how the techniques developed in Secs. III and IV can be implemented; and although the various assumptions employed are not necessarily tenable, their use should suffice to demonstrate the possible implementations of the "dependent logit models."

APPENDIX

The derivations of the maximum likelihood, minimum χ^2, and minimum logit χ^2 estimates are, for the most part, well known. Hence the various methods will merely be sketched as an aid to those unfamiliar with that body of literature.

1. MAXIMUM LIKELIHOOD

Let P represent the probability that an event E will occur, and further, let this probability depend on an exogenous variable X by means of the logistic function

(84) $$P = \frac{1}{1 + e^{-(\alpha+\beta X)}} .$$

Construct the following likelihood function from a sample of size n:

(85) $$L = \prod_{i=1}^{n} p_i^{y_i} (1 - p_i)^{1-y_i},$$

where y_i = 0 or 1 for each i.

Then

(86) $$\mathcal{L} = \ln L = \sum_{i=1}^{n} \ln p_i^{y_i} + \sum_{i=1}^{n} \ln (1 - p_i)^{1-y_i}$$

$$= \sum_{i=1}^{n} (1 - y_i) \ln \left(\frac{e^{-(\alpha+\beta X)}}{1 + e^{-(\alpha+\beta X)}} \right) - \sum_{i=1}^{n} y_i \ln (1 + e^{-(\alpha+\beta X)}),$$

and

(87) $$\frac{\partial \mathcal{L}}{\partial \alpha} = \sum_{i=1}^{n} n_i(p_i - \hat{p}_i) = 0,$$

(88) $$\frac{\partial \mathcal{L}}{\partial \beta} = \sum_{i=1}^{n} n_i X_i (p_i - \hat{p}_i) = 0,$$

where $p_i = y_i/n_i$ and $\hat{p}_i = \left[1 + e^{-(\alpha+\beta X_i)}\right]^{-1}$.

2. MINIMUM χ^2.

Let P, the probability that an event will occur, be the same as in (84) above. Consider a sample of k_i from the population and observe the successes, say y_i, in a success-fail mode. The observed probability of a success is then y_i/k_i, and the probability of a failure, $1 - y_i/k_i$. The expected probability of success is $\hat{p}_i$, where $\hat{p}_i$ is defined as above; of failure, $1 - \hat{p}_i$.

The χ^2 statistic can then be written

(89) $$\chi^2 = \sum_{i=1}^{n} \left(\frac{N(y_i/k_i - \hat{p}_i)^2}{\hat{p}_i} + \frac{(\hat{p}_i - y_i/k_i)^2}{1 - \hat{p}_i} \right),$$

and

(90) $$\chi^2 = \sum_{i=1}^{n} \frac{N}{\hat{p}_i(1 - \hat{p}_i)} (y_i/k_i - \hat{p}_i)^2.$$

Let $p_i = y_i/k_i$ and $\hat{p}_i$ be the estimate of P_i, the true probability of success. Then the χ^2 statistic is minimized when the following equations are solved:

(91) $$\frac{\partial \chi^2}{\partial \alpha} = \sum_{i=1}^{n} n_i \frac{\left[\hat{p}_i(1 - p_i) + (1 - \hat{p}_i)p_i\right]}{\hat{p}_i(1 - \hat{p}_i)} (p_i - \hat{p}_i) = 0,$$

(92) $$\frac{\partial \chi^2}{\partial \beta} = \sum_{i=1}^{n} n_i X_i \frac{\left[\hat{p}_i(1 - p_i) + (1 - \hat{p}_i)p_i\right]}{\hat{p}_i(1 - \hat{p}_i)} (p_i - \hat{p}_i) = 0$$

3. MINIMUM LOGIT χ^2

Consider n_i samples out of the population, $i = 1, \ldots, N$. Let the observed proportion of successes p_i be

(93) $$p_i = P_i + u_i,$$

where P_i is the "true proportion for the i^{th} sample, and u_i is a binomially distributed random variable such that

$$E(u_i) = 0,$$
$$Var(u_i) = \frac{P_i(1 - P_i)}{n_i}$$

and

$$E(u_i u_j) = 0 \quad \text{for } i \neq j.$$

We wish to minimize, then,

$$\sum_{i=1}^{n} w_i(p_i - \hat{p}_i)^2,$$

where w_i is inversely proportioned to the variance of p_i, and $\hat{p}_i$ is the estimate of P_i. Observe that for a rather small difference, $(p_i - \hat{p}_i)^2$ can be approximated as follows:

(94) $$(p_i - \hat{p}_i)^2 \simeq [\hat{p}_i(1 - \hat{p}_i)][p_i(1 - p_i)](\ell_i - \hat{\ell}_i)^2,$$

where $\ell_i = \ln(p_i/(1 - p_i))$, and $\hat{\ell}_i = \alpha + \beta X_i$.

Consider, then, (90) above:

(95) $$\chi^2 = \sum_{i=1}^{n} \frac{n_i}{\hat{p}_i(1 - \hat{p}_i)} (p_i - \hat{p}_i)^2 \simeq \sum_{i=1}^{n} n_i p_i(1 - p_i)(\ell_i - \hat{\ell}_i)^2.$$

Since all the terms in the approximation are linear, ordinarily least-square regressions with weights $w_i\ [p_i(1 - p_i)]^{\frac{1}{2}}$ can be used to obtain estimates of α and β, and subsequently, $\hat{p}_i$ for all i. Thus

$$\frac{\partial \chi^2}{\partial \alpha} = \sum n_i(1 - p_i)p_i(\ell_i - \hat{\ell}_i) = 0, \tag{96}$$

$$\frac{\partial \chi^2}{\partial \beta} = \sum n_i x_i(1 - P_i)p_i(\ell_i - \hat{\ell}_i) = 0. \tag{97}$$

Equations (96) and (97) are the normal equations for ordinary linear regression.

REFERENCES

1. Aitken, A. C., "On Least-squares and Linear Combination of Observations," Proceedings of the Royal Society of Edinburgh, 55, 1934-35, pp. 42-48.

2. Berkson, J., "A Statistically Precise and Relatively Simple Method of Estimating the Bioassay with Quantal Response, Based on the Logistic Function," Journal of the American Statistical Association, 48, 1953, pp. 565-599.

3. ----, "Minimum χ^2 and Maximum Likelihood Solution in Terms of a Linear Transform, with Particular Reference to Bioassay," Journal of the American Statistical Association, 44, 1949, pp. 273-278.

4. ----, "Approximation of Chi-square by 'Probits' and 'Logits'," Journal of the American Statistical Association, 41, 1946, pp. 70-74.

5. ----, "Application of the Logistic Function to Bioassay," Journal of the American Statistical Association, 39, 1944, pp. 357-365.

6. ----, "Maximum Likelihood and Minimum χ^2 Estimates of the Logistic Function," Journal of the American Statistical Association, 50, 1955, pp. 130-162.

7. ----, "Tables for the Maximum Likelihood Estimates of the Logistic Function," Biometrics, 13, 1957, pp. 28-34.

8. Cox, D. R., "The Regression Analysis of Binary Sequences," Journal of the Royal Statistical Society, Series B, 20, 1958, pp. 215-242.

9. Donaldson, T. S., and A. F. Sweetland, Trends in F-101 Aircraft Maintenance Requirements, The RAND Corporation, RM-4930-PR, April 1966.

10. Dyke, G. V., and H. D. Patterson, "Analysis of Factorial Arrangements when the Data are Proportions," Biometrics, 8, 1952, pp. 1-12.

11. McCall, J. J., and N. Wallace, Training and Retention of Air Force Airmen: An Economic Analysis, The RAND Corporation, RM-5384-PR, August 1967.

12. McGlothlin, W. H., and T. S. Donaldson, Trends in Aircraft Maintenance Requirements, The RAND Corporation, RM-4049-PR, June 1964.

13. Walker, S. H., and D. B. Duncan, "Estimation of the Probability of an Event as a Function of Several Independent Variables," Biometrics, 1967, pp. 167-180.

14. Zellner, A., "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," Journal of the American Statistical Association, June 1962, pp. 348-368.

15. ----, and T. H. Lee, "Joint Estimation of Relationships Involving Discrete Random Variables," Econometrics, April 1965, pp. 382-394.

DOCUMENT CONTROL DATA

1 ORIGINATING ACTIVITY	2a. REPORT SECURITY CLASSIFICATION
THE RAND CORPORATION	UNCLASSIFIED
	2b. GROUP

3. REPORT TITLE

ESTIMATION TECHNIQUES FOR DEPENDENT LOGIT MODELS

4. AUTHOR(S) (Last name, first name, initial)

Cook, Jr., A. A. and A. J. Gross

5. REPORT DATE	6a. TOTAL No. OF PAGES	6b. No. OF REFS.
November 1968	59	15

7. CONTRACT OR GRANT No.	8. ORIGINATOR'S REPORT No.
F44620-67-C-0045	RM-5734-PR

9a. AVAILABILITY/LIMITATION NOTICES	9b. SPONSORING AGENCY
DDC-1	United States Air Force Project RAND

10. ABSTRACT

A discussion of the estimation techniques of logit models whose response variables are not necessarily statistically independent. The techniques developed and used include the Generalized Least Squares approach of Zellner; stepwise regression and its relevance in estimating logit model parameters; and the "contingency table" concept of fixed marginal probabilities, which is used to develop a bivariate logit model that possesses certain statistical and probabilistic properties concerning the correlation coefficient and the summation of individual cell probabilities to unity. The techniques are designed for application to two Air Force problems: The first problem uses concomitant information to estimate the probability that first-term airmen reenlist; the second, information concerning flying missions to estimate the probabilities that interdependent aircraft subsystems fail during these missions. Some applications to the problems are presented, but this study is mainly directed toward development of the necessary techniques. An example illustrating the various techniques is given.

11. KEY WORDS

Reliability
Maintenance
Statistical methods and processes
Air Force
Personnel
Models