ESTIMATION WITH SAMPLES DRAWN FROM DIFFERENT BUT PARAMETRICALLY RELATED DISTRIBUTIONS

by

Raymond I. Fields, A.B., M.A.

Thesis submitted to the Graduate Faculty of the

Virginia Polytechnic Institute

in candidacy for the degree of

DOCTOR OF PHILOSPHY

in

Statistics

June, 1960

Blacksburg, Virginia

TABLE OF CONTENTS

INDEX OF TABLES

I. INTRODUCTION

1.1 The Problem

In statistical investigations we are sometimes faced with the problem of estimating a common mean by combining the individual estimates of this parameter which have been obtained from a series of experiments. The variances of the estimates may differ among the experiments due to a variety of causes: for example, variations in the experimental skill of the technicians conducting the experiments; variations in the degree of control of external conditions, and, possibly, the fact that this parameter may have been only of secondary interest in some of these experiments.

We shall not consider any laboratory bias or personnel bias, but shall consider, primarily, the problem of estimating jointly the common mean and the individual variances. In the simple situation of combining two independent estimates of a mean, a natural approach is to find the weighted mean. The reciprocals of the estimated variances of the individual sample means, from the separate experiments, are used as weights. However, this method does not take into account the fact that our independent sample means are estimates of a common mean. One way of taking this into account is to jointly estimate the common mean and the individual variances by finding the weighted mean and the individual variances about the weighted mean. Thus, the estimation of a common mean is interrelated with the estimation of variances.

The procedure which will be developed for estimating jointly the common mean and the individual variances is to take the weighted estimate, obtained by using the usual weights (the reciprocals of the estimates of the variances of the individual estimates) as an initial value. Iteration is performed by taking this estimate of the mean and recalculating weights by first obtaining sums of squares about the initial estimate and then computing new estimates of the variances. The second estimate of the weighted mean is obtained using the new weights. The limit m of this sequence is then taken as the (final) estimate of the common mean. Further, the weights used in the final stage of the iteration are functions of the (final) estimates of the individual variances.

Situations where joint estimates of a common mean and individual variances may be desired are: (1) Two laboratory groups may conduct independent experiments to determine the ratio of charge to mass of an electron. A combined estimate of this ratio is desired as well as a measure of the variation for each group which, in some way, utilizes the fact that the two groups are estimating the same parameter. (2) Two groups test independently the tensile strength of a certain lot of steel. A combined estimate of the tensile strength is desired and a measure of the variation for each group. (3) A producer of concrete employs two laboratories to test independently the strength of certain batches of his product. He requires the combined estimate of the concrete's strength and a measure of the variation for each laboratory. (4) Two laboratories independently measure the tensile strength and yielding point of an aluminum casting. The results may be combined to estimate jointly by iteration the common means for tensile strength and yielding point, as well as for each laboratory the variances for tensile strength and yielding point, and also the covariance for tensile strength and yielding point.

1.2 Review of Literature

Cochran (1937) discussed problems of estimation which arise in the analysis of a series of similar experiments. He suggested the method of maximum likelihood with the provision that a successive approximations procedure be used whenever variances must be estimated in addition to the means. Yates and Cochran (1938) gave some practical techniques for the analysis of groups of experiments. Yates (1939), in combining two estimates of a common mean, also used the maximum likelihood method, but proposed a graphical solution for the common mean.

Yates (1940) introduced the idea of recovery of information in the analysis of incomplete block designs and derived formulae for obtaining intrablock and interblock estimates of the same parameter. These estimates were then combined by weighting with the reciprocals of their variances. Since the variances were unknown, estimates of them were used in the formulae. Rao (1947) applied the method of maximum likelihood to the joint density function of the intrablock and interblock estimates and, after substituting estimates of the variances, secured results equivalent to those obtained by Yates. James (1956) used the method of Yates (1940) for combining two independent estimates of a parameter and prepared tables to be used in determining confidence limits for the parameter being estimated.

More recently, the methods of Yates and Rao for combining the intrablock and interblock estimates of a vector of means were compared by Sprott (1956) and extended by Fraser (1957). Also, Zelen (1957) developed a different type of intrablock analysis and obtained two statistically independent F-ratios for significance testing. He then combined the two F-ratios to make a single test of significance.

Except for Cochran's suggestion (1937), the methods mentioned above for combining estimates of a mean use formulae derived under the assumption that the weights are known and the estimates of these weights are substituted into the formulae. In contrast, Neyman and Scott (1948) proposed a procedure for estimating jointly the common mean

and unknown variances when there exists a set of "partially consistent" observations. The observations were called "partially consistent" if the set of unknown parameters for the population were infinite in number and could be split into two parts as follows:
(1) A finite set (called structural parameters) which appear in the probability laws of an infinity of random variables of the sequence $\{x_i\}$. With respect to the structural parameters the sequence $\{x_i\}$ is said to be consistent. (2) An infinite set (called incidental parameters) each of which appears in the probability laws of only a finite number of the random variables considered. Only structural parameters are estimated by Neyman and Scott, and an example is given for the joint estimation of the common mean and unequal variances when several groups of data are combined.

Mantel (1955) has also considered a special case of joint estimation in which each estimate is based upon a single observation. However, the general problem of joint estimation of means and unknown variances has received little attention.

1.3 Objectives of this Dissertation

In this study we shall be concerned with two random samples, each sample arising from a set of independent observations made on a normal population. The two normal populations may be either univariate or multivariate (provided the populations have the same number of variates) and are to be parametrically related in that the means are assumed to be equal. We shall endeavor to estimate jointly the common mean (or means) and all other unknown parameters of the two populations.

We shall first jointly estimate the parameters in the univariate case by the iteration procedure discussed above. Secondly, we shall set up the equations for maximum likelihood estimation and show that the iteration estimates are solutions of the likelihood equations, though they are not necessarily the maximum likelihood estimates. We shall show, however, that the iteration estimates are identical with the estimates obtained by Fisher's Method, provided that (1) the Information Matrix is recalculated at each stage of the successive approximations, and (2) the initial estimates of parameters are equivalent in the two methods. Further, we shall do empirical sampling from two univariate normal populations, using the IBM 650 Computer, and compare the iteration estimates with the maximum likelihood and other estimates. This comparison will indicate that the iteration estimates (which always comprise an unique set) either are identical with the maximum likelihood estimates or differ from them by only a small amount. We shall extend the technique of joint estimation of parameters to the multivariate case by solving simultaneously the likelihood equations by a method of iteration.

We shall also illustrate the calculations involved in obtaining joint estimates of means and variances for (1) a univariate case, Section 2.4, and (2) a bivariate case, Section 4.5.

II. JOINT ESTIMATION OF PARAMETERS IN THE UNIVARIATE CASE BY ITERATION

2.1 Objectives, Definitions, and Notation

In this chapter we shall consider independent random samples from two univariate populations. These populations are assumed to have the same mean, μ, and finite variances which may or may not be equal. We are interested in obtaining joint estimates of the unknown parameters by an iteration procedure.

Let X and Y be random variables with a common mean, μ, and respective variances σ_x^2 and σ_y^2. Let $x_1, x_2, \ldots, x_{n_x}$ and $y_1, y_2, \ldots, y_{n_y}$ be the two sets of independent observations (or measurements) made on these variables. We shall define our sample means and sample variances as follows:

$$(2.1.1) \qquad \bar{x} = \frac{1}{n_x}\sum_{i=1}^{n_x} x_i, \qquad \bar{y} = \frac{1}{n_y}\sum_{j=1}^{n_y} y_j,$$

$$(2.1.2) \qquad s_x^2 = \frac{1}{n_x}\sum_{i=1}^{n_x} (x_i - \bar{x})^2, \qquad s_y^2 = \frac{1}{n_y}\sum_{j=1}^{n_y} (y_j - \bar{y})^2.$$

The choice of the sample size as the divisor in computing the sample variance has been made because of the fact that the iteration procedure is closely related to maximum likelihood estimation as will be shown in Section 3.3. However, in all the mathematical developments to follow, the divisor n may be replaced by n - 1 if so desired.

Now, let m denote the weighted mean which is to be our estimate of the common mean μ. Also, let

$$(2.1.3) \qquad s^2_{x(I)} = \frac{1}{n_x} \sum_{i=1}^{n_x} (x_i - m)^2$$

and

$$(2.1.4) \qquad s^2_{y(I)} = \frac{1}{n_y} \sum_{j=1}^{n_y} (y_j - m)^2$$

be the respective estimates of the variances σ^2_x and σ^2_y. The subscript, I, is affixed to signify that the estimates are obtained through iteration.

2.2 Development of the Iteration Procedure

We have $\bar{x}$ and $\bar{y}$ as independent estimates of μ with $s^2_{x(I)}$ and $s^2_{y(I)}$ as estimates of σ^2_x and σ^2_y, respectively. Hence, $s^2_{x(I)}/n_x$ and $s^2_{y(I)}/n_y$ are respective estimates of the variances of $\bar{x}$ and $\bar{y}$. Consequently, our weighted mean m may be expressed as

$$(2.2.1) \qquad m = \left(\frac{n_x}{s^2_{x(I)}}.\bar{x} + \frac{n_y}{s^2_{y(I)}}.\bar{y}\right) \Bigg/ \left(\frac{n_x}{s^2_{x(I)}} + \frac{n_y}{s^2_{y(I)}}\right),$$

or

$$(2.2.2) \qquad m = (n_x s^2_{y(I)}.\bar{x} + n_y s^2_{x(I)}.\bar{y})/(n_x s^2_{y(I)} + n_y s^2_{x(I)}).$$

We may also write (2.1.3) and (2.1.4) as

(2.2.3) $$s^2_{x(I)} = s^2_x + (\bar{x} - m)^2$$

and

(2.2.4) $$s^2_{y(I)} = s^2_y + (\bar{y} - m)^2.$$

If we now let

(2.2.5) $$w_x = n_x \ s^2_{y(I)}$$

and

(2.2.6) $$w_y = n_y \ s^2_{x(I)} \ ,$$

we may write (2.2.2), by using (2.2.5) and (2.2.6), as

(2.2.7) $$m = (w_x \cdot \bar{x} + w_y \cdot \bar{y})/(w_x + w_y) \ .$$

The estimates m, $s^2_{x(I)}$, and $s^2_{y(I)}$ are, obtained from

(2.2.8) $$m_r = (w_{x:r} \cdot \bar{x} + w_{y:r} \cdot \bar{y})/(w_{x:r} + w_{y:r})$$

for r = 0, 1, 2, ... , by an iteration procedure, where, for initial values, we choose $w_{x:0} = n_x \ s^2_y$ and $w_{y:0} = n_y \ s^2_x$.

These values are substituted in (2.2.8) to give

(2.2.9) $$m_0 = (n_x s^2_y \cdot \bar{x} + n_y s^2_x \cdot \bar{y})/(n_x s^2_y + n_y s^2_x) \ .$$

We may write (2.2.9) in the equivalent form

(2.2.10) $$m_0 = \left(\frac{n_x}{s^2_x} \cdot \bar{x} + \frac{n_y}{s^2_y} \cdot \bar{y}\right)\Big/\left(\frac{n_x}{s^2_x} + \frac{n_y}{s^2_y}\right) ,$$

which depicts clearly the initial value for m mentioned in Section 1.1.

For $r = 1, 2, 3, \ldots$, we take

$$w_{x:r} = n_x [s_y^2 + (\bar{y} - m_{r-1})^2] \tag{2.2.11}$$

and

$$w_{y:r} = n_y [s_x^2 + (\bar{x} - m_{r-1})^2] . \tag{2.2.12}$$

To obtain successive estimates, the value for m_0 , from (2.2.9), is substituted in (2.2.11) and (2.2.12) to get $w_{x:1}$ and $w_{y:1}$. These values, in turn, are substituted in (2.2.8) to get m_1 . The process is continued until the desired accuracy is acquired for the estimates. The iteration estimate m of the common mean μ is the limit of m_r as r tends to infinity. In Section 2.3, we shall show that this limit exists. The iteration estimate, $s^2_{x(I)}$, of σ_x^2 is given by (2.2.3) or it may be determined from (2.2.12) as

$$s^2_{x(I)} = \lim_{r\to\infty} w_{y:r} / n_x . \tag{2.2.13}$$

Similarly, from (2.2.11),

$$s^2_{y(I)} = \lim_{r\to\infty} w_{x:r} / n_y \tag{2.2.14}$$

is the iteration estimate of σ_y^2 .

2.3 Convergence of the Iteration Procedure

To insure that the iteration procedure determines an estimate for μ which is unique, we must prove that the sequence $\{m_r\}$ converges to come constant (which we have arbitrarily called m) as r tends to infinity. First, we note that $w_{x:r}$ and $w_{y:r}$, defined by (2.2.11) and (2.2.12), are positive quantities. From (2.2.8) it is clear that m_r is bounded by $\bar{x}$ and $\bar{y}$. That is, the sequence $\{m_r\}$ as r tends to infinity, is a bounded sequence. Further, for all r, the quantities $(\bar{y} - \bar{x})$, $(\bar{y} - m_r)$, and $(m_r - \bar{x})$ have the same sign. Using (2.2.8), we obtain

$$(2.3.1) \quad m_{r+1} - m_r = (w_{x:r+1}\cdot\bar{x} + w_{y:r+1}\cdot\bar{y})/(w_{x:r+1} + w_{y:r+1})$$
$$- (w_{x:r}\cdot\bar{x} + w_{y:r}\cdot\bar{y})/(w_{x:r} + w_{y:r}) .$$

Combining the two fractions in the right member of (2.3.1), we may write

$$(2.3.2) \quad m_{r+1} - m_r = (\bar{y} - \bar{x})(w_{x:r}\cdot w_{y:r+1} - w_{x:r+1}\cdot w_{y:r})/D_r$$

where $D_r = (w_{x:r} + w_{y:r})(w_{x:r+1} + w_{y:r+1})$ is a positive quantity for all r. Using (2.2.11) and (2.2.12), we may write the numerator in the right member of (2.3.2) as

$$(2.3.3) \quad (\bar{y} - \bar{x})(w_{x:r}\cdot w_{y:r+1} - w_{x:r+1}\cdot w_{y:r})$$
$$= n_x n_y(\bar{y} - \bar{x})\left\{[s_y^2 + (\bar{y} - m_{r-1})^2]\right.$$
$$\cdot [s_x^2 + (\bar{x} - m_r)^2] - [s_y^2 + (\bar{y} - m_r)^2]$$
$$\left.\cdot [s_x^2 + (\bar{x} - m_{r-1})^2]\right\} .$$

For brevity we will write $Q = (\bar{y} - \bar{x})(w_{x:r}\cdot w_{y:r+1}$

$$- w_{x:r+1} \cdot w_{y:r}) \; .$$

By algebraic manipulation, (2.3.3) may be written as

$$(2.3.4) \quad Q = n_x n_y(\bar{y} - \bar{x}) \{ s_x^2 [(\bar{y} - m_{r-1})^2 - (\bar{y} - m_r)^2]$$
$$+ s_y^2 [(\bar{x} - m_r)^2 - (\bar{x} - m_{r-1})^2] + [(\bar{y} - m_{r-1})^2$$
$$\cdot (\bar{x} - m_r)^2 - (\bar{y} - m_r)^2(\bar{x} - m_{r-1})^2] \} \; .$$

Simplication of (2.3.4) gives

$$(2.3.5) \quad Q = n_x n_y(\bar{y} - \bar{x}) \{ s_x^2 [2\bar{y}(m_r - m_{r-1}) - (m_r^2 - m_{r-1}^2)]$$
$$+ s_y^2 [(m_r^2 - m_{r-1}^2) - 2\bar{x}(m_r - m_{r-1})]$$
$$+ (\bar{y} - \bar{x})(m_r - m_{r-1})[(m_r - \bar{x})(\bar{y} - m_{r-1})$$
$$+ (m_{r-1} - \bar{x})(\bar{y} - m_r)] \} \; .$$

We may now factor out the common term $(m_r - m_{r-1})$ in the right member of (2.3.5) and get

$$(2.3.6) \quad Q = n_x n_y(m_r - m_{r-1}) \{ s_x^2 \; (\bar{y} - \bar{x})[(\bar{y} - m_r)$$
$$+ (\bar{y} - m_{r-1})] + s_y^2 \; (\bar{y} - \bar{x})[(m_r - \bar{x})$$
$$+ (m_{r-1} - \bar{x})] + (\bar{y} - \bar{x})^2[(m_r - \bar{x})(\bar{y} - m_{r-1})$$
$$+ (m_{r-1} - \bar{x})(\bar{y} - m_r)] \} \; .$$

Recalling that $(\bar{y} - \bar{x})$, $(\bar{y} - m_r)$, and $(m_r - \bar{x})$ have the same sign for all r, we see that the coefficient of $(m_r - m_{r-1})$ in the right member of (2.3.6) is positive. We may then write (2.3.6) as $Q = (m_r - m_{r-1}) \cdot Q'$, where Q' is the aforementioned positive quantity. Returning to (2.3.2), we now have $m_{r+1} - m_r = (m_r - m_{r-1}) \cdot Q'/D_r$ where both Q' and D_r are positive quantities. Therefore, $(m_{r+1} - m_r)$ has the same sign as $(m_r - m_{r-1})$. This implies that the sequence $\{m_r\}$ as r tends to infinity, is a monotone sequence which increases if $m_1 > m_0$ and decreases if $m_1 < m_0$. Since $\{m_r\}$, as r tends to infinity, is a bounded, monotone sequence, we conclude [see Sokolnikoff (1939), p. 14] that the limit of m_r, as r tends to infinity, exists. We shall call this limit m and say that through the iteration procedure, m_r converges to m.

2.4 A Numerical Example

The Department of Public Works of the City of Louisville regularly produces concrete cylinders which it wishes to test for compressive strength after letting the concrete set for a fixed number of days. The producer assumes that the compressive strength of a particular lot of cylinders is normally distributed with mean μ and variance σ^2. From a certain lot of cylinders produced under fairly identical conditions, two random samples (usually small in size) are taken and sent to different testing laboratories. Each laboratory measures the compressive strength for each cylinder in its sample and computes the mean and variance for the sample. Heretofore, the sample means have been combined to obtain the grand mean which is taken as the estimate for μ. However, it is thought that the estimate of μ may be improved by using the method of this chapter to combine the sample means and variances to obtain jointly an estimate of μ and measures of the precision of each laboratory.

For the data* given in Table 2.1 below, the compressive strength was measured after 28 days to the nearest ten pounds per square inch. The sets of measurements from the two laboratories are denoted by X and Y, respectively, and are given in Table 2.1.

TABLE 2.1

Compressive Strength of Concrete Cylinders

Laboratory X		Laboratory Y	
2510	2560	2290	2450
2620	2640	2730	2990
2480	2790	2790	2910
2830	2650	2350	2570
2700	2730	....	....

---

* Courtesy of Professor A. H. Barnes, Department of Civil Engineering, Speed Scientific School, University of Louisville, Louisville, Kentucky

By using (2.1.1) and (2.1.2) with $n_x = 10$ and $n_y = 8$, we find $\bar{x} = 2651$, $s_x^2 = 11{,}849$, $\bar{y} = 2635$, and $s_y^2 = 59{,}175$. Using (2.2.9), we find the initial value

$$m_0 = \frac{10(59{,}175)(2651) + 8(11{,}849)(2635)}{10(59{,}175) + 8(11{,}849)} = 2{,}648.79.$$

Then, from (2.2.11) and (2.2.12),

$$w_{x:1} = 10[59{,}175 + (2635 - 2{,}648.79)^2] = 593{,}652,$$

$$w_{y:1} = 8[11{,}849 + (2651 - 2{,}648.79)^2] = 94{,}831.$$

Using (2.2.8), the second estimate is found to be

$$m_1 = \frac{593{,}652\ (2651) + 94{,}831\ (2635)}{593{,}652 + 94{,}831} = 2{,}648.80.$$

Continuing the iteration procedure, we find from (2.2.11) and (2.2.12) $w_{x:2} = 593{,}654$ and $w_{y:2} = 94{,}831$. Substitution in (2.2.8) gives $m_2 = 2{,}648.80$. Since m_2 agrees with m_1 to two decimal places, $m = 2{,}648.80$ is taken as the iteration estimate of μ. This value differs from the grand mean of 2,643.89 by approximately five pounds per square inch. From (2.2.3) and (2.2.4), $s^2_{x(I)} = 11{,}854$ and $s^2_{y(I)} = 59{,}365$ are the respective iteration estimates of σ_x^2 and σ_y^2 which are indicative of the precision of each laboratory.

III. JOINT ESTIMATION OF THE PARAMETERS IN THE UNIVARIATE CASE BY MAXIMUM LIKELIHOOD

3.1 Objectives and Assumptions

Let X and Y be two independent random variables normally distributed with a common mean μ and respective variances σ_x^2 and σ_y^2. We shall assume that μ, σ_y^2, and σ_y^2 are unknown and we wish to estimate them by the method of maximum likelihood. We shall set up the likelihood equations and consider four methods for solving them simultaneously. We shall show that the iteration procedure of Section 2.2 yields estimates of the parameters which satisfy the likelihood equations, and we shall show the conditions under which these estimates may differ from the maximum likelihood estimates. In addition, we shall show that, under certain conditions, the estimates obtained by the iteration procedure of Section 2.2 are identical with the estimates obtained by Fisher's Method where the Information Matrix is used.

To accomplish the estimation, we shall make n_x independent observations on the X-variable and n_y independent observations on the Y-variable. The resulting sample means and sample variances have been defined previously by (2.1.1) and (2.1.2).

3.2 Likelihood Equations

We shall call the equations for jointly estimating the unknown parameters by the method of mamimum likelihood the likelihood equations. Let L represent the natural logarithm of the joint likelihood function of X and Y. Thus

$$(3.2.1)\qquad \begin{aligned} L &= L(X, Y; \mu, \sigma_x^2, \sigma_y^2) \\ &= -\tfrac{1}{2}\Big[(n_x + n_y)\log 2\pi + n_x \log \sigma_x^2 + n_y \log \sigma_y^2 \\ &\quad + \sum_{i=1}^{n_x} (x_i - \mu)^2/\sigma_x^2 + \sum_{j=1}^{n_y} (y_j - \mu)^2/\sigma_y^2\Big] . \end{aligned}$$

The likelihood equations result from equating to zero the partial derivative of L with respect to each parameter.

The likelihood equations for jointly estimating the parameters of Section 3.1 are:

$$(3.2.2)\qquad \frac{n_x(\bar{x} - \hat{\mu})}{s_x^2 + (\bar{x} - \hat{\mu})^2} + \frac{n_y(\bar{y} - \hat{\mu})}{s_y^2 + (\bar{y} - \hat{\mu})^2} = 0,$$

$$(3.2.3)\qquad \hat{\sigma}_x^2 = \frac{1}{n_x}\sum_{i=1}^{n_x}(x_i - \hat{\mu})^2 = s_x^2 + (\bar{x} - \hat{\mu})^2,$$

$$(3.2.4)\qquad \hat{\sigma}_y^2 = \frac{1}{n_y}\sum_{j=1}^{n_y}(y_j - \hat{\mu})^2 = s_y^2 + (\bar{y} - \hat{\mu})^2.$$

A single circumflex will be used with any set of estimates which satisfies the likelihood equations. Whereas, a double circumflex will be used with the set of estimates which both satisfies the likelihood equations and maximizes the joint likelihood function.

The maximum likelihood estimates of μ, σ_x^2, and σ_y^2 are $\hat{\mu}$, $\hat{\sigma}_x^2$, and $\hat{\sigma}_y^2$, respectively.

3.3 Solution of the Likelihood Equations

The likelihood equation (3.2.2), when cleared of fractions, is

$$(3.3.1) \qquad n_x(\bar{x} - \hat{\mu})[s_y^2 + (\bar{y} - \hat{\mu})^2] + n_y(\bar{y} - \hat{\mu})[s_x^2 + (\bar{x} - \hat{\mu})^2] = 0,$$

or

$$(3.3.2) \qquad \{n_x [s_y^2 + (\bar{y} - \hat{\mu})^2] + n_y [s_x^2 + (\bar{x} - \hat{\mu})^2]\} \hat{\mu} - n_x [s_y^2 + (\bar{y} - \hat{\mu})^2] \bar{x} - n_y [s_x^2 + (\bar{x} - \hat{\mu})^2] \bar{y} = 0.$$

If (3.3.2) is arranged as a polynomial in $\hat{\mu}$ and then divided by the coefficient of $\hat{\mu}^3$, we obtain an equation equivalent to (3.2.2):

$$(3.3.3) \qquad f(\hat{\mu}) = \hat{\mu}^3 - A\hat{\mu}^2 + B\hat{\mu} - C = 0,$$

where $A = [n_x \bar{x} + n_y \bar{y} + 2(n_x \bar{y} + n_y \bar{x})]/(n_x + n_y)$,

$$B = [n_x(s_y^2 + \bar{y}^2) + n_y(s_x^2 + \bar{x}^2) + 2(n_x + n_y)\bar{x}\,\bar{y}]/(n_x + n_y),$$

$$C = [n_x(s_y^2 + \bar{y}^2)\bar{x} + n_y(s_x^2 + \bar{x}^2)\bar{y}]/(n_x + n_y).$$

Equation (3.3.3) is a cubic equation in $\hat{\mu}$ with real coefficients and has, therefore, at least one real root for $\hat{\mu}$.

Substitution of $\hat{\mu}$ into (3.2.3) and (3.2.4) determines the estimates $\hat{\sigma}_x^2$ and $\hat{\sigma}_y^2$. Consequently, a set of maximum likelihood estimates (not necessarily unique) will always exist.

Various methods have been proposed for securing approximate solutions for $\hat{\mu}$ from the likelihood equations. Yates (1939) suggested a graphical solution. His method consists of writing (3.2.2) in two parts. The first part is a cubic depending only on sample sizes, while the second part is linear and depends on both sample sizes and sample variances. The value of $\hat{\mu}$ is then obtained as the abscissa of the intersection of a straight line and a cubic curve.

A second method for finding $\hat{\mu}$ is to write (3.3.2) in the form

$$(3.3.4) \qquad \hat{\mu} = \frac{n_x \left[s_y^2 + (\bar{y} - \hat{\mu})^2 \right] \bar{x} + n_y \left[s_x^2 + (\bar{x} - \hat{\mu})^2 \right] \bar{y}}{n_x \left[s_y^2 + (\bar{y} - \hat{\mu})^2 \right] + n_y \left[s_x^2 + (\bar{x} - \hat{\mu})^2 \right]}.$$

It is seen that (3.3.4) is identical with (2.2.2) with m replaced by $\hat{\mu}$. Thus, $\hat{\mu}$ may be found by the iteration procedure of Section 2.2, and it follows that the iteration estimates m, $s_{x(I)}^2$, and $s_{y(I)}^2$ constitute a solution for the system of likelihood equations.

A third method of some merit for finding $\hat{\mu}$ is the Newton-Raphson Method [See Kunz (1957), p. 10]. We start with (3.3.3). $\hat{\mu}$ is obtained by successive approximations using

(3.3.5) $\quad a_{r+1} = a_r - f(a_r)/f'(a_r)$ for $r = 0, 1, 2, \ldots$.

Substitution of (3.3.3) into (3.3.5) gives

(3.3.6) $$a_{r+1} = a_r - (a_r^3 - Aa_r^2 + Ba_r - C)/(3a_r^2 - 2Aa_r + B)$$
$$= (2a_r^3 - Aa_r^2 + C)/(3a_r^2 - 2Aa_r + B).$$

A convenient choice for a_o is the m_o defined by (2.2.9).

The Newton-Raphson Method may readily be extended to the joint estimation of several parameters. Let the parameters be denoted by θ_i, $i = 1, 2, \ldots, k$, and set $s_i = \frac{\partial L}{\partial \theta_i}$ where L is defined by (3.2.1). We wish to solve the system of equations, $s_i = 0$, for $\hat{\theta}_i$.

Let $f_{ij} = \frac{\partial^2 L}{\partial \theta_i \partial \theta_j}$ and let F represent the matrix composed of the elements f_{ij}, $i,j = 1, 2, \ldots, k$. If $\theta_{i:0}$ is a set of approximate solutions of $s_i = 0$, used as first approximations to $\hat{\theta}_i$, and f^{ij} is a typical element of F^{-1}, then a set of second approximations are obtained by $\theta_{i:1} = \theta_{i:0} - \left\{\sum_{j=1}^{k} f^{ij} s_j\right\}_{\theta_i = \theta_{i:0}}$. The process may be continued and we would expect the $\theta_{i:r}$ to converge to the $\hat{\theta}_i$ as r tends to infinity. In matrix form, the equation for for the successive approximations is $\theta_{r+1} = \theta_r - [F^{-1}S]_{\theta=\theta_r}$. In practice, the matrix, F, is usually not recalculated at every stage due to the labor involved.

A fourth method, quite general in form, for finding approximate solutions to the likelihood equations, is attributed to Fisher and is illustrated by Stevens (1951) and Rao (1952, P.165). It differs from the Newton-Raphson Method only that the element f_{ij} is replaced by its expected value. Thus, if $h_{ij} = - E(f_{ij})$, we obtain

$$(3.3.7) \qquad \theta_{i:r+1} = \theta_{i:r} + \left\{ \sum_{j=1}^{k} h^{ij} s_j \right\}_{\theta_i = \theta_{i:r}},$$

or in matrix form

$$(3.3.8) \qquad \theta_{r+1} = \theta_r + \left\{ H^{-1} S \right\}_{\theta = \theta_r}.$$

The matrix, H, is commonly called the Information Matrix, and due to the labor involved, is usually not adjusted (i.e. recalculated) at every stage of the approximations. However, if H is adjusted at every stage, and if, in addition, we take $\theta_{i:0}$ equivalent to the initial estimates of the iteration procedure of Section 2.2, we shall show that the two methods produce identical estimates for the parameters of Section 3.1.

To illustrate the Fisher technique, we shall estimate the parameters of Section 3.1. Let

(3.3.9) $\quad \theta_1 = \mu, \quad \theta_2 = \sigma_x^2, \text{ and } \quad \theta_3 = \sigma_y^2.$

The elements of the column matrix, S, in (3.3.8) are $s_i = \dfrac{\partial L}{\partial \theta_i}$, where L is given by (3.2.1). Thus,

(3.3.10) $$s_1 = \frac{\partial L}{\partial \mu} = \frac{n_x\ (\bar{x} - \mu)}{\sigma_x^2} + \frac{n_y\ (\bar{y} - \mu)}{\sigma_y^2}$$

$$= [n_x\theta_3\bar{x} + n_y\theta_2\bar{y} - (n_x\theta_3 + n_y\theta_2)\theta_1]/\theta_2\theta_3,$$

(3.3.11) $$s_2 = \frac{\partial L}{\partial \sigma_x^2} = \frac{n_x}{2\sigma_x^4}[s_x^2 + (\bar{x} - \mu)^2] - \frac{n_x}{2\sigma_x^2}$$

$$= \frac{n_x}{2\theta_2^2}[s_x^2 + (\bar{x} - \theta_1)^2 - \theta_2],$$

and, similarly,

(3.3.12) $$s_3 = \frac{\partial L}{\partial \sigma_y^2} = \frac{n_y}{2\theta_3^2}[s_y^2 + (\bar{y} - \theta_1)^2 - \theta_3].$$

The elements of the matrix, H, are $h_{ij} = -E\left(\dfrac{\partial^2 L}{\partial\theta_i\,\partial\theta_j}\right)$.

Thus, $h_{11} = -E\left(\dfrac{\partial^2 L}{\partial\mu^2}\right) = -E(-n_x\ /\sigma_x^2 - n_y\ /\sigma_y^2)$

$$= n_x\ /\sigma_x^2 + n_y\ /\sigma_y^2 = (n_x\ \theta_3 + n_y\ \theta_2)/\theta_2\ \theta_3,$$

$$h_{22} = - E\left[\frac{\partial^2 L}{\partial(\sigma_x^2)^2}\right] = - E\left\{-\frac{n_x}{\sigma_x^6}\left[s_x^2 + (\bar{x} - \mu)^2\right] + \frac{n_x}{\sigma_x^4}\right\}$$

$$= n_x \ /2\sigma_x^4 = n_x \ /2\ \theta_2^2.$$

Similarly, $h_{33} = n_y \ /2\ \theta_3^2$. Since $E(\bar{x}) = E(\bar{y}) = \mu$, it is clear that $h_{ij} = 0$ for $i \neq j$, and that H is a diagonal matrix. The elements of H^{-1} are readily found to be

(3.3.13) $\quad h^{11} = \theta_2\ \theta_3 \ / \ (n_x\ \theta_3 + n_y\ \theta_2),$

(3.3.14) $\quad h^{22} = 2\ \theta_2^2 \ / \ n_x,$

(3.3.15) $\quad h^{33} = 2\ \theta_3^2 \ / \ n_y,$

(3.3.16) $\quad h^{ij} = 0$ for $i \neq j$.

Substituting (3.3.10) to (3.3.16) into (3.3.7), we obtain

(3.3.17)
$$\theta_{1:r+1} = \theta_{1:r} + \left[n_x\theta_{3:r}\bar{x} + n_y\theta_{2:r}\bar{y} - (n_x\theta_{3:r} + n_y\theta_{2:r})\theta_{1:r}\right]/(n_x\theta_{3:r} + n_y\theta_{2:r})$$
$$= (n_x\theta_{3:r}\bar{x} + n_y\theta_{2:r}\bar{y})/(n_x\theta_{3:r} + n_y\theta_{2:r}),$$

(3.3.18)
$$\theta_{2:r+1} = \theta_{2:r} + s_x^2 + (\bar{x} - \theta_{1:r})^2 - \theta_{2:r}$$
$$= s_x^2 + (\bar{x} - \theta_{1:r})^2,$$

and, similarly,

(3.3.19) $\quad \theta_{3:r+1} = s_y^2 + (\bar{y} - \theta_{1:r})^2.$

If for the first approximations of θ_i, we take $\theta_{1:0} = m_0$

[as defined by (2.2.9)], $\theta_{2:0} = s_x^2$, and $\theta_{3:0} = s_y^2$, then (3.3.17), (3.3.18), and (3.3.19) are equivalent, respectively, to (2.2.8), (2.2.12), and (2.2.11). Therefore, the estimates of the parameters obtained by letting r tend to infinity are identical. The iteration procedure proposed in this dissertation eliminates any need for calculating and inverting the H matrix at each stage of the approximations. In the example shown above, Fisher's Method requires 32 arithmetic operations at each stage without adjusting H, plus 3 additional ones to adjust H and compute a new H^{-1}. The proposed iteration procedure requires only 13 arithmetic operations to produce results identical to the estimates secured after 35 operations by Fisher's Method.

3.4 Comparison of the Iteration Estimates with Maximum Likelihood Estimates

If equation (3.3.3) has a single real root (or three identical roots) for $\hat{\mu}$, then $\hat{\mu} = m = \hat{\hat{\mu}}$. That is, when the likelihood equations have an unique solution, the iteration procedure of Section 2.2 produces the maximum likelihood estimates of the parameters. However, (3.3.3) may have three real, distinct roots. An examination of L in (3.2.1) reveals that as μ tends to either positive or negative infinity, L becomes negatively infinite.

Since L is a continuous function with respect to μ, $\hat{\hat{\mu}}$ must be one of the two extreme values. If the three real, distinct roots are denoted by $\hat{\mu}_a$, $\hat{\mu}_b$, and $\hat{\mu}_c$ where $\hat{\mu}_a < \hat{\mu}_b < \hat{\mu}_c$, then $L(\hat{\mu}_a)$ and $L(\hat{\mu}_c)$ are maxima while $L(\hat{\mu}_b)$ is a relative minimum. Thus, $\hat{\hat{\mu}}$ is equal to either $\hat{\mu}_a$ or $\hat{\mu}_c$. Numerical examples will be given in Appendix Section C to show that the iteration root m may be either $\hat{\mu}_a$, $\hat{\mu}_b$, or $\hat{\mu}_c$.

The conditions under which three real distinct roots exist will be established in Appendix Section A. In Appendix Section B some discussion will be given on the probability of the occurrence of three real distinct roots. These investigations indicate that the probability of obtaining three real distinct roots from the likelihood equation (3.3.3) is small. The results of empirical sampling to be shown in Chapter VI support this contention. They further indicate that, in the cases where three real roots exist, the iteration estimate is usually also the maximum likelihood estimate, and in the cases where the two differ, the difference is relatively small. Consequently, it appears feasible to estimate the parameters by the iteration procedure of Section 2.2 without concern for possible maximum likelihood estimates. In any case, if strict maximum likelihood estimation is desired when three real roots are thought to exist, equation (3.3.3) may be reduced to a quadratic (after m has been found) and readily solved. $\hat{\hat{\mu}}$ may then be easily identified.

IV. JOINT ESTIMATION OF PARAMETERS IN THE BIVARIATE CASE

4.1 Objectives, Definitions, and Assumptions

Let $X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ and $Y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ have independent bivariate normal distributions with a common vector of means, $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$, and respective, nonsingular covariance matrices, Σ_x and Σ_y, which are not parametrically related. Symbolically,

$$\Sigma_x = \begin{pmatrix} \sigma_{x,11} & \sigma_{x,12} \\ \sigma_{x,12} & \sigma_{x,22} \end{pmatrix} = \begin{pmatrix} \sigma^2_{x_1} & \rho_x \sigma_{x_1} \sigma_{x_2} \\ \rho_x \sigma_{x_1} \sigma_{x_2} & \sigma^2_{x_2} \end{pmatrix},$$

(4.1.1)

$$\Sigma_y = \begin{pmatrix} \sigma_{y,11} & \sigma_{y,12} \\ \sigma_{y,12} & \sigma_{y,22} \end{pmatrix} = \begin{pmatrix} \sigma^2_{y_1} & \rho_y \sigma_{y_1} \sigma_{y_2} \\ \rho_y \sigma_{y_1} \sigma_{y_2} & \sigma^2_{y_2} \end{pmatrix}.$$

We shall assume that the parameters μ_1, μ_2, $\sigma^2_{x_1}$, ρ_x, $\sigma^2_{x_2}$, $\sigma^2_{y_1}$, ρ_y, and $\sigma^2_{y_2}$ are unknown and we wish to jointly estimate them.

To accomplish the estimation, we shall make n_x independent observations on X and n_y independent observations on Y, and solve the likelihood equations using an iteration procedure.

For the particular case in which we may assume that $\sigma^2_{x_1} = \sigma^2_{x_2}$ and $\sigma^2_{y_1} = \sigma^2_{y_2}$, a linear transformation on vectors X and Y is made to obtain new vectors whose components are uncorrelated before the iteration procedure is employed.

Following the pattern of Section 2.1, we shall define our sample statistics thus:

(4.1.2) $$\overline{x}_1 = \frac{1}{n_x} \sum_{i=1}^{n_x} x_{1i}, \quad \overline{x}_2 = \frac{1}{n_x} \sum_{i=1}^{n_x} x_{2i} ,$$

(4.1.3) $$\overline{y}_1 = \frac{1}{n_y} \sum_{j=1}^{n_y} y_{1j} , \quad \overline{y}_2 = \frac{1}{n_y} \sum_{j=1}^{n_y} y_{2j} ,$$

(4.1.4) $$s_{x,\alpha B} = \frac{1}{n_x} \sum_{i=1}^{n_x} (x_{\alpha i} - \overline{x}_\alpha)(x_{Bi} - \overline{x}_B) ,$$

(4.1.5) $$s_{y,\alpha B} = \frac{1}{n_y} \sum_{j=1}^{n_y} (y_{\alpha j} - \overline{y}_\alpha)(y_{Bj} - \overline{y}_B) ,$$

where α, B = 1, 2.

We shall let the vector $m = \begin{pmatrix} m_1 \\ m_2 \end{pmatrix}$ be the iteration estimate of the common vector of means, μ. Also, we shall let

(4.1.6) $$s_{x,\alpha B(I)} = s_{x,\alpha B} + (\overline{x}_\alpha - m_\alpha)(\overline{x}_B - m_B)$$

and

(4.1.7) $$s_{y,\alpha B(I)} = s_{y,\alpha B} + (\overline{y}_\alpha - m_\alpha)(\overline{y}_B - m_B) ,$$

where $s_{x,\alpha B(I)}$ and $s_{y,\alpha B(I)}$ are the respective iteration estimates of $\sigma_{x,\alpha B}$ and $\sigma_{y,\alpha B}$.

4.2 Likelihood Equations

Employing matrix notation, we have for the natural logarithm of the joint likelihood of X and Y :

$$(4.2.1)\quad L = L(X, Y; \mu, \Sigma_x, \Sigma_y) = -(n_x + n_y)\log 2\pi - \frac{n_x}{2}\log|\Sigma_x| - \frac{n_y}{2}\log|\Sigma_y| - \frac{1}{2}\sum_{i=1}^{n_x}(x_i - \mu)' \cdot \Sigma_x^{-1}(x_i - \mu) - \frac{1}{2}\sum_{j=1}^{n_y}(y_j - \mu)'\Sigma_y^{-1}(y_j - \mu) .$$

The likelihood equations for estimating jointly the parameters [See Anderson (1958), p. 48 and Mood (1950), p. 186] are:

$$(4.2.2)\quad \hat{\sigma}_{x,\alpha B} = \frac{1}{n_x}\sum_{i=1}^{n_x}(x_{\alpha i} - \hat{\mu}_\alpha)(x_{Bi} - \hat{\mu}_B) ,$$

$$(4.2.3)\quad \hat{\sigma}_{y,\alpha B} = \frac{1}{n_y}\sum_{j=1}^{n_y}(y_{\alpha j} - \hat{\mu}_\alpha)(y_{Bj} - \hat{\mu}_B) ,$$

where α, B = 1, 2, and

$$(4.2.4)\quad n_x [(\bar{x} - \hat{\mu})'\hat{\Sigma}_x^{-1}]' + n_y [(\bar{y} - \hat{\mu})'\hat{\Sigma}_y^{-1}]' = \emptyset ,$$

where $\emptyset$ is a zero column vector, while

$$(4.2.5)\quad \hat{\Sigma}_x^{-1} = \frac{1}{1-\hat{\rho}_x^2}\begin{pmatrix} 1/\hat{\sigma}_{x_1}^2 & -\hat{\rho}_x/\hat{\sigma}_{x_1}\hat{\sigma}_{x_2} \\ -\hat{\rho}_x/\hat{\sigma}_{x_1}\hat{\sigma}_{x_2} & 1/\hat{\sigma}_{x_2}^2 \end{pmatrix},$$

$$(4.2.6)\quad \hat{\rho}_x = \hat{\sigma}_{x,12}/(\hat{\sigma}_{x,11}\hat{\sigma}_{x,22})^{1/2}.$$

A similar relationship holds for $\hat{\Sigma}_y^{-1}$.

The maximum likelihood estimates $\hat{\hat{\mu}}_\alpha$, $\hat{\hat{\sigma}}_{x,\alpha B}$, and $\hat{\hat{\sigma}}_{y,\alpha B}$ will be the set of values $\hat{\mu}_\alpha$, $\hat{\sigma}_{x,\alpha B}$, and $\hat{\sigma}_{y,\alpha B}$, respectively, which maximizes L in (4.2.1).

4.3 Solution of the Likelihood Equations

The likelihood equations (4.2.2) and (4.2.3) may be expressed as follows:

$$\hat{\sigma}_{x,11} = \hat{\sigma}_{x_1}^2 = s_{x,11} + (\bar{x}_1 - \hat{\mu}_1)^2,$$

$$(4.3.1)\qquad \hat{\sigma}_{x,22} = \hat{\sigma}_{x_2}^2 = s_{x,22} + (\bar{x}_2 - \hat{\mu}_2)^2,$$

$$\hat{\sigma}_{x,12} = \hat{\rho}_x\hat{\sigma}_{x_1}\hat{\sigma}_{x_2} = s_{x,12} + (\bar{x}_1 - \hat{\mu}_1)(\bar{x}_2 - \hat{\mu}_2),$$

where $s_{x,11}$, $s_{x,22}$, and $s_{x,12}$ are defined by (4.1.4).

Similar relationships hold for $\hat{\sigma}_{y,11}$, $\hat{\sigma}_{y,22}$, and $\hat{\sigma}_{y,12}$.

Equation (4.2.4) may also be written [See Fraser (1957), p. 815]

(4.3.2) $\hat{\mu} = (n_x \hat{\Sigma}_x^{-1} \cdot \overline{x} + n_y \hat{\Sigma}_y^{-1} \cdot \overline{y})(n_x \hat{\Sigma}_x^{-1} + n_y \hat{\Sigma}_y^{-1})^{-1}$.

If we now substitute in (4.3.2) the values for $\hat{\Sigma}_x^{-1}$ and $\hat{\Sigma}_y^{-1}$ from (4.2.5) and (4.2.6) and then perform the indicated operations, we arrive at the following expressions for $\hat{\mu}_1$ and $\hat{\mu}_2$:

$$(4.3.3) \quad \begin{cases} \hat{\mu}_1 = (w_{x,1} \cdot \overline{x}_1 + w_{y,1} \cdot \overline{y}_1 + w_{xy,1} \cdot d_2)/(w_{x,1} + w_{y,1}), \\ \hat{\mu}_2 = (w_{x,2} \cdot \overline{x}_2 + w_{y,2} \cdot \overline{y}_2 + w_{xy,2} \cdot d_1)/(w_{x,2} + w_{y,2}), \end{cases}$$

where

$$(4.3.4) \qquad d_1 = \overline{y}_1 - \overline{x}_1 \ , \qquad d_2 = \overline{y}_2 - \overline{x}_2 \ ,$$

and

$$(4.3.5) \quad \begin{cases} w_{x,1} = n_x \hat{\sigma}_{y_1} [n_x (1 - \hat{\rho}_y^2) \hat{\sigma}_{y_1} \hat{\sigma}_{y_2}^2 + n_y \hat{\sigma}_{x_2} (\hat{\sigma}_{x_2} \hat{\sigma}_{y_1} \\ \qquad\qquad - \hat{\rho}_x \hat{\rho}_y \hat{\sigma}_{x_1} \hat{\sigma}_{y_2})], \\ w_{xy,1} = n_x n_y \hat{\sigma}_{x_1} \hat{\sigma}_{y_1} (\hat{\rho}_x \hat{\sigma}_{x_2} \hat{\sigma}_{y_1} - \hat{\rho}_y \hat{\sigma}_{x_1} \hat{\sigma}_{y_2}) \ . \end{cases}$$

$w_{x,2}$, $w_{y,1}$, $w_{y,2}$, $w_{xy,2}$ are defined by the corresponding interchanges of subscripts in (4.3.5). It should be noted that $w_{x,1} + w_{y,1} = w_{x,2} + w_{y,2}$.

An extension of the iteration procedure of Section 2.2 may be used to solve, simultaneously, equations (4.3.1) and (4.3.3). In (4.1.6) and (4.1.7) we take $s_{x,\alpha B}$ and $s_{y,\alpha B}$, defined by (4.1.4) and (4.1.5), as our initial values of $s_{x,\alpha B(I)}$ and $s_{y,\alpha B(I)}$, respectively. These values, in turn, give us initial values for $w_{x,1}$, etc., in (4.3.5) by using (4.3.1). Thus, $m_{1:r}$ and $m_{2:r}$ are obtained from (4.3.3). The sequences $\{m_{1:r}\}$ and $\{m_{2:r}\}$, as r tends to infinity, may be expected to converge to the values $m_1 = \hat{\mu}_1$ and $m_2 = \hat{\mu}_2$ which satisfy equations (4.2.3). Substitutions of m_1 and m_2 for $\hat{\mu}_1$ and $\hat{\mu}_2$ in (4.3.1) supplies values for $\hat{\sigma}^2_{x_1}$, $\hat{\sigma}^2_{x_2}$, $\hat{\rho}_x$, $\hat{\sigma}^2_{y_1}$, $\hat{\sigma}^2_{y_2}$, and $\hat{\rho}_y$.

4.4 Use of a Non-Singular Linear Transformation

Rather than iterate at the stage mentioned in Section 4.3, it appears desirable to first transform the variables X and Y in such a manner as to obtain new uncorrelated variables. We would like to make an orthogonal linear transformation, such as, $U = T X$ and $V = T Y$, which produces diagonal covariance matrices for Σ_u and Σ_v. For one thing, this requires that $\Sigma_u = T \Sigma_x T'$. This equation is satisfied by $T = \begin{pmatrix} 1 & t \\ t & -1 \end{pmatrix}$, where $t = q + (q^2 + 1)^{1/2}$ and $q = (\sigma^2_{x_1} - \sigma^2_{x_2})/2\rho_x\sigma_{x_1}\sigma_{x_2}$. Since we must also have

$\Sigma_v = T\,\Sigma_y\,T'$, the transformation appears feasible only when $\sigma^2_{x_1} = \sigma^2_{x_2}$ and $\sigma^2_{y_1} = \sigma^2_{y_2}$. For this case, $T = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$.

The assumption of equal variances, though restrictive, is sometimes justified. For example, aptitude tests, commonly administered to engineering school applicants, show equal variances in the scores on Mathematics and Physics. We wish, perhaps, to combine results of two different tests to obtain jointly estimates of the Mathematics and Physics aptitude scores along with their associated variances.

For the particular case in which we may assume $\sigma^2_{x_1} = \sigma^2_{x_2}$ and $\sigma^2_{y_1} = \sigma^2_{y_2}$, differentiation of (4.2.1) with respect to the parameters to be estimated yields equation (4.2.4) for $\hat{\mu}$, plus the following:

$$(4.4.1)\quad \begin{cases} \hat{\sigma}_{x,11} = \hat{\sigma}^2_x = (1/2)[s_{x,11} + (\bar{x}_1 - \hat{\mu}_1)^2 \\ \qquad\qquad + s_{x,22} + (\bar{x}_2 - \hat{\mu}_2)^2], \\ \hat{\sigma}_{x,12} = \hat{\rho}_x\hat{\sigma}^2_x = s_{x,12} + (\bar{x}_1 - \hat{\mu}_1)(\bar{x}_2 - \hat{\mu}_2), \\ \hat{\sigma}_{y,11} = \hat{\sigma}^2_y = (1/2)[s_{y,11} + (\bar{y}_1 - \hat{\mu}_1)^2 \\ \qquad\qquad + s_{y,22} + (\bar{y}_2 - \hat{\mu}_2)^2], \\ \hat{\sigma}_{y,12} = \hat{\rho}_y\hat{\sigma}^2_y = s_{y,12} + (\bar{y}_1 - \hat{\mu}_1)(\bar{y}_2 - \hat{\mu}_2), \end{cases}$$

where $s_{x,11}$, $s_{x,12}$, $s_{y,11}$, and $s_{y,12}$ are defined by (4.1.4) and (4.1.5).

Equation (4.2.4) may be solved (as before) to obtain $\hat{\mu}_1$ and $\hat{\mu}_2$ as given by (4.3.3). However, equations (4.3.5) may now be given the simpler form:

$$
(4.4.2)\quad \begin{cases}
w_{x,1} = w_{x,2} = w_x = n_x\hat{\sigma}_y^2[n_x\hat{\sigma}_y^2 + n_y\hat{\sigma}_x^2 \\
\qquad\qquad - \hat{\rho}_y(n_x\hat{\rho}_y\hat{\sigma}_y^2 + n_y\hat{\rho}_x\hat{\sigma}_x^2)], \\
w_{xy,1} = w_{xy,2} = w_{xy} = n_xn_y\hat{\sigma}_x^2\hat{\sigma}_y^2(\hat{\rho}_x - \hat{\rho}_y), \\
w_{y,1} = w_{y,2} = w_y = n_y\hat{\sigma}_x^2[n_y\hat{\sigma}_x^2 + n_x\hat{\sigma}_y^2 \\
\qquad\qquad - \hat{\rho}_x(n_y\hat{\rho}_x\hat{\sigma}_x^2 + n_x\hat{\rho}_y\hat{\sigma}_y^2)].
\end{cases}
$$

We note that

$$
\begin{aligned}
w_{xy}/(w_x + w_y) &= [n_xn_y\hat{\sigma}_x^2\hat{\sigma}_y^2(\hat{\rho}_x - \hat{\rho}_y)]/[(n_x\hat{\sigma}_y^2 + n_y\hat{\sigma}_x^2)^2 \\
&\quad - (n_x\hat{\rho}_y\hat{\sigma}_y^2 + n_y\hat{\rho}_x\hat{\sigma}_x^2)^2] \\
&= [n_xn_y\hat{\sigma}_x^2\hat{\sigma}_y^2(\hat{\rho}_x - \hat{\rho}_y)]/[(n_x\hat{\sigma}_y^2 + n_y\hat{\sigma}_x^2 \\
&\quad + n_x\hat{\rho}_y\hat{\sigma}_y^2 + n_y\hat{\rho}_x\hat{\sigma}_x^2)(n_x\hat{\sigma}_y^2 + n_y\hat{\sigma}_x^2 \\
&\quad - n_x\hat{\rho}_y\hat{\sigma}_y^2 - n_y\hat{\rho}_x\hat{\sigma}_x^2)] \\
&= (1/4)[n_y\hat{\sigma}_x^2(1 + \hat{\rho}_x) - n_x\hat{\sigma}_y^2(1 + \hat{\rho}_y)] \\
&\quad /[n_y\hat{\sigma}_x^2(1 + \hat{\rho}_x) + n_x\hat{\sigma}_y^2(1 + \hat{\rho}_y)]
\end{aligned}
$$

$$+ (1/4)[n_x\hat{\sigma}_y^2(1 - \hat{\rho}_y) - n_y\hat{\sigma}_x^2(1 - \hat{\rho}_x)]$$

$$/[n_x\hat{\sigma}_y^2(1 - \hat{\rho}_y) + n_y\hat{\sigma}_x^2(1 - \hat{\rho}_x)].$$

Thus, we may write $w_{xy}/(w_x + w_y) = (1/4)(F_1 + F_2)$, where F_1 and F_2 are defined by their ordered parts in the preceding bracketed expression. We see that

$$F_1 = 1 - 2\, n_y\hat{\sigma}_x^2(1 + \hat{\rho}_x)/[n_y\hat{\sigma}_x^2(1 + \hat{\rho}_x) + n_x\hat{\sigma}_y^2(1 + \hat{\rho}_y)] \leq 1,$$

and also

$$F_1 = -1 + 2\, n_y\hat{\sigma}_x^2(1+\hat{\rho}_x)/[n_y\hat{\sigma}_x^2(1+\hat{\rho}_x) + n_x\hat{\sigma}_y^2(1+\hat{\rho}_y)] \geq -1 .$$

Hence, $-1 \leq F_1 \leq 1$. Similarly, we get $-1 \leq F_2 \leq 1$. Therefore, $\left| w_{xy}/(w_x + w_y) \right| \leq 1/2$. Consequently, $\left| w_{xy} \cdot d_2/(w_x + w_y) \right| \leq 1/2 \left| d_2 \right|$. Likewise, $\left| w_{xy} \cdot d_1/(w_x + w_y) \right| \leq 1/2 \left| d_1 \right|$. If we put $q_1 = (\bar{x}_1 + \bar{y}_1)/2$ and $q_2 = (\bar{x}_2 + \bar{y}_2)/2$, we have

$\left| (w_x \cdot \bar{x}_1 + w_y \cdot \bar{y}_1)/(w_x + w_y) - q_1 \right| \leq 1/2 \left| d_1 \right|$ and

$\left| (w_x \cdot \bar{x}_2 + w_y \cdot \bar{y}_2)/(w_x + w_y) - q_2 \right| \leq 1/2 \left| d_2 \right|$.

Therefore, $\left| \hat{\mu}_1 - q_1 \right| \leq 1/2 (\left| d_1 \right| + \left| d_2 \right|)$ and $\left| \hat{\mu}_2 - q_2 \right| \leq 1/2 (\left| d_1 \right| + \left| d_2 \right|)$. Consequently, values obtained for $\hat{\mu}_1$ and $\hat{\mu}_2$ from equations (4.3.3) are bounded by $q_1 \pm 1/2 (|d_1| + |d_2|)$ and $q_2 \pm 1/2 (|d_1| + |d_2|)$.

That is, $\hat{\mu}_1$ satisfies the inequality $\min (\bar{x}_1, \bar{y}_1) - (1/2) |\bar{y}_2 - \bar{x}_2| \leq \hat{\mu}_1 \leq \max (\bar{x}_1, \bar{y}_1) + (1/2) |\bar{y}_2 - \bar{x}_2|$.

Similarly, for $\hat{\mu}_2$.

If we now proceed with the iteration of Section 4.3, $\{m_{1:r}\}$ and $\{m_{2:r}\}$ are bounded sequences as r tends to infinity, and we would expect the sequences to converge to the values $m_1 = \hat{\mu}_1$ and $m_2 = \hat{\mu}_2$ which satisfy equations (4.2.4).

If we desire a transformation of the variables X and Y before iteration, we take

$$(4.4.3) \quad U = T X, \; V = T Y, \; \mu^* = T \mu, \; T = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Then,

$$(4.4.4) \quad \begin{cases} u_1 = x_1 + x_2, & u_2 = x_1 - x_2; \\ v_1 = y_1 + y_2, & v_2 = y_1 - y_2. \end{cases}$$

$$(4.4.5) \quad \begin{cases} \mu_1^* = \mu_1 + \mu_2, & \mu_2^* = \mu_1 - \mu_2, \\ \Sigma_u = \begin{pmatrix} \sigma_{u_1}^2 & 0 \\ 0 & \sigma_{u_2}^2 \end{pmatrix}, & \Sigma_v = \begin{pmatrix} \sigma_{v_1}^2 & 0 \\ 0 & \sigma_{v_2}^2 \end{pmatrix}, \end{cases}$$

where

$$\sigma_{u_1}^2 = 2\,\sigma_x^2(1 + \rho_x)\ , \quad \sigma_{u_2}^2 = 2\,\sigma_x^2(1 - \rho_x)\ ,$$

$$\sigma_{v_1}^2 = 2\,\sigma_y^2(1 + \rho_y)\ , \quad \sigma_{v_2}^2 = 2\,\sigma_y^2(1 - \rho_y)\ .$$

The likelihood equations for the means give:

$$(4.4.6) \quad \begin{cases} \mu_1^* = (w_{u_1} \cdot \bar{u}_1 + w_{v_1} \cdot \bar{v}_1)/(w_{u_1} + w_{v_2})\ , \\ \mu_2^* = (w_{u_2} \cdot \bar{u}_2 + w_{v_2} \cdot \bar{v}_2)/(w_{u_2} + w_{v_2})\ , \end{cases}$$

where $w_{u_1} = n_x\sigma_{v_1}^2$, $w_{u_2} = n_x\sigma_{v_2}^2$, $w_{v_1} = n_y\sigma_{u_1}^2$,

and $w_{v_2} = n_y\sigma_{u_2}^2$.

We may now use the iteration procedure of Section 2.2 to find $\hat{\mu}_1^*$ and $\hat{\mu}_2^*$. We shall show in Section 5.3 that the iteration procedure is invarient under a linear, orthogonal transformation. Thus, the values secured for $\hat{\mu}_1$ and $\hat{\mu}_2$ by the inverse transformation $\hat{\mu} = T^{-1}\hat{\mu}^*$ will be identical with those secured by using an iteration procedure on equations (4.3.3). Consequently, $m_1 = (1/2)(m_1^* + m_2^*)$ and $m_2 = (1/2)(m_1^* - m_2^*)$.

The iteration procedure, both with and without a transformation will be illustrated by a numerical example in Section 4.5.

4.5 A Numerical Example

A manufacturer produces a grade of steel using 0.2% carbon and rolling while hot. Usually, about 10 cylindrical pieces of steel are randomly selected from the same lot and a sample of 5 pieces submitted to each of two different testing laboratories. Each laboratory measures the yield point and ultimate strength for each piece in its sample and returns this information to the producer.

The steel manufacturer assumes that the yield point and ultimate strength of the steel have a bivariate normal distribution with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2, respectively. He wishes to combine the data from the testing laboratories to estimate the means. At the same time, the manufacturer would like to obtain measures of the precision of each laboratory with respect to the variables concerned. The joint estimation procedure of this chapter can be used.

For the data* given in Table 4.1 below, the sets of measurements (in multiples of 100 pounds per square inch of cross-sectional area) of the two laboratories are denoted as follows:

---

* Courtesy of Professor A. H. Barnes, Department of Civil Engineering, Speed Scientific School, University of Louisville, Louisville, Kentucky

First laboratory: X_1 - yield point, X_2 - ultimate strength: Second laboratory; Y_1 - yield point, Y_2 - ultimate strength.

TABLE 4.1

Yield Point and Tensile Strength of Steel

Laboratory X		Laboratory Y	
x_1	x_2	y_1	y_2
325	575	356	618
371	622	318	568
364	602	332	598
338	584	400	652
352	592	384	634

To illustrate the estimation technique of Sections 4.3 and 4.4, we shall consider two possible cases with different assumptions.

Case 1: Assume (as in Section 4.3) that the following unknown parameters are to be estimated jointly: μ_1, μ_2, $\sigma^2_{x_1}$, ρ_x, $\sigma^2_{x_2}$, $\sigma^2_{y_1}$, ρ_y, and $\sigma^2_{y_2}$.

To facilitate computation for the iteration procedure, we shall rewrite equations (4.5.1) in the following form:

$$(4.5.1) \qquad w_{x,1} = n_x^2 \left[\hat{\sigma}^2_{y_1}\hat{\sigma}^2_{y_2} - (\hat{\rho}_y\hat{\sigma}_{y_1}\hat{\sigma}_{y_2})^2\right]$$

$$+ n_x n_y \left[\hat{\sigma}^2_{y_1}\hat{\sigma}^2_{x_2} - (\hat{\rho}_x\hat{\sigma}_{x_1}\hat{\sigma}_{x_2})(\hat{\rho}_y\hat{\sigma}_{y_1}\hat{\sigma}_{y_2})\right],$$

$$(4.5.2)\quad w_{xy,1} = n_x n_y \left[\hat{\sigma}^2_{y_1}(\hat{\rho}_x \hat{\sigma}_{x_1}\hat{\sigma}_{x_2}) - \hat{\sigma}^2_{x_1}(\hat{\rho}_y \hat{\sigma}_{y_1}\hat{\sigma}_{y_2})\right].$$

$w_{x,2}$, $w_{y,1}$, $w_{y,2}$, and $w_{xy,2}$ are defined by corresponding interchanges of subscripts. We note that $w_{x,1} + w_{y,1} = w_{x,2} + w_{y,2}$. We shall designate our iteration estimates of μ_1, μ_2, $w_{x,1}$, and $w_{xy,1}$ as m_1, m_2, $w_{x,1(I)}$, and $w_{xy,1(I)}$ with $m_{1:0}$, $m_{2:0}$, $w_{x,1(0)}$, and $w_{xy,1(0)}$ as their respective initial values. An examination of (4.3.1), (4.5.1), and (4.5.2) reveals that $w_{x,1}$ and $w_{xy,1}$ are functions of m_r, the value of the vector m at the r-th iteration, and we shall emphasize this relationship by using $w_{x,1}(m_r)$ as the value of $w_{x,1}$ at the r-th iteration. $w_{x,2}$, $w_{y,1}$, $w_{y,2}$, and $w_{xy,2}$ are to be treated in like manner. We shall also modify (4.1.6) to read

$$(4.5.3)\quad s_{x,\alpha B}(m_r) = s_{x,\alpha B} + (\bar{x}_\alpha - m_{\alpha:r})(\bar{x}_B - m_{B:r}),$$

where $\lim\limits_{r\to\infty} s_{x,\alpha B}(m_r) = s_{x,\alpha B(I)} = \sigma_{x,\alpha B}$.

Similarly, we modify $s_{y,\alpha B}(m_r)$ in (4.1.7).

Applying (4.3.1) and (4.5.3), we obtain from (4.5.1) and (4.5.2)

$$(4.5.4)\quad w_{x,1}(m_r) = n_x^2 \left[s_{y,11}(m_{1:r})\cdot s_{y,22}(m_{2:4}) - s^2_{y,12}(m_{1:r}, m_{2:4})\right]$$

$$+ n_x n_y \left[s_{x,22}(m_{2:r}) \cdot s_{y,11}(m_{1:r}) - s_{x,12}(m_{1:r}, m_{2:r}) \cdot s_{y,12}(m_{1:r}, m_{2:r}) \right],$$

$$(4.5.5) \quad w_{xy,1}(m_r) = n_x n_y \left[s_{x,12}(m_{1:r}, m_{2:r}) \cdot s_{y,11}(m_{1:r}) - s_{x,11}(m_{1:r}) \cdot s_{y,12}(m_{1:r}, m_{2:r}) \right].$$

From (4.3.3) and (4.3.4), we have

$$(4.5.6) \quad m_{1:r+1} = \left[w_{x,1}(m_r) \cdot \bar{x}_1 + w_{y,1}(m_r) \cdot \bar{y}_1 + w_{xy,1}(m_r) \cdot (\bar{y}_2 - \bar{x}_2) \right] / \left[w_{x,1}(m_r) + w_{y,1}(m_r) \right].$$

An interchange of subscripts 1 and 2 produces $m_{2:r}$.

The initial values for $w_{x,1(0)}$, $w_{y,1(0)}$, $w_{xy,1(0)}$, $w_{x,2(0)}$, $w_{y,2(0)}$, and $w_{xy,2(0)}$ are obtained by using $s_{x,\alpha B}$ and $s_{y,\alpha B}$, defined by (4.1.4) and (4.1.5), for $s_{x,\alpha B}(m_r)$ and $s_{y,\alpha B}(m_r)$, respectively. Thus,

$$(4.5.7) \quad w_{x,1(0)} = n_x^2 (s_{y,11} \cdot s_{y,22} - s_{y,12}^2) + n_x n_y (s_{x,22} \cdot s_{y,11} - s_{x,12} \cdot s_{y,12}).$$

The remaining initial weights are obtained similarly.

In determining the weights, w_x and w_y, it is well to note that each weight is a sum of two terms, each term of which is repeated in the remaining weight.

For the data given in Table 4.1 above, the following statistics are found:

By (4.1.2) and (4.1.3), with $n_x = n_y = 5$, $\bar{x}_1 = 350\text{x}10^2$, $\bar{y}_1 = 358\text{x}10^2$, $\bar{x}_2 = 595\text{x}10^2$, and $\bar{y}_2 = 614\text{x}10^2$. The sample variances and covariances obtained by using (4.1.4) and (4.1.5) are $s_{x,11} = 282.0\text{x}10^4$, $s_{x,22} = 261.6\text{x}10^4$, $s_{x,12} = 258.2\text{x}10^4$, $s_{y,11} = 944.0\text{x}10^4$, $s_{y,22} = 846.4\text{x}10^4$, and $s_{y,12} = 872.8\text{x}10^4$. By (4.5.7), the initial weight for $w_{x,1}$ is

$$w_{x,1(0)} = 25\text{x}10^8[(944.0)(846.4) - (872.8)^2 + (261.6)(944.0) - (258.2)(872.8)] = 25\text{x}10^8\text{x}58{,}815.20 \;.$$

In a similar manner, we find $w_{y,1(0)} = 25\text{x}10^8\text{x}20{,}431.80$, $w_{xy,1(0)} = -\ 25\text{x}10^8\text{x}2{,}388.80$, $w_{xy,2(0)} = -\ 25\text{x}10^8\text{x}9{,}784.00$, $w_{x,2(0)} = 25\text{x}10^8\text{x}50{,}549.60$, and $w_{y,2(0)} = 25\text{x}10^8\text{x}28{,}697.40$.

From (4.5.6), the initial estimates of the means are

$$m_{1:0} = 10^2[58{,}815.20(350) + 20{,}431.80(358) - 2{,}388.80(19)] /(58{,}815.20 + 20{,}431.80) = 351.49\text{x}10^2,$$

$$m_{2:0} = 10^2[50{,}549.60(595) + 28{,}697.40(614) - 9{,}784.00(8)] /(50{,}549.60 + 28{,}697.40) = 600.89\text{x}10^2.$$

We now substitute $m_{1:0}$ and $m_{2:0}$ into (4.5.3) to get

$s_{x,\alpha B}(m_0)$, which are, in turn, used in (4.5.4) and (4.5.5) to obtain new weights. For example, from (4.5.3)

$$s_{x,11}(m_{1:0}) = 10^4[282.0 + (350 - 351.49)^2]$$

$$= 10^4(284.2201),$$

$$s_{x,12}(m_{1:0}, m_{2:0}) = 10^4[258.2 + (350-351.49)(595-600.89)]$$

$$= 10^4(266.9761),$$

$$s_{x,22}(m_{2:0}) = 10^4[261.6 + (595 - 600.89)^2]$$

$$= 10^4(296.2921),$$

$$s_{y,11}(m_{1:0}) = 10^4[944.0 + (358 - 351.49)^2]$$

$$= 10^4(986.3801),$$

$$s_{y,12}(m_{1:0}, m_{2:0}) = 10^4[872.8 + (358-351.49)(614-600.89)]$$

$$= 10^4(958.1461),$$

$$s_{y,22}(m_{2:0}) = 10^4[846.4 + (614 - 600.89)^2]$$

$$= 10^4(1018.2721).$$

Now using (4.5.4),

$$w_{x,1}(m_0) = 25\text{x}10^8[(986.3801)(1018.2721) - (958.1461)^2$$
$$+ (296.2921)(986.3801) - (266.9761)(958.1461)]$$
$$= 25\text{x}10^8(122,797.38).$$

Similarly, we find $w_{y,1}(m_0) = 25\text{x}10^8\text{x}46,539.03$,

$w_{xy,1}(m_0) = -\ 25x10^8x8{,}981.66$, $w_{xy,2}(m_0) = -\ 25x10^8$ $x12{,}032.54$, $w_{x,2}(m_0) = 25x10^8x229{,}955.55$, and $w_{y,2}(m_0) = 25x10^8x49{,}380.86$.

Now, (4.5.6) yields the new estimates of the means:

$$m_{1:1} = 10^2[122{,}797.38(350) + 46{,}539.03(358) - 8{,}981.66(19)]$$
$$/(169{,}336.41) = 351.19x10^2,$$

$$m_{2:1} = 10^2[119{,}955.55(595) + 49{,}380.86(614) - 12{,}032.54(8)]$$
$$/(169{,}336.41) = 599.97x10^2.$$

We continue the iteration procedure and find the next stage estimates to be:

$w_{x,1}(m_1) = 25x10^8x123{,}319.20$, $w_{y,1}(m_1) = 25x10^8x\ 51{,}315.85$,

$w_{x,2}(m_1) = 25x10^8x135{,}448.48$, $w_{y,2}(m_1) = 25x10^8x39{,}186.57$,

$w_{xy,1}(m_1) = -25x10^8x12{,}877.66$, $w_{xy,2}(m_1) = -25x10^8x1{,}705.63$,

and $m_{1:2} = 350.95x10^2$, $m_{2:2} = 599.19x10^2$.

Further iteration gives:

$w_{x,1}(m_2) = 25x10^8x125{,}282.82$, $w_{y,1}(m_2) = 25x10^8x55{,}528.94$,

$w_{x,2}(m_2) = 25x10^8x149{,}379.38$, $w_{y,2}(m_2) = 25x10^8x31{,}432.38$,

$w_{xy,1}(m_2) = -25x10^8x15{,}924.44$, $w_{xy,2}(m_2) = 25x10^8x6{,}617.77$,

and $m_{1:3} = 350.78x10^2$, $m_{2:3} = 598.60x10^2$.

Since the change in m_1 and m_2 through successive iteration has been small, we shall conclude with:

$w_{x,1}(m_3) = 25\text{x}10^8\text{x}127{,}796.52$, $w_{y,1}(m_3) = 25\text{x}10^8\text{x}58{,}860.84$,

$w_{x,2}(m_3) = 25\text{x}10^8\text{x}160{,}523.96$, $w_{y,2}(m_3) = 25\text{x}10^8\text{x}26{,}133.40$,

$w_{xy,1}(m_3) = -25\text{x}10^8\text{x}18{,}085.52$, $w_{xy,2}(m_3) = 25\text{x}10^8\text{x}12{,}655.70$,

and $m_{1:4} = 350.68\text{x}10^2$, $m_{2:4} = 598.20\text{x}10^2$.

If we assume that the sequence for $m_{1:r}$ (less the factor 10^2), consisting of 351.49, 351.19, 350.95, 350.78, 350.68, has approached its limit m_1, sufficiently close, then, we may accept $350.7\text{x}10^2$ as the iteration estimate of μ_1. Likewise, we may take $598.2\text{x}10^2$ as our estimate of μ_2.

Returning to (4.1.6) and (4.3.1), we find:

$$\hat{\sigma}^2_{x_1} = s_{x,11(1)} = [282.0 + (350 - 350.7)^2]\text{x}10^4$$

$$= 282.5\text{x}10^4. \text{ Similarly,}$$

$\hat{\sigma}^2_{x_2} = s_{x,22(1)} = 271.8\text{x}10^4$, $\hat{\rho}_x\hat{\sigma}_{x_1}\hat{\sigma}_{x_2} = s_{x,12(1)} = 260.4\text{x}10^4$,

$\hat{\sigma}^2_{y_1} = 997.6\text{x}10^4$, $\hat{\sigma}^2_{y_2} = 1{,}096.0\text{x}10^4$, and $\hat{\rho}_y\hat{\sigma}_{y_1}\hat{\sigma}_{y_2} = 988.5\text{x}10^4$.

In addition, $\hat{\rho}_x = 0.94$, and $\hat{\rho}_y = 0.95$.

Case 2: Assume (as in Section 4.4) that $\sigma_{x_1}^2 = \sigma_{x_2}^2 = \sigma_x^2$ and $\sigma_{y_1}^2 = \sigma_{y_2}^2 = \sigma_y^2$. We wish to estimate jointly μ_1, μ_2, σ_x^2, ρ_x, σ_y^2, and ρ_y.

First Method: We shall first estimate the parameters without a transformation of variables by applying (4.2.4) and (4.4.2).

From (4.4.1) and (4.4.2), we get

$$(4.5.8)\quad w_{x:r} = n_x^2 \ (1/4)[s_{y,11}(m_{1:r}) + s_{y,22}(m_{2:r})]^2 - s_{y,12}^2(m_{1:r}, m_{2:r}) \ + n_x n_y \ (1/4) \cdot [s_{x,11}(m_{1:r}) + s_{x,22}(m_{2:r})][s_{y,11}(m_{1:r}) + s_{y,22}(m_{2:r})] - s_{x,12}(m_{1:r}, m_{2:r}) \cdot s_{y,12}(m_{1:r}, m_{2:r}) \ ,$$

Where $s_{x,\alpha B}(m_r)$ is defined by (4.5.3). $w_{y:r}$ is obtained by interchanging x and y.

$$(4.5.9)\quad w_{xy:r} = (1/2)n_x n_y \ [s_{y,11}(m_{1:r}) + s_{y,22}(m_{2:r})] \cdot [s_{x,12}(m_{1:r}, m_{2:r})] - [s_{x,11}(m_{1:r}) + s_{x,22}(m_{2:r})][s_{y,12}(m_{1:r}, m_{2:r})] \ .$$

As in Section 4.3 we use $s_{x,\alpha B}$ and $s_{y,\alpha B}$ as initial estimates of $s_{x,\alpha B}(m_r)$ and $s_{y,\alpha B}(m_r)$, respectively, and denote the resulting initial weights as $w_{x(0)}$, $w_{y(0)}$, and $w_{xy(0)}$.

From case 1, above, $\bar{x}_1 = 350x10^2$, $\bar{y}_1 = 358x10^2$, $\bar{x}_2 = 595x10^2$, $\bar{y}_2 = 614x10^2$, $s_{x,11} = 282.0x10^4$, $s_{x,22} = 261.6x10^4$, $s_{x,12} = 258.2x10^4$, $s_{y,11} = 944.0x10^4$, $s_{y,22} = 846.4x10^4$, and $s_{y,12} = 872.8x10^4$.

Thus, from (4.5.8) and (4.5.9) with initial values for $s_{x,\alpha B}$, we find:

$$w_{x(0)} = 25x10^8[(1/4)(944.0 + 846.4)^2 - (872.8)^2$$
$$+ (1/4)(282.0 + 261.6)(944.0 + 846.4)$$
$$- (258.2)(872.8)] = 25x10^8x57,561.60.$$

Similarly, $w_{y(0)} = 25x10^8x25,166.40$ and $w_{xy(0)} = - 25x10^8x6,086.40$. Now, (4.3.3) yields:

$$m_{1:0} = 10^2[57,561.60(350) + 25,166.40(358)-6,086.40(19)]$$
$$/(82,728.00) = 351.0358x10^2,$$
$$m_{2:0} = 10^2[57,561.60(595) + 25,166.40(614)-6,086.40(8)]$$
$$/(82,728.00) = 600.1914x10^2.$$

We next substitute $m_{1:0}$ and $m_{2:0}$ into (4.5.3) to get the $s_{x,\alpha B}(m_0)$, which are, in turn, used in (4.5.8) and (4.5.9) to obtain $w_{x:0} = 25x10^8x125,574.5436$, $w_{y:0} = 25x10^8$ $x46,860.5339$, and $w_{xy:0} = - 25x10^8x9,462.9587$. Four decimal places are carried in order to compare more easily the results of the "first" and "second" methods.

From (4.3.3), we find:

$m_{1:1} = 10^2(60,547,365.1809)/(172,435.0775) = 351.1314x10^2$,

$m_{2:1} = 10^2(103,413,517.5870)/(172,435.0775) = 599.7244x10^2$.

We continue the iteration procedure and find the next stage estimates to be:

$w_{x:1} = 25x10^8x132,844.5224$, $w_{y:1} = 25x10^8x44,575.4568$,

$w_{xy:k} = - 25x10^8x6,337.3391$, and $m_{1:2} = 351.3313x10^2$,

$m_{2:2} = 599.4878x10^2$.

Further iteration gives:

$w_{x:2} = 25x10^8x138,959.9892$, $w_{y:2} = 25x10^8x43,210.8727$,

$w_{xy:2} = - 25x10^8x3,972.8942$, and $m_{1:3} = 351.4832x10^2$,

$m_{2:3} = 599.3323x10^2$.

If we assume that the sequence for $m_{1:r}$ (less the factor of 10^2), consisting of 351.0358, 351.1314, 351.3313, 351.4832 has approached its limit m_1, sufficiently close, then, we may accept $351.48x10^2$ as the iteration estimate of

μ_1. Likewise, we may take 599.33×10^2 as our estimate of μ_2.

Using (4.4.1), we find:

$\hat{\sigma}_x^2 = 282.284\times10^4$, $\hat{\rho}_x\hat{\sigma}_x^2 = 264.63\times10^4$, $\hat{\rho}_x = 0.94$,

$\hat{\sigma}_y^2 = 1{,}024.00\times10^4$, $\hat{\rho}_y\hat{\sigma}_y^2 = 968.39\times10^4$, $\hat{\rho}_y = 0.95$.

Second Method: Use the linear transformation of (4.4.3).

By (4.4.3), $U = T\,X$, $V = T\,Y$, and $\mu^* = T\,\mu$,

where $T = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$. For example, $u_1 = x_1 + x_2$ and $u_2 = x_1 - x_2$. Using the data of Table 4.1, $u_1 = 325 + 575 = 900$ and $u_2 = 325 - 575 = -250$. Table 4.2, below, gives the data of Table 4.1 after the transformation of (4.4.3).

TABLE 4.2

Yield Point and Tensile Strength of Steel
(Transformed Data)

u_1	u_2	v_1	v_2
900	-250	974	-262
993	-251	886	-250
966	-238	930	-266
922	-246	1052	-252
944	-240	1018	-250

Applying (2.1.1) to the data of Table 4.2, we obtain

$\bar{u}_1 = 945\times10^2$, $\bar{v}_1 = 972\times10^2$, $\bar{u}_2 = -245\times10^2$, $\bar{v}_2 = -256\times10^2$.

By using (2.1.2), the variances are found to be:

$s^2_{u_1} = 1{,}060.0\text{x}10^4$, $s^2_{u_2} = 27.2\text{x}10^4$, $s^2_{v_1} = 3{,}536.0\text{x}10^4$,

$s^2_{v_2} = 44.8\text{x}10^4$.

Using either (2.2.9) or (4.4.6), we obtain:

$$m^*_{1:0} = 10^2[17{,}680(945) + 5{,}300(972)]/(22{,}980)$$

$$= 951.2272\text{x}10^2,$$

$$m^*_{2:0} = 10^2[224(-245) + 136(-256)]/(360) = -\ 249.1556\text{x}10^2.$$

Using (2.2.11) and (2.2.12) with u and v replacing x and y, we get:

$w_{u_1}(m^*_0) = 5\text{x}10^4\text{x}3{,}967.5092$, $w_{u_2}(m^*_0) = 5\text{x}10^4\text{x}91.6458$,

$w_{v_1}(m^*_0) = 5\text{x}10^4\text{x}1{,}098.7780$, $w_{v_2}(m^*_0) = 5\text{x}10^4\text{x}44.4690$.

Substitution into (2.2.8) yields:

$$m^*_{1:1} = 10^2[3{,}967.5092(945) + 1{,}098.7780(972)]/(5{,}066.2872)$$

$$= 950.8558\text{x}10^2,$$

$$m^*_{2:1} = 10^2[91.6458(-245) + 44.4690(-256)]/(136.1148)$$

$$= -\ 248.5937\text{x}10^2.$$

We continue the iteration procedure and find the next stage estimates to be:

$w_{u_1}(m^*_1) = 5\text{x}10^4\text{x}3{,}983.0772$, $w_{u_2}(m^*_1) = 5\text{x}10^4\text{x}99.6533$,

$w_{v_1}(m_1^*) = 5x10^4x1{,}094.2904$, $w_{v_2}(m_1^*) = 5x10^4x40.1147$,

and $m_{1:2}^* = 950.8191x10^2$, $m_{2:2}^* = -\ 248.1571x10^2$.

Further iteration gives:

$w_{u_1}(m_2^*) = 5x10^4x3{,}984.6305$, $w_{u_2}(m_2^*) = 5x10^4x106.3111$,

$w_{v_1}(m_2^*) = 5x10^4x1{,}093.8619$, $w_{v_2}(m_2^*) = 5x10^4x37.1673$,

and $m_{1:3}^* = 950.8156x10^2$, $m_{2:3}^* = -\ 247.8495x10^2$.

We now take $\hat{\mu}_1^* = 950.82x10^2$ and $\hat{\mu}_2^* = -\ 247.85x10^2$.

From (2.2.3) and (2.2.4), we obtain:

$\hat{\sigma}_{u_1}^2 = 1{,}093.82x10^4$, $\hat{\sigma}_{u_2}^2 = 35.32x10^4$, $\hat{\sigma}_{v_1}^2 = 3{,}984.78x10^4$,

$\hat{\sigma}_{v_2}^2 = 111.23x10^4$.

We now apply the inverse transformation, $\hat{\mu} = T^{-1}\ \hat{\mu}^*$, to find:

$$\hat{\mu}_1 = (1/2)(\hat{\mu}_1^* + \hat{\mu}_2^*) = (1/2)(950.82 - 247.85)x10^2$$

$$= 351.48x10^2,$$

$$\hat{\mu}_2 = (1/2)(\hat{\mu}_1^* - \hat{\mu}_2^*) = (1/2)(950.82 + 247.85)x10^2$$

$$= 599.33x10^2.$$

From (4.4.5), we get:

$\hat{\sigma}_x^2 = 282.285x10^4$, $\hat{\sigma}_y^2 = 1{,}024.00x10^4$, $\hat{\rho}_x = 0.94$, $\hat{\rho}_y = 0.95$.

The values of the estimates of μ_1, μ_2, σ_x^2, σ_y^2, ρ_x, and ρ_y, which were obtained by iteration after the linear transformation, check with the values which were obtained by the "First Method." Since the amount of computation required is noticeably less using the "Second Method," it appears advisable to make a transformation of variables whenever $\sigma_{x_1}^2 = \sigma_{x_2}^2$ and $\sigma_{y_1}^2 = \sigma_{y_2}^2$.

V. JOINT ESTIMATION OF PARAMETERS IN THE MULTIVARIATE CASE

5.1 Objectives, Definitions, and Assumptions

Let X and Y have p-variate normal distributions with a common vector of means μ and respective covariance matrices, Σ_x and Σ_y, which are not parametrically related. We shall assume that μ is unknown, as well as vectors θ_x and θ_y which are composed of relevant functions of the elements of Σ_x and Σ_y, respectively. For example, if $\Sigma_x = \sigma_x^2 \begin{pmatrix} 1 & \rho_x \\ \rho_x & 1 \end{pmatrix}$, as in Section 4.4, we can take $\theta_x = \begin{pmatrix} \sigma_x^2 \\ \rho_x \end{pmatrix}$, whereas, if ρ_x is assumed known, we can take $\theta_x = \sigma_x^2$. We shall require a one-to-one correspondence between the vector components of θ_x and the independent parameters in Σ_x which are to be estimated. On occasion, we may write $\Sigma_x(\theta_x)$ to indicate that Σ_x can be determined when θ_x is known. We shall assume, also, that the relationships existing between θ_x and Σ_x hold, identically, for θ_y and Σ_y.

Our objective is to estimate jointly the unknown vectors μ, θ_x, and θ_y. To accomplish the estimation we shall make n_x independent observations on X and n_y independent observations on Y, and then employ an iteration procedure to solve the likelihood equations.

In this chapter we shall discuss the feasibility of making a non-singular linear transformation on the variables X and Y before iteration and we shall show that the iteration procedure is invariant when the transformation is orthogonal. In addition, two special cases for which our iteration procedure is well suited will be presented.

Our definitions of previous chapters will be augmented by letting S_x and S_y be symmetric, positive definite matrices whose elements are $s_{x,\alpha B}$ and $s_{y,\alpha B}$, defined by (4.1.4) and (4.1.5), respectively, with α, $B = 1, 2, \ldots, p$. Since m is the iteration estimate of the common vector of means, we shall take $S_x(m)$ and $S_y(m)$ to be symmetric, positive definite matrices with elements $s_{x,\alpha B(I)}$ and $s_{y,\alpha B(I)}$ defined by (4.1.6) and (4.1.7). Specifically,

$$(5.1.1) \quad \begin{cases} S_x = (1/n_x)XX' - \overline{X}\overline{X}', \quad S_y = (1/n_y)YY' - \overline{Y}\overline{Y}', \\ S_x(m) = (1/n_x)XX' - \overline{X}\overline{X}' + (\overline{X} - m)(\overline{X} - m)', \\ S_y(m) = (1/n_y)YY' - \overline{Y}\overline{Y}' + (\overline{Y} - m)(\overline{Y} - m)'. \end{cases}$$

Further, let $\theta_x(S_x(m))$ and $\theta_y(S_y(m))$ denote the iteration estimates of θ_x and θ_y based on the iterative value m.

5.2 <u>Development and Solution of the Likelihood Equations</u>

We shall let L represent the natural logarithm of the joint likelihood of the two samples, and write

$L = L(X, Y; \theta_x, \theta_y, \mu)$ to indicate, implicitly, that L is a function of the observations on X and Y, and the parameters θ_x, θ_y, and μ. Explicitly, we have

$$(5.2.1)\quad L = -(p/2)(n_x + n_y)\log 2\pi - (n_x/2)\log\left|\Sigma_x(\theta_x)\right| - (n_y/2)\log\left|\Sigma_y(\theta_y)\right| - (1/2)\sum_{i=1}^{n_x}(x_i - \mu)' \cdot \Sigma_x^{-1}(\theta_x)\cdot(x_i - \mu) - (1/2)\sum_{j=1}^{n_y}(y_j - \mu)' \cdot \Sigma_y^{-1}(\theta_y)\cdot(y_j - \mu)\ .$$

To obtain our likelihood equations, we equate to zero the partial derivatives of L with respect to the parameters θ_x, θ_y, and μ. It is to be understood that $\frac{\partial L}{\partial \theta_x} = \emptyset$, $\frac{\partial L}{\partial \theta_y} = \emptyset$, and $\frac{\partial L}{\partial \mu} = \emptyset$ are vector equations where, say, the components of $\frac{\partial L}{\partial \theta_x}$ are the partial derivatives of L with respect to the components of θ_x. The resulting likelihood equations are:

$$(5.2.2)\quad n_x \frac{\partial}{\partial \theta_x}\left|\Sigma_x(\theta_x)\right| + \frac{\partial}{\partial \theta_x}\sum_{i=1}^{n_x}(x_i - \hat{\mu})' \cdot \Sigma_x^{-1}(\hat{\theta}_x)\cdot(x_i - \hat{\mu}) = \emptyset,$$

(5.2.3) $$n_y \frac{\partial}{\partial\theta_y} \Sigma_y(\hat{\theta}_y) + \frac{\partial}{\partial\theta_y} \sum_{j=1}^{n_y} (y_j - \hat{\mu})' \cdot \Sigma_y^{-1}(\hat{\theta}_y) \cdot (y_j - \hat{\mu}) = \emptyset,$$

(5.2.4) $$n_x [(\bar{x} - \hat{\mu})' \cdot \Sigma_x^{-1}(\hat{\theta}_x)]' + n_y [(\bar{y} - \hat{\mu})' \cdot \Sigma_y^{-1}(\hat{\theta}_y)]' = \emptyset .$$

From (5.2.2) and (5.2.3), it is seen that $\hat{\theta}_x$ is composed of elements $\hat{\sigma}_{x,\alpha B}$ defined by (4.2.2) for α, B = 1, 2, ... , p. Further, if $\hat{\theta}_x$ is known, then $[\Sigma_x(\hat{\theta}_x)]^{-1}$ can be determined. Hence, we have written $\Sigma_x^{-1}(\hat{\theta}_x)$ for $[\Sigma_x(\hat{\theta}_x)]^{-1}$. Similar relationships hold for $\hat{\theta}_y$ and $\Sigma_y^{-1}(\hat{\theta})$.

Equation (5.2.4) may be written

(5.2.5) $$\hat{\mu} = [n_x \Sigma_x^{-1}(\hat{\theta}_x) + n_y \Sigma_y^{-1}(\hat{\theta}_y)]^{-1} \cdot [n_x \Sigma_x^{-1}(\hat{\theta}_x) \cdot \bar{x} + n_y \Sigma_y^{-1}(\hat{\theta}_y) \cdot \bar{y}].$$

The general iteration procedure to find jointly $\hat{\mu}$, $\hat{\theta}_x$, and $\hat{\theta}_y$ is as follows: For the initial value of m, we shall take

$$(5.2.6) \qquad m_0 = [n_x\Sigma_x^{-1}(\hat{\theta}_x(S_x)) + n_y\Sigma_y^{-1}(\hat{\theta}_y(S_y))]^{-1} \cdot [n_x\Sigma_x^{-1}(\hat{\theta}_x(S_x))\cdot\bar{x} + n_y\Sigma_y^{-1}(\hat{\theta}_y(S_y))\cdot\bar{y}].$$

Using m_0, $\hat{\theta}_x(S_x(m_0))$ and $\hat{\theta}_y(S_y(m_0))$ can be found and, subsequently, $\Sigma_x^{-1}(\hat{\theta}_x(S_x(m_0)))$ and $\Sigma_y^{-1}(\hat{\theta}_y(S_y(m_0)))$. For $r \geq 1$,

$$(5.2.7) \qquad m_r = [n_x\Sigma_x^{-1}(\hat{\theta}_x(S_x(m_{r-1})))+n_y\Sigma_y^{-1}(\hat{\theta}_y(S_y(m_{r-1})))]^{-1} \cdot [n_x\Sigma_x^{-1}(\hat{\theta}_x(S_x(m_{r-1})))\bar{x}+n_y\Sigma_y^{-1}(\hat{\theta}_y(S_y(m_{r-1})))\bar{y}].$$

If each component of the vector m_r generates a convergent sequence as r tends to infinity, we shall call the vector whose components are limits of these convergent sequences, m. The vector m satisfies the vector equation (5.2.4) and is the iteration estimate of the vector μ, defined in Section 5.1. The substitution of m in equations (5.2.2) and (5.2.3) will provide the desired iteration estimates for θ_x and θ_y.

5.3 Transformations and Iteration Procedure

In many situations, the likelihood equations (5.2.2), (5.2.3), and (5.2.4) will be combersome to solve. Sometimes, a simplification can be made by transformation. We are particularly interested in making a non-singular linear transformation which will diagonalize the covariance matrices Σ_x and Σ_y. That is, we would like to take $U = TX$ and $V = TY$ and obtain diagonal matrices for Σ_u and Σ_v. We shall elect to make an orthogonal transformation, thus making $T' = T^{-1}$. It should be noted that the transformation of variables does not have to be made on L in (5.2.1), but instead, the transformed likelihood equations may be obtained directly from the likelihood equations (5.2.2), (5.2.3), and (5.2.4) by transformation. Further, the "transformed iteration" is equivalent to the original iteration. This may be established as follows:

Let $U = TX$, $V = TY$, $n_x = n_u$, $n_y = n_v$, with $T' = T^{-1}$. Then, $\Sigma_u = T\Sigma_x T'$ and $\Sigma_v = T\Sigma_y T'$. If $\bar{U}$ and $\bar{V}$ are the vector products $(1/n_u)U$ and $(1/n_v)V$, respectively, we have also:

$$(5.3.1) \quad \begin{cases} S_u = (1/n_u)UU' - \bar{U}\bar{U}', \quad S_v = (1/n_v)VV' - \bar{V}\bar{V}', \\ S_u(m) = (1/n_u)UU' - \bar{U}\bar{U}' + (\bar{U} - m)(\bar{U} - m)', \\ S_v(m) = (1/n_v)VV' - \bar{V}\bar{V}' + (\bar{V} - m)(\bar{V} - m)'. \end{cases}$$

Let m^T be the iteration value, corresponding to m, obtained from the transformed likelihood, and let $\theta_u(S_u(m))$ and $\theta_v(S_v(m))$ be the iteration estimates of θ_u and θ_v, respectively. Thus,

$$(5.3.2) \quad m_0^T = [n_u\Sigma_u^{-1}(\theta_u(S_u)) + n_v\Sigma_v^{-1}(\theta_v(S_v))]^{-1}$$
$$\cdot[n_u\Sigma_u^{-1}(\theta_u(S_u))\cdot\bar{U} + n_v\Sigma_v^{-1}(\theta_v(S_v))\cdot\bar{V}],$$

and

$$(5.3.3) \quad m_r^T = [n_u\Sigma_u^{-1}(\theta_u(S_u(m_{r-1}^T)))+n_v\Sigma_v^{-1}(\theta_v(S_v(m_{r-1}^T)))]^{-1}$$
$$\cdot[n_u\Sigma_u^{-1}(\theta_u(S_u(m_{r-1}^T)))\cdot\bar{U}+n_v\Sigma_v^{-1}(\theta_v(S_v(m_{r-1}^T)))\cdot\bar{V}].$$

Now, $\bar{U} = T\bar{X}$ and $S_u = (1/n_x)(TX)(TX)' - (T\bar{X})(T\bar{X})' = TS_xT'$. Consequently, $\Sigma_u(\theta_u(S_u)) = T\Sigma_x(\theta_x(S_x))T'$. Similarly, $V = T\bar{Y}$, $S_v = TS_yT'$, and $\Sigma_v(\theta_v(S_v)) = T\Sigma_y(\theta_y(S_y))T'$. Substitution into (5.3.2) and simplification gives $m_0^T = Tm_0$. In like manner, $S_u(m^T) = TS_x(m)T'$ and $S_v(m^T) = TS_y(m^T)T'$. Substitution into (5.3.3) with $r = 1$ gives $m_1^T = Tm_1$. Thus, the two iteration procedures are completely equivalent. Hence, we may conclude that the iteration procedure of Section 5.2 is invariant under a linear, orthogonal transformation. Consequently, m^T, $\hat{\theta}_u$, and $\hat{\theta}_v$ may be found by iteration and then transformed to give m, $\hat{\theta}_x$, and $\hat{\theta}_y$.

5.4 Some Special Cases of Joint Estimation of Parameters

We shall now consider two special cases of joint estimation for which the iteration procedure is well suited.

Case 1. Assumptions:

$$(5.4.1)\qquad \begin{cases} \mu_{x,\alpha} = \mu_{y,\alpha} = \mu_\alpha , \\ \sigma_{x,\alpha\alpha} = \sigma^2_{x_\alpha}, \quad \sigma_{y,\alpha\alpha} = \sigma^2_{y_\alpha} , \\ \sigma_{x,\alpha B} = \sigma_{y,\alpha B} = 0 \text{ for } \alpha \neq B . \end{cases}$$

Objective: Estimate jointly μ_α, $\sigma^2_{x_\alpha}$, and $\sigma^2_{y_\alpha}$ for $\alpha = 1, 2, \ldots, p$.

Procedure: Take

$$(5.4.2)\qquad \theta_x = \begin{pmatrix} \sigma^2_{x_1} \\ \vdots \\ \sigma^2_{x_p} \end{pmatrix}, \quad \theta_y = \begin{pmatrix} \sigma^2_{y_1} \\ \vdots \\ \sigma^2_{y_p} \end{pmatrix} .$$

The iteration estimates are:

$\hat{\mu}_\alpha = m_\alpha$, $\hat{\sigma}^2_{x_\alpha} = s^2_{x_\alpha(I)}$, $\hat{\sigma}^2_{y_\alpha} = s^2_{y_\alpha(I)}$, where

$$(5.4.3)\qquad \begin{cases} s^2_{x_\alpha(I)} = s^2_{x_\alpha} + (\bar{x}_\alpha - m_\alpha)^2 , \\ s^2_{y_\alpha(I)} = s^2_{y_\alpha} + (\bar{y}_\alpha - m_\alpha)^2 , \\ n_x(\bar{x}_\alpha - m_\alpha)/s^2_{x_\alpha(I)} + n_y(\bar{y}_\alpha - m_\alpha)/s^2_{y_\alpha(I)} = 0 , \end{cases}$$

$\alpha = 1, 2, \ldots, p$.

The solution, by iteration, for each value of α, is identical with Section 2.2. The relationship

$$(5.4.4)\quad [m_{\alpha(r+1)} - m_{\alpha(r)}]/[m_{1(r+1)} - m_{1(r)}]$$
$$= (\bar{y}_\alpha - \bar{x}_\alpha)/(\bar{y}_1 - \bar{x}_1)$$

can be shown to exist. Therefore, having computed $m_{1(r+1)} - m_{1(r)}$, all succeeding $m_{\alpha(r+1)}$ can be calculated without further iteration.

Case 2. Assumptions:

$$(5.4.5)\quad \begin{cases} \mu_{x,\alpha} = \mu_{y,\alpha} = \mu_\alpha , \\ \sigma_{x,\alpha\alpha} = \sigma_x^2 , \quad \sigma_{y,\alpha\alpha} = \sigma_y^2 , \\ \sigma_{x,\alpha B} = \rho_x\sigma_x^2 , \quad \sigma_{y,\alpha B} = \rho_y\sigma_y^2 \text{ for } \alpha \neq B. \end{cases}$$

Objectives: Estimate jointly μ_α, σ_x^2, σ_y^2, ρ_x, and ρ_y.

$\alpha = 1, 2, \ldots, p.$

Procedure: Take

$$(5.4.6)\quad \theta_x = \begin{pmatrix} \sigma_x^2 \\ \rho_x \end{pmatrix}, \quad \theta_y = \begin{pmatrix} \sigma_y^2 \\ \rho_y \end{pmatrix}.$$

Make the Helmert transformation (T) on X and Y to diagonalize the matrices Σ_x and Σ_y, getting Σ_u and Σ_v of Section 5.3. We, next obtain

$$(5.4.7)\quad \theta_u = \sigma_x^2 \begin{pmatrix} 1 + (p-1)\rho_x \\ 1 - \rho_x \end{pmatrix}, \quad \theta_v = \sigma_y^2 \begin{pmatrix} 1 + (p-1)\rho_y \\ 1 - \rho_y \end{pmatrix}.$$

The iteration estimate of the transformed vector, μ^T, may be obtained from (5.3.3) by using the iteration estimates of the components of θ_u and θ_v of (5.4.7). The inverse transformation (T^{-1}) supplies the desired estimates for μ_α, σ_x^2, σ_y^2, ρ_x, and ρ_y.

VI. JOINT ESTIMATION OF PARAMETERS BY EMPIRICAL SAMPLING FROM TWO UNIVARIATE NORMAL DISTRIBUTIONS

6.1 Objectives, Definitions, and Notation

In Section 2.2, an iteration procedure was developed for joint estimation of the mean μ and the variances σ_x^2 and σ_y^2 when sampling from two univariate normal distributions with a common mean and finite variances which may or may not be equal. In Appendix Section D, the distribution of m (the iteration estimate of the mean) is discussed and particular regard is given for the case where sample sizes n_x and n_y are fairly large. Also, an attempt is made to determine the moments of the distribution of m_0, the first stage iteration value of m.

To gain further knowledge of the distribution of m whenever the sample sizes are small, empirical sampling is done with the aid of the IBM 650 electronic, digital computer. The sample values are obtained by using a modification of Neumann's program* for generating random standard normal numbers. The numbers are taken alternately in groups of n_x and n_y and multiplied by appropriate factors to simulate sampling from normal populations with a common

---

* This program for the IBM 650 was supplied by the North Carolina State University Computing Laboratory

mean, zero, and known variances, σ_x^2 and σ_y^2. For the two samples, thus obtained, the following statistics are computed:

$\bar{x}$ and $\bar{y}$, defined by (2.1.1);

s_x^2 and s_y^2, defined by (2.1.2);

m_0, m, $s_{x(I)}^2$, and $s_{y(I)}^2$, defined, respectively, by (2.2.9), (2.2.7), (2.2.3), and (2.2.4);

$\hat{\mu}$ using (C.4) from Appendix Section C;

(6.1.1) $\tilde{s}_x{}^2 = \sum_{i=1}^{n_x} (x_i - \bar{x})^2/(n_x - 1)$,

(6.1.2) $\tilde{s}_y{}^2 = \sum_{j=1}^{n_y} (y_j - \bar{y})^2/(n_y - 1)$,

(6.1.3) $\tilde{m}_0 = (n_x\tilde{s}_y{}^2\bar{x} + n_y\tilde{s}_x{}^2\bar{y})/(n_x\tilde{s}_y{}^2 + n_y\tilde{s}_x{}^2)$.

Following the selection of specific values for n_x, n_y, σ_x^2, and σ_y^2 in the two samples, described above, 1000 pairs of the two samples are generated. We shall refer to the set of 1000 pairs as a run and, for each run, additional statistics are computed as follows:

(6.1.4) $M(\mathring{Z}) = \sum_{i=1}^{1000} \mathring{Z}_i/1000$

and

(6.1.5) $V(\mathring{Z}) = \sum_{i=1}^{1000} \mathring{Z}_i{}^2/1000 - [M(\mathring{Z})]^2$.

$\mathring{z}$ is specified in each case and takes on the values $\tilde{m}_0$, m_0, m, $\hat{\mu}$, $\tilde{s}_x^2$, $\tilde{s}_y^2$, $s^2_{x(1)}$, $s^2_{y(1)}$, and k, where k is the number of iterations required to obtain m, correct to five decimal places.

We shall make Chi-Square tests of normality for the distribution of m and we shall also compare the various estimates of μ, σ_x^2, and σ_y^2.

6.2 Experimental Results on Estimating the Common Mean and Individual Variances When Sampling From Two Univariate Normal Distributions

Table 6.1. below, shows the means and variances, as defined by (6.1.4) and (6.1.5), for four different estimates of the common mean when sampling from two univariate normal distributions. Also, the mean and variance is given for the number of iterations required to obtain m with five decimal place accuracy.

For runs one and six, the pairs of samples have equal sample sizes and equal variances. Therefore, it is possible to calculate (see Appendix Section D) the population variances for m_0 in these runs. We shall show that $\text{Var}\ (m_0) = (n + 1)\sigma^2/2n^2$. Thus, the exact values are .375 and .055, which compare favorably with the table values of .384 and .055, respectively. Table 6.1 indicates that the various estimates of μ approach equality rapidly as the samples increase in size.

TABLE 6.1

Estimates Pertaining to the Common Mean

Run	1	2	3	4	5	6	7
n_x	2	2	2	2	8	10	10
n_y	2	2	2	8	2	10	10
σ_x^2	1	1	1	1	1	1	1
σ_y^2	1	4	16	4	4	1	16
$M(\tilde{\hat{m}}_0)$	.003	.065	.057	.013	.036	-.006	.013
$M(m_0)$	.003	.065	.057	.017	.034	-.006	.013
$M(m)$	.002	.074	.026	.012	.045	-.006	.013
$M(\hat{\mu})$	.002	.074	.026	.007	.033	-.006	.013
$V(\tilde{\hat{m}}_0)$	.384	.737	1.60	.361	.349	.055	.101
$V(m_0)$	.384	.737	1.60	.370	.419	.055	.101
$V(m)$	.447	.915	1.92	.403	.345	.057	.100
$V(\hat{\mu})$	.447	.915	1.92	.396	.205	.057	.100
$M(k)$	5.0	4.8	4.0	4.7	4.6	2.7	2.0
$V(k)$	38	36	20	33	30	6	2

Runs one, two, and three produced likelihood equations for μ in which $m = \hat{\mu}$ in every case, even though in 506 of the 3000 equations, there were three real, distinct roots. In the 506 cases with three real, distinct roots the maximum likelihood estimate $\hat{\mu}$ is the "best estimate" (in the sense of $M(\hat{\mu})$ being nearest zero) of μ in only 226 cases with the "best estimate" being approximately equally divided between the other two roots for the remaining cases. Runs four and five produced likelihood equations for μ with three real, distinct roots in 89 of the 2000 equations. In 42 of these 89 cases $m \neq \hat{\hat{\mu}}$, and for these 42 cases, $\hat{\hat{\mu}}$ is the "best estimate" in 39 cases when compared with m and $\tilde{m}_0$.

Table 6.1, above, also indicates that the iteration estimate m is generally closer to μ than is $\tilde{m}_0$.

Table 6.2, below, shows the means and variances, as defined by (6.1.4) and (6.1.5), of the estimates of the individual population variances when sampling from two univariate normal distributions. It appears that both types of variance estimates are virtually unbiased with the iteration estimates, $s^2_{x(I)}$ and $s^2_{y(I)}$ having less variation than $\tilde{s}_x^{\,2}$ and $\tilde{s}_y^{\,2}$.

TABLE 6.2

Estimates Pertaining to the Population Variances

Run	1	2	3	4	5	6	7
n_x	2	2	2	2	8	10	10
n_y	2	2	2	8	2	10	10
σ_x^2	1	1	1	1	1	1	1
σ_y^2	1	4	16	4	4	1	16
$M(\tilde{s}_x^2)$	1.00	1.06	1.04	.98	.99	1.01	1.03
$M(s_{x(I)}^2)$	.92	1.25	2.13	.75	1.12	.97	.94
$M(\tilde{s}_y^2)$	1.04	4.04	15.71	4.00	4.03	1.02	16.30
$M(s_{y(I)}^2)$	.97	3.67	15.37	3.93	3.72	.97	16.27
$V(\tilde{s}_x^2)$	2.21	2.06	2.22	1.97	.27	.21	.26
$V(s_{x(I)}^2)$	1.76	5.83	43.77	1.27	1.78	.20	.22
$V(\tilde{s}_y^2)$	2.27	30.80	519.36	4.66	34.14	.22	58.28
$V(s_{y(I)}^2)$	2.22	21.23	281.92	4.78	17.75	.20	53.07

The experimental results reveal that m may be useful in obtaining $s_{x(I)}^2$ and $s_{y(I)}^2$ as estimates for σ_x^2 and σ_y^2, respectively. We may be interested also in the distribution of m, itself. An examination of the cases where the likelihood equations for μ have three real roots suggests that the conditional distribution of m, when only one real root exists, is quite different from the distribution of m when the cases with three real roots are included.

We shall make Chi-square tests of normality and of symmetry about zero for m for both the conditional distribution of m, based on one real root, and the unrestricted distribution of m. To accomplish this, we shall divide the 1000 observations on m, for each run, into 2p classes which correspond, under the hypothesis of normality, to 2p classes with equal expected numbers, E. Since the expected value of m is μ, which is zero, we shall let o_i represent the observed number of m in the i-th class above zero and o_{-i} the observed number in the corresponding i-th class below zero. The variance of m is estimated for each run. The statistic

$$\sum_{i=1}^{p} (o_i + o_{-i} - 2E)^2/2E = \sum_{i=1}^{p} (o_i + o_{-i})^2/2E - 1000$$

is distributed approximately as χ^2 with p - 2 degrees of freedom, and may be used to test normality in the distribution of m. The statistic $\sum_{i=1}^{p} (o_i - o_{-i})^2/2E$ is distributed approximately as χ^2 with p degrees of freedom, and may be used to test symmetry about zero in the distribution of m.

Table 6.3, below, shows the computed values of the above statistics for runs one through seven. The value of p is taken to be ten.

TABLE 6.3

Computed Values of χ^2 for Testing Normality and Symmetry in the Distribution of m

Run	1	2	3	4	5	6	7	D.F.
Normality	14.68	34.28	289.70	10.14	160.10	6.32	18.30	8
Symmetry	3.84	18.08	12.74	10.26	9.26	2.56	14.94	10
Total	18.52	52.32	302.44	20.40	169.36	8.88	33.24	18

With a significance level of .05 we have for critical values, $\chi^2 = 15.5$ with 8 degrees of freedom and $\chi^2 = 18.3$ with 10 degrees of freedom. Hence, we do not reject the hypothesis of symmetry in any run. However, we do reject the hypothesis of normality in runs two, three, five, and seven.

In Table 6.4, below, we consider the conditional distribution of m, which is restricted to the cases where the likelihood equations for μ have only one real root. Runs six and seven resulted in cases with one real root only and are omitted.

TABLE 6.4

Computed Values of χ^2 for Testing Normality and Symmetry in the Conditional Distribution of m

Run	1 N=858	2 N=841	3 N=795	4 N=977	5 N=934	Degrees of Freedom
Normality	5.54	10.74	25.31	8.19	21.89	8
Symmetry	.48	15.26	14.03	12.29	7.58	10
Total	6.02	26.00	39.34	20.48	29.47	18

With a significance level of .05 we do not reject the hypothesis of symmetry in any runs. However, we still reject the hypothesis of normality in runs three and five. It may be noted that the computed χ^2 values for normality in runs three and five have been reduced by over 80% by restricting the distribution of m to the cases with only one real root in the likelihood equations for μ.

VII. SUMMARY

This dissertation discusses a method of estimation with random samples from two normal populations for the univariate, bivariate, and multivariate cases. Since it is assumed that the two populations have a common mean (or a common vector of means), these common parameters are jointly estimated with all the other unknown parameters. The joint estimates, called iteration estimates, are obtained by an iteration method developed for solving the likelihood equations.

A detailed study is made for the iteration estimates of the parameters when sampling from two univariate populations, and a numerical example is given to illustrate the iteration technique. It is proved that the iteration method produces an unique set of estimates which satisfy the likelihood equations. Since this set of estimates is not always identical with the set of maximum likelihood estimates, the conditions under which the two sets may possibly differ are established. Empirical sampling, with small sample sizes, was done with the aid of the IBM 650 Computer to obtain information regarding the distribution of the iteration estimates and also the maximum likelihood estimates in the cases where they differ. The experimental results indicate that the estimate of the common mean tends to be normally distributed and the estimates of the

individual variances are virtually unbiased.

The iteration procedure is compared with Fisher's Method, which uses the Information Matrix, and is shown to give identical results.

An extension of the iteration procedure is made to the case where samples are drawn from bivariate populations. For the particular case in which the individual variances within each population may be assumed equal, it is shown that a linear transformation to obtain new uncorrelated variables will materially lessen the time required for the method. A numerical example is given to illustrate the iteration technique both with and without transformation and a proof is given to show that the two methods produce identical results.

The iteration procedure is further extended to the multivariate case and joint estimates are obtained for the common vector of means and the individual covariance matrices. It is also shown that if a linear transformation can be found which gives new uncorrelated variables in each population, then transformation before iteration greatly reduces the computational labor involved in obtaining the joint estimates.

VIII. ACKNOWLEDGMENTS

I wish to express my appreciation:

To The Southern Fellowships Fund for a grant which helped make this study possible;

To President , Dean , and Professor of the University of Louisville for a generous leave of absence;

To Mrs. , my wife, for her numerous sacrifices and unlimited patience during my years of graduate study;

To Mrs. and for typing the manuscript;

To Professors R. E. Bargmann, H. A. David, R. J. Freund, and D. C. Hurst for their assistance on various phases of the work;

To Professor C. W. Clunies-Ross for suggesting the problem and directing the research;

To Professor C. Y. Kramer for his invaluable guidance in the preparation of the final draft;

To Professor Boyd Harshbarger for his continuing encouragement throughout my graduate program.

IX. BIBLIOGRAPHY

Anderson, T. W. (1958). Introduction to Multivariate Statistical Analysis. New York: John Wiley and Sons, Inc.

Aspin, A. A. (1948). An examination and further development of a formula arising in the problem of comparing two mean values. Biometrika, 35, 88.

Cochran, W. G. (1937). Problems arising in the analysis of a series of similar experiments. Journal Royal Statistical Society, Supplement 4, 102.

Dickson, L. E. (1922). First Course in the Theory of Equations. New York: John Wiley and Sons, Inc.

Fraser, D. A. S. (1957). On the combining of interblock and intrablock estimates. Annals of Mathematical Statistics, 28, 814.

Gart, J. J. (1958). Some problems in statistical inference. Oak Ridge National Laboratory.

Graybill, F. A. and Deal, R. B. (1959). Combining unbiased estimators. Biometrics, 15, 543.

James, G. S. (1956). On the accuracy of weighted means and ratios. Biometrika, 43, 304.

Kenney, J. F. and Keeping, E. S. (1951). Mathematics of Statistics, Part Two, Second Edition. New York: D. Van Nostrand Company, Inc.

Kunz, K. S. (1957). Numerical Analysis. New York: McGraw-Hill Book Company, Inc.

Mantel, N. (1955). Maximum likelihood estimation of the mean in the case of unequal precisions. Unpublished masters' thesis.

Mood, A. M. (1950). Introduction to the Theory of Statistics. New York: McGraw-Hill Book Company, Inc.

Neyman, J. and Scott, E. L. (1948). Consistent estimates based on partially consistent observations. Econometrica, 16, 1.

Pearson, E. S. and Hartley, H. O. (1954). Biometrika Tables For Statisticians. Cambridge: University Press.

Rao, C. R. (1947). General methods of analysis for incomplete block designs. Journal of the American Statistical Association, 42, 541.

Rao, C. R. (1952). Advanced Statistical Methods in Biometric Research. New York: John Wiley and Sons, Inc.

Sokolnikoff, I. S. (1939). Advanced Calculus. New York: McGraw-Hill Book Company, Inc.

Sprott, D. A. (1956). A note on combined interblock and intrablock estimation in incomplete block designs. Annals of Mathematical Statistics, 27, 633.

Stevens, W. L. (1951). Asymptotic regression. Biometrics, 7, 247.

Yates, F. and Cochran, W. G. (1938). The analysis of groups of experiments. Journal of Agricultural Science, 28, 556.

Yates, F. (1939). An apparent inconsistency arising from tests of significance based on fiducial distributions of unknown parameters. Proceedings of the Cambridge Philosophical Society, 35, 588.

Yates, F. (1940). The recovery of interblock information in balanced incomplete block designs. Annals of Eugenics, 10, 317.

Zelen, M. (1957). The analysis of incomplete block designs. Journal of the American Statistical Association, 52, 204.

The two page vita has been removed from the scanned document. Page 1 of 2

The two page vita has been removed from the scanned document. Page 2 of 2

XI. APPENDIX

A. The Conditions Under Which the Likelihood Equation for Estimating the Mean in the Univariate Case has Three Real, Distinct Roots

By (3.3.3), the likelihood equation for estimating the mean is

(A.1) $\hat{\mu}^3 - A\hat{\mu}^2 + B\hat{\mu} - C = 0,$

where

$$A = [n_x\bar{x} + n_y\bar{y} + 2(n_x\bar{y} + n_y\bar{x})]/(n_x + n_y),$$

$$B = [n_x(s_y^2 + \bar{y}^2) + n_y(s_x^2 + \bar{x}^2)$$

$$+ 2(n_x + n_y)\overline{xy}]/(n_x + n_y),$$

$$C = [n_x(s_y^2 + \bar{y}^2)\bar{x} + n_y(s_x^2 + \bar{x}^2)\bar{y}]/(n_x + n_y).$$

The discriminant of the cubic equation [see Dickson (1922), p. 47] (A.1) can be represented by

(A.2) $\Delta = A^2B^2 + 18\ ABC - 4\ A^3C - 4\ B^3 - 27\ C^2.$

Equation (A.1) has three real roots whenever Δ is positive. We shall determine the conditions under which Δ is positive.

No generality is lost if we take $\bar{x} = 0$ and $d = \bar{y} - \bar{x} = \bar{y} > 0$. Then, $A = d(2\ n_x + n_y)/(n_x + n_y)$, $B = (n_x s_y^2 + n_y s_x^2 + n_x d^2)/(n_x + n_y)$, and $C = dn_y s_x^2/(n_x + n_y)$. Now, putting A, B, and C into (A.1) and simplifying, we get

$$(A.3)\quad (n_x + n_y)^4 \cdot \Delta = d^2(2n_x + n_y)^2(n_x s_y^2 + n_y s_x^2 + n_x d^2)^2$$
$$+ 18\, d^2 n_y (n_x + n_y)(2n_x + n_y) s_x^2 (n_x s_y^2$$
$$+ n_y s_x^2 + n_x d^2) - 4\, d^4 n_y (2n_x + n_y)^3 s_x^2$$
$$- 4(n_x + n_y)(n_x s_y^2 + n_y s_x^2 + n_x d^2)^3$$
$$- 27\, d^2 n_y^2 (n_x + n_y)^2 s_x^4.$$

If we let $R = n_y/n_x$, $S = s_y^2/s_x^2$, and $n_x d^2 = k(n_y s_x^2 + n_x s_y^2)$ $= n_x s_x^2 (R + S)k$, where $k > 0$, we may substitute into (A.3) and simplify to obtain

$$(A.4)\quad [(n_x + n_y)^4/n_x^4 s_x^6 (R + S)] \cdot \Delta = (R + S)^2(k + 1)^2$$
$$\cdot [R^2 k - 4(R + 1)] + 2\, R(R + 2)(R + S)k[9(R + 1)$$
$$- (2R + 1)(R - 1)k] - 27\, R^2(R + 1)^2 k.$$

We shall write $\Delta'(R,S,k) = [(n_x + n_y)^4/n_x^4 s_x^6 (R + S)] \cdot \Delta$, so that equation (A.4) becomes

$$(A.5)\quad \Delta'(R,S,k) = (R + S)^2(k + 1)^2[R^2 k - 4(R + 1)]$$
$$+ 2R(R + 2)(R + S)k[9(R + 1)$$
$$- (2R + 1)(R - 1)k] - 27\, R^2(R + 1)^2 k.$$

For the case of equal sample sizes, we have $R = 1$, and from (A.5) we obtain

$$(A.6)\quad \Delta'(1,S,k) = (S + 1)^2(k + 1)^2(k - 8) + 108\, Sk$$
$$= (S + 1)^2 k^3 - 6(S + 1)^2 k^2 - 3(5S - 1)(S - 5)k - 8(S + 1)^2.$$

In equation (A.6), we note that

$$\Delta'(1,\ 1/S,\ k) = (1/S^2)\cdot\Delta'(1,\ S,\ k).$$

For the case where $n_y = 2\ n_x$, we have $R = 2$, and, substituting into (A.5), we get

$$\text{(A.7)}\quad (1/4)\cdot\Delta'(2,\ S,\ k) = (S + 2)^2(k + 1)^2(k - 3)$$
$$+ 4(S + 2)k(27 - 5k) - 243k$$
$$= (S + 2)^2k^3 - (S + 2)(S + 22)k^2$$
$$- (5S^2 - 88S + 47)k - 3(S + 2)^2.$$

Whenever $\Delta'(R,S,k)$ is positive, equation (A.1) will have three real, distinct roots for $\hat{\mu}$. Let $k_u(R,S)$ denote the lower bound of k (rounded to the nearest integer) for which $\Delta'(R,S,k)$ is positive. Some values of k [obtained from (A.6) and (A.7)] for various combinations of R and S are shown in Table A.1.

TABLE A.1

Approximations to $k_u(R,S)$

S	1/16	1/4	1/2	1	2	4	16
$k_u(1,S)$	7	6	4	2	4	6	7
$k_u(2,S)$	12	10	9	7	7	2	2

B. Probability of the Occurrence of Three Real, Distinct Roots in the Likelihood Equation for Estimating the Mean in the Univariate Case

We would like to know the probability that two samples, obtained under the assumptions of Section 2.1, will yield a likelihood equation (3.3.1), for estimating the mean, which has three real, distinct roots. Aspin (1948) set

(B.1) $V = (\bar{y} - \bar{x})/(\lambda_1 s_x^2 + \lambda_2 s_y^2)^{1/2}, \quad c = \lambda_1 s_x^2/(\lambda_1 s_x^2 + \lambda_2 s_y^2),$

where λ_1 and λ_2 are arbitrary, positive constants, and then sought a function $V(c)$ such that

(B.2) $P = P[|V| \geq V(c)] = 2\alpha.$

Numerical values for various combinations of c and P are given in Table 11 of Biometrika Tables for Statisticians, vol. I, by Pearson and Hartley (1954). If we take $\lambda_1 = 1/n_x$ and $\lambda_2 = 1/n_y$, we can show that V is a monotonic function of k (given in Table A.1). From Appendix Section A,

(B.3) $n_x(\bar{y} - \bar{x})^2 = k(n_y s_x^2 + n_x s_y^2).$

Substitution of (B.3) into (B.1) gives

(B.4) $V = (kn_y)^{1/2}.$

We may now write (B.2) as

(B.5) $$P[|V| \geq V(c)] = P[(kn_y)^{1/2} \geq V(c)]$$
$$= P\{k \geq [V(c)]^2/n_y\}.$$

Consequently, we may use Table A.1 in conjunction with Biometrika Tables (1954, Table 11) to estimate the probability of three real, distinct roots. For example, with samples of size ten, the probability of three real, distinct roots is less than one hundredth.

Another approach to the probability of three real, distinct roots is to investigate the distribution of Δ, defined by (A.2). This presents a formidable problem, indeed, and we shall limit ourselves to the task of determining the expected value of Δ, which we shall denote by $E(\Delta)$. We recall that X and Y are independently and normally distributed with common mean μ and respective variances σ_x^2 and σ_y^2. No generality is lost by taking $E(\bar{x}) = E(\bar{y}) = \mu = 0$. Then, $E(\bar{x}^2) = \sigma_x^2/n_x$, $E(\bar{x}^3) = 0$, $E(\bar{x}^4) = 3\sigma_x^4/n_x^2$, $E[\sum_{i=1}^{n_x} x_i] = 0$, $E[(\sum_{i=1}^{n_x} x_i^2)^2] = n_x(n_x + 2)\ \sigma_x^4$, $E[(\sum_{i=1}^{n_x} x_i^2)^3] = n_x(n_x + 2) \cdot (n_x + 4)\ \sigma_x^6$, $E[\bar{x} \sum_{i=1}^{n_x} x_i^2] = 0$, $E[\bar{x}^2 \sum_{i=1}^{n_x} x_i^2] = (n_x + 2) \cdot (n_x + 4)\ \sigma_x^6/n_x$. Similar results are obtained for Y.

If we find $E(\Delta)$ from (A.2) and let $n_x = n$, $n_y = qn$, $Z = \sigma_y^2/\sigma_x^2$, $K = q^3n^3(q + 1)^4/\sigma_x^6$, we obtain

(B.6) $K \cdot E(\Delta) = c_0 Z^3 + c_1 Z^2 + x_2 Z + c_3$, where

$$c_0 = (qn + 2)(qn + 4)[(q + 2)^2 - 4qn(q + 1)],$$

$$c_1 = q\{(qn + 2)[3qn(11q^2 + 22q + 12) - 4(5q^3 + 13q^2 + 13q + 2) - 12q^2n^2(q + 1)] + 12(q + 2)^2[(q + 1)^2 - qn(q + 2)]\},$$

$$c_2 = q^2\{q(n + 2)[3qn(12q^2 + 22q + 11) - 4(2q^3 + 13q^2 + 13q + 5) - 12q^2n^2(q + 1)] + 12(2q + 1)^2[(q + 1)^2 - qn(2q + 1)]\},$$

$$c_3 = q^5(n + 2)(n + 4)[(2q + 1)^2 - 4qn(q + 1)].$$

For the case of $n_x = n_y$, we have $q = 1$, and

$$c_0 = c_3 = -(n + 2)(n + 4)(8n - 9), \quad c_1 = c_2 = -3(8n^3 - 29n^2 + 62n - 56), \quad K = 16n^3/\sigma_x^2.$$

We substitute into (B.6) to get

$$\text{(B.7)} \quad K \cdot E(\Delta) = -(n + 2)(n + 4)(8n - 9)(Z^3 + 1) - 3[8(n - 2)^3 + 19(n - 2)^2 + 42(n - 2) + 16](Z^2 + Z).$$

For $n \geq 2$, $E(\Delta) < 0$ for all Z. Further, for fixed n, $E(\Delta)$ decreases as Z (the ratio of variances) increases. We also note from (B.6) that when q (the ratio of sample sizes) increases, $E(\Delta)$ remains negative for $n \geq 2$. Since a negative Δ indicates that equation (A.1) has only one real root, we would expect the situations, arising from sampling, which yield three real roots to be in the minority.

C. Some Observations on the Iteration Estimate of the Mean in the Univariate Case

It has been noted in Section 3.4 that if the likelihood equation (3.3.3) for estimating μ has only one real root, then the iteration estimate m is identical with the maximum likelihood estimate $\hat{\mu}$. It was also shown in Section 3.4 that when (3.3.3) has three real, distinct roots $\hat{\mu}$ must be one of the extreme roots, $\hat{\mu}_a$ or $\hat{\mu}_c$, where the roots are $\hat{\mu}_a < \hat{\mu}_b < \hat{\mu}_c$.

We shall next show that it is possible for m, the iteration estimate of μ, to be equal to $\hat{\mu}_b$ the middle root. We shall first transform the likelihood equation for μ (3.3.1) by putting $\hat{\mu} = m_0 + u/(n_x s_y^2 + n_y s_x^2)$, where m_0, the initial value of the iteration estimate m, is defined by (2.2.9). Taking $d = \bar{y} - \bar{x} > 0$, we can write $\bar{x} - \hat{\mu}$ $= -(n_y d s_x^2 + u)/(n_x s_y^2 + n_y s_x^2)$ and $\bar{y} - \hat{\mu} = (n_x d s_y^2 - u)/(n_x s_y^2 + n_y s_x^2)$. Substituting into (3.3.1) and simplifying, we get

$$(C.1)\quad (n_x + n_y)u^3 + d[n_x n_y(s_x^2 - s_y^2) + 2(n_y^2 s_x^2 - n_x^2 s_y^2)]u^2$$
$$+ \{(n_x s_y^2 + n_y s_x^2)^3 - d^2[2n_x n_y(n_x + n_y)s_x^2 s_y^2$$
$$- (n_x^3 s_y^4 - n_y^3 s_x^4)]\}u + n_x n_y d^3 s_x^2 s_y^2 (n_x^2 s_y^2 - n_y^2 s_x^2) = 0.$$

If $n_x = n_y$ and $s_x^2 = s_y^2$, equation (C.1) reduces to

$$(C.2)\quad u[u^2 - 2n_x^2 s_x^4 (d^2 - 2s_x^2)] = 0.$$

The roots of (C.2) are $u = 0$ and $u = \pm\, n_x s_x^2 (2d^2 - 4s_x^2)^{1/2}$, and the roots of equation (3.3.2) are, therefore, $\hat{\mu}_b = m = m_0 = (1/2)(\bar{x} + \bar{y})$, $\hat{\mu}_a = m_0 - (1/2)(2d^2 - 4s_x^2)^{1/2} = (1/2)[\bar{x} + \bar{y} - (2d^2 - 4s_x^2)^{1/2}]$, and $\hat{\mu}_c = m_0 + (1/2)(2d^2 - 4s_x^2)^{1/2} = (1/2)[\bar{x} + \bar{y} + (2d^2 - 4s_x^2)^{1/2}]$. Consequently, whenever $n_x = n_y$, $s_x^2 = s_y^2$, and $d^2 - 2s_x^2 > 0$, m does not equal $\hat{\hat{\mu}}$. That is, the iteration estimate and the maximum likelihood estimate are not identical.

To determine the maximum likelihood estimate $\hat{\hat{\mu}}$ when three real, distinct roots exist, we must evaluate either the joint likelihood function or some other function suitably related to it. We shall take L, defined by (3.2.1), and replace the parameters μ, σ_x^2, and σ_y^2 by their respective estimates $\hat{\mu}$, $\hat{\sigma}_x^2$, and $\hat{\sigma}_y^2$, which were obtained from the likelihood equations (3.2.2), (3.2.3), and (3.2.4). Equation (3.2.1) becomes

$$(C.3) \quad L(X, Y; \hat{\mu}, \hat{\sigma}_x^2, \hat{\sigma}_y^2) = -(1/2)\{(n_x + n_y)(1 + \log 2\pi) + n_x \log[s_x^2 + (\bar{x} - \hat{\mu})^2] + n_y \log[s_y^2 + (\bar{y} - \hat{\mu})^2]\}.$$

It will suffice to consider only

$$(C.4) \quad L^*(\hat{\mu}) = n_x \log[s_x^2 + (\bar{x} - \hat{\mu})^2] + n_y \log[s_y^2 + (\bar{y} - \hat{\mu})^2]$$

and note that L in (C.3) is a maximum whenever $L^*(\hat{\mu})$ is a minimum.

In the illustration, above, with $n_x = n_y$ and $s_x^2 = s_y^2$, $L^*(\hat{\mu})$ is symmetric with respect to $\hat{\mu} = m_0$, and $\hat{\hat{\mu}} \neq m$. We shall now show that whenever three real, distinct roots exist in the asymmetric case, it is possible for $\hat{\hat{\mu}}$ to be equal to m. Consider the example:

$$n_x = n_y = n, \ \bar{x} = -1.0, \ \bar{y} = 0.5, \ s_x^2 = 0.1, \ s_y^2 = 0.3.$$

Equation (A.1) becomes $\hat{\mu}^3 + 0.75\hat{\mu}^2 - 0.175\hat{\mu} = 0$, and $m = \hat{\mu}_a = -0.937$ (obtained by the method of Section 2.2), $\hat{\mu}_b = 0$, and $\hat{\mu}_c = 0.187$. Substituting into (C.4), we find $L^*(-0.937) = -1.4n$, $L^*(0) = -0.5n$, $L^*(0.187) = 0.4n$.

Since $L^*(m)$ is a minimum here, we conclude that L in (C.3) will be maximized by m. Therefore, $\hat{\hat{\mu}} = m$, where m is one of the extreme roots of the likelihood equation.

As a concluding example, let

$n_x = 2$, $n_y = 8$, $\bar{x} = 0.783$, $\bar{y} = -1.383$, $s_x^2 = 0.174$, $s_y^2 = 3.047$. If we substitute these values into (A.1) and solve the resulting cubic equation, we find $m = 0.475 = \hat{\mu}_c$, $\hat{\hat{\mu}} = \hat{\mu}_a = -0.937$, $\hat{\mu}_b = 0.213$, where (C.4) has been used to identify $\hat{\hat{\mu}}$.

To summarize, three examples have been taken where the likelihood equations have had three real, distinct roots represented by $\hat{\mu}_a < \hat{\mu}_b < \hat{\mu}_c$. In the first example, $m = \hat{\mu}_b \neq \hat{\hat{\mu}}$. For the second case, $m = \hat{\mu}_a = \hat{\hat{\mu}}$. In the third case, $m = \hat{\mu}_c \neq \hat{\hat{\mu}}$.

D. An Approach to the Distribution of m in the Univariate Case

The results of empirical sampling, shown in Chapter VI, and theoretical investigation, shown in Appendix Sections A and B, indicate that the likelihood equation for estimating μ in the univariate case has only one real root when the sample sizes are appreciably large, say ten or more. In such cases, the iteration estimate m is identical with the maximum likelihood estimate $\hat{\mu}$. Gart (1958) has shown that $\hat{\mu}$ is approximately normally distributed when the sample sizes are large. As an approach to the problem of determining the distribution of m, we shall investigate the distribution of m_0, the first stage iteration value of m. For equal sample sizes, Graybill and Deal (1959) have shown that m_0, defined by (2.2.9), is a uniformly better unbiased estimator of μ than is either $\bar{x}$ or $\bar{y}$, provided that the sample size is greater than ten.

To determine the moments of m_0, we shall proceed as follows: Let

(D.1) $$r = n_x s_y^2/(n_x s_y^2 + n_y s_x^2).$$

We substitute (D.1) into (2.2.9) to get

(D.2) $$m_0 = r\bar{x} + (1 - r)\bar{y}.$$

Now let $s = n_y s_y^2/\sigma_y^2$, $t = n_x s_x^2/\sigma_x^2$, $a = n_x^2/\sigma_x^2$ and

$b = n_y^2/\sigma_y^2$.

From (D.1), we have

(D.3) $$r = as/(as + bt),$$

where s and t are distributed independently as χ^2 with $n_x - 1$ and $n_y - 1$ degrees of freedom, respectively. The joint probability function of s and t is

$$(D.4)\quad h(s,t) = k\cdot e^{-(s + t)/2}\cdot s^{(n_y - 3)/2}\cdot t^{(n_x - 3)/2},$$

where k is a constant expressible as a gamma function in the form $1/k = \Gamma[(n_x - 1)/2]\cdot\Gamma[(n_y - 1)/2]$

$$\cdot 2^{(n_x + n_y - 2)/2}.$$

Also, $s = btr/a(1 - r)$, and for the Jacobian of this transformation, $J(s,t/r,t) = bt/a(1-r)^2$. Hence (D.4) yields

$$(D.5)\quad h(r,t) = k\cdot e^{-[btr/a(1 - r) + t]/2}$$
$$\cdot [btr/a(1 - r)]^{(n_y - 3)/2}$$
$$\cdot t^{(n_x - 3)/2}\cdot bt/a(1 - r)^2$$
$$= k(b/a)^{(n_x - 1)/2}\cdot e^{-[br/a(1-r) + 1]/2}$$
$$\cdot t^{(n_x + n_y - 4)/2}\cdot r^{(n_y - 3)/2}/[(1 - r)^{(n_y-3)/2}$$
$$- 1].$$

Upon integrating out the t, we get

$$(D.6)\quad f(r) = (k/2)(b/a)^{(n_y - 1)/2}[br/a(1 - r)$$
$$+ 1]^{(n_x + n_y - 2)/2}\cdot\Gamma[(n_x + n_y - 4)/2]$$
$$\cdot r^{(n_y - 3)/2}/[(1 - r)^{(n_y - 3)/2} - 1].$$

Simplifying, we obtain the probability function of r:

$$(D.7)\quad f(r) = (b/a)^{(n_y - 1)/2}\cdot r^{(n_y - 3)/2}\cdot (1 - r)^{(n_x - 3)/2}$$
$$/B[(n_x - 1)/2,\ (n_y - 1)/2]\cdot [br/a$$
$$+ (1 - r)]^{(n_x + n_y - 4)/2},$$

where B signifies the beta function.

For the special case of $a = b$ (i.e. $n_x^2\sigma_y^2 = n_y^2\sigma_x^2$), we have [See Kenney and Keeping (1951), p. 95] the beta distribution:

(D.8) $f(r) = r^{(n_y - 3)/2} \cdot (1 - r)^{(n_x - 3)/2} / B[(n_x - 1)/2, (n_y - 1)/2]$ with $0 < r < 1$.

From (D.8), we find:

$E(r) = (n_y - 1)/(n_x + n_y - 2)$, $E(1 - r) = (n_x - 1)/(n_x + n_y - 2)$, $E(r^2) = (n_y^2 - 1)/(n_x + n_y)(n_x + n_y - 2)$, and $E[(1 - r)^2] = (n_x^2 - 1)/(n_x + n_y)(n_x + n_y - 2)$.

Using (D.2) and noting that r is independent of $\bar{x}$ and $\bar{y}$, we obtain $E(m_0) = E[r\bar{x} + (1 - r)\bar{y}] = E(\bar{y}) = \mu$. In addition, $E(m_0^2) = \sigma_{\bar{x}}^2 E(r^2) + \sigma_{\bar{y}}^2 E[(1 - r)^2] + \mu^2$. So,

$$\text{(D.9)} \quad \text{Var }(m_0) = E(m_0^2) - \mu^2 = \sigma_{\bar{x}}^2 E(r^2) + \sigma_{\bar{y}}^2 E[(1 - r)^2]$$
$$= \sigma_x^2(n_y^2 - 1)/n_x(n_x + n_y)(n_x + n_y - 2)$$
$$+ \sigma_y^2(n_x^2 - 1)/n_y(n_x + n_y)(n_x + n_y - 2).$$

Since for our special case, $n_x^2\sigma_y^2 = n_y^2\sigma_x^2$, (D.9) simplifies:

$$\text{(D.10)} \quad \text{Var }(m_0) = \sigma_x^2(n_x n_y - 1)/n_x^2(n_x + n_y - 2).$$

If we impose the additional restriction that $n_x = n_y = n$

and $\sigma_x^2 = \sigma_y^2 = \sigma^2$, we have

(D.11) $\quad \mathrm{Var}\ (m_0) = \sigma^2(n + 1)/2n^2.$

For the case of $n_x^2\sigma_y^2 = n_y^2\sigma_x^2$, r has the beta distribution of (D.8). Consequently, the higher moments for m_0 may be found without difficulty for this conditional distribution. However, in the general case, exact moments for r from (D.7) are quite difficult (if not impossible) to obtain.

James (1956) has prepared tables which may be used for some half dozen cases of equal sample sizes in working with the distribution of m_0. James used $\widetilde{m}_0$, as defined by (6.1.3), and set

$$u = (\widetilde{m}_0 - \mu)/[s_x^2/(n_x - 1) + s_y^2/(n_y - 1)]^{1/2}$$

and

$$\gamma = [s_x^2/(n_x - 1)]/[s_x^2/(n_x - 1) + s_y^2/(n_y - 1)].$$

He then found a function of γ, $u(\gamma)$, such that

$$P[|u| \geq u(\gamma)] = 2\alpha.$$

Since m_0 and $\widetilde{m}_0$ are identical whenever the sample sizes are equal, the tables of James may be used to place confidence limits on μ after computing m_0.

The empirical sampling of Chapter VI throws further light on the distribution of m and m_0.

A B S T R A C T

This dissertation discusses a method of estimation with random samples drawn from two different normal populations. The two populations may be either univariate or multivariate (provided the populations have the same number of variates) and they are to be parametrically related in that the means (or vector of means) are equal. Since it is assumed that the two normal populations have a common mean (or a common vector of means), these common parameters are jointly estimated along with all the other unknown parameters. The joint estimates, called iteration estimates, are obtained by an iteration method developed for solving the likelihood equations.

A detailed study is made for the joint estimation of parameters when sampling from two univariate normal populations. The iteration procedure is based on jointly estimating the common mean and the individual variances by finding a weighted mean and the individual variances about the weighted mean. The initial weighted mean is found by taking as weights the reciprocals of the estimates of the variances of the individual estimates of the mean. It is proved that the iteration method produces an unique set of estimates which satisfies the likelihood equations. Since this set of estimates is not always identical with the set of maximum likelihood estimates, the conditions under which

the two sets may possibly differ are established. Numerical examples are given to illustrate the iteration technique and to compare the iteration estimates with maximum likelihood estimates in the cases where they differ.

Empirical sampling, with small sizes, is done with the aid of the I B M 650 Computer to obtain information regarding the distribution of the iteration estimates and also the maximum likelihood estimates in the cases where they differ. The experimental results indicate that the iteration estimate of the common mean tends to be normally distributed and the iteration estimates of the individual variances are virtually unbiased.

The iteration procedure is compared with Fisher's Method, which uses the Information Matrix, and is shown to give identical results while requiring less computation.

An extension of the iteration procedure is made to the case where the samples are drawn from two bivariate normal populations with the components of the common vector of means and the elements of the individual covariance matrices being estimated jointly. For the particular case in which the individual variances within each population may be assumed equal, it is shown that a linear transformation to obtain new uncorrelated variables will materially lessen the time required for the iteration method. A numerical example is given to illustrate the iteration technique both with and without a transformation of variables and a proof is given to show that the two methods produce identical results.

The iteration procedure is further extended to the case where the samples are drawn from two multivariate normal populations which have the same number of variates and joint estimates are obtained for the common vector of means and the individual covariance matrices. It is also shown that if a linear transformation can be found which gives new uncorrelated variables in each population, then transformation before iteration greatly reduces the computational labor involved in obtaining the joint estimates.