THE LINEAR STRUCTURAL RELATION FOR SEVERAL SETS OF DATA

by

NEIL RALPH COX

Thesis submitted for the degree of Doctor of Philosophy in the University of London and for the D.I.C. at Imperial College.

Imperial College of Science and Technology, August, 1975.

ABSTRACT

The structural relationship is studied between variables with observations coming from two or more populations.

Initially we examine in detail the structural relationship between two variables with observations from two populations, or groups. The maximum likelihood estimates of the unknown parameters are derived, tests of the adequacy of the simplest model against various more general alternatives are found, and, finally, tests of hypotheses about the parameter of principal interest, the slope of the relationship, are considered.

An extension of the two-group, two-variable case to permit an arbitrary number of groups is then considered. All the aspects studied in the two-group case are again considered in this generalization.

Finally, a multivariate generalization is considered. In this model an arbitrary number of parameters and groups are permitted though identifiability considerations restrict the number of variates that can be included in relation to the number of groups. Only the difficult problem of finding the maximum likelihood estimates of the unknown parameters is considered in this model.

The theoretical results are applied as part of the analysis of an experiment on apple trees.

ACKNOWLEDGEMENTS

I am extremely grateful to my supervisor, Professor D.R. Cox, for suggesting the research described in this thesis and for his constant invaluable guidance. I would also like to thank Mr M. Pearce for providing computer routines for the calculation of univariate and bivariate normal probabilities, and the other members of the staff and fellow students whom I consulted on various topics.

This research was supported throughout by a New Zealand National Research Advisory Council Fellowship and I am most grateful to the Council and to the New Zealand Ministry of Agriculture and Fisheries for granting me the necessary study leave.

My thanks also go to Dr S.C. Pearce for providing the data used in Chapter 5.

CONTENTS

LIST OF TABLES AND ILLUSTRATIONS

Chapter 1

INTRODUCTION

This thesis deals with structural models for the relationships between several variables. Various models have been considered for studying such relationships; these were outlined in Moran's (1971) review paper on functional and structural relationships. In all these models we wish to study something about the relationship between a dependent variable, Y, and one or more independent variables, X_i.

The most commonly studied models are the regression model, which relates a random variable to several fixed variables, and the multivariate normal regression in which the variables are assumed to have a joint multivariate normal distribution. A third model, studied by Berkson (1950), assumes that the observations on the independent variables are target values for some quantity in an experiment but because of random variation these target values are not actually attained. Finally, there are the errors-in-variables models; that is, the functional and structural relationships. In these both the dependent and the independent variables are measured with error and it is the relationship between the true, unobserved, variables which is supposed of interest.

Suppose that we have observations (Y_i,X_i) (for simplicity we consider just one independent variable, X) which are derived from the variables (V_i,U_i) by the addition of random errors (ε_i,δ_i). That is,

$$Y_i = V_i + \varepsilon_i \quad \text{and} \quad X_i = U_i + \delta_i \ .$$

We assume that the true variables, V and U, are connected by the linear relation

$$V = \alpha + \beta U \ .$$

What analysis is appropriate depends on the reasons for which we are studying the relation. If we are really interested only in studying the relationship between the observed variables (Y_i,X_i), for example for the purposes of the prediction of a value of Y corresponding to an observed value of X, then we should use a regression analysis. If, however, we are interested in the underlying linear relationship itself then we should use one of the other models. If the errors, ε and δ, can be assumed to be independent of U then we use either the functional or the structural relationship while if, on the other hand, we can assume that the errors are independent of the observed values, X_i, then we use the Berkson model. The first case is of interest here. By assuming that the variables U_i are fixed we get the functional relationship model, and by assuming that they are independent random variables we get the structural relationship model. We shall now concentrate on the structural relationship.

Let us assume that the errors, ε and δ, have normal distributions with zero means and unknown variances, that U has a normal distribution with unknown mean and variance and, further, that the ε_i and δ_i are independent of each other and of the U_i. These are the usual assumptions in the structural relationship model but it is well known (see, for instance, Moran, 1971 or Kendall and Stuart, 1973, Ch.29) that the parameters are unidentifiable in this model. If we can assume non-normality of U or the errors, or if we have additional knowledge about some of the parameters (for instance, knowledge of one or both of the error variances or their ratio), then we may overcome this unidentifiability; Moran (1971) reviews this topic, so that no additional references need be given here.

The additional information can also take the form of an instrumental variable; if this variable, W, is correlated with U but not with ε or δ then we can find a consistent estimator for the slope, β (see, for instance, Moran, 1971). When the instrumental variable takes only the values ± 1 and so groups the observations then the relationship is estimated by the line joining the sample means of the two groups; see, for instance, Kendall and Stuart (1973), Ch.29. This estimate was suggested also by Wald (1940) when the additional information takes the form of particular knowledge of the ranking of the unobserved U_i. Richardson and Wu (1970) also study this model when the observations are known to belong to several groups and they suggest using the ordinary regression estimate for β but based on the group means rather than the individual observations. For two groups this is the same as Wald's estimate. In this thesis the whole question of the analysis, via the structural relationship model, of data from several groups is examined in much more detail.

In Chapter 2 we assume that there are just two groups. We estimate the parameters by maximum likelihood (ML) in §2.2 where it is noted that, while Wald's estimate (ie the between-means slope) for the slope is found as the solution of the likelihood equations, it is not always the ML estimate and it is not, by itself, a satisfactory estimate of the slope. We find that the slope estimator is bounded by the two regression slopes (of Y on X and of X on Y) so that the limiting behaviour as the two groups merge into one is the same as that for the one-group case as described by Moran (1956). Asymptotic estimates of the variances of the ML estimators are derived in §2.3. In §2.4 we consider tests of the adequacy of the model; in particular we want to know whether the within-group variances are equal and if it is reasonable to assume that the observations are scattered about a common line rather than about a different line for each group. The final section of Chapter 2 is devoted to the question of testing hypotheses about the slope.

The results of Chapter 2 are extended, in Chapter 3, to the more general case in which the observations belong to $k \geq 2$ groups. Many of the results carry over with only minor modifications.

In Chapter 4 we consider a further extension of the model by studying the relationship between $p+1$ $(p \geq 1)$ random variables. Only the question of estimating the parameters has been considered in this model and we are led to the interesting result that the vector of slopes is estimated by the last canonical variate of a between-and-within canonical regression analysis.

In Chapter 5 the theoretical results from the earlier chapters are applied in the analysis of an experiment on apple trees. Two measurements were made on each tree in the experiment, one providing a measurement of cambial activity and the other providing a measurement of apical activity. The trees were grouped according to thirteen different rootstocks.

A new procedure that enables relatively easy determination of the nature of turning points of likelihoods is discussed in Appendix 1. When a solution has been found to likelihood equations it is then necessary to see if it is at a maximum; this is particularly true when the likelihood equations have more than one solution. Virtually all the likelihoods studied in this thesis have more than one turning point and the method described in Appendix 1 has considerably simplified the problem of finding the maxima.

In the final two appendices we outline the procedures for finding the moments of various sums of squares and products and various regression slopes that are required in different parts of the thesis.

Chapter 2

THE TWO-GROUP, TWO-VARIABLE MODEL

2.1 Specification of the model

Through the structural relationship model we wish to investigate the underlying relationship between two or more variables observed with error. For instance, in the example that we shall analyse in Chapter 5, we want to study the relationship between cambial and apical activity in apple trees but we can only get imprecise measurements of these variables through the measurement of the trunk girth and the weight of the tree above ground. In this chapter we consider the simplest case, the linear structural relationship between two variables sampled from two groups or populations.

Let the two variables of interest be U and V. We assume that these variables are linearly related,

$$V = \alpha + \beta U,$$

and it is this relationship which is of interest. However, we actually observe the variables X and Y which differ from their "true" values, U and V, by random errors;

$$X = U + \delta$$

and $$Y = V + \varepsilon.$$

Let X_{ij} and Y_{ij} be random variables representing observations on the j'th member of the i'th group and let the number of members of group i be n_i. Then

$$Y_{ij} = \alpha + \beta U_{ij} + \varepsilon_{ij}$$

$$X_{ij} = U_{ij} + \delta_{ij} \quad (j=1,..,n_i;\ i=1,2),$$

where we assume that the U_{ij} have independent normal distributions with means μ_i and variance σ^2, $N(\mu_i,\sigma^2)$, the δ_{ij} are independently distributed as $N(0,\sigma_\delta^2)$, the ε_{ij} are independently distributed as $N(0,\sigma_\varepsilon^2)$, and the errors, ε_{ij} and δ_{ij}, are independent of each other and of the U_{ij}. We shall later refer to this model as model I.

This model assumes that the observations in the two groups are scattered about a common line rather than about a different line for each group, the groups differing only in the mean. In other words, the between-group slope is the same as the common within-group slope. In §2.4 we relax some of the restrictions on the model, mainly for the purpose of testing the adequacy of the model for a particular set of data against certain more general models.

The assumptions of model I imply that the observations (Y_{ij},X_{ij}) have independent normal distributions with means $(\alpha+\beta\mu_i,\mu_i)$ and a common covariance matrix

$$\underset{\sim}{\Sigma} = \begin{pmatrix} \beta^2\sigma^2+\sigma_\varepsilon^2 & \beta\sigma^2 \\ \beta\sigma^2 & \sigma^2+\sigma_\delta^2 \end{pmatrix}.$$

Let $$n = n_1 + n_2,$$

$$S_1 = \sum_{i=1}^{2}\sum_{j=1}^{n_i}(y_{ij} - \alpha - \beta\mu_i)^2,$$

$$S_2 = \sum\sum(y_{ij} - \alpha - \beta\mu_i)(x_{ij} - \mu_i),$$

and $$S_3 = \sum\sum(x_{ij} - \mu_i)^2.$$

The log-likelihood of the observations $\{(y_{ij},x_{ij});j=1,..,n_i;\ i=1,2\}$ is thus

$$\ell = -n.\log(2\pi) - \tfrac{1}{2}n.\log|\underset{\sim}{\Sigma}| - \frac{1}{2|\underset{\sim}{\Sigma}|}\{(\sigma^2+\sigma_\delta^2)S_1 - 2\beta\sigma^2 S_2 + (\beta^2\sigma^2+\sigma_\varepsilon^2)S_3\}, \qquad (1)$$

where $$|\underset{\sim}{\Sigma}| = \beta^2\sigma^2\sigma_\delta^2 + \sigma^2\sigma_\varepsilon^2 + \sigma_\delta^2\sigma_\varepsilon^2.$$

Note that the parameter space is restricted by the inequalities $\sigma^2\geq 0$, $\sigma_\delta^2\geq 0$, $\sigma_\varepsilon^2\geq 0$ and $|\underset{\sim}{\Sigma}|>0$. The last inequality ensures that no two variances can be zero simultaneously; this would make the model trivial and the above expression for the log-likelihood would be undefined.

2.2 Maximum likelihood estimation of the parameters

To facilitate the determination of the nature of the solutions of the likelihood equations, the equations will be solved sequentially. That is, the first likelihood equation will be solved to get an estimate of the first parameter in terms of the remaining parameters, this estimate will be substituted into the second equation which will then be solved for the second parameter in terms of the remaining parameters, and so on. We will then be able to calculate the pivots of the matrix of double derivatives of the log-likelihood. In fact, the j'th pivot is found from the derivative of the log-likelihood with respect to the j'th parameter by substituting the estimates obtained for the previous parameters and then differentiating again with respect to the j'th parameter. If all the pivots are negative when evaluated at a solution of the likelihood equations then that solution is at a local maximum of the likelihood, if all are positive then the solution is at a local minimum, and if any two have opposite signs then the solution is at a saddle-point. The method is described in more detail in Appendix 1.

So that complicated notation involving circumflexes, tildes, etc, can be avoided, the notation used below will be a little unconventional but it should not lead to any confusion. For instance, the expression in equation (2) below is to be interpreted as an estimate of μ_i when α, β, σ_δ^2 and σ_ε^2 are known. If α, β, σ_δ^2 and σ_ε^2 are replaced by their ML estimates, then the expression in equation (2) becomes the ML estimate of μ_i.

The log-likelihood of the observations is given by equation (1). Differentiating it with respect to μ_i, we get

$$\partial \ell/\partial \mu_i = n_i\{\beta\sigma_\delta^2(\bar{y}_{i.} - \alpha - \beta\mu_i) + \sigma_\varepsilon^2(\bar{x}_{i.} - \mu_i)\}/|\underset{\sim}{\Sigma}|.$$

Equating this to zero we find

$$\mu_i = \{\beta\sigma_\delta^2(\bar{y}_{i.} - \alpha) + \sigma_\varepsilon^2\bar{x}_{i.}\}/(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2) \quad (i=1,2). \qquad (2)$$

The first two pivots are

$$\partial^2\ell/\partial\mu_i^2 = -n_i(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)/|\underset{\sim}{\Sigma}| \quad (i=1,2)$$

which are both negative for any choice of β, σ^2, σ_δ^2 and σ_ε^2 in the parameter space.

Differentiating the log-likelihood with respect to α we get

$$\partial\ell/\partial\alpha = n\{(\sigma^2+\sigma_\delta^2)(\bar{y}_{..} - \alpha - \beta\bar{\mu}) - \beta\sigma^2(\bar{x}_{..} - \bar{\mu})\}/|\underset{\sim}{\Sigma}|,$$

where $\bar{\mu} = \Sigma n_i\mu_i/n$. Substituting for μ_i from equation (2), we find that $\partial\ell/\partial\alpha$ becomes simply

$$n(\bar{y}_{..} - \alpha - \beta\bar{x}_{..})/(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2). \qquad (3)$$

Equating (3) to zero gives us

$$\alpha = \bar{y}_{..} - \beta\bar{x}_{..} \qquad (4)$$

and, on substitution into equation (2), we get

$$\mu_i = \{\beta\sigma_\delta^2(\bar{y}_{i.} - \bar{y}_{..} + \beta\bar{x}_{..}) + \sigma_\varepsilon^2\bar{x}_{i.}\}/(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2) \quad (i=1,2).(5)$$

The third pivot is the derivative of (3) with respect to α; this is just $-n/(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)$, which is again negative for any choice of β, σ_δ^2 and σ_ε^2 in the parameter space.

Differentiating the log-likelihood with respect to σ^2 we get

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{n(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)}{2|\underset{\sim}{\Sigma}|} + \frac{1}{2|\underset{\sim}{\Sigma}|^2}\left(\beta^2\sigma_\delta^4 S_1 + 2\beta\sigma_\delta^2\sigma_\varepsilon^2 S_2 + \sigma_\varepsilon^4 S_3\right).$$

When α and μ_i are replaced by the expressions in equations (4) and (5) we find that

$$\bar{y}_{i.} - \alpha - \beta\mu_i = \sigma_\varepsilon^2\{(\bar{y}_{i.} - \bar{y}_{..}) - \beta(\bar{x}_{i.} - \bar{x}_{..})\}/(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)$$

and $\bar{x}_{i.} - \mu_i = -\beta\sigma_\delta^2\{(\bar{y}_{i.} - \bar{y}_{..}) - \beta(\bar{x}_{i.} - \bar{x}_{..})\}/(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)$,

and hence that

$$S_1 = ns_{yy} + n\sigma_\varepsilon^4 B(\beta)/(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)^2,$$

$$S_2 = ns_{yx} - n\beta\sigma_\delta^2\sigma_\varepsilon^2 B(\beta)/(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)^2,$$

and $$S_3 = ns_{xx} + n\beta^2\sigma_\delta^4 B(\beta)/(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)^2,$$

where $$B(\beta) = b_{yy} - 2\beta b_{yx} + \beta^2 b_{xx},$$

$$b_{yy} = \sum_i\sum_j (\bar{y}_{i.} - \bar{y}_{..})^2/n ,$$

$$s_{yy} = \sum_i\sum_j (y_{ij} - \bar{y}_{i.})^2/n ,$$

and similarly for b_{yx}, b_{xx}, s_{yx} and s_{xx}.

Hence, after substitution of the expressions in equations (4) and (5) for α and μ_i, $\partial\ell/\partial\sigma^2$ becomes

$$-\frac{n(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)}{2|\underset{\sim}{\Sigma}|} + \frac{n}{2|\underset{\sim}{\Sigma}|^2}\left(\beta^2\sigma_\delta^4 s_{yy} + 2\beta\sigma_\delta^2\sigma_\varepsilon^2 s_{yx} + \sigma_\varepsilon^4 s_{xx}\right). \qquad (6)$$

Equating this to zero gives us

$$\sigma^2 = \frac{\beta^2\sigma_\delta^4 s_{yy} + 2\beta\sigma_\delta^2\sigma_\varepsilon^2 s_{yx} + \sigma_\varepsilon^4 s_{xx} - \sigma_\delta^2\sigma_\varepsilon^2(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)}{(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)^2}. \qquad (7)$$

The fourth pivot is the derivative of (6) with respect to σ^2. Since, at any turning point, (6) is itself equal to zero, this pivot, when evaluated at any turning point, can be shown to equal

$$-n(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)^2/(2|\underset{\sim}{\Sigma}|^2),$$

which is negative for any choice of β, σ^2, σ_δ^2 and σ_ε^2 in the parameter space.

Differentiating the log-likelihood with respect to σ_δ^2 we get

$$\frac{\partial\ell}{\partial\sigma_\delta^2} = -\frac{n(\beta^2\sigma^2+\sigma_\varepsilon^2)}{2|\underset{\sim}{\Sigma}|} + \frac{1}{2|\underset{\sim}{\Sigma}|^2}\{\beta^2\sigma^4 S_1 - 2\beta\sigma^2(\beta^2\sigma^2+\sigma_\varepsilon^2)S_2$$

$$+ (\beta^2\sigma^2+\sigma_\varepsilon^2)^2 S_3\}.$$

When α, μ_i and σ^2 are replaced by the expressions in equations (4), (5) and (7), $\partial\ell/\partial\sigma_\delta^2$ becomes

$$\frac{n\{\beta^2 T(\beta)h(\beta,\sigma_\delta^2,\sigma_\varepsilon^2) + \sigma_\varepsilon^4(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)W(\beta) - s_{yy}(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)^3\}}{2(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)^2 h(\beta,\sigma_\delta^2,\sigma_\varepsilon^2)}, \qquad (8)$$

where $W(\beta) = s_{yy} - 2\beta s_{yx} + \beta^2 s_{xx}$, $T(\beta) = W(\beta) + B(\beta)$, and $\quad h(\beta,\sigma_\delta^2,\sigma_\varepsilon^2) = \beta^2\sigma_\delta^4 s_{yy} + 2\beta\sigma_\delta^2\sigma_\varepsilon^2 s_{yx} + \sigma_\varepsilon^4 s_{xx}$.

Equating (8) to zero we find that σ_δ^2 is a solution of the cubic equation

$$\beta^2 T(\beta)h(\beta,\sigma_\delta^2,\sigma_\varepsilon^2) + \sigma_\varepsilon^4(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)W(\beta) - s_{yy}(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)^3 = 0. \quad (9)$$

While this equation can, in priniple, be solved explicitly, the solutions are too complicated to be of any use. However, we can avoid the problem. We shall carry on to the next stage of the solution of the equations.

Differentiation of the log-likelihood with respect to σ_ε^2 gives us

$$\frac{\partial \ell}{\partial \sigma_\varepsilon^2} = -\frac{n(\sigma^2+\sigma_\delta^2)}{2|\underset{\sim}{\Sigma}|} + \frac{1}{2|\underset{\sim}{\Sigma}|^2}\{(\sigma^2+\sigma_\delta^2)^2 S_1 - 2\beta\sigma^2(\sigma^2+\sigma_\delta^2)S_2 + \beta^2\sigma^4 S_3\}.$$

Rather than attempting to solve this by direct substitution, let us consider

$$-\frac{\sigma_\delta^2}{(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)}\frac{\partial \ell}{\partial \sigma^2} + \frac{\sigma_\delta^2}{\sigma_\varepsilon^2}\frac{\partial \ell}{\partial \sigma_\delta^2} + \frac{\partial \ell}{\partial \sigma_\varepsilon^2} = -\frac{n}{2\sigma_\varepsilon^2} + \frac{(S_1 - 2\beta S_2 + \beta^2 S_3)}{2\sigma_\varepsilon^2(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)}.$$

If we substitute for α and μ_i from equations (4) and (5) this expression becomes

$$-\frac{n}{2\sigma_\varepsilon^2} + \frac{nT(\beta)}{2\sigma_\varepsilon^2(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)}. \quad (10)$$

Equating (10) to zero gives us $\beta^2\sigma_\delta^2+\sigma_\varepsilon^2 = T(\beta)$. In a similar manner we can show that $\beta^2\sigma^2+\sigma_\varepsilon^2 = S_1/n$,

$\beta\sigma^2 = S_1/n$, and $\sigma^2+\sigma_\delta^2 = S_3/n$. Substituting for α and μ_i from equations (4) and (5) we can easily solve the resulting equations to get, for $\beta \neq 0$,

$$\sigma^2 = \frac{s_{yx}}{\beta} - \frac{(\beta s_{xx} - s_{yx})(s_{yy} - \beta s_{yx})B(\beta)}{\beta W^2(\beta)}, \qquad (11)$$

$$\sigma_\delta^2 = \frac{(\beta s_{xx} - s_{yx})T(\beta)}{\beta W(\beta)} = \hat{\sigma}_\delta^2(\beta), \text{ say,} \qquad (12)$$

and

$$\sigma_\varepsilon^2 = \frac{(s_{yy} - \beta s_{yx})T(\beta)}{W(\beta)} = \hat{\sigma}_\varepsilon^2(\beta), \text{ say.} \qquad (13)$$

Substitution of $\hat{\sigma}_\delta^2(\beta)$ and $\hat{\sigma}_\varepsilon^2(\beta)$ for σ_δ^2 and σ_ε^2 in equation (5) gives us

$$\mu_i = \frac{(\beta s_{xx} - s_{yx})(\bar{y}_{i.} - \bar{y}_{..} + \beta \bar{x}_{..}) + (s_{yy} - \beta s_{yx})\bar{x}_{i.}}{W(\beta)} \quad (i=1,2) \qquad (14)$$

We now return to the problem of finding the fifth pivot. It is the derivative of (8) with respect to σ_δ^2. Hence the fifth pivot, evaluated when $\sigma_\delta^2 = \hat{\sigma}_\delta^2(\beta)$ and $\sigma_\varepsilon^2 = \hat{\sigma}_\varepsilon^2(\beta)$, is

$$-\frac{n\beta^2\{\beta^2(s_{yy}s_{xx} - s_{yx}^2) + 2(s_{yy} - \beta s_{yx})^2\}}{2T^2(\beta)(s_{yy}s_{xx} - s_{yx}^2)},$$

which is negative for any real β.

It is the sixth pivot which is more tricky, because we were unable to get an estimate for σ_δ^2 in terms of σ_ε^2 and β without using the likelihood equation for σ_ε^2. But let us suppose that $\tilde{\sigma}_\delta^2(\sigma_\varepsilon^2,\beta)$ is the solution of equation (9) for which $\tilde{\sigma}_\delta^2(\hat{\sigma}_\varepsilon^2(\beta),\beta) = \hat{\sigma}_\delta^2(\beta)$.

Note that when $\sigma_\varepsilon^2 = \hat{\sigma}_\varepsilon^2(\beta)$, equation (9) has just one real root, $\sigma_\delta^2 = \hat{\sigma}_\delta^2(\beta)$. Substitution for α, μ_i and σ^2 from equations (4), (5) and (7) and substitution of $\tilde{\sigma}_\delta^2(\sigma_\varepsilon^2,\beta)$ for σ_δ^2 in $\partial\ell/\partial\sigma_\varepsilon^2$ gives us expression (10) with σ_δ^2 replaced by $\tilde{\sigma}_\delta^2(\sigma_\varepsilon^2,\beta)$; the sixth pivot is the derivative of this with respect to σ_ε^2. We thus find that the sixth pivot, evaluated when $\sigma_\varepsilon^2 = \hat{\sigma}_\varepsilon^2(\beta)$, is

$$-\frac{nW(\beta)\{\beta^2 g(\beta) + 1\}}{2(s_{yy} - \beta s_{yx})T^2(\beta)}, \qquad (15)$$

where $$g(\beta) = \left\{ \frac{\partial}{\partial\sigma_\varepsilon^2}\tilde{\sigma}_\delta^2(\sigma_\varepsilon^2,\beta) \right\}_{\sigma_\varepsilon^2=\hat{\sigma}_\varepsilon^2(\beta)}.$$

To find $g(\beta)$ we replace σ_δ^2 by $\tilde{\sigma}_\delta^2(\sigma_\varepsilon^2,\beta)$ in the cubic equation (9) and differentiate with respect to σ_ε^2. From this we get a linear equation for $\partial\tilde{\sigma}_\delta^2(\sigma_\varepsilon^2,\beta)/\partial\sigma_\varepsilon^2$ in terms of $\tilde{\sigma}_\delta^2(\sigma_\varepsilon^2,\beta)$, σ_ε^2 and β. Hence, since we know that $\tilde{\sigma}_\delta^2\{\hat{\sigma}_\varepsilon^2(\beta),\beta\} = \hat{\sigma}_\delta^2(\beta)$, we can easily find $g(\beta)$. Substituting $g(\beta)$ into expression (15), we find that the sixth pivot, evaluated at any turning point, is

$$\frac{-nW^2(\beta)}{T^2(\beta)\{\beta^2(s_{yy}s_{xx} - s_{yx}^2) + 2(s_{yy} - \beta s_{yx})^2\}},$$

which is negative for any real β.

Differentiating the log-likelihood with respect to β we get

$$\frac{\partial \ell}{\partial \beta} = -\frac{n\beta\sigma^2\sigma_\delta^2}{|\underset{\sim}{\Sigma}|} + \frac{\sigma^2}{|\underset{\sim}{\Sigma}|^2}\{\beta\sigma_\delta^2(\sigma^2+\sigma_\delta^2)S_1 + (-\beta^2\sigma^2\sigma_\delta^2 + \sigma^2\sigma_\varepsilon^2 + \sigma_\delta^2\sigma_\varepsilon^2)S_2$$

$$- \beta\sigma^2\sigma_\varepsilon^2 S_3\} + \frac{1}{|\underset{\sim}{\Sigma}|}\{(\sigma^2+\sigma_\delta^2)\sum_i n_i\mu_i(\bar{y}_{i.} - \alpha - \beta\mu_i)$$

$$- \beta\sigma^2\sum_i n_i\mu_i(\bar{x}_{i.} - \mu_i)\} \quad . \qquad (16)$$

When α and μ_i are replaced by the expressions in equations (4) and (5) we find that

$$\sum_i n_i\mu_i(\bar{y}_{i.} - \alpha - \beta\mu_i) = \frac{n\sigma_\varepsilon^2\{\beta\sigma_\delta^2(b_{yy} - \beta b_{yx}) + \sigma_\varepsilon^2(b_{yx} - \beta b_{xx})\}}{(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)^2}$$

and

$$\sum_i n_i\mu_i(\bar{x}_{i.} - \mu_i) = -\frac{n\beta\sigma_\delta^2\{\beta\sigma_\delta^2(b_{yy} - \beta b_{yx}) + \sigma_\varepsilon^2(b_{yx} - \beta b_{xx})\}}{(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)^2}.$$

We showed previously that when $\partial\ell/\partial\sigma^2 = \partial\ell/\partial\sigma_\delta^2 = \partial\ell/\partial\sigma_\varepsilon^2 = 0$, then $S_1 = n(\beta^2\sigma^2+\sigma_\varepsilon^2)$, $S_2 = n\beta\sigma^2$ and $S_3 = n(\sigma^2+\sigma_\delta^2)$. Substituting all these into equation (16), we find that $\partial\ell/\partial\beta$ becomes simply

$$n\{\beta\sigma_\delta^2(b_{yy} - \beta b_{yx}) + \sigma_\varepsilon^2(b_{yx} - \beta b_{xx})\}/(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)^2, \qquad (17)$$

which, on substitution of $\hat{\sigma}_\delta^2(\beta)$ and $\hat{\sigma}_\varepsilon^2(\beta)$ for σ_δ^2 and σ_ε^2 becomes

$$\frac{n\{(\beta s_{xx} - s_{yx})(b_{yy} - \beta b_{yx}) + (s_{yy} - \beta s_{yx})(b_{yx} - \beta b_{xx})\}}{T(\beta)W(\beta)}. \qquad (18)$$

Hence β is chosen as a solution of the quadratic equation

$$f(\beta) = (\beta s_{xx} - s_{yx})(b_{yy} - \beta b_{yx}) + (s_{yy} - \beta s_{yx})(b_{yx} - \beta b_{xx}) = 0.$$

Before finding the roots of this equation, let us consider the last pivot; it is the derivative of (18) with respect to β. Hence the last pivot, evaluated at the turning points (ie when $f(\beta) = 0$), is

$$nf'(\beta)/\{T(\beta)W(\beta)\} ,$$

where $f'(\beta)$ is the derivative of $f(\beta)$. To get a maximum we want a solution of $f(\beta) = 0$ for which $f'(\beta)$ is negative. The two roots of $f(\beta) = 0$ are

$$\beta = \frac{s_{yy}b_{xx} - s_{xx}b_{yy} \pm d(y,x)}{2(s_{yx}b_{xx} - s_{xx}b_{yx})} ,$$

where $d(y,x) =$

$$\{(s_{xx}b_{yy} - s_{yy}b_{xx})^2 - 4(s_{yx}b_{xx} - s_{xx}b_{yx})(s_{yy}b_{yx} - s_{yx}b_{yy})\}^{\frac{1}{2}}.$$

If we choose the minus sign we can show that f' is negative and so we get a maximum while if we choose the plus sign f' is positive and so we get a saddle point. As there are only two groups, $b_{yy}b_{xx} = b_{yx}^2$ and the two solutions of $f(\beta) = 0$ simplify to

$$\beta = \tilde{\beta} = b_{yx}/b_{xx}$$

at the maximum and

$$\beta = (b_{xx}s_{yy} - b_{yx}s_{yx})/(b_{xx}s_{yx} - b_{yx}s_{xx})$$

at the saddle point.

Since

$$\bar{y}_{i.} - \bar{y}_{..} = \tilde{\beta}(\bar{x}_{i.} - \bar{x}_{..}) \quad (i=1,2) \quad \text{and} \quad B(\tilde{\beta}) = 0,$$

the estimates for σ^2, σ_δ^2, σ_ε^2 and μ_i from (11), (12),

(13) and (14) simplify considerably. In fact, the unique local maximum of the likelihood occurs at

$$\mu_i = \tilde{\mu}_i = \bar{x}_{i.} \qquad (i=1,2),$$

$$\alpha = \tilde{\alpha} = \bar{y}_{..} - \tilde{\beta}\bar{x}_{..} ,$$

$$\sigma^2 = \tilde{\sigma}^2 = s_{yx}/\tilde{\beta} ,$$

$$\sigma_\delta^2 = \tilde{\sigma}_\delta^2 = s_{xx} - s_{yx}/\tilde{\beta} ,$$

$$\sigma_\varepsilon^2 = \tilde{\sigma}_\varepsilon^2 = s_{yy} - \tilde{\beta}s_{yx} ,$$

and

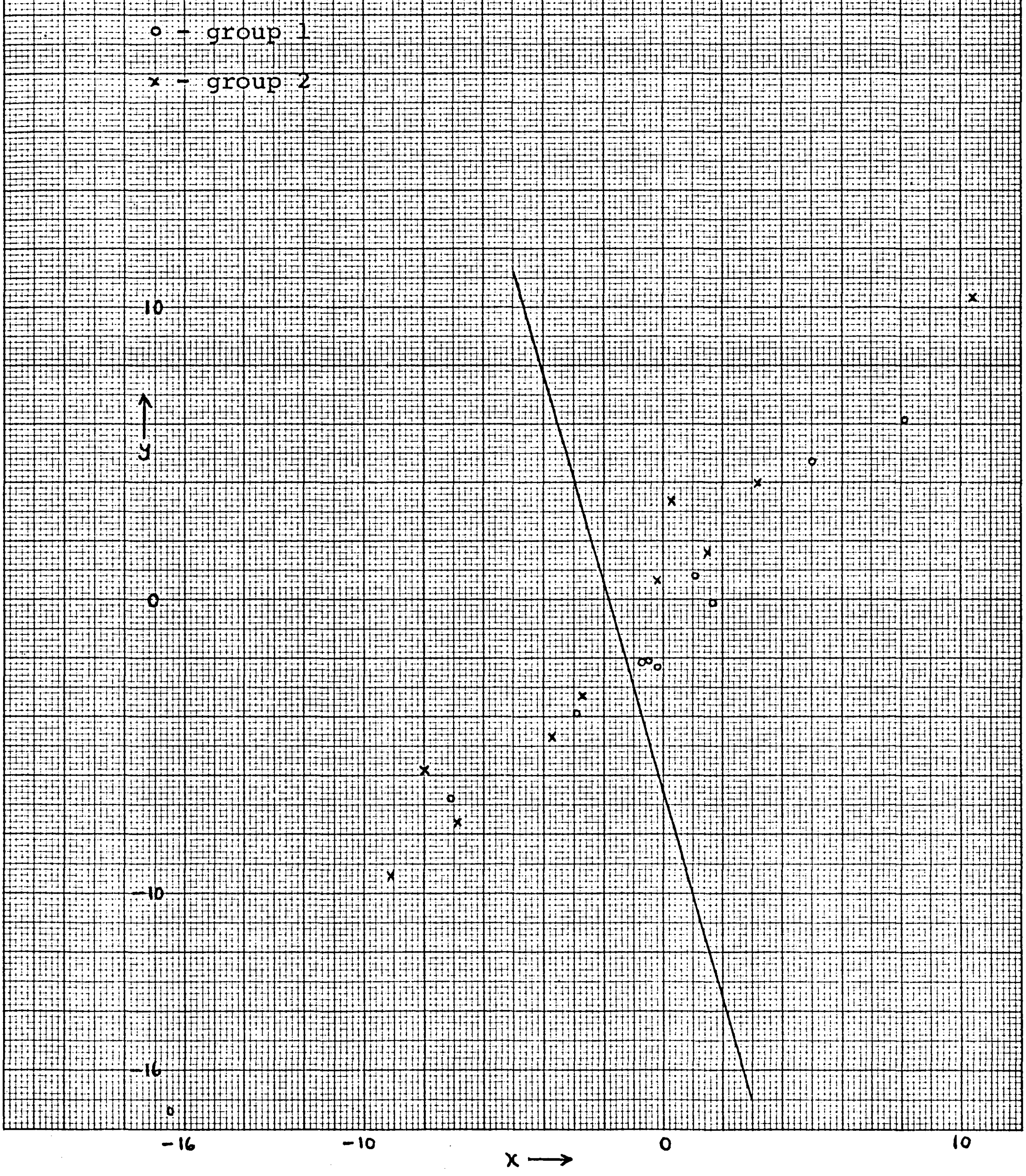
$$\beta = \tilde{\beta} = b_{yx}/b_{xx} .$$

Note that the estimates of σ^2 and σ_δ^2 are undefined when $\tilde{\beta} = 0$; during the derivation of these estimates we had to assume that $\tilde{\beta} \neq 0$. However, we shall see later that the above estimates are inappropriate when $\tilde{\beta}$ is near zero. We have now completed the solution of the equations.

It is interesting to note that our estimate, $\tilde{\beta}$, of β depends only on the between-group scatter and in no way on the within-group scatter. This is rather surprising because the model assumes that the scatter within groups tends to be along the same line as that between groups, and so the within-group scatter should also contain some information about the slope. This information is clearly going to be important when the group means are close together. Figure I illustrates such a situation. The points plotted are from a simulation of the model with group means 0 and 1,

FIGURE I

ESTIMATED RELATIONSHIP FOR SIMULATED DATA



$\alpha = 0$, $\beta = 1$, $\sigma^2 = 25$ and the error variances, σ^2_δ and σ^2_ε, both 1. The between-group slope, $\tilde{\beta}$, for these data is -3.56 and the corresponding fitted line is clearly quite ridiculous. This underlines the fact that the estimates we have found are not necessarily the ML estimates.

All we have done so far is to locate the local maximum of the likelihood. But we can easily see that this point is not always in the parameter space, which is restricted by the inequalities $\sigma^2 \geq 0$, $\sigma^2_\delta \geq 0$ and $\sigma^2_\varepsilon \geq 0$ (and $|\underset{\sim}{\Sigma}| > 0$). Depending on the value of $\tilde{\beta}$, any one of $\tilde{\sigma}^2$, $\tilde{\sigma}^2_\delta$ and $\tilde{\sigma}^2_\varepsilon$ could be negative. When this occurs, the maximum within the parameter space (which is what we require for a ML estimate) will be on one of the boundaries.

It is not difficult to see that $\tilde{\beta}$ must lie between the two within-group estimates of the slope, s_{yx}/s_{xx} and s_{yy}/s_{yx}, for all three of $\tilde{\sigma}^2$, $\tilde{\sigma}^2_\delta$ and $\tilde{\sigma}^2_\varepsilon$ to be non-negative. When this condition holds (for convenience we shall call it the "internal condition"), the local maximum of the likelihood lies inside the parameter space and the above estimates (ie $\tilde{\alpha}$, $\tilde{\mu}_i$, etc), which we shall call the "internal estimates", are the ML estimates. Note that when the internal condition holds the saddle point is outside the parameter space so that the local maximum is the only stationary point, and hence the absolute maximum, inside the parameter space. When the internal condition does not hold, the internal estimates are no longer the ML estimates. The maximum will be on one of the boundaries and so we must now find the maximum of the likelihood on each of the three boundaries.

To find the maximum on the (σ_δ^2=0)-boundary, we set σ_δ^2 to zero and find the solution of the other likelihood equations. The derivative $\partial\ell/\partial\mu_i$ becomes

$$n_i(\bar{x}_{i.} - \mu_i)/\sigma^2$$

so that we get $\mu_i = \bar{x}_{i.}$ (i=1,2). The first two pivots are $-n_i/\sigma^2$ (i=1,2), which are both necessarily negative. The third derivative, $\partial\ell/\partial\alpha$, becomes

$$n(\bar{y}_{..} - \alpha - \beta\bar{x}_{..})/\sigma_\varepsilon^2$$

so that we get $\alpha = \bar{y}_{..} - \beta\bar{x}_{..}$ and a third pivot of $-n/\sigma_\varepsilon^2$. The fourth derivative, $\partial\ell/\partial\sigma^2$, becomes

$$(-n\sigma^2 + S_3)/(2\sigma^4),$$

which, on substitution of $\bar{x}_{i.}$ for μ_i, equals

$$n(-\sigma^2 + s_{xx})/(2\sigma^4).$$

Equating this to zero gives us $\sigma^2 = s_{xx}$ and the fourth pivot, evaluated when $\partial\ell/\partial\sigma^2 = 0$, is $-n/(2\sigma^4)$. The derivative of ℓ with respect to σ_ε^2, $\partial\ell/\partial\sigma_\varepsilon^2$ becomes

$$-n/(2\sigma_\varepsilon^2) + (S_1 - 2\beta S_2 + \beta^2 S_3)/(2\sigma_\varepsilon^4),$$

which, on substitution of $\bar{y}_{..} - \beta\bar{x}_{..}$ for α and $\bar{x}_{i.}$ for μ_i, equals

$$n\{-\sigma_\varepsilon^2 + T(\beta)\}/(2\sigma_\varepsilon^4).$$

Hence we get $\sigma_\varepsilon^2 = T(\beta)$ and a fifth pivot, evaluated when $\partial\ell/\partial\sigma_\varepsilon^2 = 0$, of $-n/(2\sigma_\varepsilon^4)$. Finally, $\partial\ell/\partial\beta$ becomes

$$(S_2 - \beta S_3)/\sigma_\varepsilon^2 + \{\sum_i n_i\mu_i(\bar{y}_{i.} - \alpha - \beta\mu_i) - \beta\sum_i n_i\mu_i(\bar{x}_{i.} - \mu_i)\}/\sigma_\varepsilon^2,$$

which, on substitution of $\bar{y}_{..} - \beta\bar{x}_{..}$ for α and $\bar{x}_{i.}$ for μ_i, equals

$$n(t_{yx} - \beta t_{xx})/\sigma_\varepsilon^2 .$$

We thus get $\beta = t_{yx}/t_{xx}$ and a final pivot of $-t_{xx}/\sigma_\varepsilon^2$, which is again negative. We have thus established that the maximum on the $(\sigma_\delta^2{=}0)$-boundary occurs when

$$\mu_i = \bar{x}_{i.} \quad (i=1,2),$$

$$\alpha = \bar{y}_{..} - \bar{x}_{..}\, t_{yx}/t_{xx},$$

$$\sigma^2 = s_{xx},$$

$$\sigma_\delta^2 = 0,$$

$$\sigma_\varepsilon^2 = t_{yy} - t_{yx}^2/t_{xx},$$

and

$$\beta = t_{yx}/t_{xx}.$$

We shall call these the "$(\sigma_\delta^2{=}0)$-boundary estimates". The estimate of t_{yx}/t_{xx} for β is hardly surprising for when $\sigma_\delta^2{=}0$ there is no error in the variable X and so we have the familiar Y on X regression problem. The maximum of the log-likelihood on the $(\sigma_\delta^2{=}0)$-boundary is

$$\ell_\delta = -n.\log(2\pi) - \tfrac{1}{2}n.\log\{s_{xx}(t_{yy} - t_{yx}^2/t_{xx})\} - n .$$

Finding the maximum on the $(\sigma_\varepsilon^2=0)$-boundary is very similar to finding the maximum on the $(\sigma_\delta^2=0)$-boundary and so we shall omit the details. The unique turning point, a maximum, on the $(\sigma_\varepsilon^2=0)$-boundary occurs at

$$\mu_i = \bar{x}_{..} + (\bar{y}_{i.} - \bar{y}_{..})t_{yx}/t_{yy} \quad (i=1,2),$$

$$\alpha = \bar{y}_{..} - \bar{x}_{..}t_{yy}/t_{yx} ,$$

$$\sigma^2 = s_{yy}t_{yx}^2/t_{yy}^2 ,$$

$$\sigma_\delta^2 = t_{xx} - t_{yx}^2/t_{yy} ,$$

$$\sigma_\varepsilon^2 = 0 ,$$

and

$$\beta = t_{yy}/t_{yx} .$$

These estimates are the $(\sigma_\varepsilon^2=0)$-boundary estimates. When $\sigma_\varepsilon^2=0$ there is no error in the variable Y so that the estimate t_{yy}/t_{yx}, the inverse of the overall regression slope of X on Y, is again not unexpected. The maximum of the log-likelihood on the $(\sigma_\varepsilon^2=0)$-boundary is

$$\ell_\varepsilon = -n.\log(2\pi) - \tfrac{1}{2}n.\log\{s_{yy}(t_{xx} - t_{yx}^2/t_{yy})\} - n .$$

We now find the maximum of the log-likelihood on the $(\sigma^2=0)$-boundary. Note that when $\sigma^2=0$,

$$X_{ij} = \mu_i + \delta_{ij} \quad \text{and} \quad Y_{ij} = \alpha + \beta\mu_i + \varepsilon_{ij},$$

so that we have replicated observations on just two points. From the derivative of the log-likelihood with respect to μ_i, we get

$$\mu_i = \{\beta\sigma_\delta^2(\bar{y}_{i.} - \alpha) + \sigma_\varepsilon^2\bar{x}_{i.}\}/(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2) \quad (i=1,2),$$

as in equation (2). The first two pivots are

$$-n_i(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)/(\sigma_\delta^2\sigma_\varepsilon^2) \quad (i=1,2),$$

which are both negative. When $\sigma^2=0$, $\partial\ell/\partial\alpha$ becomes

$$n(\bar{y}_{..} - \alpha - \beta\bar{\mu})/\sigma_\varepsilon^2 ,$$

which, on substitution of the above expression for μ_i, becomes

$$n(\bar{y}_{..} - \alpha - \beta\bar{x}_{..})/(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2).$$

Hence $\alpha = \bar{y}_{..} - \beta\bar{x}_{..}$ and the third pivot is $-n/(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)$. It is easiest if we introduce $\partial\ell/\partial\beta$ next. With $\sigma^2=0$, $\partial\ell/\partial\beta$ becomes

$$\sum n_i\mu_i(\bar{y}_{i.} - \alpha - \beta\mu_i)/\sigma_\varepsilon^2 ,$$

which, on substitution for α and μ_i, equals

$$n\{\beta\sigma_\delta^2(b_{yy} - \beta b_{yx}) + \sigma_\varepsilon^2(b_{yx} - \beta b_{xx})\}/(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)^2,$$

as in equation (17). We thus get two estimates for β, the solutions of the quadratic equation

$$g(\beta) = \beta\sigma_\delta^2(b_{yy} - \beta b_{yx}) + \sigma_\varepsilon^2(b_{yx} - \beta b_{xx}) = 0.$$

Since $b_{yy}b_{xx} = b_{yx}^2$,

$$g(\beta) = (\beta\sigma_\delta^2 b_{yx}/b_{xx} + \sigma_\varepsilon^2)(b_{yx} - \beta b_{xx})$$

and the solution of $g(\beta) = 0$ that makes $g'(\beta)$, and hence also the fourth pivot, negative is $\beta = b_{yx}/b_{xx}$. The other solution makes $g'(\beta)$ positive and so it is at a saddle point. When $\beta = b_{yx}/b_{xx}$ the above expression for μ_i simplifies to $\bar{x}_{i.}$. When $\sigma^2=0$, $\partial\ell/\partial\sigma_\delta^2$ equals

$$(-n\sigma_\delta^2 + S_3)/(2\sigma_\delta^4),$$

which, on substitution of $\bar{x}_{i.}$ for μ_i, becomes

$$n(-\sigma_\delta^2 + s_{xx})/(2\sigma_\delta^4).$$

Hence $\sigma_\delta^2 = s_{xx}$ and the fifth pivot is $-n/(2\sigma_\delta^4)$. Similarly, we get $\sigma_\varepsilon^2 = s_{yy}$ and a final pivot of $-n/(2\sigma_\varepsilon^4)$. Hence the $(\sigma^2=0)$-boundary estimates are

$$\mu_i = \bar{x}_{i.} \quad (i=1,2),$$

$$\alpha = \bar{y}_{..} - \bar{x}_{..}b_{yx}/b_{xx},$$

$$\sigma^2 = 0,$$

$$\sigma_\delta^2 = s_{xx},$$

$$\sigma_\varepsilon^2 = s_{yy},$$

and

$$\beta = b_{yx}/b_{xx}.$$

The maximum of the log-likelihood on the $(\sigma^2=0)$-boundary is

$$\ell_0 = -n.\log(2\pi) - \tfrac{1}{2}n.\log(s_{yy}s_{xx}) - n .$$

We can show that this is the absolute maximum on the $(\sigma^2=0)$-boundary despite the fact that there is also a saddle point on the boundary.

The various sets of estimates are summarized in

Table I. The ML estimate of a parameter is its internal estimate when b_{yx}/b_{xx} lies between s_{yx}/s_{xx} and s_{yy}/s_{yx}, its $(\sigma_\delta^2=0)$-boundary estimate when b_{yx}/b_{xx} does not lie between s_{yx}/s_{xx} and s_{yy}/s_{yx} but $s_{xx}(t_{yy}-t_{yx}^2/t_{xx})$ is less than both $s_{yy}(t_{xx}-t_{yx}^2/t_{yy})$ and $s_{yy}s_{xx}$, its $(\sigma_\varepsilon^2=0)$-boundary estimate when b_{yx}/b_{xx} does not lie between s_{yx}/s_{xx} and s_{yy}/s_{yx} but $s_{yy}(t_{xx}-t_{yx}^2/t_{yy})$ is less than both $s_{xx}(t_{yy}-t_{yx}^2/t_{xx})$ and $s_{yy}s_{xx}$, and its $(\sigma^2=0)$-boundary estimate otherwise.

TABLE I

MAXIMUM LIKELIHOOD ESTIMATES OF THE PARAMETERS

Parameter	Internal estimate	$(\sigma_\delta^2=0)$-bdy estimate	$(\sigma_\varepsilon^2=0)$-bdy estimate	$(\sigma^2=0)$-bdy estimate
μ_i	$\bar{x}_{i.}$	$\bar{x}_{i.}$	$\bar{x}_{..}+(\bar{y}_{i.}-\bar{y}_{..})\frac{t_{yx}}{t_{yy}}$	$\bar{x}_{i.}$
α	$\bar{y}_{..}-\frac{b_{yx}}{b_{xx}}\bar{x}_{..}$	$\bar{y}_{..}-\frac{t_{yx}}{t_{xx}}\bar{x}_{..}$	$\bar{y}_{..}-\frac{t_{yy}}{t_{yx}}\bar{x}_{..}$	$\bar{y}_{..}-\frac{b_{yx}}{b_{xx}}\bar{x}_{..}$
σ^2	$s_{yx}b_{xx}/b_{yx}$	s_{xx}	$s_{yy}t_{yx}^2/t_{yy}^2$	0
σ_δ^2	$s_{xx}-s_{yx}\frac{b_{xx}}{b_{yx}}$	0	$t_{xx}-t_{yx}^2/t_{yy}$	s_{xx}
σ_ε^2	$s_{yy}-s_{yx}\frac{b_{yx}}{b_{xx}}$	$t_{yy}-t_{yx}^2/t_{xx}$	0	s_{yy}
β	b_{yx}/b_{xx}	t_{yx}/t_{xx}	t_{yy}/t_{yx}	b_{yx}/b_{xx}

With some algebraic manipulation we can describe the regions of the sample space in which the different sets of estimates are appropriate in terms of inequalities involving b_{yx}/b_{xx}. Define region A to be the region of the sample space in which b_{yx}/b_{xx} lies between s_{yx}/s_{xx} and s_{yy}/s_{yx}. Then the internal estimates are the ML estimates when the sample lies in region A. Define

$$r_1 = (s_{yy}/s_{xx})^{\frac{1}{2}},$$

$$r_2 = \{1 - (t_{xx}/b_{xx})^{\frac{1}{2}}\} s_{yx}/s_{xx}$$

and

$$r_3 = \tfrac{1}{2}(s_{yy}/s_{yx} - s_{yx}/b_{xx}).$$

Let region B be the region of the sample space in which b_{yx}/b_{xx} lies in the interval

$$(\ \max(-r_1, r_2)\ ,\ s_{yx}/s_{xx}\ ) \quad \text{when } s_{yx} > 0$$

or in the interval

$$(\ s_{yx}/s_{xx}\ ,\ \min(r_1, r_2)\ ) \quad \text{when } s_{yx} < 0.$$

Let region C be the region of the sample space in which b_{yx}/b_{xx} lies in either of the intervals

$(s_{yy}/s_{yx}\ , \infty)$ or, if $r_3 < -r_1$, $(r_3\ , -r_1)$ when $s_{yx} > 0$,

or in either of the intervals

$(-\infty\ , s_{yy}/s_{yx})$ or, if $r_1 < r_3$, $(r_1\ , r_3)$ when $s_{yx} < 0$.

Let region D be the remainder of the space; that is, the region in which b_{yx}/b_{xx} lies in the interval

$$\left(-\infty \, , \min(r_2, r_3) \right) \quad \text{when } s_{yx} > 0,$$

or in the interval

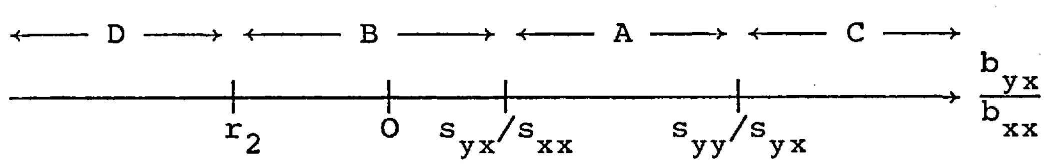
$$\left(\max(r_2, r_3) \, , \infty \right) \quad \text{when } s_{yx} < 0.$$

Then the $(\sigma_\delta^2=0)$-boundary estimates are the ML estimates when the sample lies in region B, the $(\sigma_\varepsilon^2=0)$-boundary estimates are the ML estimates when the sample lies in region C, and the $(\sigma^2=0)$-boundary estimates are the ML estimates when the sample lies in region D. When $s_{yx} > 0$, we can show that either $-r_1 \le r_2 \le r_3$ or $r_3 < r_2 < -r_1$. In the first case the regions are illustrated in Figure II (a) and in the second case they are illustrated in Figure II (b).

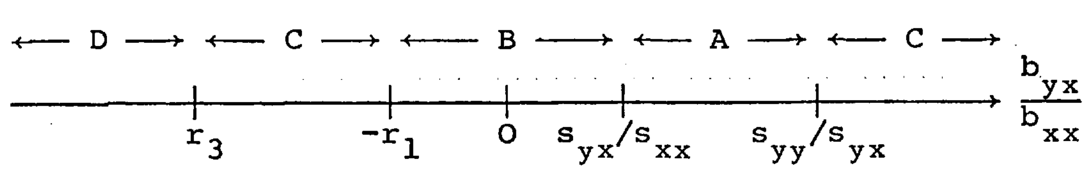
FIGURE II

THE REGIONS OF APPLICABILITY OF THE PARAMETER ESTIMATES WHEN $s_{yx} > 0$

(a) $-r_1 \le r_2 \le r_3$



(b) $r_3 < r_2 < -r_1$



A : internal B : $\sigma_\delta^2 = 0$ C : $\sigma_\varepsilon^2 = 0$ D : $\sigma^2 = 0$

Except when the group means are extremely close together, the probability that b_{YX}/b_{XX} lies well outside the range of s_{YX}/s_{XX} and s_{YY}/s_{YX} will be very small. So, except in this extreme, we can make the simplifying approximation of ignoring the $(\sigma^2=0)$-boundary estimates and using either the $(\sigma_\delta^2=0)$-boundary estimates or the $(\sigma_\varepsilon^2=0)$-boundary estimates in region D. But the maximum on the $(\sigma_\delta^2=0)$-boundary is greater than the maximum on the $(\sigma_\varepsilon^2=0)$-boundary when

$$| b_{yx}/b_{xx} | < (s_{yy}/s_{xx})^{\frac{1}{2}}.$$

Hence, we can approximate the ML estimates by using the internal estimates when b_{yx}/b_{xx} lies between s_{yx}/s_{xx} and s_{yy}/s_{yx}, the $(\sigma_\delta^2=0)$-boundary estimates when b_{yx}/b_{xx} lies between s_{yx}/s_{xx} and $-\text{sign}(s_{yx})(s_{yy}/s_{xx})^{\frac{1}{2}}$ and the $(\sigma_\varepsilon^2=0)$- boundary estimates otherwise.

If we re-examine the numerical example discussed earlier in this section (and illustrated in Figure I), we find that the maximum occurs on the $(\sigma_\varepsilon^2=0)$-boundary and so the ML estimate of β is $t_{yy}/t_{yx} = 1.030$. This clearly makes a good deal more sense than the internal estimate of -3.56. The ML estimates are clearly not going to give the sort of ridiculous fit that the internal estimates can give. The final estimate of β, emerging only after a careful study of the likelihood function, is sensible on general grounds.

2.3 Asymptotic expected values and variances of the estimators

In order to estimate the precision of the parameter estimates we need their estimated variances. In this section we obtain asymptotic expected values and variances. These should be interpreted as moments of the asymptotic distribution of the ML estimators rather than as asymptotic expansions of the exact moments. The asymptotic variances could be found from the information matrix but it is easier to calculate them directly. In using the estimates, it must be borne in mind that they are only asymptotic estimates and that when the group means are close together the asymptotic properties of the ML estimators will be approached very slowly. The asymptotic properties of the ML estimators are, in fact, just those of the internal estimators.

The moments of the within-group and between-group sums of squares are derived in Appendix 2 and the asymptotic moments of b_{YX}/b_{XX} are found in Appendix 3 but no further details of the calculations will be given here. The asymptotic expected values and variances of the internal estimators are

$$E(\tilde{\bar{\mu}}_i) = \mu_i \quad (i=1,2),$$

$$E(\tilde{\alpha}) = \alpha - \beta\sigma_\delta^2\bar{\mu}/(nb_{\mu\mu}) + O(1/n^2),$$

where $\bar{\mu} = (n_1\mu_1 + n_2\mu_2)/n$ and $b_{\mu\mu} = n_1 n_2(\mu_1-\mu_2)^2/n^2$,

$$E(\tilde{\sigma}^2) = \sigma^2 + \sigma^2(\sigma_\varepsilon^2 - 2\beta^2 b_{\mu\mu})/(n\beta^2 b_{\mu\mu}) + O(1/n^2),$$

$$E(\tilde{\sigma}_\delta^2) = \sigma_\delta^2 - (\sigma^2\sigma_\varepsilon^2 + 2\beta^2\sigma_\delta^2 b_{\mu\mu})/(n\beta^2 b_{\mu\mu}) + O(1/n^2),$$

$$E(\tilde{\sigma}_\varepsilon^2) = \sigma_\varepsilon^2 - (\beta^2\sigma^2\sigma_\delta^2 + 2\sigma_\varepsilon^2 b_{\mu\mu})/(nb_{\mu\mu}) + O(1/n^2),$$

$$E(\tilde{\beta}) = \beta + \beta\sigma_\delta^2/(nb_{\mu\mu}) + O(1/n^2),$$

$$\mathrm{var}(\tilde{\mu}_i) = (\sigma^2+\sigma_\delta^2)/n_i \quad (i=1,2),$$

$$\mathrm{var}(\tilde{\alpha}) = (\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)\textstyle\sum n_i\mu_i^2/(n^2 b_{\mu\mu}) + O(1/n^2),$$

$$\mathrm{var}(\tilde{\sigma}^2) = \{\sigma^4(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2) + b_{\mu\mu}(|\underset{\sim}{\Sigma}|+2\beta^2\sigma^4)\}/(n\beta^2 b_{\mu\mu}) + O(1/n^2),$$

$$\mathrm{var}(\tilde{\sigma}_\delta^2) = \{\sigma^4(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2) + b_{\mu\mu}(|\underset{\sim}{\Sigma}|+2\beta^2\sigma_\delta^4)\}/(n\beta^2 b_{\mu\mu}) + O(1/n^2),$$

$$\mathrm{var}(\tilde{\sigma}_\varepsilon^2) = \{\beta^2\sigma^4(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2) + b_{\mu\mu}(\beta^2|\underset{\sim}{\Sigma}|+2\sigma_\varepsilon^4)\}/(nb_{\mu\mu}) + O(1/n^2),$$

and
$$\mathrm{var}(\tilde{\beta}) = (\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)/(nb_{\mu\mu}) + O(1/n^2).$$

2.4 Tests of the adequacy of the simplest model

Before we fit the simple model (ie model I, the model discussed in the earlier sections of this chapter) to data we will usually want to check that the data are consistent with the model. As we normally wish to fit the simplest model which adequately describes the data, we want to fit model I whenever possible rather than a model with more parameters. In this section we shall consider three such more general models and find procedures for testing whether they fit the data significantly better than the simplest model. Of course many other departures from the simplest model could be considered but it is hoped that the ones discussed here will be the most useful.

In the first generalized model we allow the intercepts on the y-axis of the two groups to differ. In other words, the common within-group slope of the two groups is allowed to be different from the between-group slope. This model, which we shall call model II, has the same specifications as model I except that α is replaced by α_i. We now find a test for the hypothesis that $\alpha_1 = \alpha_2$ in this model.

But first let us find the maximum of the likelihood for model II and the ML estimates for α_1 and α_2. Leaving the parameter β till last, we can solve the remaining likelihood equations in terms of β in the same way that we did in model I. In fact, the algebra is very similar and we get as solutions $\mu_i = \bar{x}_{i.}$ (i=1,2), $\alpha_i = \bar{y}_{i.} - \beta\bar{x}_{i.}$ (i=1,2), $\sigma^2 = s_{yx}/\beta$, $\sigma_\delta^2 = s_{xx} - s_{yx}/\beta$ and $\sigma_\varepsilon^2 = s_{yy} - \beta s_{yx}$, with all the corresponding pivots of the matrix of double derivatives of the log-likelihood being negative. However, on substitution of these estimates, $\partial \ell/\partial\beta$ becomes identically zero. Hence the likelihood is maximized in a subspace of the parameter space, any choice of the parameters in that subspace being equally good. The reason for this lack of identifiability is that while we have increased the number of parameters in the likelihood by one to eight, the dimension of the minimal sufficient statistic has remained at seven. Note, however, that we cannot choose β completely arbitrarily for we still require any choice of the estimates to be in the parameter space. To ensure that all the variance estimates are non-negative we must choose β between s_{yx}/s_{xx} and s_{yy}/s_{yx}. Hence the log-likelihood achieves its maximum,

$$\hat{\ell} = -n\{1+\log(2\pi)\} - \tfrac{1}{2}n \log(s_{yy}s_{xx} - s_{yx}^2),$$

for any choice of the parameters such that $\mu_i = \bar{x}_{i.}$ $(i=1,2)$, $\alpha_i = \bar{y}_{i.} - \beta\bar{x}_{i.}$ $(i=1,2)$, $\sigma^2 = s_{yx}/\beta$, $\sigma_\delta^2 = s_{xx} - s_{yx}/\beta$, $\sigma_\varepsilon^2 = s_{yy} - \beta s_{yx}$ and β lies between s_{yx}/s_{xx} and s_{yy}/s_{yx}.

Now let us consider the likelihood ratio test. When the null hypothesis that $\alpha_1=\alpha_2$ is true, model II reduces to model I. The maximum of the log-likelihood for model I equals $\hat{\ell}$ when b_{yx}/b_{xx} lies between s_{yx}/s_{xx} and s_{yy}/s_{yx} and otherwise it equals the maximum of

$$\ell_\delta = -n\{1+\log(2\pi)\} - \tfrac{1}{2}n \log\{s_{xx}(t_{yy} - t_{yx}^2/t_{xx})\},$$

$$\ell_\varepsilon = -n\{1+\log(2\pi)\} - \tfrac{1}{2}n \log\{s_{yy}(t_{xx} - t_{yx}^2/t_{yy})\}$$

and $$\ell_o = -n\{1+\log(2\pi)\} - \tfrac{1}{2}n \log(s_{yy}s_{xx}).$$

Making the approximation suggested near the end of §2.2 (ie of ignoring ℓ_o), we get as twice the logarithm of the likelihood ratio a statistic which has the value 0 when b_{yx}/b_{xx} lies between s_{yx}/s_{xx} and s_{yy}/s_{yx}, the value

$$n \log(s_{yy}s_{xx} - s_{yx}^2) - n \log\{s_{xx}(t_{yy} - t_{yx}^2/t_{xx})\}$$

when b_{yx}/b_{xx} lies between s_{yx}/s_{xx} and $-\text{sign}(s_{yx})(s_{yy}/s_{xx})^{\frac{1}{2}}$, and the value

$$n \log(s_{yy}s_{xx} - s_{yx}^2) - n \log\{s_{yy}(t_{xx} - t_{yx}^2/t_{yy})\}$$

otherwise. The problem is finding the distribution of this statistic. Certainly it would be very unreasonable to think that it might be approximately distributed as a chi-squared variable for there is a finite probability that it takes the value zero. Instead we find another test procedure.

Moran (1956) gave an example of a test of significance in a similar unidentifiable model and a very similar argument can be used to get a test here. To achieve the maximum of the likelihood in model II we must choose the estimate of $\alpha_1 - \alpha_2$ between the bounds

$$a(y,x) = (\bar{y}_{1.} - \bar{y}_{2.}) - \frac{s_{yy}}{s_{yx}}(\bar{x}_{1.} - \bar{x}_{2.})$$

and $$b(y,x) = (\bar{y}_{1.} - \bar{y}_{2.}) - \frac{s_{yx}}{s_{xx}}(\bar{x}_{1.} - \bar{x}_{2.}).$$

We get a conservative test of the hypothesis that $\alpha_1 = \alpha_2$ by considering tests based on these bounds.

First we find the distribution of $b(Y,X)$ conditional on $\underset{\sim}{X} = \underset{\sim}{x}$, where $\underset{\sim}{X}^T = (X_{11},..,X_{1n_1},X_{21},..,X_{2n_2})$ and similarly for $\underset{\sim}{x}$. Conditional on $\underset{\sim}{X} = \underset{\sim}{x}$, the Y_{ij} have independent normal distributions with means

$$m_{ij} = \alpha_i + \beta\sigma_\delta^2\mu_i/(\sigma^2+\sigma_\delta^2) + \beta\sigma^2 x_{ij}/(\sigma^2+\sigma_\delta^2)$$

and variance

$$\sigma_{YX}^2 = \beta^2\sigma^2+\sigma_\varepsilon^2 - \beta^2\sigma^4/(\sigma^2+\sigma_\delta^2).$$

Hence we find that, conditional on $\underset{\sim}{X} = \underset{\sim}{x}$, b(Y,X) has a normal distribution with mean

$$\alpha_1 - \alpha_2 + \beta\sigma_\delta^2(\mu_1 - \mu_2)/(\sigma^2+\sigma_\delta^2)$$

and variance $g(x)\sigma_{YX}^2$, where

$$g(x) = \frac{1}{n_1} + \frac{1}{n_2} + \frac{(\bar{x}_{1.} - \bar{x}_{2.})^2}{ns_{xx}} .$$

Hence, conditional on $\underset{\sim}{X} = \underset{\sim}{x}$,

$$B(Y,X) = \frac{b(Y,X) - (\alpha_1 - \alpha_2) - \beta\sigma_\delta^2(\mu_1 - \mu_2)/(\sigma^2+\sigma_\delta^2)}{g^{\frac{1}{2}}(X)\sigma_{YX}}$$

has an N(0,1) distribution.

We now find an independent estimate of σ_{YX} which, on substitution into B(Y,X), gives us a Student t distribution. Let

$$Z_{ij} = (Y_{ij} - m_{ij})/\sigma_{YX}.$$

Then, conditional on $\underset{\sim}{X} = \underset{\sim}{x}$, the Z_{ij} are independently distributed as N(0,1). Define the vector $\underset{\sim}{L}_p$ to be a column of p ones. Let $\underset{\sim}{Z}^T = (Z_{11},..,Z_{1n_1},Z_{21},..,Z_{2n_2})$ and

$$\underset{\sim}{e}^T = \frac{1}{g^{\frac{1}{2}}(X)}\left\{\left(\frac{1}{n_1}\underset{\sim}{L}_{n_1}^T, -\frac{1}{n_2}\underset{\sim}{L}_{n_2}^T\right) - \frac{(\bar{X}_{1\cdot}-\bar{X}_{2\cdot})}{ns_{XX}}(X_{11}-\bar{X}_{1\cdot},\ldots,\bar{X}_{2n_2}-\bar{X}_{2\cdot})\right\}.$$

Then $\underset{\sim}{e}^T\underset{\sim}{Z} = B(Y,X)$. Since, conditional on $\underset{\sim}{X} = \underset{\sim}{x}$, $B(Y,X)$ is $N(0,1)$ then

$$B^2(Y,X) = \underset{\sim}{Z}^T\underset{\sim}{e}\underset{\sim}{e}^T\underset{\sim}{Z}$$

has a chi-squared distribution with one degree of freedom, which we shall write as χ_1^2. Let

$$\underset{\sim}{c}_1^T = \frac{1}{n_1^{\frac{1}{2}}}(\underset{\sim}{L}_{n_1}^T, \underset{\sim}{O}_{n_2}^T) \quad \text{and} \quad \underset{\sim}{c}_2^T = \frac{1}{n_2^{\frac{1}{2}}}(\underset{\sim}{O}_{n_1}^T, \underset{\sim}{L}_{n_2}^T)$$

where $\underset{\sim}{O}_p$ is a vector of p zeros. Then

$$\underset{\sim}{Z}^T\underset{\sim}{c}_i\underset{\sim}{c}_i^T\underset{\sim}{Z} = n_i\bar{Z}_{i\cdot}^2.$$

is distributed as χ_1^2 for i=1,2. We now wish to choose a vector $\underset{\sim}{d}$ so that

$$\underset{\sim}{e}\underset{\sim}{e}^T(\underset{\sim}{I}_n - \underset{\sim}{c}_1\underset{\sim}{c}_1^T - \underset{\sim}{c}_2\underset{\sim}{c}_2^T - \underset{\sim}{d}\underset{\sim}{d}^T) = \underset{\sim}{O}$$

and $(\underset{\sim}{I}_n - \underset{\sim}{c}_1\underset{\sim}{c}_1^T - \underset{\sim}{c}_2\underset{\sim}{c}_2^T - \underset{\sim}{d}\underset{\sim}{d}^T)$ is idempotent. For then $B^2(Y,X)$ and

$$\underset{\sim}{Z}^T(\underset{\sim}{I}_n - \underset{\sim}{c}_1\underset{\sim}{c}_1^T - \underset{\sim}{c}_2\underset{\sim}{c}_2^T - \underset{\sim}{d}\underset{\sim}{d}^T)\underset{\sim}{Z}$$

will be independent χ^2 variables. The choice

$$\underset{\sim}{d}^T = \frac{1}{(ns_{XX})^{\frac{1}{2}}}(X_{11} - \bar{X}_{1\cdot},\ldots,\bar{X}_{2n_2} - \bar{X}_{2\cdot})$$

achieves this aim. That is, conditional on $\underset{\sim}{X} = \underset{\sim}{x}$, $B^2(Y,X)$ is distributed as χ_1^2 and is independent of

$$\underset{\sim}{z}^T(\underset{\sim}{I}_n - \underset{\sim}{c}_1\underset{\sim}{c}_1^T - \underset{\sim}{c}_2\underset{\sim}{c}_2^T - \underset{\sim}{d}\underset{\sim}{d}^T)\underset{\sim}{z} = n(s_{YY} - s_{YX}^2/s_{XX})/\sigma_{YX}^2,$$

which is distributed as χ^2_{n-3}.

Hence we have established that, conditional on $\underset{\sim}{X} = \underset{\sim}{x}$,

$$V(Y,X) = B(Y,X)\left\{\frac{(n-3)\sigma_{YX}^2}{n(s_{YY} - s_{YX}^2/s_{XX})}\right\}$$

is distributed as t_{n-3}, the Student t distribution with n-3 degrees of freedom. But this distribution does not depend on $\underset{\sim}{x}$ so that, unconditionally,

$$V(Y,X) = \left\{b(Y,X) - (\alpha_1-\alpha_2) - \frac{\beta\sigma_\delta^2(\mu_1-\mu_2)}{\sigma^2+\sigma_\delta^2}\right\}\left\{\frac{n-3}{ng(X)(s_{YY}-s_{YX}/s_{XX})}\right\}^{\frac{1}{2}},$$

is distributed as t_{n-3}.

Defining

$$c(Y,X) = -\frac{s_{YX}}{s_{YY}}\,a(Y,X) = (\bar{X}_{1.} - \bar{X}_{2.}) - \frac{s_{YX}}{s_{YY}}(\bar{Y}_{1.} - \bar{Y}_{2.}),$$

we can similarly show that

$$U(Y,X) = \left\{c(Y,X) - \frac{\beta\sigma^2(\alpha_2-\alpha_1)}{\beta^2\sigma^2+\sigma_\varepsilon^2} - \frac{\sigma_\varepsilon^2(\mu_1-\mu_2)}{\beta^2\sigma^2+\sigma_\varepsilon^2}\right\}\left\{\frac{n-3}{ng(Y)(s_{XX}-s_{YX}^2/s_{YY})}\right\}^{\frac{1}{2}}$$

is also distributed as t_{n-3} though not, of course, independently of $V(Y,X)$.

Suppose, temporarily, that $\beta > 0$ and $\mu_1 - \mu_2 \geq 0$ and consider the statistic

$$S(Y,X) = c(Y,X)\left\{ \frac{n-3}{ng(Y)(s_{XX} - s_{YX}^2/s_{YY})} \right\}^{\frac{1}{2}}.$$

Then, when the null hypothesis that $\alpha_1 = \alpha_2$ is true, $S(Y,X)$ is greater than the t_{n-3} variable $U(Y,X)$. Consider first the alternative hypothesis that $\alpha_2 - \alpha_1 < 0$. When this alternative is true, the factor

$$\beta\sigma^2(\alpha_2 - \alpha_1)/(\beta^2\sigma^2 + \sigma_\varepsilon^2)$$

will increase $U(Y,X)$ relative to $S(Y,X)$. Hence a small value of $S(Y,X)$ indicates a possible departure from the null hypothesis in the direction of $\alpha_2 - \alpha_1 < 0$. But since $S(Y,X)$ is larger than a t_{n-3} variable when the null hypothesis is true, then the true tail probability is less than that of the t_{n-3} distribution, namely

$$\Pr\{W < S(y,x) \mid W \text{ has a } t_{n-3} \text{ distribution}\} = F_{n-3}\{S(y,x)\},$$

where F_{n-3} is the distribution function of the t_{n-3} distribution. In other words, $S(y,x)$ is more significant than is indicated by the t_{n-3} distribution.

When we consider the alternative hypothesis that $\alpha_2 - \alpha_1 > 0$, the unknown factor

$$\beta\sigma^2(\alpha_2 - \alpha_1)/(\beta^2\sigma^2 + \sigma_\varepsilon^2)$$

has the opposite effect and so a large value of $S(Y,X)$ is indicative of a departure from the null hypothesis in the direction of $\alpha_2 - \alpha_1 > 0$. The value $S(y,x)$ will be less significant than is indicated by the t_{n-3} distribution since the true tail probability is greater than

$$\Pr\{W > S(y,x) \mid W \text{ has a } t_{n-3} \text{ distribution}\} = F_{n-3}\{-S(y,x)\}.$$

Now consider the statistic

$$T(Y,X) = b(Y,X)\left\{\frac{n-3}{ng(X)(s_{YY} - s_{YX}^2/s_{XX})}\right\}^{\frac{1}{2}}.$$

Assuming still that $\beta > 0$ and $\mu_1 - \mu_2 \geq 0$, $T(Y,X)$ is greater than the t_{n-3} variable $V(Y,X)$ when the null hypothesis is true. This is the same as the relationship between $S(Y,X)$ and $U(Y,X)$ but now a large value of $T(Y,X)$ indicates a possible departure from the null hypothesis in the direction of $\alpha_2 - \alpha_1 < 0$ while a small value of $T(Y,X)$ indicates a possible departure in the direction of $\alpha_2 - \alpha_1 > 0$. The true tail area greater than $T(y,x)$ is greater than

$$\Pr\{W > T(y,x) \mid W \text{ has a } t_{n-3} \text{ distribution}\} = F_{n-3}\{-T(y,x)\}$$

while the true tail area less than $T(y,x)$ is less than

$$\Pr\{W < T(y,x) \mid W \text{ has a } t_{n-3} \text{ distribution}\} = F_{n-3}\{T(y,x)\}.$$

Combining the tests based on $S(Y,X)$ and $T(Y,X)$, we see that, when $\beta > 0$ and $\mu_1 - \mu_2 \geq 0$, a departure from the null hypothesis that $\alpha_1 = \alpha_2$ in the direction of $\alpha_2 - \alpha_1 < 0$ is indicated by the values $S(y,x)$ and $T(y,x)$

with significance somewhere between $F_{n-3}\{-T(y,x)\}$ and $F_{n-3}\{S(y,x)\}$. A departure from the null hypothesis in the direction of $\alpha_2 - \alpha_1 > 0$, on the other hand, is indicated by the values $S(y,x)$ and $T(y,x)$ with significance somewhere between $F_{n-3}\{-S(y,x)\}$ and $F_{n-3}\{T(y,x)\}$. If we wish to do a significance test we use the least significant value for each one-sided test. That is, we reject the null hypothesis in favour of the alternative that $\alpha_2 - \alpha_1 < 0$ if

$$S(y,x) < -t_{n-3}(\gamma)$$

and we reject in favour of the alternative that $\alpha_2 - \alpha_1 > 0$ if

$$T(y,x) < -t_{n-3}(\gamma).$$

Each of these tests has size less than γ.

The above tests are only appropriate when $\beta > 0$ and $\mu_1 - \mu_2 \geq 0$. The other cases can be treated in a similar manner and the results are summarized in Table II, where, for simplicity of notation, we have written $F(S)$ in place of $F_{n-3}\{S(y,x)\}$, etc, and t in place of $t_{n-3}(\gamma)$. In the absence of prior knowledge of the signs of β and $\mu_1 - \mu_2$ we must choose our test on the basis of the signs of s_{yx} and $\bar{x}_{1.} - \bar{x}_{2.}$ instead. But if we have strong prior knowledge of the signs of β and $\mu_1 - \mu_2$ we should use this knowledge to choose the appropriate test even if the signs of s_{yx} and $\bar{x}_{1.} - \bar{x}_{2.}$ indicate the use of another test. However, any contradiction with the prior belief of the signs must be weighed against the degree of belief and a cautious approach might be to use both tests.

TABLE II

BOUNDS ON THE SIGNIFICANCE OF THE INTERCEPT TEST

Sign of β	Sign of $\mu_1 - \mu_2$	Alternative hypothesis	Significance of sample Lower bound	Significance of sample Upper bound	Conservative critical region
+	+	$\alpha_1 > \alpha_2$	F(-T)	F(S)	$S < -t$
+	+	$\alpha_1 < \alpha_2$	F(-S)	F(T)	$T < -t$
+	-	$\alpha_1 > \alpha_2$	F(S)	F(-T)	$T > t$
+	-	$\alpha_1 < \alpha_2$	F(T)	F(-S)	$S > t$
-	+	$\alpha_1 > \alpha_2$	F(-S)	F(-T)	$T > t$
-	+	$\alpha_1 < \alpha_2$	F(T)	F(S)	$S < -t$
-	-	$\alpha_1 > \alpha_2$	F(-T)	F(-S)	$S > t$
-	-	$\alpha_1 < \alpha_2$	F(S)	F(T)	$T < -t$

We now consider extensions of the simplest model that allow the variances in the two groups to differ. In a practical situation the objective will normally be to find a model reasonably consistent with the data and reasonably economical in parameters. The number of possible models of the present type is great, of course, but we shall consider just two. In the first, model III, we allow the variances in the second group to be proportional to those in the first group, the constant of proportionality being unknown but the same for σ^2, σ_δ^2 and σ_ε^2. In the second extended model, model IV, the variances in one group are assumed unrelated to those in the other group. We use the likelihood ratio test for testing the appropriate hypothesis in each model. We now find the maximum of the likelihood in the first of these models.

If the variances in the second group are λ times those in the first group, the log-likelihood of model III is

$$\ell = -n\ \log(2\pi) - \tfrac{1}{2}n\ \log|\underset{\sim}{\Sigma}| - n_2\log\lambda - \{(\sigma^2+\sigma_\delta^2)(S_{11}+S_{12}/\lambda)$$

$$- 2\beta\sigma^2(S_{21}+S_{22}/\lambda) + (\beta^2\sigma^2+\sigma_\varepsilon^2)(S_{31}+S_{32}/\lambda)\}/(2|\underset{\sim}{\Sigma}|),$$

where $$|\underset{\sim}{\Sigma}| = \beta^2\sigma^2\sigma_\delta^2 + \sigma^2\sigma_\varepsilon^2 + \sigma_\delta^2\sigma_\varepsilon^2,$$

as before, and

$$S_{1i} = \sum_{j=1}^{n_i}(y_{ij} - \alpha - \beta\mu_i)^2, \quad S_{2i} = \sum_{j=1}^{n_i}(y_{ij} - \alpha - \beta\mu_i)(x_{ij} - \mu_i)$$

and $$S_{3i} = \sum_{j=1}^{n_i}(x_{ij} - \mu_i)^2 \quad (i=1,2).$$

Maximization of the likelihood is very similar to the maximization in model I. The likelihood equations for μ_1 and μ_2 yield

$$\mu_i = \{\beta\sigma_\delta^2(\bar{y}_{i.} - \alpha) + \sigma_\varepsilon^2\bar{x}_{i.}\}/(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2) \quad (i=1,2),$$

as in the simplest model, and the first two pivots are

$$-n_1(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)/|\underset{\sim}{\Sigma}| \quad \text{and} \quad -n_2(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)/(\lambda|\underset{\sim}{\Sigma}|).$$

The likelihood equation for α, on substitution of the above expressions for μ_i, yields

$$\alpha = \bar{y}_{..}(\lambda) - \beta\bar{x}_{..}(\lambda),$$

where $\quad \bar{y}_{..}(\lambda) = (\lambda n_1\bar{y}_{1.} + n_2\bar{y}_{2.})/(\lambda n_1 + n_2)$, etc.

The third pivot is

$$-(\lambda n_1 + n_2)/\{\lambda(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)\}.$$

On substitution of the above expressions for α and μ_i, $S_{11} + S_{12}/\lambda$ becomes

$$ns_{yy}(\lambda) + n\sigma_\varepsilon^4 B(\beta,\lambda)/(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)^2,$$

where $\quad ns_{yy}(\lambda) = \sum_j(y_{1j} - \bar{y}_{1.})^2 + \sum_j(y_{2j} - \bar{y}_{2.})^2/\lambda,$

$$nb_{yy}(\lambda) = n_1\{\bar{y}_{1.}-\bar{y}_{..}(\lambda)\}^2 + \frac{n_2}{\lambda}\{\bar{y}_{2.}-\bar{y}_{..}(\lambda)\}^2 = \frac{n_1 n_2(\bar{y}_{1.}-\bar{y}_{2.})^2}{\lambda n_1 + n_2},$$

etc, and $B(\beta,\lambda) = b_{yy}(\lambda) - 2\beta b_{yx}(\lambda) + \beta^2 b_{xx}(\lambda)$.

Similarly, $S_{21} + S_{22}/\lambda$ becomes

$$ns_{yx}(\lambda) - n\beta\sigma_\delta^2\sigma_\varepsilon^2 B(\beta,\lambda)/(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)^2$$

and $S_{31} + S_{32}/\lambda$ becomes

$$ns_{xx}(\lambda) + n\beta^2\sigma_\delta^4 B(\beta,\lambda)/(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)^2.$$

Thus maximization with respect to σ^2, σ_δ^2, σ_ε^2 and β becomes the same as in model I but with $s_{yy}(\lambda)$ replacing s_{yy}, $b_{yy}(\lambda)$ replacing b_{yy}, and so on. We thus get

$$\sigma^2(\lambda) = s_{yx}(\lambda)/\beta,$$

$$\sigma_\delta^2(\lambda) = s_{xx}(\lambda) - s_{yx}(\lambda)/\beta,$$

$$\sigma_\varepsilon^2(\lambda) = s_{yy}(\lambda) - \beta s_{yx}(\lambda),$$

and $$\beta = \frac{b_{yx}(\lambda)}{b_{xx}(\lambda)} = \frac{\bar{y}_{1.} - \bar{y}_{2.}}{\bar{x}_{1.} - \bar{x}_{2.}} = \frac{b_{yx}}{b_{xx}} = \hat{\beta}, \text{ say.}$$

Since $B(\hat{\beta},\lambda) = 0$, then, on substitution of the above estimates, $S_{11} + S_{12}/\lambda$ becomes $ns_{yy}(\lambda)$, $S_{21} + S_{22}/\lambda$ becomes $ns_{yx}(\lambda)$ and $S_{31} + S_{32}/\lambda$ becomes $ns_{xx}(\lambda)$. Hence, we find that the log-likelihood when maximized with respect to all parameters except λ is

$$\hat{\ell}(\lambda) = -n\{1 + \log(2\pi)\} - \tfrac{1}{2}n \log|\hat{\underset{\sim}{\Sigma}}(\lambda)| - n_2\log\lambda$$

where $$|\hat{\underset{\sim}{\Sigma}}(\lambda)| = s_{yy}(\lambda)s_{xx}(\lambda) - s_{yx}^2(\lambda).$$

Maximization of $\hat{\ell}(\lambda)$ with respect to λ yields the ML estimate of λ, namely

$$\hat{\lambda} = \{t_{12} + (t_{12}^2 + 16t_{11}t_{22})^{\frac{1}{2}}\}/(4t_{11}),$$

where $$t_{ii} = n_i(s_{yyi}s_{xxi} - s_{yxi}^2) \quad (i=1,2),$$

$$t_{12} = (n_1-n_2)(s_{yy1}s_{xx2} - 2s_{yx1}s_{yx2} + s_{yy2}s_{xx1}),$$

and $$s_{yyi} = \sum_j (y_{ij} - \bar{y}_{i.})^2/n_i, \text{ etc.}$$

The maximum of the log-likelihood in model I is

$$\ell_o = -n\{1 + \log(2\pi)\} - \tfrac{1}{2}n \log(s_{yy}s_{xx} - s_{yx}^2),$$

and so we use

$$2\{\hat{\ell}(\hat{\lambda}) - \ell_o\} = -n \log|\hat{\Sigma}(\hat{\lambda})| + n \log(s_{yy}s_{xx} - s_{yx}^2) + 2n_2\log\hat{\lambda}$$

as a test statistic for testing the hypothesis that $\lambda = 1$ against the alternative that $\lambda \neq 1$. This statistic is asymptotically distributed as χ_1^2 when $\lambda = 1$.

As in the simplest model, the maximum of the likelihood may not be in the parameter space. If there were a significant probability of this occuring we could do better by taking account of it through the use of the ratio of the maxima of the likelihoods in the respective parameter spaces rather than simply the ratio of the likelihoods at the respective internal estimates. However, this complicated task will not be attempted here.

We now turn to the other extension of the simplest model in which the variances in one group are unrelated to the variances in the other group. While the extension we have just considered is included as a special case of the present model, we will get a more sensitive test against the particular departure from the simplest model by using the above test rather the test we are about to derive. The specifications of the model, model IV, are the same as for model I but with σ^2 replaced by σ_i^2, σ_δ^2 replaced by $\sigma_{\delta i}^2$ and σ_ε^2 replaced by $\sigma_{\varepsilon i}^2$. Thus, for example, $\sigma_{\delta 1}^2$ is the variance of the errors in the X variable for group 1 while $\sigma^2_{\delta 2}$ is the error variance in the X variable for group 2. Define the new parameters

$$\phi_i = \beta^2\sigma_i^2 + \sigma^2_{\varepsilon i} \quad , \quad \gamma_i = \beta\sigma_i^2 \quad \text{and} \quad \chi_i = \sigma_i^2 + \sigma^2_{\delta i} .$$

Then the log-likelihood of the observations is

$$\ell = -n \log(2\pi) - \tfrac{1}{2}\sum n_i \log|\underset{\sim}{\Sigma}_i| - \tfrac{1}{2}\sum (\chi_i S_{1i} - 2\gamma_i S_{2i} + \phi_i S_{3i})/|\underset{\sim}{\Sigma}_i| ,$$

where
$$|\underset{\sim}{\Sigma}_i| = \phi_i \chi_i - \gamma_i^2$$

and S_{1i}, S_{2i} and S_{3i} are defined as before.

The solution of the likelihood equations is again very similar to the solution in model I, though in this case it is best to introduce the equation for β before the equations for χ_i, γ_i and ϕ_i. We again find that there are two turning points, the maximum occuring when $\mu_i = \bar{x}_{i.}$ (i=1,2), $\alpha = \bar{y}_{..} - \beta\bar{x}_{..}$, $\beta = b_{yx}/b_{xx}$, $\chi_i = s_{xxi}$, $\gamma_i = s_{yxi}$ and $\phi_i = s_{yyi}$. The maximum of the log-likelihood for model IV is thus

$$\tilde{\ell} = -n\{1 + \log(2\pi)\} - \tfrac{1}{2}\sum n_i \log(s_{yyi}s_{xxi} - s^2_{yxi}) .$$

Hence, for testing the null hypothesis that $\sigma_1^2 = \sigma_2^2$, $\sigma^2_{\delta 1} = \sigma^2_{\delta 2}$ and $\sigma^2_{\varepsilon 1} = \sigma^2_{\varepsilon 2}$ against the general alternative we use the statistic

$$2(\tilde{\ell} - \ell_o) = n \log(s_{yy}s_{xx} - s^2_{yx}) - \sum_i n_i \log(s_{yyi}s_{xxi} - s^2_{yxi}) ,$$

which, when the null hypothesis is true, is asymptotically distributed as χ^2_3.

2.5 Tests of hypotheses about the slope

In this section we consider tests of the null hypothesis that $\beta = \beta_o$ against the alternative that $\beta \neq \beta_o$, all other parameters being unspecified in both hypotheses. From such tests we can obtain also confidence intervals for β by defining the $(1-\gamma)$ confidence interval for β to be the set of all β_o for which the null hypothesis that $\beta = \beta_o$ would not be rejected at significance level γ. First we find a test statistic, based on the ML estimate of β, which is asymptotically normally distributed.

As the number of observations, n, tends to infinity, the probability that the local maximum of the likelihood lies inside the parameter space tends to one, unless $\beta = 0$ or one of the error variances, σ_δ^2 or σ_ε^2, is zero. Hence the asymptotic properties of the ML estimator of β are just those of the internal estimator, b_{YX}/b_{XX}. Now

$$\tilde{\beta} = b_{YX}/b_{XX} = (\bar{Y}_{1.} - \bar{Y}_{2.})/(\bar{X}_{1.} - \bar{X}_{2.})$$

is the ratio of two correlated normal variables. The numerator, $(\bar{Y}_{1.} - \bar{Y}_{2.})$, has mean $\beta_o(\mu_1 - \mu_2)$ and variance $n(\beta_o^2\sigma^2+\sigma_\varepsilon^2)/(n_1 n_2)$ when $\beta = \beta_o$, the denominator, $\bar{X}_{1.} - \bar{X}_{2.}$, has mean $\mu_1 - \mu_2$ and variance $n(\sigma^2+\sigma_\delta^2)/(n_1 n_2)$, and their covariance is $n\beta_o\sigma^2/(n_1 n_2)$. The exact distribution of $\tilde{\beta}$ is given, for example, by Hinkley (1969) where it is shown also that

$$U^{\dagger} = (\tilde{\beta} - \beta_o)\left\{ \frac{nb_{\mu\mu}}{(\sigma^2+\sigma_\delta^2)\tilde{\beta}^2 - 2\beta_o\tilde{\beta}\sigma^2 + \beta_o^2\sigma^2+\sigma_\varepsilon^2} \right\}^{\frac{1}{2}},$$

where $$b_{\mu\mu} = \sum n_i(\mu_i - \bar{\mu})^2/n,$$

is approximately distributed as N(0,1). In fact, the distribution of $U^{\dagger}$ tends to the N(0,1) distribution as the probability that $\bar{x}_{1.} - \bar{x}_{2.}$ has constant sign tends to one.

In Appendix 3 we show that

$$E(\tilde{\beta}) = \beta + O(1/n)$$

and $$\text{var}(\tilde{\beta}) = (\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)/(nb_{\mu\mu}) + O(1/n^2).$$

Hence, under the null hypothesis that $\beta = \beta_o$,

$$U^* = (\tilde{\beta} - \beta_o)\left\{ \frac{nb_{\mu\mu}}{\beta_o^2\sigma_\delta^2 + \sigma_\varepsilon^2} \right\}^{\frac{1}{2}}$$

will also be approximately distributed as N(0,1). The choice between the statistics found from $U^{\dagger}$ and U^* by replacing the unknown parameters by their ML estimators is not obvious but a limited numerical study suggested that the statistic based on U^* was the better. So we shall concentrate just on U^*. We now need the ML estimators of the parameters in the null hypothesis model in which $\beta = \beta_o$ is known.

In the solution of the likelihood equations in §2.2, the likelihood equation for β was not used until the end. Equations (4), (11), (12), (13), and (14) thus give us the ML estimates for the model with $\beta = \beta_o$ known. These are, for $\beta_o \neq 0$,

$$\tilde{\mu}_i = \frac{(\beta_o s_{xx} - s_{yx})(\bar{y}_{i.} - \bar{y}_{..} + \beta_o \bar{x}_{..}) + (s_{yy} - \beta_o s_{yx})\bar{x}_{i.}}{W(\beta_o)} \quad (i=1,2),$$

$$\tilde{\alpha} = \bar{y}_{..} - \beta_o \bar{x}_{..} ,$$

$$\tilde{\sigma}^2 = \frac{s_{yx}}{\beta_o} - \frac{(\beta_o s_{xx} - s_{yx})(s_{yy} - \beta_o s_{yx})B(\beta_o)}{\beta_o W^2(\beta_o)} ,$$

$$\tilde{\sigma}_\delta^2 = \frac{(\beta_o s_{xx} - s_{yx})T(\beta_o)}{\beta_o W(\beta_o)} ,$$

and

$$\tilde{\sigma}_\varepsilon^2 = \frac{(s_{yy} - \beta_o s_{yx})T(\beta_o)}{W(\beta_o)} .$$

Of course these estimates are only internal estimates, being appropriate when $\tilde{\sigma}^2$, $\tilde{\sigma}_\delta^2$ and $\tilde{\sigma}_\varepsilon^2$ are all non-negative. The various boundary estimates apply otherwise. The estimates on the $\sigma_\delta^2 = 0$ and $\sigma_\varepsilon^2 = 0$ boundaries are found later but for the meantime we assume that the boundary estimates are not important.

Substitution of the internal estimates of the parameters into U^* gives us, for $\beta_o \neq 0$, the statistic

$$U(\beta_o) = \frac{n^{\frac{1}{2}}(b_{YX}/b_{XX} - \beta_o)\{(\beta_o s_{XX} - s_{YX})b_{YX} + (s_{YY} - \beta_o s_{YX})b_{XX}\}}{\{W^2(\beta_o)T(\beta_o)b_{XX}\}^{\frac{1}{2}}} ,$$

which we use as an approximately N(0,1) distributed test statistic. This test will be asymptotically efficient. But if we have a moderate sample size and the group means are close together, then the test will be far from efficient for the same reasons that b_{YX}/b_{XX} is a poor estimator of β in this sort of situation. Essentially the problem is that b_{YX}/b_{XX}, the internal estimator of β, will have a much larger variance than the ML estimator. Improving the estimate of the variance of b_{YX}/b_{XX} by taking account of the boundary estimates or obtaining the exact distribution of b_{YX}/b_{XX} will not overcome this problem. Instead we must look for a quite different test statistic.

One way of dealing with problems involving nuisance parameters is to condition on a sufficient statistic for them. By doing so we will get a similar test. But, unfortunately, the minimal sufficient statistic for the nuisance parameters in the present problem is the same as the minimal sufficient statistic for all the parameters (ie β and the nuisance parameters) when $\beta \neq 0$, so that this approach is not possible. For $\beta = 0$, however, the dimension of the minimal sufficient statistic for the nuisance parameters is two less than the dimension of the minimal sufficient statistic for all the parameters. Hence we can find the joint distribution of the observations conditional on this sufficient statistic. The derivation of this distribution is rather long and the distribution itself is complex and so they will not be given here. The test based on the derivative of the conditional log-likelihood with respect to β, $\partial \ell_c/\partial \beta$ say, evaluated at $\beta = 0$, would be locally most powerful among all tests based on the conditional distribution if it were not for the fact that it is a function of the unknown nuisance parameters. But as the derivative does contain these nuisance parameters we must replace them by the ML estimators in the null hypothesis model. We now derive these estimators.

When $\beta = 0$ the log-likelihood is

$$\ell = -n \log(2\pi) - \tfrac{1}{2}n \log\sigma_\varepsilon^2 - \tfrac{1}{2}n \log(\sigma^2+\sigma_\delta^2)$$

$$- \sum\sum(y_{ij}-\alpha)^2/(2\sigma_\varepsilon^2) - \sum\sum(x_{ij}-\mu_i)^2/\{2(\sigma^2+\sigma_\delta^2)\},$$

from which it is immediately obvious that σ^2 and σ_δ^2 cannot be estimated separately. We must treat $\sigma^2+\sigma_\delta^2$ as a single parameter. We again solve the likelihood equations sequentially and check on the signs of the pivots of the matrix of double derivatives of the log-likelihood. From the derivative of ℓ with respect to μ_i we get $\bar{x}_{i.}$ as the estimate of μ_i. The first two pivots are both $-1/(\sigma^2+\sigma_\delta^2)$. From $\partial\ell/\partial\alpha$ we get $\bar{y}_{..}$ as the estimate of α and a third pivot equal to $-1/\sigma_\varepsilon^2$. The derivative with respect to $\sigma^2+\sigma_\delta^2$ with μ_i replaced by $\bar{x}_{i.}$ gives us s_{xx} as the estimate of $\sigma^2+\sigma_\delta^2$ and $-\tfrac{1}{2}n/s_{xx}^2$ as the fourth pivot (evaluated at $\sigma^2+\sigma_\delta^2 = s_{xx}$). Similarly we get t_{yy} as the estimate of σ_ε^2 and $-\tfrac{1}{2}n/t_{yy}^2$ as the final pivot. All the pivots are negative so that the unique turning point is a local maximum.

Dividing $\partial \ell_c/\partial\beta$, evaluated at $\beta = 0$, by its conditional standard deviation (conditional on the sufficient statistic for the nuisance parameters) and then replacing the unknown nuisance parameters by the estimators we have just found, we get the statistic

$$Q^* = (n-1)^{\frac{1}{2}} t_{YX}/(t_{YY}t_{XX})^{\frac{1}{2}}.$$

Actually we have to estimate σ^2 by the estimate of $\sigma^2+\sigma_\delta^2$, s_{xx}, because σ^2 and σ_δ^2 cannot be estimated separately. We could use Q^* as an approximately $N(0,1)$ distributed statistic but we can actually get an exact distribution.

Let $V = Q^{*2}/(n-1)$. Then

$$Q^2 = \frac{(n-2)V}{1-V} = \frac{(n-2)t_{YX}^2/t_{XX}}{t_{YY} - t_{YX}^2/t_{XX}}$$

can easily be shown to have an F distribution with 1 and n-2 degrees of freedom. First we condition on $X_{ij} = x_{ij}$ for all i and j, and then we use Cochran's theorem to show that

$$\frac{nt_{YX}^2/t_{XX}}{|\underset{\sim}{\Sigma}|/(\sigma^2+\sigma_\delta^2)} \quad \text{and} \quad \frac{n(t_{YY} - t_{YX}^2/t_{XX})}{|\underset{\sim}{\Sigma}|/(\sigma^2+\sigma_\delta^2)}$$

have independent (non-central) χ^2 distributions and so their ratio has a non-central F distribution. When $\beta = 0$ the non-centrality parameters both become zero so that the ratio has a central F distribution. But this distribution does not depend on the x_{ij} so that the result holds unconditionally. If we let $R = t_{YX}/(t_{YY}t_{XX})^{\frac{1}{2}}$ then

$$Q = (n-2)^{\frac{1}{2}}R/(1-R^2)^{\frac{1}{2}}$$

has a Student t distribution with n-2 degrees of freedom when $\beta = 0$; we use this for testing the hypothesis that $\beta = 0$.

But we still do not have a satisfactory test of the hypothesis that $\beta = \beta_0$ when $\beta_0 \neq 0$, so let us try the Neyman $C(\alpha)$ test (Neyman, 1959). Letting ℓ be the log-likelihood function for the full model and $\underset{\sim}{\gamma}$ represent the nuisance parameters, we calculate the function

$$\partial\ell/\partial\beta - (\partial\ell/\partial\underset{\sim}{\gamma})^T\underset{\sim}{b} ,$$

where the vector $\underset{\sim}{b}$ is defined by

$$E\left(\frac{\partial^2\ell}{\partial\beta\partial\underset{\sim}{\gamma}}\right) - E\left(\frac{\partial^2\ell}{\partial\underset{\sim}{\gamma}^2}\right)\underset{\sim}{b} = \underset{\sim}{0}.$$

Dividing this function by its standard deviation and replacing the unknown nuisance parameters by their ML estimators in the model where $\beta = \beta_0$ is known, we get the Neyman $C(\alpha)$ test statistic which we use as an approximately $N(0,1)$ distributed variable. The calculations required to get this statistic are long and tedious and so no details will be given here. We get the statistic from

$$\sum\sum(\mu_i - \bar{\mu})(Y_{ij} - \alpha - \beta_o X_{ij})/\{(\beta_o^2\sigma_\delta^2 + \sigma_\varepsilon^2)\sum n_i(\mu_i - \bar{\mu})^2\}^{\frac{1}{2}} \quad (\beta_o \neq 0)$$

by replacing the unknown parameters by their ML estimators. We now derive these ML estimators.

The internal estimates were found earlier in this section. We saw that they had virtually been obtained in §2.2 and the same is true of the $(\sigma_\delta^2=0)$- and $(\sigma_\varepsilon^2=0)$-boundary estimates. The $(\sigma_\delta^2=0)$-boundary estimates for μ_i (i=1,2), α, σ^2, σ_δ^2 and σ_ε^2 are $\bar{x}_{i.}$ (i=1,2), $\bar{y}_{..} - \beta_o\bar{x}_{..}$, s_{xx}, 0 and $T(\beta_o)$, respectively, and the $(\sigma_\varepsilon^2=0)$-boundary estimates of the same parameters are $\bar{x}_{..} + (\bar{y}_{i.} - \bar{y}_{..})/\beta_o$ (i=1,2), $\bar{y}_{..} - \beta_o\bar{x}_{..}$, s_{yy}/β_o^2, $T(\beta_o)/\beta_o^2$ and 0, respectively. Explicit expressions for the $(\sigma^2=0)$-boundary estimates (ie, the estimates when the true X-variance is zero) could not, however, be obtained. So let us make the approximation of ignoring the $(\sigma^2=0)$-boundary estimates and then examine the $C(\alpha)$ statistic. It turns out that the statistics we get by using the internal estimates, the $(\sigma_\delta^2=0)$-boundary estimates and the $(\sigma_\varepsilon^2=0)$-boundary estimates are all the same, namely

$$C(\beta_o) = n^{\frac{1}{2}}(b_{YX} - \beta_o b_{XX})/\{T(\beta_o)b_{XX}\}^{\frac{1}{2}}.$$

But it is clear that this is really no improvement on the test based on the ML estimator of β.

The above tests are inadequate because they fail to take account of the fact that if β_o is well outside the range of the within-group slopes though not significantly far from the between-group slope then a departure from the null hypothesis is suggested. The information about the slope comes from two sources, the between-group slope and the range of the within-group slopes. The tests we have obtained so far, except the test for the null hypothesis that $\beta = 0$, use only the first source but it is not difficult to construct a test which uses both sources of information.

First suppose that we have a sample in which β_o lies within the range of the within-group slopes. Then no departure from the null hypothesis is indicated by these within-group slopes and so we can use only the first source of information and test for such a departure with a statistic based on the between-group slope. That is, we use the statistic $U(\beta_o)$.

But suppose that β_o lies outside the range of the two within-group slopes. Then both sources of information can be used. The appropriate set of estimates of the parameters is one of the sets of boundary estimates. The regions of the sample space in which the different sets of estimates are appropriate cannot be specified just in terms of inequalities involving β_o (like we did in §2.2 where the regions could be specified by inequalities involving b_{yx}/b_{xx}). However, as in §2.2, we can get a good approximation by choosing the regions so that the internal estimates are used when β_o lies between s_{yx}/s_{xx} and s_{yy}/s_{yx}, the

$(\sigma_\delta^2=0)$-boundary estimates are used when β_o lies between s_{yx}/s_{xx} and $-\text{sign}(s_{yx})(s_{yy}/s_{xx})^{\frac{1}{2}}$, and the $(\sigma_\varepsilon^2=0)$-boundary estimates are used when β_o lies elsewhere. First suppose that the sample is such that β_o lies between s_{yx}/s_{xx} and $-\text{sign}(s_{yx})(s_{yy}/s_{xx})^{\frac{1}{2}}$; then we use the $(\sigma_\delta^2=0)$- boundary estimates. In particular, our estimate of σ_δ^2 is zero. Now, if we knew that $\sigma_\delta^2=0$ then we would be in the familiar Y on X regression situation and would use the overall regression slope for testing the null hypothesis that $\beta=\beta_o$. So it seems reasonable to do the same here. We now derive a test based on the overall regression slope of Y on X, t_{YX}/t_{XX}.

We shall reject the null hypothesis if

$$Z = |t_{YX}/t_{XX} - E(t_{YX}/t_{XX}\,;\,\beta=\beta_o)| > k(\gamma),$$

where $k(\gamma)$ is chosen so that the probability of rejection, given that β_o lies between s_{YX}/s_{XX} and $-\text{sign}(s_{YX})(s_{YY}/s_{XX})^{\frac{1}{2}}$, is γ. We assume that

$$\Pr\{-(s_{YY}/s_{XX})^{\frac{1}{2}} > \beta_o \mid s_{YX}>0\,;\,\beta=\beta_o\}$$

and $$\Pr\{(s_{YY}/s_{XX})^{\frac{1}{2}} < \beta_o \mid s_{YX}<0\,;\,\beta=\beta_o\}$$

are both negligible. Hence, when $s_{YX}>0$, we condition on

$$s_{YX}/s_{XX} > \max(0,\beta_o)$$

and when $s_{YX}<0$ we condition on

$$s_{YX}/s_{XX} < \min(0,\beta_o).$$

Suppose first that $s_{YX} > 0$. Then we reject the null hypothesis if

$$Z > k(\gamma)$$

where $k(\gamma)$ is chosen so that

$$\Pr\{Z > k(\gamma) \mid s_{YX}/s_{XX} > \max(0,\beta_o) \; ; \; \beta = \beta_o\} = \gamma.$$

In other words, we reject the null hypothesis if

$$\Pr\{Z > z \mid s_{YX}/s_{XX} > \max(0,\beta_o) \; ; \; \beta = \beta_o\} < \gamma.$$

To evaluate this conditional probability we approximate the joint distribution of t_{YX}/t_{XX} and s_{YX}/s_{XX} by the bivariate normal distribution. Expressions for the asymptotic moments of t_{YX}/t_{XX} and s_{YX}/s_{XX} are given in Appendix 3 and we approximate the means and variances in the normal distribution by these expressions with the unknown parameters replaced by their ML estimates when $\beta = \beta_o$. Note that the ML estimates are always the $(\sigma_\delta^2=0)$-boundary estimates when we use this test. Define

$$W = (t_{YX}/t_{XX} - \beta_o)\{nt_{XX}/T(\beta_o)\}^{\frac{1}{2}},$$

$$V = (s_{YX}/s_{XX} - \beta_o)\{(n-4)s_{XX}/T(\beta_o)\}^{\frac{1}{2}}$$

and

$$b = \min(0,\beta_o)\{(n-4)s_{xx}/T(\beta_o)\}^{\frac{1}{2}}.$$

Then we reject the null hypothesis if

$$\Pr\{|W| > |w| \mid V > -b \; ; \; \beta = \beta_o\} < \gamma$$

where W and V have a bivariate normal distribution with zero means, unit variances and correlation

$$\rho = (s_{xx}/t_{xx})^{\frac{1}{2}}.$$

Let $\Phi(.,.;\rho)$ be the distribution function of this distribution and let $\Phi(.)$ be the distribution function of the standard univariate normal distribution. Then we find that the null hypothesis is to be rejected if

$$\frac{\Phi(b) - \Phi(|w|,b;\rho) + \Phi(-|w|,b;\rho)}{\Phi(b)} < \gamma.$$

For $s_{YX} < 0$ we get exactly the same result except that b is replaced by

$$c = \min(0,-\beta_o)\{(n-4)s_{xx}/T(\beta_o)\}^{\frac{1}{2}}.$$

Define

$$e = \min\{0 , \text{sign}(s_{yx})\beta_o\}\{(n-4)s_{xx}/T(\beta_o)\}^{\frac{1}{2}}.$$

Then if β_o lies between s_{yx}/s_{xx} and $-\text{sign}(s_{yx})(s_{yy}/s_{xx})^{\frac{1}{2}}$ we reject the null hypothesis that $\beta = \beta_o$ if

$$\frac{\Phi(e) - \Phi(|w|,e;\rho) + \Phi(-|w|,e;\rho)}{\Phi(e)} < \gamma. \qquad (19)$$

Before we go on to find a test for the case when $1/\beta_o$ lies between s_{yx}/s_{yy} and $-\text{sign}(s_{yx})(s_{xx}/s_{yy})^{\frac{1}{2}}$, let us see whether or not the above test really does have size approximately equal to γ. The doubt over the achievement of this aim stems from the fact that we always estimate σ_δ^2 by zero, which is a lower bound for

σ_δ^2. This means that, when σ_δ^2 is not zero, we tend to under-estimate the variances of s_{YX}/s_{XX} and t_{YX}/t_{XX} and to over-estimate the absolute values of their expectations. To see the effect of this we first note that e will most commonly be zero and the numerator in (19) will usually be dominated by $\Phi(-|w|,e;\rho)$. So let us consider the effect of the under-estimation of σ_δ^2 on $\Phi(-|w|,0;\rho)$. Clearly $|w|$ will tend to be larger than

$$\frac{|t_{yx}/t_{xx} - E(t_{YX}/t_{XX})|}{\{var(t_{YX}/t_{XX})\}^{\frac{1}{2}}}$$

so that the rejection criterion (19) will tend to reject fewer samples than planned. That is, the test size will tend to be less than γ.

It can be seen on general grounds that the test will have size approximately γ when σ_δ^2 is zero but will become more conservative as σ_δ^2 increases. Of course, as σ_δ^2 increases (relative to σ^2 and σ_ε^2), it becomes less likely that this test will be used, which is an important compensation. However, given only that this test is used on a particular sample, we cannot say how conservative it is. There seems to be no (non-Bayesian) way of satisfactorily adjusting for this conservatism. We might consider replacing γ in (19) by some compromise value greater than γ but this would mean that for some values of the parameters the test would have size greater than γ while for other values it would have size less than γ; it seems better to leave the test as it is.

Now let us consider the case when $1/\beta_o$ lies between s_{yx}/s_{yy} and $-\mathrm{sign}(s_{yx})(s_{xx}/s_{yy})^{\frac{1}{2}}$. Then we always estimate σ_ε^2 by zero and so, as we would if we knew σ_ε^2 were zero, we base our test statistic on the overall regression slope of X on Y. Rewriting the null hpothesis as $1/\beta = 1/\beta_o$ and interchanging the roles of Y and X, we can use the previous results. Define

$$w^* = (t_{yx}/t_{yy} - 1/\beta_o)\{n\beta_o^2 t_{yy}/T(\beta_o)\}^{\frac{1}{2}},$$

$$e^* = \min\{0\ , \mathrm{sign}(s_{yx})1/\beta_o\}\{(n-4)\beta_o^2 s_{yy}/T(\beta_o)\}^{\frac{1}{2}},$$

and

$$\rho^* = (s_{yy}/t_{yy})^{\frac{1}{2}}.$$

Then, when $1/\beta_o$ lies between s_{yx}/s_{yy} and $-\mathrm{sign}(s_{yx})(s_{xx}/s_{yy})^{\frac{1}{2}}$, we reject the null hypothesis that $\beta = \beta_o$ if

$$\frac{\Phi(e^*) - \Phi(|w^*|,e^*;\rho^*) + \Phi(-|w^*|,e^*;\rho^*)}{\Phi(e^*)} < \gamma.$$

To summarize, then, our test procedure is as follows. If β_o $(\neq 0)$ lies between s_{yx}/s_{xx} and s_{yy}/s_{yx}, we reject the null hypothesis that $\beta = \beta_o$ if

$$|u(\beta_o)| > k_{\frac{1}{2}\gamma}\ ,$$

where $k_{\frac{1}{2}\gamma}$ is the upper $\frac{1}{2}\gamma$ point of the standard univariate normal distribution. If β_o lies between s_{yx}/s_{xx} and $-\mathrm{sign}(s_{yx})(s_{yy}/s_{xx})^{\frac{1}{2}}$, we reject the null hypothesis if

$$\frac{\Phi(e) - \Phi(|w|,e;\rho) + \Phi(-|w|,e;\rho)}{\Phi(e)} < \gamma.$$

And if $1/\beta_o$ lies between s_{yx}/s_{yy} and $-\text{sign}(s_{yx})(s_{xx}/s_{yy})^{\frac{1}{2}}$, we reject the null hypothesis if

$$\frac{\Phi(e^*) - \Phi(|w^*|,e^*;\rho^*) + \Phi(-|w^*|,e^*;\rho^*)}{\Phi(e^*)} < \gamma.$$

The test based on $u(\beta_o)$ has size approximately γ while the other tests have size approximately equal to or less than γ.

For confirmation of the conclusions about the sizes of the different tests, ten thousand samples were drawn from a simulated model with $n_1 = n_2 = 20$, $\beta = \beta_o = 1$, $\alpha = 0$, $\mu_1 = 0$, $\mu_2 = 10$, $\sigma^2 = 25$ and $\sigma_\delta^2 = \sigma_\varepsilon^2 = 1$. The size, γ, was set at 10% and the proportions rejected by the tests were calculated after each 1000 samples. The results are summarized in Table III. The number of times a particular test is used does, of course, vary slightly from one set of 1000 samples to the next but this in no way obscures the general picture. On average the test based on b_{YX}/b_{XX} was used on about 60% of the samples while each of the other two tests were used on about 20% of the samples.

Obtaining confidence intervals from the above tests is not a trivial matter but with routines available for computing bivariate normal probabilities it is not too difficult to do this on a computer.

TABLE III

REJECTION PERCENTAGES OF THE SLOPE TESTS

No. of samples	Rejection percentages, at 10% significance level, of test based on		
	b_{YX}/b_{XX}	t_{YX}/t_{XX}	t_{YY}/t_{YX}
1000	10.9	2.7	6.2
1000	12.5	7.6	6.3
1000	11.6	5.3	6.1
1000	10.7	5.9	8.9
1000	9.8	3.5	5.6
1000	8.6	10.3	4.8
1000	10.3	5.7	6.6
1000	10.6	9.1	5.8
1000	10.5	7.0	9.6
1000	9.2	6.4	5.3
Average rejection percentage	10.45	6.38	6.55
Total no. of times test used	6056	1976	1968

The tests we have derived in the latter part of this section are for testing the null hypothesis that

$\beta = \beta_o$ only when $\beta_o \neq 0$. When $\beta_o = 0$ we use the statistic Q defined earlier in the section.

Chapter 3

THE K-GROUP, TWO-VARIABLE MODEL

3.1 Specification of the model

In this chapter we consider an extension of the two-group model to allow for the inclusion of more than two groups of data. The model has the same specifications as the two-group model of Chapter 2 except that now the data come from $k \geq 2$ groups. That is, the observations $\{(y_{ij},x_{ij}) ; j=1,..,n_i , i=1,..,k\}$ come from the model

$$Y_{ij} = \alpha + \beta U_{ij} + \varepsilon_{ij}$$

and
$$X_{ij} = U_{ij} + \delta_{ij} ,$$

where the U_{ij} are independently distributed as $N(\mu_i,\sigma^2)$, the ε_{ij} have independent $N(0,\sigma_\varepsilon^2)$ distributions, the δ_{ij} have independent $N(0,\sigma_\delta^2)$ distributions, and the errors, ε_{ij} and δ_{ij}, are independent of each other and of the U_{ij}. We shall call this model V.

As in the two-group model, the (Y_{ij},X_{ij}) have independent normal distributions with means $(\alpha+\beta\mu_i,\mu_i)$ and variance

$$\underset{\sim}{\Sigma} = \begin{pmatrix} \beta^2\sigma^2+\sigma_\varepsilon^2 & \beta\sigma^2 \\ \beta\sigma^2 & \sigma^2+\sigma_\delta^2 \end{pmatrix},$$

and the log-likelihood of the observations is

$$\ell = -n\log(2\pi) - \tfrac{1}{2}n\log|\underset{\sim}{\Sigma}| - \tfrac{1}{2}\{(\sigma^2+\sigma_\delta^2)S_1 - 2\beta\sigma^2 S_2 + (\beta^2\sigma^2+\sigma_\varepsilon^2)S_3\}/|\underset{\sim}{\Sigma}| ,$$

where $n = \sum n_i$,

$$|\underset{\sim}{\Sigma}| = \beta^2\sigma^2\sigma_\delta^2 + \sigma^2\sigma_\varepsilon^2 + \sigma_\delta^2\sigma_\varepsilon^2 ,$$

$$S_1 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \alpha - \beta\mu_i)^2,$$

$$S_2 = \sum\sum(y_{ij} - \alpha - \beta\mu_i)(x_{ij} - \mu_i),$$

and

$$S_3 = \sum\sum(x_{ij} - \mu_i)^2.$$

3.2 Maximum likelihood estimation of the parameters

In the solution of the likelihood equations for the two-group model in §2.2, the special property that holds when there are two groups but not when there are more than two groups, namely $b_{yy}b_{xx} = b_{yx}^2$, was deliberately not used until near the end, when it was used to simplify the expression for the estimate of β. Hence, as in the two-group model, there is one saddle point and one local maximum, the local maximum occuring at

$$\mu_i = \tilde{\mu}_i = \frac{(\tilde{\beta}s_{xx} - s_{yx})(\bar{y}_{i.} - \bar{y}_{..} + \tilde{\beta}\bar{x}_{..}) + (s_{yy} - \tilde{\beta}s_{yx})\bar{x}_{i.}}{W(\tilde{\beta})} ,$$

$$\alpha = \tilde{\alpha} = \bar{y}_{..} - \tilde{\beta}\bar{x}_{..} ,$$

$$\sigma^2 = \tilde{\sigma}^2 = \frac{s_{yx}}{\tilde{\beta}} - \frac{(\tilde{\beta}s_{xx} - s_{yx})(s_{yy} - \tilde{\beta}s_{yx})B(\tilde{\beta})}{\tilde{\beta}W^2(\tilde{\beta})} ,$$

$$\sigma_\delta^2 = \tilde{\sigma}_\delta^2 = \frac{(\tilde{\beta}s_{xx} - s_{yx})T(\tilde{\beta})}{\tilde{\beta}W(\tilde{\beta})} ,$$

$$\sigma_\varepsilon^2 = \tilde{\sigma}_\varepsilon^2 = \frac{(s_{yy} - \tilde{\beta}s_{yx})T(\tilde{\beta})}{W(\tilde{\beta})} ,$$

and
$$\beta = \tilde{\beta} = \frac{s_{yy}b_{xx} - s_{xx}b_{yy} - d(y,x)}{2(s_{yx}b_{xx} - s_{xx}b_{yx})} ,$$

where $d(y,x) =$

$$\{(s_{xx}b_{yy} - s_{yy}b_{xx})^2 - 4(s_{yx}b_{xx} - s_{xx}b_{yx})(s_{yy}b_{yx} - s_{yx}b_{yy})\}^{\frac{1}{2}},$$

$$B(\tilde{\beta}) = b_{yy} - 2\tilde{\beta}b_{yx} + \tilde{\beta}^2 b_{xx} , \quad W(\tilde{\beta}) = s_{yy} - 2\tilde{\beta}s_{yx} + \tilde{\beta}^2 s_{xx}$$

and
$$T(\tilde{\beta}) = W(\tilde{\beta}) + B(\tilde{\beta}).$$

Again this local maximum can lie outside the parameter space so that the above estimates are only internal estimates. When $\tilde{\sigma}^2$, $\tilde{\sigma}_\delta^2$ or $\tilde{\sigma}_\varepsilon^2$ is negative the appropriate estimates will be on one of the boundaries $\sigma^2 = 0$, $\sigma_\delta^2 = 0$, or $\sigma_\varepsilon^2 = 0$. We now find the maximum on each of these boundaries.

The special property of the two-group model was not used at all in obtaining the $(\sigma_\delta^2 = 0)$-boundary estimates in §2.2 and so the results apply equally to the more general k-group model. That is, the only turning point on the $(\sigma_\delta^2=0)$-boundary, a local maximum, occurs at

$$\mu_i = \bar{x}_{i.} \quad (i=1,\ldots,k),$$

$$\alpha = \bar{y}_{..} - (t_{yx}/t_{xx})\bar{x}_{..} ,$$

$$\sigma^2 = s_{xx} ,$$

$$\sigma_\delta^2 = 0,$$

$$\sigma_\varepsilon^2 = t_{yy} - t_{yx}^2/t_{xx} ,$$

and

$$\beta = t_{yx}/t_{xx} .$$

As for the $(\sigma_\delta^2=0)$-boundary estimates, the $(\sigma_\varepsilon^2=0)$-boundary estimates are unchanged in the k-group model. That is, the maximum on the $(\sigma_\varepsilon^2=0)$-boundary occurs at

$$\mu_i = \bar{x}_{..} + (t_{yx}/t_{yy})(\bar{y}_{i.} - \bar{y}_{..}) \quad (i=1,..,k),$$

$$\alpha = \bar{y}_{..} - (t_{yy}/t_{yx})\bar{x}_{..} ,$$

$$\sigma^2 = s_{yy}t_{yx}^2/t_{yy}^2 ,$$

$$\sigma_\delta^2 = t_{xx} - t_{yx}^2/t_{yy} ,$$

$$\sigma_\varepsilon^2 = 0 ,$$

and

$$\beta = t_{yy}/t_{yx} .$$

Finding the $(\sigma^2=0)$-boundary estimates is more difficult because in the two-group case we had to use the fact that $b_{yy}b_{xx} = b_{yx}^2$. We still get

$$\alpha = \bar{y}_{..} - \beta\bar{x}_{..}$$

and

$$\mu_i = \{\beta\sigma_\delta^2(\bar{y}_{i.} - \bar{y}_{..} + \beta\bar{x}_{..}) + \sigma_\varepsilon^2\bar{x}_{i.}\}/(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2) \quad (i=1,..,k)$$

from the likelihood equations for α and μ_i, and from the equation for β we get that β must be a solution of

$$\beta\sigma_\delta^2(b_{yy} - \beta b_{yx}) + \sigma_\varepsilon^2(b_{yx} - \beta b_{xx}) = 0 . \qquad (20)$$

From the likelihood equations for σ_δ^2 and σ_ε^2 we get

$$\sigma_\delta^2 = \sum\sum(x_{ij} - \mu_i)^2/n \quad \text{and} \quad \sigma_\varepsilon^2 = \sum\sum(y_{ij} - \alpha - \beta\mu_i)^2/n,$$

which, on substitution of the above expressions for α and μ_i, become

$$\sigma_\delta^2 = s_{xx} + \beta^2\sigma_\delta^4 B(\beta)/(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)^2 \qquad (21)$$

and $$\sigma_\varepsilon^2 = s_{yy} + \sigma_\varepsilon^4 B(\beta)/(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)^2. \qquad (22)$$

From equation (20) we can show that

$$\beta\sigma_\delta^2/(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2) = (\beta b_{xx} - b_{yx})/B(\beta)$$

and $$\sigma_\varepsilon^2/(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2) = (b_{yy} - \beta b_{yx})/B(\beta).$$

Substituting these into equations (21) and (22), we get

$$\sigma_\delta^2 = s_{xx} + (\beta b_{xx} - b_{yx})^2/B(\beta)$$

and $$\sigma_\varepsilon^2 = s_{yy} + (b_{yy} - \beta b_{yx})^2/B(\beta).$$

Hence, from equation (20), we get a quartic equation for β,

$$\{\beta s_{xx}(b_{yy} - \beta b_{yx}) - s_{yy}(\beta b_{xx} - b_{yx})\}B(\beta)$$
$$- (b_{yy} - \beta b_{yx})(\beta b_{xx} - b_{yx})(b_{yy} - \beta^2 b_{xx}) = 0, \qquad (23)$$

which requires numerical solution. Having found a numerical solution, we should, of course, check that it corresponds to the global maximum. This will not be particularly easy, in general, but we saw in the two-group model that the $(\sigma^2=0)$-boundary estimates are the least likely to be used and that we could safely ignore them in all but the most extreme cases. (By "extreme" we mean that the group means are very close together and β is near zero.)

To be strictly correct we should use the internal estimates when all the internal estimates of the variances are non-negative; we should use the $(\sigma_\delta^2=0)$-boundary estimates when one of the internal estimates of the variances is negative and a_{yx} is less than c_{yx} and $d(\beta^*)$, where

$$a_{yx} = s_{xx}(t_{yy} - t_{yx}^2/t_{xx}),$$

$$c_{yx} = s_{yy}(t_{xx} - t_{yx}^2/t_{yy}),$$

$$d(\beta^*) = \{s_{xx} + (\beta^* b_{xx} - b_{yx})^2/B(\beta^*)\}\{s_{yy} + (b_{yy} - \beta^* b_{yx})^2/B(\beta^*)\},$$

and β^* is a solution of equation (23); we should use the $(\sigma_\varepsilon^2=0)$-boundary estimates when one of the internal variance estimates is negative and c_{yx} is less than a_{yx} and $d(\beta^*)$; and we should use the $(\sigma^2=0)$-boundary estimates otherwise. However, as in the two-group case, we will usually be able to approximate the ML estimates by choosing the internal estimates when $\tilde{\beta}$, the internal estimate of β, lies between s_{yx}/s_{xx} and s_{yy}/s_{yx}, the $(\sigma_\delta^2=0)$-boundary estimates when $\tilde{\beta}$ lies outside the range of s_{yx}/s_{xx} and s_{yy}/s_{yx} but $|\tilde{\beta}| < (s_{yy}/s_{xx})^{\frac{1}{2}}$, and the $(\sigma_\varepsilon^2=0)$-boundary estimates otherwise.

3.3 Asymptotic variances of the estimators

In §2.3 we found the asymptotic variances of the ML estimators in the two-group model directly but in the more general k-group model this method would be too difficult because of the much more complicated expression for the estimator of β. Instead, we find the asymptotic variances by inverting the information matrix.

Let ℓ be the log-likelihood of the observations and define

$$i_{\alpha\alpha} = -E(\partial^2\ell/\partial\alpha^2), \quad \underset{\sim}{i}_{\alpha\mu} = -E(\partial^2\ell/\partial\alpha\partial\underset{\sim}{\mu}),$$

$$\underset{\sim}{i}_{\sigma\sigma} = -E(\partial^2\ell/\partial\underset{\sim}{\sigma}^2),$$

and so on, where $\underset{\sim}{\mu}^T = (\mu_1,\ldots,\mu_k)$ and $\underset{\sim}{\sigma}^T = (\sigma^2,\sigma_\delta^2,\sigma_\varepsilon^2)$. Let $\underset{\sim}{\gamma}^T = (\alpha, \underset{\sim}{\mu}^T, \underset{\sim}{\sigma}^T, \beta)$; then the information matrix is

$$\underset{\sim}{i}_{\gamma\gamma} = \begin{pmatrix} i_{\alpha\alpha} & \underset{\sim}{i}_{\alpha\mu}^T & \underset{\sim}{i}_{\alpha\sigma}^T & i_{\alpha\beta} \\ \underset{\sim}{i}_{\alpha\mu} & \underset{\sim}{i}_{\mu\mu} & \underset{\sim}{i}_{\mu\sigma} & \underset{\sim}{i}_{\mu\beta} \\ \underset{\sim}{i}_{\alpha\sigma} & \underset{\sim}{i}_{\mu\sigma}^T & \underset{\sim}{i}_{\sigma\sigma} & \underset{\sim}{i}_{\sigma\beta} \\ i_{\alpha\beta} & \underset{\sim}{i}_{\mu\beta}^T & \underset{\sim}{i}_{\sigma\beta}^T & i_{\beta\beta} \end{pmatrix}.$$

We can easily show that

$$i_{\alpha\alpha} = n(\sigma^2+\sigma_\delta^2)/|\boldsymbol{\Sigma}|, \quad \mathbf{i}_{\alpha\mu}^T = (\beta\sigma_\delta^2/|\boldsymbol{\Sigma}|)(n_1,\ldots,n_k), \quad \mathbf{i}_{\alpha\sigma} = \mathbf{0},$$

$$i_{\alpha\beta} = n\bar{\mu}(\sigma^2+\sigma_\delta^2)/|\boldsymbol{\Sigma}|, \quad \mathbf{i}_{\mu\mu} = \{(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)/|\boldsymbol{\Sigma}|\}\text{diag}(n_1,\ldots,n_k),$$

$$\mathbf{i}_{\mu\sigma} = \mathbf{0}, \quad \mathbf{i}_{\mu\beta}^T = (\beta\sigma_\delta^2/|\boldsymbol{\Sigma}|)(n_1\mu_1,\ldots,n_k\mu_k),$$

$$\mathbf{i}_{\sigma\beta}^T = (n\beta\sigma^2/|\boldsymbol{\Sigma}|)\left(\sigma_\delta^2(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2), -\sigma^2\sigma_\varepsilon^2, \sigma_\delta^2(\sigma^2+\sigma_\delta^2)\right),$$

$$i_{\beta\beta} = n\sigma^4(|\boldsymbol{\Sigma}| + 2\beta^2\sigma_\delta^4)/|\boldsymbol{\Sigma}|^2 + (\sigma^2+\sigma_\delta^2)\sum n_i\mu_i^2/|\boldsymbol{\Sigma}|,$$

and finally that

$$\mathbf{i}_{\sigma\sigma} = \frac{n}{2|\boldsymbol{\Sigma}|^2}\begin{pmatrix} (\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)^2 & \sigma_\varepsilon^4 & \beta^2\sigma_\delta^4 \\ \sigma_\varepsilon^4 & (\beta^2\sigma^2+\sigma_\varepsilon^2)^2 & \beta^2\sigma^4 \\ \beta^2\sigma_\delta^4 & \beta^2\sigma^4 & (\sigma^2+\sigma_\delta^2)^2 \end{pmatrix}.$$

Let $v_{\alpha\alpha}$, $\mathbf{v}_{\alpha\mu}^T$, etc be the corresponding terms in

$$\mathbf{v}_{\gamma\gamma} = \mathbf{i}_{\gamma\gamma}^{-1}.$$

We wish to find $v_{\alpha\alpha}$, $v_{\beta\beta}$ and the diagonal elements of $\mathbf{v}_{\mu\mu}$ and $\mathbf{v}_{\sigma\sigma}$. Let

$$\mathbf{A} = \begin{pmatrix} i_{\alpha\alpha} & \mathbf{i}_{\alpha\mu}^T \\ \mathbf{i}_{\alpha\mu} & \mathbf{i}_{\mu\mu} \end{pmatrix} \quad \text{and} \quad \mathbf{b}^T = (i_{\alpha\beta}, \mathbf{i}_{\mu\beta}^T).$$

Then we can show that

$$v_{\beta\beta} = 1/(i_{\beta\beta} - \mathbf{b}^T\mathbf{A}^{-1}\mathbf{b} - \mathbf{i}_{\sigma\beta}^T\mathbf{i}_{\sigma\sigma}^{-1}\mathbf{i}_{\sigma\beta}),$$

$$\text{and} \quad \mathbf{B} = \begin{pmatrix} v_{\alpha\alpha} & \mathbf{v}_{\alpha\mu}^T \\ \mathbf{v}_{\alpha\mu} & \mathbf{v}_{\mu\mu} \end{pmatrix} = \{\mathbf{A} - \mathbf{b}\mathbf{b}^T/(i_{\beta\beta} - \mathbf{i}_{\sigma\beta}^T\mathbf{i}_{\sigma\sigma}^{-1}\mathbf{i}_{\sigma\beta})\}^{-1}.$$

Hence, using the result that

$$(\underset{\sim}{A} + \underset{\sim}{u}\underset{\sim}{v}^T)^{-1} = \underset{\sim}{A}^{-1} - \underset{\sim}{A}^{-1}\underset{\sim}{u}\underset{\sim}{v}^T\underset{\sim}{A}^{-1}/(1 + \underset{\sim}{v}^T\underset{\sim}{A}^{-1}\underset{\sim}{u}) ,$$

where $\underset{\sim}{u}$ and $\underset{\sim}{v}$ are p-dimensional vectors and $\underset{\sim}{A}$ is p×p and non-singular (see, for example, Press, 1972,p23), we find that

$$\underset{\sim}{B} = \underset{\sim}{A}^{-1} + v_{\beta\beta}\underset{\sim}{A}^{-1}\underset{\sim}{b}\underset{\sim}{b}^T\underset{\sim}{A}^{-1}$$

and, similarly, that

$$\underset{\sim}{v}_{\sigma\sigma} = \underset{\sim}{i}_{\sigma\sigma}^{-1} + v_{\beta\beta}\underset{\sim}{i}_{\sigma\sigma}^{-1}\underset{\sim}{i}_{\sigma\beta}\underset{\sim}{i}_{\sigma\beta}^{T}\underset{\sim}{i}_{\sigma\sigma}^{-1} \quad .$$

It is not difficult to find the inverse of $\underset{\sim}{A}$,

$$\underset{\sim}{A}^{-1} = \begin{pmatrix} \underset{\sim}{c} & \underset{\sim}{e}^T \\ \underset{\sim}{e} & \underset{\sim}{F} \end{pmatrix} ,$$

say, because $\underset{\sim}{i}_{\mu\mu}$ is diagonal. In fact,

$$c = 1/(i_{\alpha\alpha} - \underset{\sim}{i}_{\alpha\mu}^T\underset{\sim}{i}_{\mu\mu}^{-1}\underset{\sim}{i}_{\alpha\mu}), \quad \underset{\sim}{e}^T = -c\underset{\sim}{i}_{\alpha\mu}^T\underset{\sim}{i}_{\mu\mu}^{-1}$$

$$\underset{\sim}{F} = (\underset{\sim}{i}_{\mu\mu} - \underset{\sim}{i}_{\alpha\mu}\underset{\sim}{i}_{\alpha\mu}^T/i_{\alpha\alpha})^{-1} = \underset{\sim}{i}_{\mu\mu}^{-1} + c\underset{\sim}{i}_{\mu\mu}^{-1}\underset{\sim}{i}_{\alpha\mu}\underset{\sim}{i}_{\alpha\mu}^T\underset{\sim}{i}_{\mu\mu}^{-1},$$

from which we get

$$c = (\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)/n \ , \quad \underset{\sim}{e}^T = -(\beta\sigma_\delta^2/n)\underset{\sim}{L}_k^T \ ,$$

where $\underset{\sim}{L}_k$ is a vector of k ones, and

$$\underset{\sim}{F} = \frac{|\underset{\sim}{\Sigma}|}{\beta^2\sigma_\delta^2+\sigma_\varepsilon^2}\,\mathrm{diag}(1/n_1,\ldots,1/n_k) + \frac{\beta^2\sigma_\delta^4}{n(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)}\underset{\sim}{L}_k\underset{\sim}{L}_k^T \ .$$

Hence we find

$$\underset{\sim}{A}^{-1}\underset{\sim}{b} = \left(\bar{\mu} \ , \ \{\beta\sigma_\delta^2/(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)\}(\mu_1-\bar{\mu},\ldots,\mu_k-\bar{\mu}) \right)$$

and so

$$\underset{\sim}{b}^T \underset{\sim}{A}^{-1} \underset{\sim}{b} = n\bar{\mu}^2/(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2) + \beta^2\sigma_\delta^4 \sum n_i\mu_i^2/\{|\underset{\sim}{\Sigma}|(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)\}$$

and

$$\underset{\sim}{A}^{-1}\underset{\sim}{b}\underset{\sim}{b}^T\underset{\sim}{A}^{-1} = \begin{pmatrix} \bar{\mu}^2 & \underset{\sim}{g}^T \\ \underset{\sim}{g} & \underset{\sim}{H} \end{pmatrix},$$

where

$$\underset{\sim}{g}^T = \{\beta\sigma_\delta^2\bar{\mu}/(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)\}(\underset{\sim}{\mu} - \bar{\mu}\underset{\sim}{L}_k)^T$$

and

$$\underset{\sim}{H} = \{\beta^2\sigma_\delta^4/(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)^2\}(\underset{\sim}{\mu} - \bar{\mu}\underset{\sim}{L}_k)(\underset{\sim}{\mu} - \bar{\mu}\underset{\sim}{L}_k)^T.$$

The 3×3 matrix $\underset{\sim}{i}_{\sigma\sigma}$ can easily be inverted to give

$$\underset{\sim}{i}_{\sigma\sigma}^{-1} = \frac{1}{n\beta^2}\begin{bmatrix} (|\underset{\sim}{\Sigma}|+2\beta^2\sigma^4) & -(|\underset{\sim}{\Sigma}|-2\beta^2\sigma^2\sigma_\delta^2) & -\beta^2(|\underset{\sim}{\Sigma}|-2\sigma^2\sigma_\varepsilon^2) \\ -(|\underset{\sim}{\Sigma}|-2\beta^2\sigma^2\sigma_\delta^2) & (|\underset{\sim}{\Sigma}|+2\beta^2\sigma_\delta^4) & -\beta^2(|\underset{\sim}{\Sigma}|-2\sigma_\delta^2\sigma_\varepsilon^2) \\ -\beta^2(|\underset{\sim}{\Sigma}|-2\sigma^2\sigma_\varepsilon^2) & -\beta^2(|\underset{\sim}{\Sigma}|-2\sigma_\delta^2\sigma_\varepsilon^2) & \beta^2(\beta^2|\underset{\sim}{\Sigma}|+2\sigma_\varepsilon^4) \end{bmatrix}.$$

Hence

$$\underset{\sim}{i}_{\sigma\beta}^T\underset{\sim}{i}_{\sigma\sigma}^{-1} = (\sigma^2/\beta)(1, -1, \beta^2),$$

$$\underset{\sim}{i}_{\sigma\beta}^T\underset{\sim}{i}_{\sigma\sigma}^{-1}\underset{\sim}{i}_{\sigma\beta} = n\sigma^4(|\underset{\sim}{\Sigma}| + 2\beta^2\sigma_\delta^4)/|\underset{\sim}{\Sigma}|^2$$

and

$$\underset{\sim}{i}_{\sigma\sigma}^{-1}\underset{\sim}{i}_{\sigma\beta}\underset{\sim}{i}_{\sigma\beta}^T\underset{\sim}{i}_{\sigma\sigma}^{-1} = \frac{\sigma^4}{\beta^2}\begin{pmatrix} 1 & -1 & \beta^2 \\ -1 & 1 & -\beta^2 \\ \beta^2 & -\beta^2 & \beta^4 \end{pmatrix}.$$

We are now in a position to use the formulae obtained earlier for evaluating $v_{\beta\beta}$, $v_{\alpha\alpha}$, $\underset{\sim}{v}_{\mu\mu}$ and $\underset{\sim}{v}_{\sigma\sigma}$. We get

$$v_{\beta\beta} = (\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)/(nb_{\mu\mu}), \quad v_{\alpha\alpha} = (\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)\sum n_i\mu_i^2/(n^2 b_{\mu\mu}),$$

where $b_{\mu\mu} = \sum n_i(\mu_i - \bar{\mu})^2/n$, the j'th diagonal element of $\underset{\sim}{v}_{\mu\mu}$ is

$|\underset{\sim}{\Sigma}|/\{n_j(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)\} + \beta^2\sigma_\delta^4\{b_{\mu\mu} + (\mu_j - \bar{\mu})^2\}/\{nb_{\mu\mu}(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)\}$,

and the diagonal elements of $\underset{\sim}{v}_{\sigma\sigma}$ are

$$(|\underset{\sim}{\Sigma}| + 2\beta^2\sigma^4)/(n\beta^2) + \sigma^4(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)/(n\beta^2 b_{\mu\mu}),$$

$$(|\underset{\sim}{\Sigma}| + 2\beta^2\sigma_\delta^4)/(n\beta^2) + \sigma^4(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)/(n\beta^2 b_{\mu\mu}),$$

and $$(\beta^2|\underset{\sim}{\Sigma}| + 2\sigma_\varepsilon^4)/n + \beta^2\sigma^4(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)/(nb_{\mu\mu}).$$

Let $\hat{\alpha}$ be the ML estimator of α, and so on. Then we have proved that

$$\text{var}(\hat{\alpha}) = (\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)\sum n_i\mu_i^2/(n^2 b_{\mu\mu}) + O(1/n^2),$$

$$\text{var}(\hat{\mu}_i) = \frac{|\underset{\sim}{\Sigma}|}{n_i(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)} + \frac{\beta^2\sigma_\delta^4\{b_{\mu\mu} + (\mu_i - \bar{\mu})^2\}}{nb_{\mu\mu}(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)} + O(1/n^2), \quad (i=1,..,k),$$

$$\text{var}(\hat{\sigma}^2) = \{\sigma^4(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2) + b_{\mu\mu}(|\underset{\sim}{\Sigma}|+2\beta^2\sigma^4)\}/(n\beta^2 b_{\mu\mu}) + O(1/n^2),$$

$$\text{var}(\hat{\sigma}_\delta^2) = \{\sigma^4(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2) + b_{\mu\mu}(|\underset{\sim}{\Sigma}|+2\beta^2\sigma_\delta^4)\}/(n\beta^2 b_{\mu\mu}) + O(1/n^2),$$

$$\text{var}(\hat{\sigma}_\varepsilon^2) = \{\beta^2\sigma^4(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2) + b_{\mu\mu}(\beta^2|\underset{\sim}{\Sigma}|+2\sigma_\varepsilon^4)\}/(nb_{\mu\mu}) + O(1/n^2),$$

and $$\text{var}(\hat{\beta}) = (\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)/(nb_{\mu\mu}) + O(1/n^2).$$

These expressions are the same as we got in the two-group model with the exception of $\text{var}(\hat{\mu}_i)$ $(i=1,..,k)$. However, the above expression for $\text{var}(\hat{\mu}_i)$ reduces to that given in §2.3 when k=2, for when k=2 we can show that

$$b_{\mu\mu} + (\mu_i - \bar{\mu})^2 = nb_{\mu\mu}/n_i \quad (i=1,2).$$

3.4 Tests of the adequacy of the simplest model

As in the two-group case, we shall be interested in testing whether the data are adequately described by the simplest model, model V. We shall define extensions of this model and then find tests of the hypothesis that model V is the true model.

First let us consider model VI which is the same as model V except that the intercepts, α_i, of the groups differ from each other. We wish to find a test of the null hypothesis that $\alpha_1 = \ldots = \alpha_k$ in model VI. Before we maximize the likelihood for this model we note that the model has 2k+4 parameters but that the dimension of the minimal sufficient statistic is only 2k+3. Hence, as in the two-group model, we expect that the likelihood will be maximized over a subspace of the parameter space rather than at just one point. In fact nothing essential is changed in going from the two-group case to the k-group case and so the log-likelihood for model VI attains its maximum,

$$\hat{\ell} = -n\{1 + \log(2\pi)\} - \tfrac{1}{2}n \log(s_{yy}s_{xx} - s_{yx}^2),$$

whenever $\mu_i = \bar{x}_{i.}$ $(i=1,..,k)$, $\alpha_i = \bar{y}_{i.} - \beta\bar{x}_{i.}$ $(i=1,..,k)$, $\sigma^2 = s_{yx}/\beta$, $\sigma_\delta^2 = s_{xx} - s_{yx}/\beta$, $\sigma_\varepsilon^2 = s_{yy} - \beta s_{yx}$ and β lies between s_{yx}/s_{xx} and s_{yy}/s_{yx}.

Let $\tilde{\ell}$ be the maximum of the log-likelihood in model V. Then

$$\tilde{\ell} = -n\{1 + \log(2\pi)\} - \tfrac{1}{2}n \log\{(s_{yy}s_{xx} - s_{yx}^2)T(\tilde{\beta})/W(\tilde{\beta})\},$$

where $W(\tilde{\beta}) = s_{yy} - 2\tilde{\beta}s_{yx} + \tilde{\beta}^2 s_{xx}$, $T(\tilde{\beta}) = t_{yy} - 2\tilde{\beta}t_{yx} + \tilde{\beta}^2 t_{xx}$,

$$\tilde{\beta} = \frac{s_{yy}b_{xx} - s_{xx}b_{yy} - d(y,x)}{2(s_{yx}b_{xx} - s_{xx}b_{yx})},$$

and $d(y,x) =$

$$\{(s_{xx}b_{yy} - s_{yy}b_{xx})^2 - 4(s_{yx}b_{xx} - s_{xx}b_{yx})(s_{yy}b_{yx} - s_{yx}b_{yy})\}^{\frac{1}{2}},$$

provided $\tilde{\beta}$ lies between s_{yx}/s_{xx} and s_{yy}/s_{yx}, and otherwise $\tilde{\ell}$ is the maximum of

$$\ell_\delta = -n\{1 + \log(2\pi)\} - \tfrac{1}{2}n \log\{s_{xx}(t_{yy} - t_{yx}^2/t_{xx})\}$$

and $\ell_\varepsilon = -n\{1 + \log(2\pi)\} - \tfrac{1}{2}n \log\{s_{yy}(t_{xx} - t_{yx}^2/t_{yy})\}.$

Note that this is actually only an approximation because we have ignored the maximum on the $(\sigma^2=0)$-boundary which we saw in §3.2 could not be found explicitly.

We would like to use $2(\hat{\ell} - \tilde{\ell})$ as a test statistic but does it have an asymptotic χ^2 distribution? Certainly the regularity conditions (see, for example, Cox and Hinkley, 1974, p281) do not hold, for model VI, because the nonidentifiability means that the probability distributions defined by two different values of the vector of parameters are not necessarily distinct. However, we can still show that $2(\hat{\ell} - \tilde{\ell})$ is asymptotically distributed as χ^2_{k-2}.

Consider placing a restriction on the parameters in model VI in such a way that the new model, model VII, is identifiable. We also choose the constraint to be consistent with the null hypothesis. For instance, we could choose any constraint of the form $\sum g_i\alpha_i = 0$ where $\sum g_i = 0$. In the usual way we can show that the log-likelihood for model VII attains a maximum of $\hat{\ell}$ when $\mu_i = \bar{x}_{i.}$ $(i=1,..,k)$, $\alpha_i = \bar{y}_{i.} - \beta\bar{x}_{i.}$ $(i=1,..,k)$, $\sigma^2 = s_{yx}/\beta$, $\sigma_\delta^2 = s_{xx} - s_{yx}/\beta$, $\sigma_\varepsilon^2 = s_{yy} - \beta s_{yx}$ and $\beta = \beta_g = \sum g_i\bar{y}_{i.}/\sum g_i\bar{x}_{i.}$ when β_g lies between s_{yx}/s_{xx} and s_{yy}/s_{yx}. Otherwise, the log-likelihood attains its maximum on one of the three boundaries, $\sigma^2 = 0$, $\sigma_\delta^2 = 0$ or $\sigma_\varepsilon^2 = 0$, and the maximum attained will be less than $\hat{\ell}$. Let ℓ_g be the maximum of the log-likelihood for model VII. We showed that $\ell_g = \hat{\ell}$ when β_g lies between s_{yx}/s_{xx} and s_{yy}/s_{yx}. But since $g_1,..,g_k$ were chosen to make the constraint consistent with the null hypothesis then, when the null hypothesis is true, the probability that β_g lies between s_{yx}/s_{xx} and s_{yy}/s_{yx} tends to one as $n \to \infty$. Hence, when the null hypothesis is true,

$$\Pr(\ell_g = \hat{\ell}) \to 1 \quad \text{as} \quad n \to \infty.$$

Model VII satisfies the regularity conditions and so $2(\ell_g - \tilde{\ell})$ is asymptotically distributed as χ^2_{k-2} when the null hypothesis is true. Hence $2(\hat{\ell} - \tilde{\ell})$ is also asymptotically distributed as χ^2_{k-2} when the null hypothesis is true. The result does not depend on any particular choice of $g_1,..,g_k$ but, in the event of a rejection of the null hypothesis, the possible choices of $g_1,..,g_k$ which satisfy the condition that β_g lies between s_{yx}/s_{xx} and s_{yy}/s_{yx} give information on what possible extensions of model V might adequately describe the data.

Our test of the adequacy of model V against the alternative that allows the intercepts of the groups to differ is thus to reject model V if $2(\hat{\ell} - \tilde{\ell})$ is greater than the selected significance point of the χ^2_{k-2} distribution. This test is clearly of no use when $k = 2$ but in §2.4 we described a test which can be used in this special case.

Now let us consider extensions of model V in which the covariance matrices differ from group to group. We shall study the same extensions that we considered in the two-group case, first letting the matrices be proportional and then letting them be completely heterogeneous.

Define model VIII to be the same as the simplest k-group model except that the covariance matrix of (Y_{ij}, X_{ij}) is $\lambda_i \underset{\sim}{\Sigma}$ instead of just $\underset{\sim}{\Sigma}$. We shall need to impose some constraint on $\lambda_1,..,\lambda_k$, for instance $\sum \lambda_i = k$ or the restriction used in the two-group case, $\lambda_1 = 1$. And, of course, we require $\lambda_i > 0$ $(i=1,..,k)$. The log-likelihood of the observations is

$$\ell = -n \log(2\pi) - \tfrac{1}{2}n \log|\underset{\sim}{\Sigma}| - \sum_i n_i \log\lambda_i - \frac{1}{2|\underset{\sim}{\Sigma}|}\sum_i \frac{1}{\lambda_i}\{(\sigma^2+\sigma_\delta^2)S_{1i}$$

$$- 2\beta\sigma^2 S_{2i} + (\beta^2\sigma^2+\sigma_\varepsilon^2)S_{3i}\},$$

where $$S_{1i} = \sum_j (y_{ij} - \alpha - \beta\mu_i)^2,$$

$$S_{2i} = \sum_j (y_{ij} - \alpha - \beta\mu_i)(x_{ij} - \mu_i)$$

and $$S_{3i} = \sum_j (x_{ij} - \mu_i)^2 \quad (i=1,..,k).$$

In maximizing the log-likelihood we proceed much as in the two-group case. The likelihood equations for $\mu_1,..,\mu_k$ yield the solutions

$$\mu_i = \{\beta\sigma_\delta^2(\bar{y}_{i.} - \alpha) + \sigma_\varepsilon^2\bar{x}_{i.}\}/(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2) \quad (i=1,..,k),$$

and the corresponding pivots of the matrix of double derivatives of the log-likelihood are all negative. The likelihood equation for α yields, on substitution of the above expressions for the μ_i,

$$\alpha = \bar{y}_{..}(\lambda) - \beta\bar{x}_{..}(\lambda)$$

where $\bar{y}_{..}(\lambda) = (\sum n_i\bar{y}_{i.}/\lambda_i)/n$ and $\bar{x}_{..}(\lambda) = (\sum n_i\bar{x}_{i.}/\lambda_i)/n$,

and the corresponding pivot is negative. When μ_i and α equal the above expressions,

$$S_1(\lambda) = \sum S_{1i}/\lambda_i = ns_{yy}(\lambda) + n\sigma_\varepsilon^4 B(\beta,\lambda)/(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)^2,$$

$$S_2(\lambda) = \sum S_{2i}/\lambda_i = ns_{yx}(\lambda) - n\beta\sigma_\delta^2\sigma_\varepsilon^2 B(\beta,\lambda)/(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)^2,$$

and $S_3(\lambda) = \sum S_{3i}/\lambda_i = ns_{xx}(\lambda) + n\beta^2\sigma_\delta^4 B(\beta,\lambda)/(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)^2$,

where $B(\beta,\lambda) = b_{yy}(\lambda) - 2\beta b_{yx}(\lambda) + \beta^2 b_{xx}(\lambda)$,

$$ns_{yy}(\lambda) = \sum\sum(y_{ij} - \bar{y}_{i.})^2/\lambda_i,$$

$$nb_{yy}(\lambda) = \sum n_i\{\bar{y}_{i.} - \bar{y}_{..}(\lambda)\}^2/\lambda_i,$$

and so on.

Hence, after maximization with respect to $\mu_1,\ldots,\mu_k$ and α, the log-likelihood becomes

$$\ell^* = -n\log(2\pi) - \sum n_i\log\lambda_i - \tfrac{1}{2}n\log|\underset{\sim}{\Sigma}| - \frac{1}{2|\underset{\sim}{\Sigma}|}\{(\sigma^2+\sigma_\delta^2)S_1(\lambda) - 2\beta\sigma^2 S_2(\lambda) + (\beta^2\sigma^2+\sigma_\varepsilon^2)S_3(\lambda)\},$$

which is the same as in the simplest k-group model except for the term $\sum n_i\log\lambda_i$ and with $s_{yy}(\lambda)$ replacing s_{yy}, $b_{yy}(\lambda)$ replacing b_{yy}, and so on. Hence maximization with respect to σ^2, σ_δ^2, σ_ε^2 and β becomes the same as in the simplest model and we get

$$\sigma^2 = \frac{s_{yx}(\lambda)}{\beta} - \frac{\{\beta s_{xx}(\lambda) - s_{yx}(\lambda)\}\{s_{yy}(\lambda) - \beta s_{yx}(\lambda)\}B(\beta,\lambda)}{\beta W^2(\beta,\lambda)},$$

$$\sigma_\delta^2 = \frac{\{\beta s_{xx}(\lambda) - s_{yx}(\lambda)\}T(\beta,\lambda)}{\beta W(\beta,\lambda)},$$

$$\sigma_\varepsilon^2 = \frac{\{s_{yy}(\lambda) - \beta s_{yx}(\lambda)\}T(\beta,\lambda)}{W(\beta,\lambda)},$$

and $$\beta = \tilde{\beta}(\lambda),$$

where $W(\beta,\lambda) = s_{yy}(\lambda) - 2\beta s_{yx}(\lambda) + \beta^2 s_{xx}(\lambda)$,

$$T(\beta,\lambda) = W(\beta,\lambda) + B(\beta,\lambda),$$

and $\tilde{\beta}(\lambda)$ is the same as the $\tilde{\beta}$ defined earlier in this section (ie, the ML estimate of β in the simplest k-group model) but with s_{yy} replaced by $s_{yy}(\lambda)$, b_{yy} replaced by $b_{yy}(\lambda)$ and so on.

The maximum relative likelihood of $\lambda_1,..,\lambda_k$ is thus

$$\hat{\ell}(\lambda) = -n\{1+\log(2\pi)\} - \sum n_i \log\lambda_i$$

$$- \tfrac{1}{2}n \log\left(\frac{\{s_{yy}(\lambda)s_{xx}(\lambda) - s_{yx}^2(\lambda)\}T(\tilde{\beta}(\lambda),\lambda)}{W(\tilde{\beta}(\lambda),\lambda)} \right) .$$

If we choose the constraint on $\lambda_1,..,\lambda_k$ to be $\sum\lambda_i = k$, then we need to maximize $\hat{\ell}(\lambda) - \theta(\sum\lambda_i - k)$ with respect to $\lambda_1,..,\lambda_k$ and θ. This is too complicated to be done analytically and so the maximization must be done numerically. Let $\hat{\lambda}_1,..,\hat{\lambda}_k$ be the values of $\lambda_1,..,\lambda_k$ which maximize $\hat{\ell}(\lambda)$ subject to the constraint $\sum\lambda_i = k$. Then for testing the null hypothesis that $\lambda_1 = ... = \lambda_k = 1$ we use the statistic $2\{\hat{\ell}(\hat{\lambda}) - \tilde{\ell}\}$, where $\tilde{\ell}$, the maximum of the log-likelihood for the simplest k-group model, is defined earlier in this section. This statistic is asymptotically distributed as χ_{k-1}^2 when the null hypothesis is true.

We now turn to the other extension of the simplest model in which the variances in each group are unrelated to the variances in the other groups. As in the two-group case, this model includes the previous one as a special case and so the test derived below could be used in place of the previous one. However, the previous test will be more sensitive in detecting the presence of proportional covariance matrices when they exist and the simpler model provides a more economical description of the data.

The log-likelihood for the new model, model IX, is

$$\ell = -n\log(2\pi) - \tfrac{1}{2}\sum_{i=1}^{k} n_i \log|\underset{\sim}{\Sigma}_i| - \tfrac{1}{2}\sum_{i=1}^{k}\{(\sigma_i^2+\sigma_{\delta i}^2)S_{1i} - 2\beta\sigma_i^2 S_{2i} + (\beta^2\sigma_i^2+\sigma_{\varepsilon i}^2)S_{3i}\}/|\underset{\sim}{\Sigma}_i|,$$

where $|\underset{\sim}{\Sigma}_i| = \beta^2\sigma_i^2\sigma_{\delta i}^2 + \sigma_i^2\sigma_{\varepsilon i}^2 + \sigma_{\delta i}^2\sigma_{\varepsilon i}^2 \quad (i=1,..,k)$

and S_{1i}, S_{2i} and S_{3i} are as defined for model VIII. The solution of the likelihood equations is much the same as before but it is best to leave the equation for α until the end. The likelihood equations for $\mu_1,..,\mu_k$ yield

$$\mu_i = \{\beta\sigma_{\delta i}^2(\bar{y}_{i.} - \alpha) + \sigma_{\varepsilon i}^2\bar{x}_{i.}\}/(\beta^2\sigma_{\delta i}^2+\sigma_{\varepsilon i}^2) \quad (i=1,..,k),$$

and the corresponding pivots of the matrix of double derivatives of the log-likelihood are all negative. If we substitute these estimates of $\mu_1,..,\mu_k$ back into the log-likelihood we get $\ell^* = \sum \ell_i^*$, where

$$\ell_i^* = -n_i\log(2\pi) - \tfrac{1}{2}n_i\log|\underset{\sim}{\Sigma}_i| - \tfrac{1}{2}n_i\{(\sigma_i^2+\sigma_{\delta i}^2)s_{yyi} - 2\beta\sigma_i^2 s_{yxi} + (\beta^2\sigma_i^2+\sigma_{\varepsilon i}^2)s_{xxi}\}/|\underset{\sim}{\Sigma}_i| - \tfrac{1}{2}n_i B_i(\beta,\alpha)/(\beta^2\sigma_{\delta i}^2+\sigma_{\varepsilon i}^2),$$

$B_i(\beta,\alpha) = (\bar{y}_{i.} - \alpha - \beta\bar{x}_{i.})^2$ and $n_i s_{yyi} = \sum(y_{ij} - \bar{y}_{i.})^2$, etc.

Each ℓ_i^* is the same as the log-likelihood for the simplest k-group model (model V) maximized over $\mu_1,..,\mu_k$ and α but with $B_i(\beta,\alpha)$ replacing $B(\beta)$, n_i replacing n, σ_i^2 replacing σ^2, and so on. As ℓ_i^* does not contain σ_j^2, $\sigma_{\delta j}^2$ or $\sigma_{\varepsilon j}^2$ for $j \neq i$, each ℓ_i^* can be maximized separately with respect to σ_i^2, $\sigma_{\delta i}^2$ and $\sigma_{\varepsilon i}^2$ and so we get (as in model V)

$$\sigma_i^2 = s_{yxi}/\beta - (\beta s_{xxi}-s_{yxi})(s_{yyi}-\beta s_{yxi})B_i(\beta,\alpha)/\{\beta W_i^2(\beta)\},$$

$$\sigma_{\delta i}^2 = (\beta s_{xxi} - s_{yxi})T_i(\beta,\alpha)/\{\beta W_i(\beta)\},$$

and $$\sigma_{\varepsilon i}^2 = (s_{yyi} - \beta s_{yxi})T_i(\beta,\alpha)/W_i(\beta),$$

where $$W_i(\beta) = s_{yyi} - 2\beta s_{yxi} + \beta^2 s_{xxi}$$

and $$T_i(\beta,\alpha) = W_i(\beta) + B_i(\beta,\alpha) \quad (i=1,..,k).$$

On substitution of these expressions into ℓ^*, we see that the maximum relative likelihood of α and β is

$$\hat{\ell}(\alpha,\beta) = -n\{1 + \log(2\pi)\} - \tfrac{1}{2}\sum n_i \log|\underset{\sim}{\Sigma}_i^*|,$$

where $$|\underset{\sim}{\Sigma}_i^*| = (s_{yyi}s_{xxi} - s_{yxi}^2)T_i(\beta,\alpha)/W_i(\beta) \quad (i=1,..,k).$$

We can easily show that

$$\partial\hat{\ell}(\alpha,\beta)/\partial\alpha = \sum n_i(\bar{y}_{i.} - \alpha - \beta\bar{x}_{i.})/T_i(\beta,\alpha)$$

and $\partial\hat{\ell}(\alpha,\beta)/\partial\beta =$

$$\sum_i \frac{n_i(\bar{y}_{i.}-\alpha-\beta\bar{x}_{i.})\{(\beta s_{xxi}-s_{yxi})(\bar{y}_{i.}-\alpha) + (s_{yyi}-\beta s_{yxi})\bar{x}_{i.}\}}{T_i(\beta,\alpha)W_i(\beta)}.$$

No analytical solution has been found to these two likelihood equations so that $\hat{\ell}(\alpha,\beta)$ must be maximized numerically. When $k=2$, however, we can solve the equations by choosing α and β to satisfy $\bar{y}_{i.} - \alpha - \beta\bar{x}_{i.} = 0$ $(i=1,2)$, and we saw in §2.4 that this solution maximizes the likelihood. Let $\hat{\alpha}$ and $\hat{\beta}$ be the values of α and β that maximize $\hat{\ell}(\alpha,\beta)$. Then for testing the null hypothesis that $\sigma_1^2=\ldots=\sigma_k^2$, $\sigma_{\delta 1}^2=\ldots=\sigma_{\delta k}^2$ and $\sigma_{\varepsilon 1}^2=\ldots=\sigma_{\varepsilon k}^2$ we use the statistic

$$2\{\hat{\ell}(\hat{\alpha},\hat{\beta}) - \tilde{\ell}\},$$

where $\tilde{\ell}$ is the maximum of the log-likelihood for model V. When the null hypothesis is true this statistic has an asymptotic $\chi^2_{3(k-1)}$ distribution.

Note that in the last two models the maximum of the likelihood may lie outside the parameter space which means that we found only the internal estimates of the parameters. However, this does not affect the asymptotic results.

If we have rejected the simplest k-group model in favour of one of the extended models we may then be interested in knowing whether this extended model adequately describes the data. For this purpose we now consider an even more general model in which both the intercepts and the variances vary from group to group. That is, the model has the same specifications as the simplest model but with α replaced by α_i, σ^2 replaced by σ_i^2, σ_δ^2 replaced by $\sigma_{\delta i}^2$ and σ_ε^2 replaced by $\sigma_{\varepsilon i}^2$.

In much the same way as in the previous models discussed in this section, we can show that the log-likelihood achieves its maximum,

$$\ell^\dagger = -n\{1+\log(2\pi)\} - \tfrac{1}{2}\sum n_i \log(s_{yyi}s_{xxi} - s_{yxi}^2),$$

whenever $\mu_i = \bar{x}_{i.}$, $\alpha_i = \bar{y}_{i.} - \beta\bar{x}_{i.}$, $\sigma_i^2 = s_{yxi}/\beta$, $\sigma_{\delta i}^2 = s_{xxi} - s_{yxi}/\beta$, $\sigma_{\varepsilon i}^2 = s_{yyi} - \beta s_{yxi}$ $(i=1,..,k)$ and β lies between s_{yxi}/s_{xxi} and s_{yyi}/s_{yxi} for all i.

The nonidentifiability problem is the same as we encountered in model VI and so we can still derive asymptotic χ^2 results. For testing the null hypothesis that $\alpha_1 = \ldots = \alpha_k$ we use the statistic

$$2\{\ell^\dagger - \hat{\ell}(\hat{\alpha},\hat{\beta})\},$$

which has an asymptotic χ_{k-2}^2 distribution when the null hypothesis is true. For testing the null hypothesis that $\sigma_1^2 = \ldots = \sigma_k^2$, $\sigma_{\delta 1}^2 = \ldots = \sigma_{\delta k}^2$ and $\sigma_{\varepsilon 1}^2 = \ldots = \sigma_{\varepsilon k}^2$ we use the statistic

$$2(\ell^\dagger - \hat{\ell}),$$

which has an asymptotic $\chi_{3(k-1)}^2$ distribution when the null hypothesis is true.

3.5 Tests of hypotheses about the slope

We now consider tests of the hypothesis that $\beta = \beta_o$ in the simplest k-group model. Most of the work done on the two-group model applies equally to the k-group model with little or no modification. As in §2.5 we first find a test statistic based on the ML estimate of β.

The asymptotic properties of the ML estimator of β are just those of the internal estimator,

$$\tilde{\beta} = \frac{s_{yy}b_{xx} - s_{xx}b_{yy} - d(y,x)}{2(s_{yx}b_{xx} - s_{xx}b_{yx})},$$

where $d(y,x)$ =

$$\{(s_{xx}b_{yy} - s_{yy}b_{xx})^2 - 4(s_{yx}b_{xx} - s_{xx}b_{yx})(s_{yy}b_{yx} - s_{yx}b_{yy})\}^{\frac{1}{2}}.$$

Subtracting from $\tilde{\beta}$ its asymptotic expectation, β_o, and dividing by its asymptotic standard deviation,

$$\{(\beta_o^2\sigma_\delta^2+\sigma_\varepsilon^2)/(nb_{\mu\mu})\}^{\frac{1}{2}} \quad \text{(see Appendix 3)},$$

we get

$$U^* = (\tilde{\beta} - \beta_o)\left\{\frac{nb_{\mu\mu}}{\beta_o^2\sigma_\delta^2+\sigma_\varepsilon^2}\right\}^{\frac{1}{2}},$$

as in §2.5. We wish to replace the unknown parameters in U^* by their ML estimators when $\beta = \beta_o$ is known. The expressions for these estimators are exactly the same as in the two-group case since their derivation is unaffected by the change from 2 to k groups. Hence we get the statistic

$$U(\beta_o) = \frac{n^{\frac{1}{2}}(\tilde{\beta} - \beta_o)}{W(\beta_o)T^{\frac{1}{2}}(\beta_o)} \{ (\beta_o s_{XX} - s_{YX})^2 b_{YY} +$$

$$2(\beta_o s_{XX} - s_{YX})(s_{YY} - \beta_o s_{YX}) b_{YX} + (s_{YY} - \beta_o s_{YX})^2 b_{XX} \}^{\frac{1}{2}},$$

which reduces to the statistic in §2.5 when $b_{YY} b_{XX} = b_{YX}^2$. We use $U(\beta_o)$ as an approximately $N(0,1)$ distributed statistic for testing the hypothesis that $\beta = \beta_o \neq 0$.

The criticisms given in §2.5 of the use of this test alone apply equally here. Furthermore, the construction of the test that overcame these criticisms depended in no significant way on there being only two groups. In fact the only changes we need to make to the test in §2.5 are to change $U(\beta_o)$ to the expression given above and to alter the term (n-4) to (n-k-2). The latter term comes from the expression in Appendix 3 for the variance of s_{YX}/s_{XX}. Our test of the null hypothesis that $\beta = \beta_o \neq 0$ against the alternative that $\beta \neq \beta_o$ is thus as follows. If β_o lies between s_{yx}/s_{xx} and s_{yy}/s_{yx} we reject the null hypothesis if

$$|u(\beta_o)| > k_{\frac{1}{2}\gamma} \, ,$$

where $k_{\frac{1}{2}\gamma}$ is the upper $\frac{1}{2}\gamma$ point of the $N(0,1)$ distribution. If β_o lies between s_{yx}/s_{xx} and $-\mathrm{sign}(s_{yx})(s_{yy}/s_{xx})^{\frac{1}{2}}$ we reject the null hypothesis if

$$\frac{\Phi(e) - \Phi(|w|, e; \rho) + \Phi(-|w|, e; \rho)}{\Phi(e)} < \gamma,$$

where $\Phi(.,.;\rho)$ is the distribution function of the bivariate normal distribution with means zero, variances one and correlation ρ, $\Phi(.)$ is the distribution function of the standard univariate normal distribution,

$$e = \min\{0\,,\ \mathrm{sign}(s_{yx})\beta_o\}\{(n-k-2)s_{xx}/T(\beta_o)\}^{\frac{1}{2}},$$

$$w = (t_{yx}/t_{xx} - \beta_o)\{nt_{xx}/T(\beta_o)\}^{\frac{1}{2}},$$

$$\rho = (s_{xx}/t_{xx})^{\frac{1}{2}},$$

and $$T(\beta_o) = t_{yy} - 2\beta_o t_{yx} + \beta_o^2 t_{xx}.$$

If $1/\beta_o$ lies between s_{yx}/s_{yy} and $-\mathrm{sign}(s_{yx})(s_{xx}/s_{yy})^{\frac{1}{2}}$ we reject the null hypothesis if

$$\frac{\Phi(e^*) - \Phi(|w^*|\,,e^*\,;\rho^*) + \Phi(-|w^*|\,,e^*\,;\rho^*)}{\Phi(e^*)} < \gamma,$$

where $$e^* = \min\{0\,,\ \mathrm{sign}(s_{yx})1/\beta_o\}\{(n-k-2)\beta_o^2 s_{yy}/T(\beta_o)\}^{\frac{1}{2}},$$

$$w^* = (t_{yx}/t_{yy} - 1/\beta_o)\{n\beta_o^2 t_{yy}/T(\beta_o)\}^{\frac{1}{2}},$$

and $$\rho^* = (s_{yy}/t_{yy})^{\frac{1}{2}}.$$

The test based on $u(\beta_o)$ has size approximately equal to γ while the other two tests have size approximately equal to or less than γ.

For testing the null hypothesis that $\beta = 0$ we can use the same statistic as in the two-group case, namely

$$Q = (n-2)^{\frac{1}{2}}R/(1-R^2)^{\frac{1}{2}},$$

where $R = t_{YX}/(t_{YY}t_{XX})^{\frac{1}{2}}$. This has a Student t distribution with n-2 degrees of freedom when $\beta = 0$.

Chapter 4

MULTIVARIATE GENERALIZATION

In this chapter we outline an extension of the model of previous chapters to allow for the inclusion of several independent variables $\underset{\sim}{X}^T = (X_1,..,X_p)$. We assume that we have observations $\{(Y_{ij},\underset{\sim}{X}_{ij}^T);j=1,..,n_i , i=1,..,k\}$ which come from the model

$$Y_{ij} = \alpha + \underset{\sim}{\beta}^T\underset{\sim}{U}_{ij} + \varepsilon_{ij},$$

$$\underset{\sim}{X}_{ij} = \underset{\sim}{U}_{ij} + \underset{\sim}{\delta}_{ij},$$

where the $\underset{\sim}{U}_{ij}$ have independent multivariate normal distributions with means $\underset{\sim}{\mu}_i$ and common covariance matrix $\underset{\sim}{\tau}$, the ε_{ij} are independently normally distributed with means O and variances σ_ε^2, the $\underset{\sim}{\delta}_{ij}$ are independently multivariate normally distributed with means $\underset{\sim}{O}$ and diagonal covariance matrices $\underset{\sim}{\tau}_\delta = \text{diag}(\tau_{\delta 1},..,\tau_{\delta p})$, and the errors, ε_{ij} and $\underset{\sim}{\delta}_{ij}$, are independent of each other and of the $\underset{\sim}{U}_{ij}$.

If we let $\underset{\sim}{Z}_{ij}^T = (Y_{ij},\underset{\sim}{X}_{ij}^T)$, then we can easily show that the $\underset{\sim}{Z}_{ij}$ have independent normal distributions with means

$$\underset{\sim}{\theta}_i = \begin{pmatrix} \alpha + \underset{\sim}{\beta}^T\underset{\sim}{\mu}_i \\ \underset{\sim}{\mu}_i \end{pmatrix}$$

and covariance matrix

$$\underset{\sim}{\Sigma} = \begin{pmatrix} \underset{\sim}{\beta}^T \underset{\sim}{\tau}\underset{\sim}{\beta} + \sigma_\varepsilon^2 & \underset{\sim}{\beta}^T \underset{\sim}{\tau} \\ \underset{\sim}{\tau}\underset{\sim}{\beta} & \underset{\sim}{\tau} + \underset{\sim}{\tau}_\delta \end{pmatrix}.$$

Note that we could not expect to be able to estimate more unknowns in $\underset{\sim}{\Sigma}$ than the number of distinct elements $(p+1)(p+2)/2$, of the sample covariance matrix. Apart from $\underset{\sim}{\beta}$, which also appears in the mean of the distribution, there are $p(p+1)/2$ distinct unknowns in the symmetric matrix $\underset{\sim}{\tau}$, p unknowns in $\underset{\sim}{\tau}_\delta$, and with the other unknown, σ_ε^2, this makes exactly $(p+1)(p+2)/2$ unknowns. Hence we could not have allowed $\underset{\sim}{\tau}_\delta$ to be a general symmetric positive semidefinite matrix and still have been able to estimate all the parameters. Of course, in some circumstances it may be more appropriate to restrict the form of $\underset{\sim}{\tau}$ and reduce the assumptions about the form of $\underset{\sim}{\tau}_\delta$. We must also have

$$k \geq p + 1$$

in order to be able to estimate all the parameters in the means. The number of sample means is $k(p+1)$ and the number of unknowns, including $\underset{\sim}{\beta}$, is $pk + p + 1$.

The log-likelihood of the observations is

$$\ell = -\tfrac{1}{2}n(p+1)\log(2\pi) - \tfrac{1}{2}n \log|\underset{\sim}{\Sigma}| - \tfrac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(\underset{\sim}{z}_{ij} - \underset{\sim}{\theta}_i)^T \underset{\sim}{\Sigma}^{-1}(\underset{\sim}{z}_{ij} - \underset{\sim}{\theta}_i),$$

where $n = \sum n_i$, as before, and we assume that $\underset{\sim}{\tau}$ is positive semidefinite, $\tau_{\delta 1},\ldots,\tau_{\delta p}$ and σ_ε^2 are all non-negative and $\underset{\sim}{\Sigma}$ is positive definite. We now find a solution of the likelihood equations.

Since $\underset{\sim}{\mu}_i$ appears only in $\underset{\sim}{\theta}_i$, we first consider

$$\partial \ell / \partial \underset{\sim}{\theta}_i = n_i \underset{\sim}{\Sigma}^{-1}(\bar{\underset{\sim}{z}}_{i\cdot} - \underset{\sim}{\theta}_i),$$

where $\bar{\underset{\sim}{z}}_{i\cdot} = \sum \underset{\sim}{z}_{ij}/n_i$. Since $\underset{\sim}{\Sigma}$ is positive definite then it is also nonsingular and we can show that its inverse may be written as

$$\underset{\sim}{\Sigma}^{-1} = \begin{pmatrix} 0 & \underset{\sim}{0}^T \\ \underset{\sim}{0} & (\underset{\sim}{\tau}+\underset{\sim}{\tau}_\delta)^{-1} \end{pmatrix} + \frac{1}{\gamma}\underset{\sim}{\phi}\underset{\sim}{\phi}^T,$$

where $$\gamma = \underset{\sim}{\beta}^T \underset{\sim}{\tau}(\underset{\sim}{\tau}+\underset{\sim}{\tau}_\delta)^{-1}\underset{\sim}{\tau}_\delta \underset{\sim}{\beta} + \sigma_\varepsilon^2 = \frac{|\underset{\sim}{\Sigma}|}{|\underset{\sim}{\tau}+\underset{\sim}{\tau}_\delta|} > 0$$

and $$\underset{\sim}{\phi}^T = \left(1 , -\underset{\sim}{\beta}^T \underset{\sim}{\tau}(\underset{\sim}{\tau}+\underset{\sim}{\tau}_\delta)^{-1}\right).$$

Note that the positive definiteness, and hence non-singularity, of $\underset{\sim}{\tau}+\underset{\sim}{\tau}_\delta$ follows from that of $\underset{\sim}{\Sigma}$. It is easily seen that

$$\left(\partial \underset{\sim}{\theta}_i / \partial \underset{\sim}{\mu}_i\right)^T = \left(\underset{\sim}{\beta} , \underset{\sim}{I}_p\right),$$

where $\underset{\sim}{I}_p$ is the p×p identity matrix. Hence

$$\frac{\partial \ell}{\partial \underset{\sim}{\mu}_i} = \left(\frac{\partial \underset{\sim}{\theta}_i}{\partial \underset{\sim}{\mu}_i}\right)^T \frac{\partial \ell}{\partial \underset{\sim}{\theta}_i} = n_i \left(\underset{\sim}{\beta} , \underset{\sim}{I}_p\right)\underset{\sim}{\Sigma}^{-1}(\bar{\underset{\sim}{z}}_{i\cdot} - \underset{\sim}{\theta}_i) \quad (i=1,..,k),$$

from which we find

$$\frac{\partial \ell}{\partial \underset{\sim}{\mu}_i} = \frac{n_i(\underset{\sim}{\tau}+\underset{\sim}{\tau}_\delta)^{-1}}{\gamma}\{\underset{\sim}{\tau}_\delta\underset{\sim}{\beta}(\bar{y}_{i.}-\alpha) + [(\underset{\sim}{\tau}+\underset{\sim}{\tau}_\delta)\gamma - \underset{\sim}{\tau}_\delta\underset{\sim}{\beta}\underset{\sim}{\beta}^T\underset{\sim}{\tau}](\underset{\sim}{\tau}+\underset{\sim}{\tau}_\delta)^{-1}\bar{\underset{\sim}{x}}_{i.}$$

$$- \underset{\sim}{\Delta}(\underset{\sim}{\tau}+\underset{\sim}{\tau}_\delta)^{-1}\underset{\sim}{\mu}_i\} \quad (i=1,..,k),$$

where $\underset{\sim}{\Delta} = (\underset{\sim}{\tau}+\underset{\sim}{\tau}_\delta)\gamma + \underset{\sim}{\tau}_\delta\underset{\sim}{\beta}\underset{\sim}{\beta}^T\underset{\sim}{\tau}_\delta$. We can easily show that $\underset{\sim}{\Delta}$ is positive definite and hence that

$$\underset{\sim}{\mu}_i = (\underset{\sim}{\tau}+\underset{\sim}{\tau}_\delta)\underset{\sim}{\Delta}^{-1}\{\underset{\sim}{\tau}_\delta\underset{\sim}{\beta}(\bar{y}_{i.}-\alpha) + [(\underset{\sim}{\tau}+\underset{\sim}{\tau}_\delta)\gamma - \underset{\sim}{\tau}_\delta\underset{\sim}{\beta}\underset{\sim}{\beta}^T\underset{\sim}{\tau}](\underset{\sim}{\tau}+\underset{\sim}{\tau}_\delta)^{-1}\bar{\underset{\sim}{x}}_{i.}\} \quad (i=1,..,k)$$

and $\partial^2\ell/\partial\underset{\sim}{\mu}_i^2 = -(n_i/\gamma)(\underset{\sim}{\tau}+\underset{\sim}{\tau}_\delta)^{-1}\underset{\sim}{\Delta}(\underset{\sim}{\tau}+\underset{\sim}{\tau}_\delta)^{-1} \quad (i=1,..,k),$

which is clearly negative definite. Hence the above solution for the $\underset{\sim}{\mu}_i$ in terms of the other parameters is at a maximum (with respect to the $\underset{\sim}{\mu}_i$) of the likelihood. We can expand $\underset{\sim}{\Delta}^{-1}$ as

$$\underset{\sim}{\Delta}^{-1} = \frac{1}{\gamma}(\underset{\sim}{\tau}+\underset{\sim}{\tau}_\delta)^{-1} - \frac{1}{\gamma\psi}(\underset{\sim}{\tau}+\underset{\sim}{\tau}_\delta)^{-1}\underset{\sim}{\tau}_\delta\underset{\sim}{\beta}\underset{\sim}{\beta}^T\underset{\sim}{\tau}_\delta(\underset{\sim}{\tau}+\underset{\sim}{\tau}_\delta)^{-1},$$

where $\psi = \gamma + \underset{\sim}{\beta}^T\underset{\sim}{\tau}_\delta(\underset{\sim}{\tau}+\underset{\sim}{\tau}_\delta)^{-1}\underset{\sim}{\tau}_\delta\underset{\sim}{\beta} > 0,$

from which we can derive

$$\underset{\sim}{\mu}_i = \bar{\underset{\sim}{x}}_{i.} + \underset{\sim}{\tau}_\delta\underset{\sim}{\beta}(\bar{y}_{i.} - \alpha - \underset{\sim}{\beta}^T\bar{\underset{\sim}{x}}_{i.})/\psi \quad (i=1,..,k). \quad (24)$$

Let us now consider the parameter α. The derivative of ℓ with respect to α is

$$\frac{\partial \ell}{\partial \alpha} = \sum_{i=1}^{k} \frac{\partial \underset{\sim}{\theta}_i^T}{\partial \alpha}\frac{\partial \ell}{\partial \underset{\sim}{\theta}_i} = \sum_{i=1}^{k} n_i(1, \underset{\sim}{0}^T)\underset{\sim}{\Sigma}^{-1}(\bar{\underset{\sim}{z}}_{i.} - \underset{\sim}{\theta}_i).$$

But $$(1, \underset{\sim}{0}^T)\underset{\sim}{\Sigma}^{-1} = \frac{1}{\gamma}\underset{\sim}{\phi}^T,$$

so that

$$\partial \ell / \partial \alpha = n\{(\bar{y}_{..} - \alpha - \underset{\sim}{\beta}^T \underset{\sim}{\bar{\mu}}) - \underset{\sim}{\beta}^T \underset{\sim}{\tau}(\underset{\sim}{\tau}+\underset{\sim}{\tau}_\delta)^{-1}(\underset{\sim}{\bar{x}}_{..} - \underset{\sim}{\bar{\mu}})\}/\gamma .$$

When $\underset{\sim}{\mu}_i$ is given by the expression in (24), then

$$\bar{y}_{i.} - \alpha - \underset{\sim}{\beta}^T \underset{\sim}{\mu}_i = (1 - \underset{\sim}{\beta}^T \underset{\sim}{\tau}_\delta \underset{\sim}{\beta}/\psi)(\bar{y}_{i.} - \alpha - \underset{\sim}{\beta}^T \underset{\sim}{\bar{x}}_{i.})$$

and $$\underset{\sim}{\bar{x}}_{i.} - \underset{\sim}{\mu}_i = -\underset{\sim}{\tau}_\delta \underset{\sim}{\beta}(\bar{y}_{i.} - \alpha - \underset{\sim}{\beta}^T \underset{\sim}{\bar{x}}_{i.})/\psi .$$

Hence, with the $\underset{\sim}{\mu}_i$ replaced by the expressions in (24), $\partial \ell / \partial \alpha$ becomes simply

$$n(\bar{y}_{..} - \alpha - \underset{\sim}{\beta}^T \underset{\sim}{\bar{x}}_{..})/\psi , \quad (25)$$

from which we get

$$\alpha = \bar{y}_{..} - \underset{\sim}{\beta}^T \underset{\sim}{\bar{x}}_{..} . \quad (26)$$

To show that we have now found a maximum with respect to the $\underset{\sim}{\mu}_i$ and α we need just to show that the derivative of (25) with respect to α is negative (see Appendix 1). But this derivative is just $-n/\psi$ and so the result is proved.

Define

$$b_{yy} = \sum n_i(\bar{y}_{i.} - \bar{y}_{..})^2/n ,$$

$$\underset{\sim}{b}_{yx} = \sum n_i(\bar{y}_{i.} - \bar{y}_{..})(\underset{\sim}{\bar{x}}_{i.} - \underset{\sim}{\bar{x}}_{..})/n ,$$

$$\underset{\sim}{b}_{xx} = \sum n_i(\underset{\sim}{\bar{x}}_{i.} - \underset{\sim}{\bar{x}}_{..})(\underset{\sim}{\bar{x}}_{i.} - \underset{\sim}{\bar{x}}_{..})^T/n ,$$

$$s_{yy} = \sum\sum(y_{ij} - \bar{y}_{i.})^2/n , \text{ etc,}$$

$$B(\underset{\sim}{\beta}) = b_{yy} - 2\underset{\sim}{\beta}^T \underset{\sim}{b}_{yx} + \underset{\sim}{\beta}^T \underset{\sim}{b}_{xx} \underset{\sim}{\beta} ,$$

and $$\underset{\sim}{S} = \begin{pmatrix} s_{yy} & \underset{\sim}{s}_{yx}^T \\ \underset{\sim}{s}_{yx} & \underset{\sim}{s}_{xx} \end{pmatrix}.$$

Then, with the $\underset{\sim}{\mu}_i$ and α replaced by the expressions in (24) and (26), we find that

$$\sum\sum(\underset{\sim}{z}_{ij} - \underset{\sim}{\theta}_i)^T\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{z}_{ij} - \underset{\sim}{\theta}_i)$$

becomes

$$n\ \text{trace}(\underset{\sim}{S}\underset{\sim}{\Sigma}^{-1}) + nB(\underset{\sim}{\beta})/(\underset{\sim}{\beta}^T\underset{\sim}{\tau}_\delta\underset{\sim}{\beta} + \sigma_\varepsilon^2).$$

Hence the maximum relative log-likelihood of $\underset{\sim}{\tau}$, $\underset{\sim}{\tau}_\delta$, σ_ε^2 and $\underset{\sim}{\beta}$ is

$$\ell^* = -\tfrac{1}{2}n(p+1)\log(2\pi) - \tfrac{1}{2}n\log|\underset{\sim}{\Sigma}| - \tfrac{1}{2}n\ \text{trace}(\underset{\sim}{S}\underset{\sim}{\Sigma}^{-1}) - \tfrac{1}{2}nB(\underset{\sim}{\beta})/(\underset{\sim}{\beta}^T\underset{\sim}{\tau}_\delta\underset{\sim}{\beta} + \sigma_\varepsilon^2).$$

By analogy with the univariate case we might anticipate that the maximum of ℓ^* with respect to $\underset{\sim}{\tau}$, $\underset{\sim}{\tau}_\delta$ and σ_ε^2 occurs when $\underset{\sim}{\tau}$, $\underset{\sim}{\tau}_\delta$ and σ_ε^2 are chosen so that

$$\underset{\sim}{\Sigma} = \underset{\sim}{S}^* = \underset{\sim}{S} + \frac{B(\underset{\sim}{\beta})}{W^2(\underset{\sim}{\beta})}\underset{\sim}{\xi}\underset{\sim}{\xi}^T,$$

where $$W(\underset{\sim}{\beta}) = s_{yy} - 2\underset{\sim}{\beta}^T\underset{\sim}{s}_{yx} + \underset{\sim}{\beta}^T\underset{\sim}{s}_{xx}\underset{\sim}{\beta}$$

and $$\underset{\sim}{\xi}^T = (s_{yy} - \underset{\sim}{\beta}^T\underset{\sim}{s}_{yx},\ \underset{\sim}{s}_{yx}^T - \underset{\sim}{\beta}^T\underset{\sim}{s}_{xx}).$$

So let us now check to see if this choice satisfies the likelihood equations. Let

$$\ell^\dagger = -\tfrac{1}{2}n\log|\underset{\sim}{\Sigma}| - \tfrac{1}{2}n\ \text{trace}(\underset{\sim}{S}\underset{\sim}{\Sigma}^{-1}).$$

Then

$$\frac{\partial \ell^*}{\partial \underset{\sim}{\tau}} = (\underset{\sim}{\beta}, \underset{\sim}{I}_p) \frac{\partial \ell^\dagger}{\partial \underset{\sim}{\Sigma}} \begin{pmatrix} \underset{\sim}{\beta}^T \\ \underset{\sim}{I}_p \end{pmatrix},$$

$$\frac{\partial \ell^*}{\partial \underset{\sim}{\tau}_\delta} = D\left\{ (\underset{\sim}{O}, \underset{\sim}{I}_p) \frac{\partial \ell^\dagger}{\partial \underset{\sim}{\Sigma}} \begin{pmatrix} \underset{\sim}{O}^T \\ \underset{\sim}{I}_p \end{pmatrix} \right\} + \frac{nB(\underset{\sim}{\beta})}{2(\underset{\sim}{\beta}^T \underset{\sim}{\tau}_\delta \underset{\sim}{\beta} + \sigma_\varepsilon^2)^2} \text{diag}(\beta_1^2, \ldots, \beta_p^2),$$

where $\text{diag}(a_1, \ldots, a_p)$ is a diagonal matrix with $a_1, \ldots, a_p$ as diagonal elements and $D(\underset{\sim}{A}) = \text{diag}(a_{11}, a_{22}, \ldots, a_{pp})$, where $\underset{\sim}{A} = (a_{ij})$, and

$$\frac{\partial \ell^*}{\partial \sigma_\varepsilon^2} = (1, \underset{\sim}{O}^T) \frac{\partial \ell^\dagger}{\partial \underset{\sim}{\Sigma}} \begin{pmatrix} 1 \\ \underset{\sim}{O} \end{pmatrix} + \tfrac{1}{2} nB(\underset{\sim}{\beta}) / (\underset{\sim}{\beta}^T \underset{\sim}{\tau}_\delta \underset{\sim}{\beta} + \sigma_\varepsilon^2)^2,$$

where $$\frac{\partial \ell^\dagger}{\partial \underset{\sim}{\Sigma}} = -\tfrac{1}{2} n \underset{\sim}{\Sigma}^{-1} (\underset{\sim}{\Sigma} - \underset{\sim}{S}) \underset{\sim}{\Sigma}^{-1}.$$

To evaluate $\partial \ell^\dagger / \partial \underset{\sim}{\Sigma}$ we first expand $\underset{\sim}{S}^{*-1}$ as

$$\underset{\sim}{S}^{*-1} = \underset{\sim}{S}^{-1} - B(\underset{\sim}{\beta}) \underset{\sim}{S}^{-1} \underset{\sim}{\xi} \underset{\sim}{\xi}^T \underset{\sim}{S}^{-1} / \{W^2(\underset{\sim}{\beta}) + B(\underset{\sim}{\beta}) \underset{\sim}{\xi}^T \underset{\sim}{S}^{-1} \underset{\sim}{\xi}\}$$

(see Press, 1972, p23). Using the fact that

$$\underset{\sim}{I}_{p+1} = \underset{\sim}{S} \underset{\sim}{S}^{-1} = \begin{bmatrix} (s_{yy}, \underset{\sim}{s}_{yx}^T) \underset{\sim}{S}^{-1} \\ (\underset{\sim}{s}_{yx}, \underset{\sim}{S}_{xx}) \underset{\sim}{S}^{-1} \end{bmatrix},$$

we see immediately that

$$\underset{\sim}{\xi}^T \underset{\sim}{S}^{-1} = (1, -\underset{\sim}{\beta}^T)$$

and hence that

$$\underset{\sim}{S}^{*-1} = \underset{\sim}{S}^{-1} - \frac{B(\underset{\sim}{\beta})}{W(\underset{\sim}{\beta}) T(\underset{\sim}{\beta})} \begin{bmatrix} 1 \\ -\underset{\sim}{\beta} \end{bmatrix} (1, -\underset{\sim}{\beta}^T), \qquad (27)$$

where $T(\beta) = W(\beta) + B(\beta)$, and finally we see that

$$\left.\frac{\partial \ell^{\dagger}}{\partial \Sigma}\right|_{\Sigma=S^{*}} = -\frac{nB(\beta)}{2W(\beta)T(\beta)}\begin{pmatrix}1\\-\beta\end{pmatrix}(1\ ,-\beta^{T}).$$

When $\Sigma = S^{*}$ we can easily show that $\beta^{T}\tau_{\delta}\beta + \sigma_{\varepsilon}^{2}$ becomes $T(\beta)$. We can now verify that the likelihood equations for τ, τ_{δ} and σ_{ε}^{2} are satisfied and hence that a solution of the likelihood equations is

$$\tau = s_{xx} + \{B(\beta)/W^{2}(\beta)\}(s_{xx}\beta - s_{yx})(\beta^{T}s_{xx} - s_{yx}^{T}) - \tau_{\delta}\ ,$$

$$\tau_{\delta i} = [T(\beta)/\{W(\beta)\beta_{i}\}](s_{xx}\beta - s_{yx})_{i} \quad (i=1,..,p),$$

and

$$\sigma_{\varepsilon}^{2} = \{T(\beta)/W(\beta)\}(s_{yy} - \beta^{T}s_{yx}).$$

We have not, of course, proved that this solution maximizes ℓ^{*} and, so far, attempts to do this have failed. However, in the univariate case we know that this solution is the only solution of the likelihood equations and that it maximizes ℓ^{*}; it would be rather surprising if this were not also the case here. So, bearing in mind that the assertion that the above solutions maximize ℓ^{*} with respect to τ, τ_{δ} and σ_{ε}^{2} is unproved, let us continue with the maximization.

Using the expression in (27) for S^{*-1}, we can show that

$$\text{trace}(SS^{*-1}) = (p+1) - B(\beta)/T(\beta).$$

Hence, after maximization with respect to τ, τ_{δ} and σ_{ε}^{2}, ℓ^{*} becomes

$$\hat{\ell}(\underset{\sim}{\beta}) = -\tfrac{1}{2}n(p+1)\{1+\log(2\pi)\} - \tfrac{1}{2}n \log|\underset{\sim}{S}^*|.$$

Consider

$$\begin{vmatrix} W^2(\underset{\sim}{\beta}) & -B(\underset{\sim}{\beta})\underset{\sim}{\xi}^T \\ \underset{\sim}{\xi} & \underset{\sim}{S} \end{vmatrix} .$$

This determinant may be written both as

$$W^2(\underset{\sim}{\beta})|\underset{\sim}{S} + \{B(\underset{\sim}{\beta})/W^2(\underset{\sim}{\beta})\}\underset{\sim}{\xi}\underset{\sim}{\xi}^T| = W^2(\underset{\sim}{\beta})|\underset{\sim}{S}^*|$$

and as $|\underset{\sim}{S}|\{W^2(\underset{\sim}{\beta}) + B(\underset{\sim}{\beta})\underset{\sim}{\xi}^T\underset{\sim}{S}^{-1}\underset{\sim}{\xi}\} = |\underset{\sim}{S}|W(\underset{\sim}{\beta})T(\underset{\sim}{\beta})$.

Hence $|\underset{\sim}{S}^*| = |\underset{\sim}{S}|T(\underset{\sim}{\beta})/W(\underset{\sim}{\beta})$,

and so we can write the maximum relative log-likelihood of $\underset{\sim}{\beta}$ as

$$\hat{\ell}(\underset{\sim}{\beta}) = -\tfrac{1}{2}n(p+1)\{1+\log(2\pi)\} - \tfrac{1}{2}n \log|\underset{\sim}{S}| + \tfrac{1}{2}n \log\{W(\underset{\sim}{\beta})/T(\underset{\sim}{\beta})\}.$$

To maximize $\hat{\ell}$ we need only to maximize $W(\underset{\sim}{\beta})/T(\underset{\sim}{\beta})$.

Define

$$\underset{\sim}{T} = \begin{pmatrix} t_{yy} & \underset{\sim}{t}_{yx}^T \\ \underset{\sim}{t}_{yx} & \underset{\sim}{t}_{xx} \end{pmatrix} ,$$

where $t_{yy} = s_{yy} + b_{yy}$, and let $\underset{\sim}{b}^T = (1, -\underset{\sim}{\beta}^T)$. Then we wish to maximize

$$\underset{\sim}{b}^T\underset{\sim}{S}\underset{\sim}{b} / \underset{\sim}{b}^T\underset{\sim}{T}\underset{\sim}{b} .$$

But this is maximized over arbitrary $\underset{\sim}{b}$ by choosing any $\underset{\sim}{b}$ such that

$$(\underset{\sim}{S} - \lambda_m\underset{\sim}{T})\underset{\sim}{b} = \underset{\sim}{0} , \qquad (28)$$

where λ_m is the largest root of

$$|\underset{\sim}{S} - \lambda \underset{\sim}{T}| = 0$$

(see, for example, Rao, 1965, p59). The vector $\underset{\sim}{b}$ that satisfies (28) is unique only up to changes of scale and so we can scale $\underset{\sim}{b}$ to make the first element equal to one. A slight difficulty could occur in practice if the first element of $\underset{\sim}{b}$ turned out to be zero or very nearly so. This means that the line of best fit is parallel to the Y-axis. As the problem is essentially symmetric in the Y and X's we could interchange the Y with one of the X's to remove the difficulty. We have thus found that $\hat{\ell}$ achieves its maximum of

$$-\tfrac{1}{2}n(p+1)\{1+\log(2\pi)\} - \tfrac{1}{2}n \log|\underset{\sim}{S}| + \tfrac{1}{2}n \log\lambda_m$$

when $$\hat{\underset{\sim}{\beta}} = (\underset{\sim}{s}_{xx} - \lambda_m \underset{\sim}{t}_{xx})^{-1}(\underset{\sim}{s}_{yx} - \lambda_m \underset{\sim}{t}_{yx}).$$

The vector $\underset{\sim}{b}$ is the last canonical variate (ie, the canonical variate associated with the smallest canonical correlation) of a between-and-within canonical regression analysis. Essentially this means that the slope of the estimated line is chosen so that the values $y_{ij} - \underset{\sim}{\beta}^T \underset{\sim}{x}_{ij}$ are least correlated with group differences, as might have been expected on general grounds.

We have now completed the solution of the likelihood equations having found the solution

$$\hat{\underset{\sim}{\mu}}_i = \bar{\underset{\sim}{x}}_{i.} + \{(\bar{y}_{i.} - \bar{y}_{..}) - \hat{\underset{\sim}{\beta}}^T(\bar{\underset{\sim}{x}}_{i.} - \bar{\underset{\sim}{x}}_{..})\}(\underset{\sim}{s}_{xx}\hat{\underset{\sim}{\beta}} - \underset{\sim}{s}_{yx})/W(\hat{\underset{\sim}{\beta}}) \quad (i=1,..,k),$$

$$\hat{\alpha} = \bar{y}_{..} - \hat{\underset{\sim}{\beta}}^T \bar{\underset{\sim}{x}}_{..},$$

$$\hat{\underset{\sim}{\tau}} = \underset{\sim}{s}_{xx} + \{B(\hat{\underset{\sim}{\beta}})/W^2(\hat{\underset{\sim}{\beta}})\}(\underset{\sim}{s}_{xx}\hat{\underset{\sim}{\beta}} - \underset{\sim}{s}_{yx})(\hat{\underset{\sim}{\beta}}^T\underset{\sim}{s}_{xx} - \underset{\sim}{s}_{yx}^T) - \hat{\underset{\sim}{\tau}}_\delta ,$$

$$\hat{\tau}_{\delta i} = T(\hat{\underset{\sim}{\beta}})(\underset{\sim}{s}_{xx}\hat{\underset{\sim}{\beta}} - \underset{\sim}{s}_{yx})_i/\{\hat{\beta}_i W(\hat{\underset{\sim}{\beta}})\} \quad (i=1,..,p),$$

$$\hat{\sigma}_\varepsilon^2 = \{T(\hat{\underset{\sim}{\beta}})/W(\hat{\underset{\sim}{\beta}})\}(s_{yy} - \hat{\underset{\sim}{\beta}}^T\underset{\sim}{s}_{yx}) ,$$

and
$$\hat{\underset{\sim}{\beta}} = (\underset{\sim}{s}_{xx} - \lambda_m\underset{\sim}{t}_{xx})^{-1}(\underset{\sim}{s}_{yx} - \lambda_m\underset{\sim}{t}_{yx}),$$

where λ_m is the largest root of

$$|\underset{\sim}{S} - \lambda\underset{\sim}{T}| = 0.$$

This solution is thought to correspond to the absolute maximum of the likelihood but there is one unproved step in the argument. Of course, we know from the univariate special case that these solutions can sometimes lie outside the parameter space, in which case the maximum inside the parameter space will probably lie on one of the boundaries. No attempt will be made here to find any of the boundary estimators.

Chapter 5

ANALYSIS OF AN EXPERIMENT ON APPLE TREES

In this chapter we apply some of the theory of previous chapters as part of the analysis of an experiment on apple trees. The data were provided by Dr S.C. Pearce, East Malling Research Station, for a discussion at the British Region of the Biometric Society in 1969.

A commercial apple tree consists of two parts grafted together, the upper part, or scion, and the rootstock. In this experiment different rootstocks, which largely determine the size and development of the trees, were being compared. From each of thirteen rootstocks a clone (a set of plants raised asexually from a single parent) of eight was raised and trees of the scion Worcester Pearmain were grafted on rootstocks from these clones. After 4 and 15 years measurements of cambial and apical activity were taken. We shall concentrate solely on the data collected after 15 years. Activity of the cambium, the meristematic tissue beneath the bark, may be measured by the girth of the trunk while activity of the meristematic tissue at the apex of shoots may be measured by the weight of the part of the tree above ground. The data for these two measurements at 15 years are given in Table IV. In Figure III we have plotted the natural logs of weight versus girth. Logarithms were taken so that the data would at least approximately meet the assumptions of the models we wish to fit, for instance to make the relationship linear and to comply with the assumption that the errors are independent of the true values of the X variable. Also the slope, β, being now dimensionless, is easier to interpret.

The log-log plots in Figure III suggest that the simplest k-group model studied in Chapter 3 might fit the data reasonably well; the within-group scatters do not look inconsistent with bivariate normal samples and the means of the groups appear to be scattered along the same line as the within-group scatters. The three sets of estimates (internal, $(\sigma_\delta^2=0)$-boundary and $(\sigma_\varepsilon^2=0)$-boundary) obtained from fitting this model are given in Table V. The estimated standard deviation (S.D.) is based on the asymptotic standard deviation of the internal estimator. The ML estimates are, in this case, the $(\sigma_\delta^2=0)$-boundary estimates, though the internal estimates are only slightly different.

The 95% confidence interval for the slope, β, is (2.15 , 2.38). Since we took logs and the weight is essentially a volume measurement, the slope represents a proportional difference in growth between the volume

TABLE IV

MEASUREMENTS ON THE APPLE TREES AT 15 YEARS

Rootstock 1

Trunk girth (mm)	358	375	393	394	360	351	398	362
Weight above ground (lbs)	760	821	928	1009	766	726	1209	750

Rootstock 2

Trunk girth (mm)	409	406	487	498	438	465	469	440
Weight above ground (lbs)	1036	1094	1635	1517	1197	1244	1495	1026

Rootstock 3

Trunk girth (mm)	376	444	438	467	448	478	457	456
Weight above ground (lbs)	912	1398	1197	1613	1476	1571	1506	1458

Rootstock 4

Trunk girth (mm)	398	405	405	392	327	395	427	385
Weight above ground (lbs)	944	1241	1023	1067	693	1085	1242	1017

Rootstock 5

Trunk girth (mm)	404	416	479	442	347	441	464	457
Weight above ground (lbs)	1084	1151	1381	1242	673	1137	1455	1325

Rootstock 6

Trunk girth (mm)	376	314	375	399	334	321	363	395
Weight above ground (lbs)	800	606	790	853	610	562	707	952

TABLE IV (continued)

Rootstock 7

Trunk girth (mm)	266	241	380	401	296	315	358	343
Weight above ground (lbs)	414	335	885	1012	489	616	788	733

Rootstock 8

Trunk girth (mm)	231	250	219	275	205	213	266	226
Weight above ground (lbs)	375	410	335	560	251	272	478	278

Rootstock 9

Trunk girth (mm)	299	381	362	372	369	368	408	410
Weight above ground (lbs)	506	882	737	772	827	821	1149	1035

Rootstock 10

Trunk girth (mm)	431	465	484	527	463	412	514	522
Weight above ground (lbs)	1609	1658	1789	2375	1556	1418	2266	2508

Rootstock 11

Trunk girth (mm)	387	414	387	390	327	424	421	382
Weight above ground (lbs)	1052	1167	981	944	737	1392	1326	1052

Rootstock 12

Trunk girth (mm)	448	435	451	450	428	424	482	469
Weight above ground (lbs)	1258	1304	1290	1288	1176	1177	1331	1490

Rootstock 13

Trunk girth (mm)	452	412	425	460	464	457	463	473
Weight above ground (lbs)	1499	1412	1488	1751	1937	1823	1838	1817

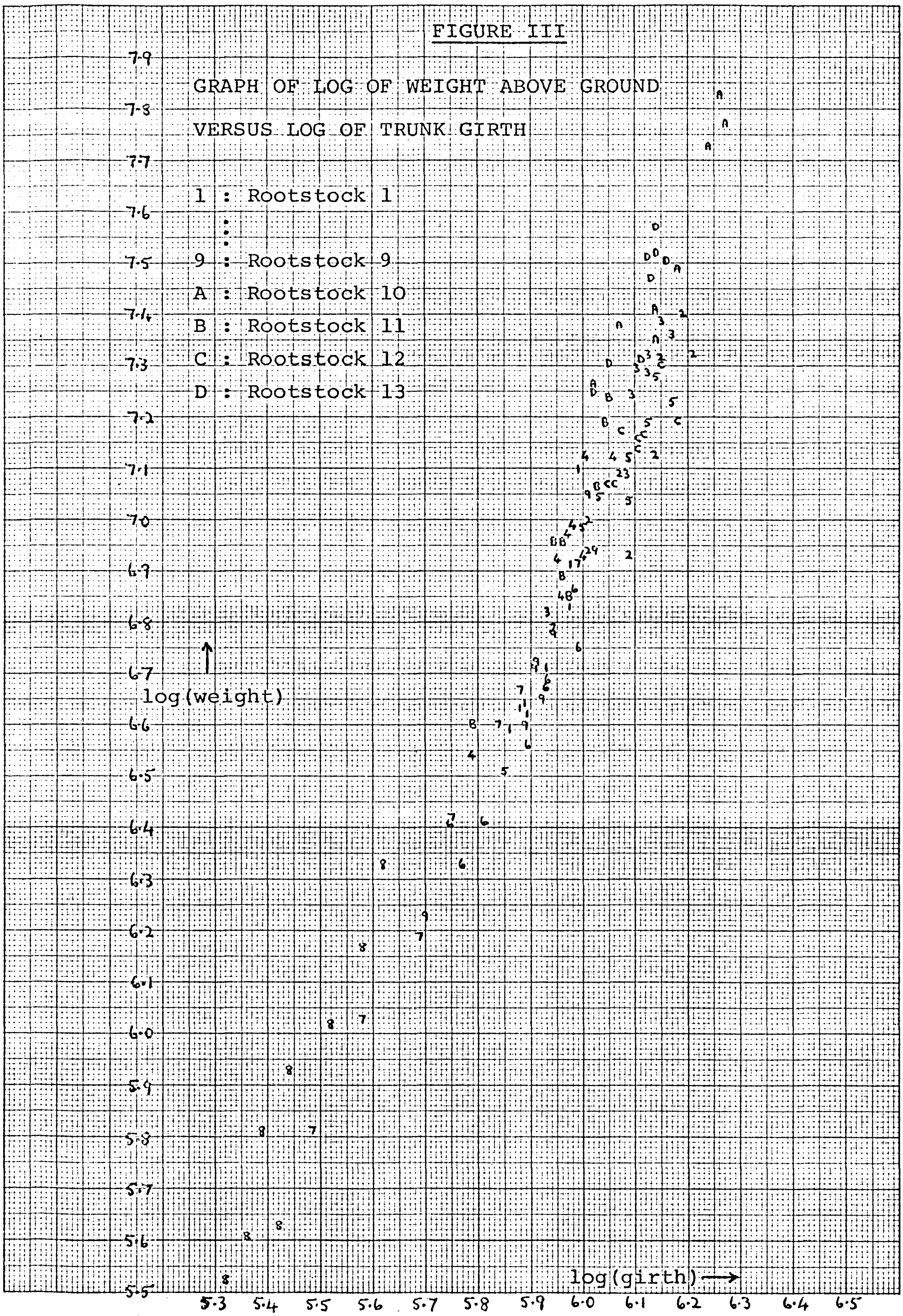


FIGURE III
GRAPH OF LOG OF WEIGHT ABOVE GROUND
VERSUS LOG OF TRUNK GIRTH
1 : Rootstock 1
9 : Rootstock 9
A : Rootstock 10
B : Rootstock 11
C : Rootstock 12
D : Rootstock 13
log(weight)
log(girth)
7·9
7·8
7·7
7·6
7·5
7·4
7·3
7·2
7·1
7·0
6·9
6·8
6·7
6·6
6·5
6·4
6·3
6·2
6·1
6·0
5·9
5·8
5·7
5·6
5·5
5·3
5·4
5·5
5·6
5·7
5·8
5·9
6·0
6·1
6·2
6·3
6·4
6·5

TABLE V

PARAMETER ESTIMATES IN THE SIMPLEST MODEL

Parameter	Internal estimate	$(\sigma_\delta^2=0)$-bdy estimate	$(\sigma_\varepsilon^2=0)$-bdy estimate	Estimated S.D.
μ_1	5.926	5.923	5.901	0.030
μ_2	6.115	6.110	6.060	0.030
μ_3	6.096	6.097	6.095	0.030
μ_4	5.964	5.965	5.974	0.030
μ_5	6.066	6.062	6.023	0.030
μ_6	5.888	5.881	5.830	0.030
μ_7	5.773	5.770	5.776	0.030
μ_8	5.452	5.457	5.537	0.030
μ_9	5.917	5.913	5.882	0.030
μ_{10}	6.156	6.165	6.220	0.030
μ_{11}	5.964	5.967	5.989	0.030
μ_{12}	6.108	6.105	6.068	0.030
μ_{13}	6.100	6.110	6.180	0.030
α	-6.49	-6.59	-7.50	0.38
σ^2	0.0077	0.0074	0.0072	0.0011
σ_δ^2	-0.00031	0.0000	0.0026	0.00049
σ_ε^2	0.0157	0.0142	0.0000	0.0032
β	2.246	2.263	2.416	0.064

and linear measurements. If the growth in all three dimensions were proportional we would expect a value of about 3 for β while if the trees tended to be of uniform height but growth in the two dimensions were proportional we would expect a value of about 2. The confidence interval includes neither 2 nor 3, suggesting that the true situation is somewhere between these two special cases.

It would be interesting to test the assumption that the within-group scatters of observations are parallel. If we are going to allow these slopes to vary with group it would be unreasonable to assume a common intercept. Hence we want to test the hypothesis that the within-group slopes are equal, the intercepts being allowed to vary with group. We have not derived a test procedure for this but we would not expect to get a substantially different answer by assuming that either $\sigma_\delta^2 = 0$ or $\sigma_\varepsilon^2 = 0$ and using the appropriate regression test. In this case the ML estimate of σ_δ^2 is exactly zero which suggests that the errors in the X variable, trunk girth, are small. Hence it seems sensible to test the hypothesis assuming that $\sigma_\delta^2 = 0$. The F-statistic for this purpose has a value of about 0.96 which is very insignificant. Hence there is no evidence of non-parallelism of the within-group scatters.

When we consider other tests of the adequacy of the simplest model we see, however, that it does not fit the data satisfactorily. The χ^2_{11} statistic for testing the null hypothesis that the intercepts of the thirteen groups are all equal against the general alternative, with the variances assumed not to differ with group, has a value of approximately 131 which is highly significant. If the variances are now allowed to vary with group, the test is still highly significant. So let us now assume that the group intercepts are not equal and test to see if the variances differ with group.

The test of the null hypothesis that the variances do not vary with group against the general alternative, the group intercepts being assumed to differ in both hypotheses, has a value of 55.4 which has a significance of about 2% on the χ^2_{36} distribution. So while there is still some evidence of heterogeneity amongst the variances this evidence is not overwhelming and the model that allows the intercepts to vary with group but which assumes that the variances are constant from group to group fits the data reasonably well. Unfortunately we are unable to identify the parameters in this model so that it is of little direct use. We are thus forced to make an additional assumption if we are to proceed.

In §3.4 we saw that if we could impose some linear constraint on the intercepts then we could estimate all the parameters. However, in this case there is no obvious constraint that could be used and it seems much more reasonable instead to set $\sigma_\delta^2 = 0$. This is what we now do.

We have already accepted that the model which allows the intercepts to vary with group but which assumes that the variances do not vary with group provides an adequate fit to the data. With the additional assumption that $\sigma_\delta^2 = 0$ we have the model

$$Y_{ij} = \alpha_i + \beta X_{ij} + \varepsilon_{ij} \quad (j=1,..,n_i\ ;\ i=1,..,k),$$

where the X_{ij} have independent normal distributions with means μ_i and common variance σ^2, the ε_{ij} have independent normal distributions with means 0 and variances σ_ε^2, and the ε_{ij} are independent of the X_{ij}. In this model we estimate β by s_{yx}/s_{xx}, μ_i by $\bar{x}_{i.}$, α_i by $\bar{y}_{i.} - (s_{yx}/s_{xx})\bar{x}_{i.}$, σ^2 by s_{xx} and σ_ε^2 by $s_{yy} - s_{yx}^2/s_{xx}$. For this set of data, then, we estimate β by 2.273 and its standard deviation by 0.078. The estimates of σ^2 and σ_ε^2 are 0.0074 and 0.0040 respectively. The estimates of the means, μ_i, are the same as those in the "$(\sigma_\delta^2=0)$-boundary estimates" column of Table V and their estimated standard deviations are again 0.030 in each case. The estimates of the intercepts, α_i, and their standard deviations are summarized in Table VI.

Note that, while the standard deviations of the individual intercept estimates are quite large, the standard deviations of differences between them are much smaller; in fact, they vary from 0.032 to 0.064. There are no obvious groupings of the intercepts and, in the absence of further information, the assumption of a random effect for the intercepts seems a reasonable one. In fact a plot of the estimates, $\hat{\alpha}_i$, of the intercepts against the expected order statistics from a normal distribution shows that the values $\hat{\alpha}_i$ are reasonably consistent with a sample from a normal distribution. We can estimate the variance, σ_α^2, of the α_i by subtracting the average of the estimated variances of the $\hat{\alpha}_i - \hat{\hat{\alpha}}$ from the mean sum of squares of the $\hat{\alpha}_i - \hat{\hat{\alpha}}$. This gives us 0.0095 as an estimate of σ_α^2.

TABLE VI

INTERCEPT ESTIMATES IN THE REGRESSION MODEL

Parameter	α_1	α_2	α_3	α_4	α_5	α_6	α_7
Estimate	-6.71	-6.75	-6.63	-6.63	-6.73	-6.78	-6.69
Estimated S.D.	0.46	0.48	0.48	0.47	0.47	0.46	0.45

Parameter	α_8	α_9	α_{10}	α_{11}	α_{12}	α_{13}
Estimate	-6.53	-6.73	-6.48	-6.59	-6.72	-6.46
Estimated S.D.	0.43	0.46	0.48	0.47	0.48	0.48

The standard deviation of differences between the estimates, $\bar{x}_{i.}$, of the means, μ_i, is approximately 0.043.

From the ranked estimates of the X means, 5.457, 5.770, 5.881, 5.913, 5.923, 5.965, 5.967, 6.062, 6.097, 6.105, 6.110, 6.110 and 6.165, we can see that the mean, 5.457, of group 8 is clearly significantly different from the others. But are the others consistent with a common mean? A between and within analysis of variance on the x_{ij} for all groups except group 8 gives an F ratio with a significance of about 8%, showing that there is, in fact, little evidence that these group means are unequal.

Our overall summary of the data is thus as follows. The observations on the variable log of trunk girth are measured with small errors which we assume to be zero, and within each group the observations are consistent with random samples from a normal distribution with variance 0.0074 (S.D. = 0.0011); the mean of the distribution for group 8 is estimated by 5.457 (S.D. = 0.030) and it is significantly less than the common mean, 6.006 (S.D. = 0.0087), of the other 12 groups. The observations on log of weight above ground, Y_{ij}, are related to the observations on log of trunk girth, X_{ij}, by the equation

$$Y_{ij} = \alpha_i + \beta X_{ij} + \varepsilon_{ij} \quad (j=1,..,8\,;\, i=1,..,13),$$

where β is estimated by 2.273 (S.D. = 0.078), the α_i are consistent with a random sample from a normal distribution with mean -6.65 and variance 0.0095 (the individual estimates of the α_i are given in Table VI), and the ε_{ij} are consistent with a random sample from a normal distribution with mean zero and variance 0.0040. This gives a quite concise representation of a potentially quite complicated situation. Data were also collected when the trees were four years old but the next step of analysing the four-year-old data and relating the results from the two ages will not be done here.

Appendix 1

THE DETERMINATION OF THE NATURE OF TURNING POINTS OF LIKELIHOODS

In relatively complicated problems the likelihood equations may have several solutions and it is then desirable to examine which solutions correspond to local maxima; of course the same problem arises in optimization calculations in other contexts. In principle, this examination can be done by calculating the matrix of second derivatives of the log-likelihood and then looking at the signs of the principal minors. If the equations have been solved analytically, however, this can be very tedious and we present here a method that, when applicable, avoids these calculations. Virtually all the likelihoods studied in this thesis have more than one stationary value and the amount of work saved by the use of this method is considerable.

In the method described here the pivots of the matrix of double derivatives of the log-likelihood are calculated. The pivots can be interpreted as the factors by which one principal minor is multiplied in order to get the next principal minor. Hence the method is equivalent to that described above. In the theorem stated below we find expressions for these pivots. It is these expressions that hold the key to the method because we shall see that they can be determined with very little effort during the solution of the likelihood equations.

The essential idea is related to pivotal reduction in linear systems (see, for instance, Fraser, 1959, pp. 170-3) and to a result of Richards (1961) for finding the covariance matrix in nonlinear problems. In particular, the interpretation of the pivots that we give here, roughly speaking as second derivatives of the log-likelihood after optimization with respect to previous parameters, is closely related to the interpretation given by Jowett (1963) in the linear regression model.

The likelihood equations are solved in a sequential manner that enables easy calculation of the pivots of the matrix of double derivatives of the log-likelihood. The signs of these pivots determine the nature of the turning points; if all are negative when evaluated at a turning point then that point is at a local maximum; if all are positive the point is at a local minimum; if any two have opposite signs the point is at a saddle point; and if one of the pivots is zero we may need to investigate higher order derivatives or to reparameterize.

The method depends on the following result.

Theorem. Let $\ell(\underset{\sim}{\theta}, \phi)$ be a twice differentiable function of $\underset{\sim}{\theta}$ and ϕ, and let $\underset{\sim}{\gamma}^T = (\underset{\sim}{\theta}^T, \phi)$. Suppose that

$$\left.\frac{\partial \ell}{\partial \underset{\sim}{\theta}}\right|_{\underset{\sim}{\theta}=\underset{\sim}{g}(\phi)} = \underset{\sim}{0} \quad \text{for all } \phi,$$

where $\underset{\sim}{g}$ is a differentiable vector function of ϕ, and suppose also that the matrix $\partial^2\ell/\partial\underset{\sim}{\theta}^2$ is nonsingular at $(\underset{\sim}{\theta}, \phi) = (\underset{\sim}{g}(\phi_o), \phi_o)$. Then, when $\phi = \phi_o$,

$$\left|\frac{\partial^2 \ell}{\partial \underset{\sim}{\gamma}^2}\right|_{\underset{\sim}{\theta}=\underset{\sim}{g}(\phi)} = \left|\frac{\partial^2 \ell}{\partial \underset{\sim}{\theta}^2}\right|_{\underset{\sim}{\theta}=\underset{\sim}{g}(\phi)} \frac{d}{d\phi}\left\{\left.\frac{\partial \ell}{\partial \phi}\right|_{\underset{\sim}{\theta}=\underset{\sim}{g}(\phi)}\right\}.$$

The proof of this theorem is not difficult; essentially it is a matter of showing that the pivot,

$$\frac{d}{d\phi}\left\{\left.\frac{\partial \ell}{\partial \phi}\right|_{\underset{\sim}{\theta}=\underset{\sim}{g}(\phi)}\right\},$$

evaluated at $\phi = \phi_o$, is equal to

$$\frac{\partial^2 \ell}{\partial \phi^2} - \left(\frac{\partial^2 \ell}{\partial \underset{\sim}{\theta} \partial \phi}\right)^T \left(\frac{\partial^2 \ell}{\partial \underset{\sim}{\theta}^2}\right)^{-1} \left(\frac{\partial^2 \ell}{\partial \underset{\sim}{\theta} \partial \phi}\right)$$

evaluated at $(\underset{\sim}{\theta}, \phi) = (\underset{\sim}{g}(\phi_o), \phi_o)$. A more general result was essentially proved by Richards (1961). The pivots of the matrix of double derivatives of a log-likelihood are terms like the last term in the equation in the statement of the theorem and their calculation is made easy by solving the equations sequentially as we now see.

Let $\ell(\underset{\sim}{\theta})$ be a log-likelihood function of $\underset{\sim}{\theta}^T =$

$(\theta_1,..,\theta_p)$. Then the first step in the sequential procedure is to solve the first likelihood equation, $\partial\ell/\partial\theta_1 = 0$, for θ_1 in terms of the remaining parameters, $\theta_2,..,\theta_p$. Let such a solution be $\tilde{\theta}_1(\theta_2,..,\theta_p)$. The first pivot is just $\partial^2\ell/\partial\theta_1^2$.

The next step is to substitute $\tilde{\theta}_1(\theta_2,..,\theta_p)$ for θ_1 in the likelihood equation for θ_2 and then to solve the resulting equation,

$$\left.\frac{\partial\ell}{\partial\theta_2}\right|_{\theta_1=\tilde{\theta}_1(\theta_2,..,\theta_p)} = 0,$$

for θ_2 in terms of the remaining parameters, $\theta_3,..,\theta_p$. Let such a solution be $\tilde{\theta}_2(\theta_3,..,\theta_p)$ and let $\theta_1^*(\theta_3,..,\theta_p)$ be the result of replacing θ_2 by $\tilde{\theta}_2$ in $\tilde{\theta}_1(\theta_2,..,\theta_p)$. The second pivot is the derivative of the left-hand-side of the above equation with respect to θ_2.

At the third step we substitute θ_1^* for θ_1 and $\tilde{\theta}_2$ for θ_2 in the derivative of the log-likelihood with respect to θ_3. By equating this derivative to zero we get an equation which we solve for θ_3 in terms of $\theta_4,..,\theta_p$, and by differentiating again with respect to θ_3 we obtain the third pivot.

Continuing in this manner, we eventually get the sets of solutions of the likelihood equations and the p pivots whose signs must be investigated at each of these sets in order to determine which are at local maxima, which are at local minima, and so on.

This method is useful only if the likelihood equations can be easily solved sequentially. However, the functions $\underset{\sim}{g}(\phi)$ are not necessarily needed explicitly so that the pivots can sometimes still be found without too much difficulty even when the equations have not been solved strictly sequentially; §2.2 of this thesis provides an example of this. Of course it is not necessary to adhere strictly to the method presented here or to the direct method; in some problems it may be best to use the direct method to get the first few minors and then to use the method described here for determining the higher order determinants where its advantages can be most pronounced.

We can prove a multivariate generalization of the above theorem which states, under similar restrictions, that if

$$\frac{\partial}{\partial \underset{\sim}{\theta}} \ell(\underset{\sim}{\theta},\underset{\sim}{\phi})\Big|_{\underset{\sim}{\theta}=\underset{\sim}{g}(\underset{\sim}{\phi})} = \underset{\sim}{0} \quad \text{for all } \underset{\sim}{\phi},$$

then the matrix

$$\frac{\partial^2 \ell}{\partial \underset{\sim}{\gamma}^2}\Big|_{\underset{\sim}{\theta}=\underset{\sim}{g}(\underset{\sim}{\phi})}$$

is positive (negative) definite if the two matrices

$$\frac{\partial^2 \ell}{\partial \underset{\sim}{\theta}^2}\Big|_{\underset{\sim}{\theta}=\underset{\sim}{g}(\underset{\sim}{\phi})} \quad \text{and} \quad \frac{d}{d\underset{\sim}{\phi}}\left\{\frac{\partial \ell}{\partial \underset{\sim}{\phi}}\Big|_{\underset{\sim}{\theta}=\underset{\sim}{g}(\underset{\sim}{\phi})}\right\}$$

are both positive (negative) definite. This theorem is used in Chapter 4, though there it was convenient to frame the problem as the sequential maximization of $\ell(\underset{\sim}{\theta}_1, \underset{\sim}{\theta}_2, \underset{\sim}{\theta}_3)$ with respect to $\underset{\sim}{\theta}_1$, $\ell\{\hat{\underset{\sim}{\theta}}_1(\underset{\sim}{\theta}_2, \underset{\sim}{\theta}_3), \underset{\sim}{\theta}_2, \underset{\sim}{\theta}_3\}$ with respect to $\underset{\sim}{\theta}_2$, and so on.

Appendix 2

MOMENTS OF THE SUMS OF SQUARES AND PRODUCTS IN THE K-GROUP MODEL

We outline here a derivation of the moments of the sums of squares and products of the observations in the k-group model studied in Chapter 3. These moments could, of course, be found from first principles but this would be long and tedious; here we use the noncentral Wishart distribution to get the moments much more economically.

Since the observations, Y_{ij} and X_{ij}, are independently normally distributed with a common covariance matrix, then the matrices of between-group, within-group and overall sums of squares and products have noncentral Wishart distributions, the first two being independent. Consider first the between-group sums of squares and products. Since the $\sqrt{n_i}(\bar{Y}_{i.},\bar{X}_{i.})$ $(i=1,..,k)$ are independently normally distributed with means $\sqrt{n_i}(\alpha+\beta\mu_i\ ,\ \mu_i)$ and common covariance matrix

$$\underset{\sim}{\Sigma} = \begin{pmatrix} \beta^2\sigma^2+\sigma_\varepsilon^2 & \beta\sigma^2 \\ \beta\sigma^2 & \sigma^2+\sigma_\delta^2 \end{pmatrix} ,$$

then

$$\underset{\sim}{B} = n\begin{pmatrix} b_{YY} & b_{YX} \\ b_{YX} & b_{XX} \end{pmatrix} ,$$

where $b_{YY} = \sum n_i(\bar{Y}_{i.} - \bar{Y}_{..})^2/n$, etc, has a noncentral Wishart distribution with k-1 degrees of freedom, covariance matrix $\underset{\sim}{\Sigma}$ and noncentrality matrix

$$\sum_{\ell=1}^{k} (\underset{\sim}{\theta}_\ell - \bar{\underset{\sim}{\theta}})(\underset{\sim}{\theta}_\ell - \bar{\underset{\sim}{\theta}})^T = nb_{\mu\mu}\begin{pmatrix} \beta^2 & \beta \\ \beta & 1 \end{pmatrix},$$

where $\underset{\sim}{\theta}_\ell^T = \sqrt{n_\ell}(\alpha+\beta\mu_\ell, \mu_\ell)$ and $b_{\mu\mu} = \sum n_i(\mu_i - \bar{\mu})^2/n$; see, for example, Press (1972, p112).

The characteristic function of $a_{11} = nb_{YY}$, $2a_{12} = 2nb_{YX}$ and $a_{22} = nb_{XX}$ was shown by Anderson (1946) to be

$$c(\underset{\sim}{\Phi}) = E[\exp\{i \text{ trace}(\underset{\sim}{A}\underset{\sim}{\Phi})\}]$$

$$= \frac{|\underset{\sim}{\Sigma}|^{-\frac{1}{2}(k-1)}\exp\{-\frac{1}{2}\sum(\underset{\sim}{\theta}_\ell-\bar{\underset{\sim}{\theta}})^T\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\theta}_\ell-\bar{\underset{\sim}{\theta}})\}}{|\underset{\sim}{\Omega}|^{-\frac{1}{2}(k-1)}\exp\{-\frac{1}{2}\sum(\underset{\sim}{\theta}_\ell-\bar{\underset{\sim}{\theta}})^T\underset{\sim}{\Sigma}^{-1}\underset{\sim}{\Omega}\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\theta}_\ell-\bar{\underset{\sim}{\theta}})\}},$$

where $\underset{\sim}{\Omega} = (\underset{\sim}{\Sigma}^{-1} - 2i\underset{\sim}{\Phi})^{-1}$, $\underset{\sim}{A} = (a_{ij})$ and $\underset{\sim}{\Phi} = (\phi_{ij})$. Let

$$a(\underset{\sim}{\Phi}) = 2i\{(\beta^2\sigma^2+\sigma_\varepsilon^2)\phi_{11} + 2\beta\sigma^2\phi_{12} + (\sigma^2+\sigma_\delta^2)\phi_{22}\}$$
$$+ 4|\underset{\sim}{\Sigma}|(\phi_{11}\phi_{22} - \phi_{12}^2),$$

and

$$b(\underset{\sim}{\Phi}) = i(\beta^2\phi_{11} + 2\beta\phi_{12} + \phi_{22}) + 2(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)(\phi_{11}\phi_{22} - \phi_{12}^2).$$

Then we can show that

$$c(\underset{\sim}{\Phi}) = \{1 - a(\underset{\sim}{\Phi})\}^{-\frac{1}{2}(k-1)}\exp[nb_{\mu\mu}b(\underset{\sim}{\Phi})/\{1 - a(\underset{\sim}{\Phi})\}]. \qquad (29)$$

From this it is easy to show that the cumulant generating function is

$$\log\{c(\underset{\sim}{\Phi})\} = nb_{\mu\mu}b(\underset{\sim}{\Phi})\{1 + a(\underset{\sim}{\Phi}) + a^2(\underset{\sim}{\Phi}) + \ldots\}$$

$$+ \tfrac{1}{2}(k-1)\{a(\underset{\sim}{\Phi}) + \tfrac{1}{2}a^2(\underset{\sim}{\Phi}) + \tfrac{1}{3}a^3(\underset{\sim}{\Phi}) + \ldots\}.$$

We can easily pick out the first and second order terms in the ϕ_{ij} from $\log\{c(\underset{\sim}{\Phi})\}$ and hence get the first and second order moments of the between-group sums of squares and products. These are

$$E(b_{YY}) = \beta^2 b_{\mu\mu} + (k-1)(\beta^2\sigma^2+\sigma_\varepsilon^2)/n\ ,$$

$$E(b_{YX}) = \beta b_{\mu\mu} + (k-1)\beta\sigma^2/n\ ,$$

$$E(b_{XX}) = b_{\mu\mu} + (k-1)(\sigma^2+\sigma_\delta^2)/n\ ,$$

$$\mathrm{var}(b_{YY}) = 2(\beta^2\sigma^2+\sigma_\varepsilon^2)\{2\beta^2 b_{\mu\mu}/n + (k-1)(\beta^2\sigma^2+\sigma_\varepsilon^2)/n^2\}\ ,$$

$$\mathrm{cov}(b_{YY},b_{YX}) = 2\beta b_{\mu\mu}(2\beta^2\sigma^2+\sigma_\varepsilon^2)/n + 2(k-1)\beta\sigma^2(\beta^2\sigma^2+\sigma_\varepsilon^2)/n^2\ ,$$

$$\mathrm{var}(b_{YX}) = b_{\mu\mu}(4\beta^2\sigma^2+\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)/n + (k-1)\left(|\underset{\sim}{\Sigma}| + 2\beta^2\sigma^4\right)/n^2\ ,$$

$$\mathrm{cov}(b_{YY}\ , b_{XX}) = 2\beta^2\sigma^2\{2b_{\mu\mu}/n + (k-1)\sigma^2/n^2\}\ ,$$

$$\mathrm{cov}(b_{YX}\ , b_{XX}) = 2\beta b_{\mu\mu}(2\sigma^2+\sigma_\delta^2)/n + 2(k-1)\beta\sigma^2(\sigma^2+\sigma_\delta^2)/n^2\ ,$$

and $$\mathrm{var}(b_{XX}) = 2(\sigma^2+\sigma_\delta^2)\{2b_{\mu\mu}/n + (k-1)(\sigma^2+\sigma_\delta^2)/n^2\}\ .$$

Higher order moments can also be obtained easily.

Let us now consider the within-group sums of squares and products. Define $s_{YY} = \sum\sum(Y_{ij} - \bar{Y}_{i.})^2/n$, etc. Then, since the (Y_{ij}, X_{ij}) $(j=1,..,n_i\ ; i=1,..,k)$ are independently normally distributed with means $(\alpha+\beta\mu_i\ , \mu_i)$ and common covariance matrix $\underset{\sim}{\Sigma}$,

$$\underset{\sim}{S} = n\begin{pmatrix} s_{YY} & s_{YX} \\ s_{YX} & s_{XX} \end{pmatrix}$$

has a (central) Wishart distribution with n-k degrees of freedom and covariance matrix $\underset{\sim}{\Sigma}$. The characteristic function of $a_{11} = ns_{YY}$, $2a_{12} = 2ns_{YX}$ and $a_{22} = ns_{XX}$ can be found from equation (29) by replacing $b_{\mu\mu}$ by 0 and k-1 by n-k. Hence the moments of the within-group sums of squares and products can be found from the above expressions for the moments of the between-group sums of squares and products by replacing $b_{\mu\mu}$ by 0 and k-1 by n-k.

We can get the moments of the overall sums of squares and products in a similar fashion. Of course we could also get these from the above results using the fact that $\underset{\sim}{B}$ and $\underset{\sim}{S}$ are independent. Define $t_{YY} = \sum\sum(Y_{ij} - \bar{Y}_{..})^2/n$, etc. Then

$$\underset{\sim}{T} = n\begin{pmatrix} t_{YY} & t_{YX} \\ t_{YX} & t_{XX} \end{pmatrix}$$

has a noncentral Wishart distribution with n-1 degrees of freedom, covariance matrix $\underset{\sim}{\Sigma}$ and with the same noncentrality matrix as in the distribution of $\underset{\sim}{B}$. Hence we can obtain the moments of the overall sums of squares and products from the expressions for the moments of the between-group sums of squares and products by replacing k-1 by n-1.

Appendix 3

MOMENTS OF THE REGRESSION SLOPES IN THE K-GROUP MODEL

Here we outline the derivation of the first and second moments of the between-group, within-group and overall regression slopes in the k-group model discussed in Chapter 3. We can get exact moments for the within-group slope but for the other two slopes we can get only asymptotic moments.

To find the asymptotic expected value of the between-group regression slope of Y on X we use

$$E\left(\frac{b_{YX}}{b_{XX}}\right) = \frac{1}{E(b_{XX})} E\left[\{E(b_{YX}) + U\}\left\{1 - \frac{V}{E(b_{XX})} + \frac{V^2}{E(b_{XX})^2} - \dots\right\}\right]$$

where $U = b_{YX} - E(b_{YX})$ and $V = b_{XX} - E(b_{XX})$. We get

$$E(b_{YX}/b_{XX}) = \beta - \beta\sigma_\delta^2(k-3)/(nb_{\mu\mu}) + O(1/n^2).$$

We use a similar expression to find $E(b_{YX}^2/b_{XX}^2)$, from which we can show that

$$\text{var}(b_{YX}/b_{XX}) = (\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)/(nb_{\mu\mu}) + O(1/n^2).$$

We can get the asymptotic moments (by which we mean the moments of the asymptotic distribution) of the overall regression slope of Y on X in exactly the same manner. The first two such moments are

$$E\left(\frac{t_{YX}}{t_{XX}}\right) = \beta\left(\frac{\sigma^2 + b_{\mu\mu}}{\sigma^2+\sigma_\delta^2+b_{\mu\mu}}\right) + \frac{\beta\sigma_\delta^2 b_{\mu\mu}(\sigma^2+\sigma_\delta^2+3b_{\mu\mu})}{n(\sigma^2+\sigma_\delta^2+b_{\mu\mu})^3} + O(1/n^2)$$

and

$$\mathrm{var}(t_{YX}/t_{XX}) = \{(\sigma^2+\sigma_\delta^2)^2|\underset{\sim}{\Sigma}| + b_{\mu\mu}(\sigma^2+\sigma_\delta^2)(3|\underset{\sim}{\Sigma}| + \beta^2\sigma_\delta^4) +$$

$$3b_{\mu\mu}^2|\underset{\sim}{\Sigma}| + b_{\mu\mu}^3(\beta^2\sigma_\delta^2+\sigma_\varepsilon^2)\}/\{n(\sigma^2+\sigma_\delta^2+b_{\mu\mu})^4\} + O(1/n^2).$$

The asymptotic moments of the within-group regression slope of Y on X could also be found in the same way, but in this case we can find the exact moments. The joint density of s_{YY}, s_{YX} and s_{XX} is

$$f(s_{yy}, s_{yx}, s_{xx}) = \frac{(\frac{1}{2}n)^{n-k}(s_{yy}s_{xx} - s_{yx}^2)^{\frac{1}{2}(n-k-3)}}{\pi^{\frac{1}{2}}|\underset{\sim}{\Sigma}|^{\frac{1}{2}(n-k)}\Gamma\{\frac{1}{2}(n-k)\}\Gamma\{\frac{1}{2}(n-k-1)\}}$$

$$\times \exp[-\tfrac{1}{2}n\{(\sigma^2+\sigma_\delta^2)s_{yy} - 2\beta\sigma^2 s_{yx} + (\beta^2\sigma^2+\sigma_\varepsilon^2)s_{xx}\}/|\underset{\sim}{\Sigma}|] ;$$

see, for example, Press (1972, p101). Using the fact, easily proved by induction, that

$$\int_a^\infty (x-a)^{\frac{1}{2}m}\exp(-cx)\,dx = \Gamma(\tfrac{1}{2}m+1)\exp(-ca)/c^{\frac{1}{2}m+1} ,$$

where m is an integer, we can evaluate

$$\int_0^\infty\int_{-\infty}^\infty\int_{s_{yx}^2/s_{xx}}^\infty s_{yx}^i/s_{xx}^i f(s_{yy}, s_{yx}, s_{xx})\,ds_{yy}ds_{yx}ds_{xx}$$

directly and so find the moments of s_{YX}/s_{XX}. The first two such moments are

$$E(s_{YX}/s_{XX}) = \beta\sigma^2/(\sigma^2+\sigma_\delta^2)$$

and $$\mathrm{var}(s_{YX}/s_{XX}) = \frac{|\tilde{\Sigma}|}{(n-k-2)(\sigma^2+\sigma_\delta^2)^2} \quad .$$

Much algebra is required in the evaluation of $\mathrm{cov}(s_{YX}/s_{XX}\ ,\ t_{YX}/t_{XX})$ but the method is straight-forward. We already have $E(s_{YX}/s_{XX})E(t_{YX}/t_{XX})$ and we calculate the other term, $E\{(s_{YX}t_{YX})/(s_{XX}t_{XX})\}$, with an expansion like that used to find $E(b_{YX}/b_{XX})$. The independence of the between-group and within-group sums of squares and products makes the calculation possible with the moments found in Appendix 2. We find

$$\mathrm{cov}(s_{YX}/s_{XX}\ ,\ t_{YX}/t_{XX}) = \frac{|\tilde{\Sigma}|}{n(\sigma^2+\sigma_\delta^2)(\sigma^2+\sigma_\delta^2+b_{\mu\mu})} + O(1/n^2).$$

Similar expressions for the moments of the regression slopes of X on Y can be determined from the symmetry of the model. By replacing Y by X, X by Y, $b_{\mu\mu}$ by $\beta^2 b_{\mu\mu}$, β by $1/\beta$, σ^2 by $\beta^2\sigma^2$, σ_δ^2 by σ_ε^2 and σ_ε^2 by σ_δ^2 we can use the above expressions for the moments of the regression slopes of Y on X to get the moments of the regression slopes of X on Y.

REFERENCES

ANDERSON, T.W. (1946). The non-central Wishart distribution and certain problems of multivariate statistics. Ann. Math. Statist. 17, 409-31.

BERKSON, J. (1950). Are there two regressions? J. Amer. Statist. Assoc. 45, 164-80.

COX, D.R. and HINKLEY, D.V. (1974). Theoretical Statistics. London: Chapman and Hall.

FRASER, D.A.S. (1958). Statistics, An Introduction. New York: Wiley.

HINKLEY, D.V. (1969). On the ratio of two correlated normal random variables. Biometrika 56, 635-9.

JOWETT, G.H. (1963). Applications of Jordan's procedure for matrix inversion in multiple regression and multivariate distance analysis. J.R. Statist. Soc. B 25, 352-7.

KENDALL, M.G. and STUART, A. (1973). The Advanced Theory of Statistics, Vol. 2. London: Griffin.

MORAN, P.A.P. (1956). A test of significance for an unidentifiable relation. J.R. Statist. Soc. B 18, 61-4.

MORAN, P.A.P. (1971). Estimating structural and functional relationships. J. Multi. Anal. 1, 232-55.

NEYMAN, J. (1959). Optimal asymptotic tests of composite statistical hypotheses. In Probability and Statistics, pp. 213-34, ed. Grenander, U. Stockholm: Almqvist and Wiksell.

PRESS, S.J. (1972). Applied Multivariate Analysis. New York: Holt, Rinehart and Winston.

RAO, C.R. (1965). Linear Statistical Inference and Its Applications. New York: Wiley.

RICHARDS, F.S.G. (1961). A method of maximum-likelihood estimation. J.R. Statist. Soc. B 23, 469-75.

RICHARDSON, D.H. and WU, DE-MIN. (1970). Least squares and grouping method estimators in the errors in variables model. J. Amer. Statist. Assoc. 65, 724-48.

WALD, A. (1940). The fitting of straight lines if both variables are subject to error. Ann. Math. Statist. 11, 284-300.