NBSIR 74-587

The Use of the Method of Least Squares in Calibration

J. M. Cameron

Institute for Basic Standards
National Bureau of Standards
Washington, D. C. 20234

September 1974

NBSIR 74-587

THE USE OF THE METHOD OF
LEAST SQUARES IN CALIBRATION

J. M. Cameron

Institute for Basic Standards
National Bureau of Standards
Washington, D. C. 20234

September 1974

THE USE OF THE METHOD OF LEAST SQUARES IN CALIBRATION

by

J. M. Cameron

1. Introduction

When more than one measurement is made on the same quantity, we are accustomed to taking an average and we have the feeling that the result is "better" than any single value that might be chosen from the set. Exactly why the average should be better needs some justification and the fundamental step toward a general approach to the problem of measurement was taken by Thomas Simpson in 1755. In showing the advantage of taking an average of values arising from a number of probability distributions, "he took the bold step of regarding errors, not as individual unrelated happenings, but as properties of the measurement process itself . . . He thus opened the way to a mathematical theory of measurement based on the mathematical theory of probability" [3, page 29].

The taking of an average is a special case of the method of least squares for which the original justification by Lengendre in 1805 did not involve any probability considerations but was advanced as a convenient method for the combination of observations. It was Gauss who recognized that one could not arrive at a "best" value unless the probability distribution of the measurement errors were known. In 1798 he showed the optimality of the least squares values when the underlying distribution is normal and in 1821 showed that the method of least squares leads to values of the parameters which have minimum variance among all possible unbiased linear functions* of the observations regardless of the underlying distribution. It is this property that gives the method of least squares its position of dominance among methods of combination of observations.

In this paper the statistical concepts needed for the method of least squares will be stated as a prelude to the usual modern version of the Gauss theorem. The formation of the observational equations and the derivation of the normal equations are illustrated for several situations arising in calibration. The role of restraints in the solution of systems which are not of full rank is discussed. The results are presented in a form designed to facilitate computation.

---

*An example of a nonlinear function with smaller variance than the average (the "best" linear estimator) is given by the midrange for the rectangular distribution. The midrange (average of the largest and smallest observation) has variance $1/[2(N+1)(N+2)]$ when based on n measurements, whereas the average has variance $1/12N$. Thus if $N \geq 3$, the midrange is to be preferred.

2. The Physical and Statistical Model of an Experiment

In physics, one is familiar with the construction and interpretation of the physical model of an experiment. One has a substantial body of theory on which to base such a model and one need only consider the determination of length by interferometric measurements to remind oneself of the various elements involved: a defined unit, the apparatus, the procedure, the corrections for environmental factors, etc. One realization of the experiment leads to values for the quantities of interest.

But one realizes that a repetition of the experiment will lead to different values--differences for which the physical model does not provide corrections. One is thus confronted with the need for a statistical model to account for the variations encountered in a sequence of measurements. In building the statistical model, one is first faced with the issue of what is meant by a repetition of the experiment--many readings within a few minutes or *ab initio* determinations a week apart.

The objective is to describe the output of the physical process not only in terms of the physical quantities involved but also in terms of the random variation and systematic influences due to environmental, procedural, or instrumental factors in the experiment.

3. Equation of Expected Values of the Observation

If one measured the same quantity again and again to obtain the sequence

$$y_1, y_2, \ldots y_n \ldots$$

then if the process that generates these numbers is "in control," the long run average or *limiting mean*, μ, will exist. By "in control" one means that the values of y behave as random variables from a probability distribution (for a discussion of this topic, see Eisenhart [1]). This limiting mean, μ, is usually called the *expected value* of y designated by the operator $E(\ )$ so that the statement becomes in symbols $E(y) = \mu$. Because y is regarded as a random variable one can represent it as

$$y = \mu + \varepsilon$$

where ε is the random component that follows some probability distribution with a limiting mean of zero, i.e., $E(\varepsilon) = 0$.

The quantity μ may involve one or more parameters. Consider the measurement of the difference in length of all distinct pairings of

four gage blocks, A, B, C, D. Denote the 6 measurements by y_1, y_2, . . . y_6, then one may write

$$E(y_1) = A-B$$

$$E(y_2) = A-C$$

$$E(y_3) = A-D$$

$$E(y_4) = B-C$$

$$E(y_5) = B-D$$

$$E(y_6) = C-D$$

Other representations are useful.

Observation	Expected Value: $E(y)$	Matrix Form: $X\beta$
y_1	A - B	$\begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} A \\ B \\ C \\ D \end{bmatrix}$
y_2	A - C	
y_3	A - D	
y_4	B - C	
y_5	B - D	
y_6	C - D	

Consider a sequence of measurements of the same quantity in the presence of a linear drift of Δ per observation. The expected values are thus:

Observation	Matrix Form: $X\beta$
$E(y_1) = \mu$	$\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ & \\ & \\ & \\ 1 & (n-1) \end{bmatrix} \begin{bmatrix} \mu \\ \Delta \end{bmatrix}$
$E(y_2) = \mu + \Delta$	
$E(y_3) = \mu + 2\Delta$	
.	
.	
.	
$E(y_n) = \mu + (n-1)\Delta$	

There is an alternative representation that measures the drift from the central point of the experiment so that the drift is represented by . . . -3Δ, -2Δ, $-\Delta$, 0, Δ, 2Δ, 3Δ . . . for an odd number of observations and by . . . $\frac{-5\Delta}{2}$, $\frac{-3\Delta}{2}$, $\frac{-\Delta}{2}$, $\frac{\Delta}{2}$, $\frac{3\Delta}{2}$, $\frac{5\Delta}{2}$ . . . for an even number of observations.

If, as for example with some gage blocks, the value changes approximately linearly with time; then one can represent the observation as follows:

$$\underline{\text{Expected Value } E(y)} \qquad \underline{\text{Matrix Form: } X\beta}$$

$$E(y_1) = \alpha + \beta x_1$$
$$E(y_2) = \alpha + \beta x_2$$
$$\vdots$$
$$E(y_n) = \alpha + \beta x_n$$

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ & \\ & \\ & \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$$

The sequence of measurements for the intercomparison of 4 gage blocks is as follows:

Observation	Expected Value: $E(y)$	Matrix Form: $X\beta$					
y_1	S. $-$ S.. $- 7\Delta/2$	1	-1	0	0	-7	S.
y_2	Y $-$ S. $- 5\Delta/2$	-1	0	0	1	-5	S..
y_3	X $-$ Y $- 3\Delta/2$	0	0	1	-1	-3	X
y_4	S..$-$ X $- \Delta/2$	0	1	-1	0	-1	Y
y_5	S..$-$ Y $+ \Delta/2$	0	1	0	-1	1	$\Delta/2$
y_6	Y $-$ S. $+ 3\Delta/2$	-1	0	0	1	3	
y_7	S. $-$ X $+ 5\Delta/2$	1	0	-1	0	5	
y_8	X $-$ S.. $+ 7\Delta/2$	0	-1	1	0	7	

(Note that for simplicity, $\Delta/2$ is regarded as the parameter.)
For a detailed analysis of this and related experimental arrangements, see J. M. Cameron and G. E. Hailes [1]. The notation is that used in [1] where S. and S.. refer to reference standards and X and Y are the objects being calibrated.

4

If, as often occurs in the intercomparison of electrical standards, the comparator has a left-right polarity effect, this can be represented as an additive effect, α, as shown below for the intercomparison of 5 standards.

Observation	Expected Value: E(y)	Matrix Form: $X\beta$

$$
\begin{array}{lll}
y_1 & A - B \qquad\qquad + \alpha \\
y_2 & \quad B - C \qquad\quad + \alpha \\
y_3 & \qquad C - D \quad\; + \alpha \\
y_4 & \qquad\quad D - E + \alpha \\
y_5 & -A \qquad\qquad + E + \alpha \\
y_6 & -A \qquad + D \qquad + \alpha \\
y_7 & \quad B \quad - D \qquad + \alpha \\
y_8 & \quad - B \qquad\quad + E + \alpha \\
y_9 & \qquad C \quad\; - E + \alpha \\
y_{10} & A \quad - C \qquad\quad + \alpha
\end{array}
\qquad
\begin{bmatrix}
1 & -1 & 0 & 0 & 0 & 1 \\
0 & 1 & -1 & 0 & 0 & 1 \\
0 & 0 & 1 & -1 & 0 & 1 \\
0 & 0 & 0 & 1 & -1 & 1 \\
-1 & 0 & 0 & 0 & 1 & 1 \\
-1 & 0 & 0 & -1 & 0 & 1 \\
0 & 1 & 0 & -1 & 0 & 1 \\
0 & -1 & 0 & 0 & 1 & 1 \\
0 & 0 & 1 & 0 & -1 & 1 \\
1 & 0 & -1 & 0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
A \\ B \\ C \\ D \\ E \\ \alpha
\end{bmatrix}
$$

4. Statistical Independence

The sequence of differences from a zero measurement, y_o,

A: $y_1 - y_0, \; y_2 - y_0, \; y_3 - y_0, \ldots, y_n - y_0, \ldots$

are clearly dependent because an error in y_o will be common to all. Similarly, the successive differences

B: $y_2 - y_1, \; y_3 - y_2, \ldots, y_n - y_{n-1}, \ldots$

will be correlated in pairs because an error in y_n affects both the (n-1)st and n-th difference.

5

If it is assumed in both cases that each y_i has the form $\mu_i = \mu_i + \epsilon_i$ where $E(\epsilon_i) = 0$, Var $(\epsilon_i) = \sigma^2$ and cov $(\epsilon_i, \epsilon_j) = 0$, then the variance of the differences for sequence A is, as one would expect,

$$V(y_i - y_0) = 2\sigma^2$$

and the covariance of two differences is

$$\text{cov } (y_i - y_0, \ y_j - y_0) = E[(\epsilon_i - \epsilon_0)(\epsilon_j - \epsilon_0)] = E(\epsilon_0^2) = \sigma^2$$

because terms of the form $E(\epsilon_i, \epsilon_j) = 0$

For sequence B the variance is also $V(y_i - y_{i-1}) = 2\sigma^2$ and the covariance terms are

$$\text{cov}(y_i - y_{i-1}, \ y_j - y_{j-1}) = E[(\epsilon_i - \epsilon_{i-1})(\epsilon_j - \epsilon_{j-1})] = \begin{cases} 0 & \text{if } |i-j| \geq 2 \\ -\sigma^2 & \text{if } |i-j| = 1 \end{cases}$$

These variance-covariance relationships can be represented in matrix form:

Sequence A: $V = \begin{bmatrix} 2 & 1 & 1 & . & . & . & 1 \\ 1 & 2 & 1 & . & . & . & 1 \\ & & . & & & & \\ & & . & & & & \\ & & . & & & & \\ 1 & 1 & 1 & . & . & . & 2 \end{bmatrix} \sigma^2$ Sequence B: $V = \begin{bmatrix} 2 & -1 & 0 & 0 & . & . & . & 0 \\ -1 & 2 & -1 & 0 & . & . & . & 0 \\ & . & . & & & & & \\ & & . & & & & & \\ & & . & & & & & \\ 0 & 0 & 0 & 0 & . & . & . & 2 \end{bmatrix} \sigma^2$

All are familiar with the phenomenon of much closer agreement among measurements taken immediately after each other when compared to a sequence of values taken days or weeks apart. The simplest statistical model for this case is that each day has its own limiting mean, $\mu_i = \mu + \delta_i$, where $E(\delta_i) = 0$, $\text{Var}(\delta_i) = \sigma_B^2$, $\text{Cov}(\delta_i, \delta_j) = 0$, and the successive values on each day have the form

$$y_{ij} = \mu_i + \epsilon_{ij} = \mu + \delta_i + \epsilon_{ij}$$

where $E(\epsilon_{ij}) = 0$, $\text{Var}(\epsilon_{ij}) = \sigma_W^2$, $\text{Cov}(\epsilon_{ij}, \epsilon_{k\ell}) = 0$, and $\text{Cov}(\epsilon_{ij}, \delta_k) = 0$.

These three examples serve to illustrate the point that the physical conduct of the experiment is the essential element in dictating the appropriate statistical analysis. In all three cases the correlation among the variables vitiates the usual formula: standard deviation of the mean = $(1/\sqrt{n})$ standard deviation. (See Appendix, Section 1(b).)

It is in the physical conduct of the experiment that one has to build in the independence of the measurements. For Sequence A one could remeasure the zero setting each time or in Sequence B, make an independent duplicate measurement. Ordinarily this is too much of an expense to pay to achieve uncorrelated variables just for a simpler analysis.

Statistical independence is to be desired in the sense that if the successive measurements are highly correlated, then many measurements are only slightly better than a single one. The really important issue is that the proper statistical model be used so that the results are valid.

5. Normal Equations For the Method of Least Squares (independent random variables)

When there are more observations than parameters, the "best" (in the sense of minimum variance) linear unbiased estimates for the parameters are given by the so-called least squares estimators. For example, assume one has the problem of deriving values for A, B, C, and D from the following measurements.

Measurements	Expected Value: E(y)	Matrix Form: $X\beta$
y_1	A	
y_2	B	
y_3	C	
y_4	D	
y_5	A + B	
y_6	B + C	
y_7	C + D	
y_8	D + A	

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} A \\ B \\ C \\ D \end{bmatrix}$$

7

An obvious estimator, $\tilde{A}$, is the average of the three values,

	Expected Value
y_1	A
$y_5 - y_2$	$(A+B)-B$
$y_8 - y_4$	$(A+D)-D$

so that, assuming independent measurements with variance, σ^2,

$$\tilde{A} = \tfrac{1}{3}(y_1 + y_5 - y_2 + y_8 - y_4)$$

$$Var(\tilde{A}) = \tfrac{5}{9}\sigma^2$$

The least squares estimator is obtained by forming the normal equations (see Appendix, Section 2).

$$3A + B + \qquad + D = y_1 + y_5 + y_8$$

$$A + 3B + C \qquad = y_2 + y_6 + y_5$$

$$B + 3C + D = y_3 + y_7 + y_6$$

$$A \qquad + C + 3D = y_4 + y_8 + y_7$$

The solution gives the following estimators for the parameters.

$$\hat{A} = (7y_1 - 3y_2 + 2y_3 - 3y_4 + 4y_5 - y_6 - y_7 + 4y_8)/15$$

$$\hat{B} = (-3y_1 + 7y_2 - 3y_3 + 2y_4 + 4y_5 + 4y_6 - y_7 - y_8)/15$$

$$\hat{C} = (2y_1 - 3y_2 + 7y_3 - 3y_4 - y_5 + 4y_6 + 4y_7 - y_8)/15$$

$$\hat{D} = (-3y_1 + 2y_2 - 3y_3 + 7y_4 - y_5 - y_6 + 4y_7 + 4y_8)/15$$

Using formula (1.11) of Appendix, gives

$$Var(\hat{A}) = 105\sigma^2/225 = 21\sigma^2/45 = 7\sigma^2/15$$

which can be compared to the variance of $\tilde{A}$ which was $25\sigma^2/45$. The Gauss theorem on least squares guarantees that no other linear unbiased estimator will have smaller variance.

In matrix form one has

$$(X'X)\hat{\beta} = \begin{bmatrix} 3 & 1 & 0 & 1 \\ 1 & 3 & 1 & 0 \\ 0 & 1 & 3 & 1 \\ 1 & 0 & 1 & 3 \end{bmatrix} \begin{bmatrix} A \\ B \\ C \\ D \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_8 \end{bmatrix}$$

$$\hat{\beta} = \frac{1}{15} \begin{bmatrix} 7 & -3 & 2 & -3 \\ -3 & 7 & -3 & 2 \\ 2 & -3 & 7 & -3 \\ -3 & 2 & -3 & 7 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \end{bmatrix} y$$

$$\hat{\beta} = \frac{1}{15} \begin{bmatrix} 7 & -3 & 2 & -3 & 4 & -1 & -1 & 4 \\ -3 & 7 & -3 & 2 & 4 & 4 & -1 & -1 \\ 2 & -3 & 7 & -3 & -1 & 4 & 4 & -1 \\ -3 & 2 & -3 & 7 & -1 & -1 & 4 & 4 \end{bmatrix} y$$

When only differences among a group of objects (such as gage blocks, voltage cells, etc.) are measured the normal equation will not be of full rank so that a unique solution will not exist. For the design involving differences between all distinct pairings of objects the normal equations are, for the case of 4 objects discussed in Section 3,

$$3A - B - C - D = y_1 + y_2 + y_3 = q_1$$

$$-A + 3B - C - D = -y_1 + y_4 + y_5 = q_2$$

$$-A - B + 3C - D = -y_2 - y_4 + y_6 = q_3$$

$$-A - B - C + 3D = -y_3 - y_5 - y_6 = q_4$$

Or in matrix form:

$$X'X\hat{\beta} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 & 1 & 0 \\ 0 & -1 & 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 0 & -1 & -1 \end{bmatrix} \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix} \beta = \begin{bmatrix} 3 & -1 & -1 & -1 \\ -1 & 3 & -1 & -1 \\ -1 & -1 & 3 & -1 \\ -1 & -1 & -1 & 3 \end{bmatrix} \beta = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 & 1 & 0 \\ 0 & -1 & 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 0 & -1 & -1 \end{bmatrix} y$$

which can be seen not to be of full rank because the sum of the four equations is zero.

One needs a baseline to which the differences can be referred--a restraint to bring the system of equations up to full rank. If one of the objects were designated as the standard, or if a number (or all) of them were regarded as a reference group whose value was known, values for the items could be obtained.

If the restraint $A = K_0$ is invoked, the normal equations become (using the methods of Appendix, Section 3)

$$\begin{cases} 3A - B - C - D + \lambda = q_1 \\ -A + 3B - C - D = q_2 \\ -A - B + 3C - D = q_3 \\ -A - B - C + 3D = q_4 \\ A = K_0 \end{cases}$$

$$\begin{bmatrix} 3 & -1 & -1 & -1 & 1 \\ -1 & 3 & -1 & -1 & 0 \\ -1 & -1 & 3 & -1 & 0 \\ -1 & -1 & -1 & 3 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{\lambda} \end{bmatrix} = \begin{bmatrix} X'y \\ K_0 \end{bmatrix}$$

The solution is given by

$\hat{A} = K$

$\hat{B} = K + (-2y_1 - y_2 - y_3 + y_4 + y_5)/4$

$\hat{C} = K + (-y_1 - 2y_2 - y_3 - y_4 + y_6)/4$

$\hat{D} = K + (-y_1 - y_2 - 2y_3 - y_5 - y_6)/4$

$\lambda = 0$

$$\begin{bmatrix} \hat{\beta} \\ \hat{\lambda} \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 0 & 0 & 0 & 0 & 4 \\ 0 & 2 & 1 & 1 & 4 \\ 0 & 1 & 2 & 1 & 4 \\ 0 & 1 & 1 & 2 & 4 \\ 4 & 4 & 4 & 4 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & -1 & 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y \\ K_0 \end{bmatrix}$$

$$\begin{bmatrix} \hat{\beta} \\ \hat{\lambda} \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 4 \\ -2 & -1 & -1 & 1 & 1 & 0 & 4 \\ -1 & -2 & -1 & -1 & 0 & 1 & 4 \\ -1 & -1 & -2 & 0 & -1 & -1 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} y \\ K_0 \end{bmatrix}$$

The variances of the values are $V(\hat{A}) = 0$; $V(\hat{B}) = V(\hat{C}) = V(\hat{D}) = \sigma^2/2$.

If the restraint $A + B + C + D = K_1$ is invoked, the normal equations become

$$\begin{array}{rcl}
3A - B - C - D + \lambda &=& q_1 \\
-A + 3B - C - D + \lambda &=& q_2 \\
-A - B + 3C - D + \lambda &=& q_3 \\
-A - B - C + 3D + \lambda &=& q_4 \\
A + B + C + D &=& K_1
\end{array}
\qquad
\begin{bmatrix} 3 & -1 & -1 & -1 & 1 \\ -1 & 3 & -1 & -1 & 1 \\ -1 & -1 & 3 & -1 & 1 \\ -1 & -1 & -1 & 3 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{\lambda} \end{bmatrix} = \begin{bmatrix} X'y \\ K_1 \end{bmatrix}$$

and the solution is given by

$\hat{A} = (y_1 + y_2 + y_3 + K_1)/4$

$\hat{B} = (-y_1 + y_4 + y_5 + K_1)/4$

$\hat{C} = (-y_2 - y_4 + y_6 + K_1)/4$

$\hat{D} = (-y_3 - y_5 - y_6 + K_1)/4$

$\lambda = 0$

$$\begin{bmatrix} \hat{\beta} \\ \hat{\lambda} \end{bmatrix} = \frac{1}{16} \begin{bmatrix} 3 & -1 & -1 & -1 & 4 \\ -1 & 3 & -1 & -1 & 4 \\ -1 & -1 & 3 & -1 & 4 \\ -1 & -1 & -1 & 3 & 4 \\ 4 & 4 & 4 & 4 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & -1 & 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y \\ K_1 \end{bmatrix}$$

$$= \frac{1}{16} \begin{bmatrix} 4 & 4 & 4 & 0 & 0 & 0 & 4 \\ -4 & 0 & 0 & 4 & 4 & 0 & 4 \\ 0 & -4 & 0 & -4 & 0 & 4 & 4 \\ 0 & 0 & -4 & 0 & -4 & -4 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} y \\ K_1 \end{bmatrix}$$

The variances of the values are $V(\hat{A}) = V(\hat{B}) = V(\hat{C}) = V(\hat{D}) = 3\sigma^2/16$.

11

Although it is a simple matter to change the reference point for the parameters (i.e., change the restraint) after one solution has been found, the corresponding change of variances for the parameter values should not be ignored. These variances are given by the diagonal terms of the inverse of the matrix of normal equation, the inverse being indicated by double brackets in these examples. The difference in variance for β in the last example, arises from the fact that in the first case one is concerned only with the difference between A (the standard) and B, whereas in the second case it is the difference between B and the average of the others that is involved.

For completeness, the matrices of normal equations and their inverses for the examples of Section 3 are shown below.

Linear Drift

$$X = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ \cdot \\ \cdot \\ 1 & (n-1) \end{bmatrix} \qquad X'X = \begin{bmatrix} n & n(n-1)/2 \\ n(n-1)/2 & n(n-1)(2n-1)/6 \end{bmatrix}$$

$$(X'X)^{-1} = \frac{12}{n^2(n^2-1)} \begin{bmatrix} n(n-1)(2n-1)/6 & -n(n-1)/2 \\ -n(n-1)/2 & n \end{bmatrix}$$

y a linear function of x

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \cdot \\ \cdot \\ 1 & x_n \end{bmatrix} \qquad X'X = \begin{bmatrix} n & \Sigma x \\ \Sigma x & \Sigma x^2 \end{bmatrix} \qquad (X'X)^{-1} = \frac{1}{n\Sigma x^2 - (\Sigma x)^2} \begin{bmatrix} \Sigma x^2 & -\Sigma x \\ -\Sigma x & n \end{bmatrix}$$

Gage block design

$$X = \begin{bmatrix} 1 & -1 & 0 & 0 & -7 \\ -1 & 0 & 0 & 1 & -5 \\ 0 & 0 & 1 & -1 & -3 \\ 0 & 1 & -1 & 0 & -1 \\ 0 & 1 & 0 & -1 & 1 \\ -1 & 0 & 0 & 1 & 3 \\ 1 & 0 & -1 & 0 & 5 \\ 0 & -1 & 1 & 0 & 7 \end{bmatrix}$$

$$\begin{bmatrix} X'X & B \\ B' & 0 \end{bmatrix} = \begin{bmatrix} 4 & -1 & -1 & -2 & 0 & 1 \\ -1 & 4 & -2 & -1 & 0 & 1 \\ -1 & -2 & 4 & -1 & 0 & 0 \\ -2 & -1 & -1 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 168 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} X'X & B \\ B' & 0 \end{bmatrix}^{-1} = \frac{1}{336} \begin{bmatrix} 35 & -35 & -7 & 7 & 0 & 168 \\ -35 & 35 & 7 & -7 & 0 & 168 \\ -7 & 7 & 91 & 21 & 0 & 168 \\ 7 & -7 & 21 & 91 & 0 & 168 \\ 0 & 0 & 0 & 0 & 2 & 0 \\ 168 & 168 & 168 & 168 & 0 & 0 \end{bmatrix}$$

Intercomparison of 5 standards (Sum of all used as restraint)

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

$$\begin{bmatrix} X'X & B \\ B' & 0 \end{bmatrix} = \begin{bmatrix} 4 & -1 & -1 & -1 & -1 & 0 & 1 \\ -1 & 4 & -1 & -1 & -1 & 0 & 1 \\ -1 & -1 & 4 & -1 & -1 & 0 & 1 \\ -1 & -1 & -1 & 4 & -1 & 0 & 1 \\ -1 & -1 & -1 & -1 & 4 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 10 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} X'X & B \\ B' & 0 \end{bmatrix}^{-1} = \frac{1}{25}\begin{bmatrix} 4 & -1 & -1 & -1 & 1 & 0 & 5 \\ -1 & 4 & -1 & -1 & -1 & 0 & 5 \\ -1 & -1 & 4 & -1 & -1 & 0 & 5 \\ -1 & -1 & -1 & 4 & -1 & 0 & 5 \\ -1 & -1 & -1 & -1 & 4 & 0 & 5 \\ 0 & 0 & 0 & 0 & 0 & 5/2 & 0 \\ 5 & 5 & 5 & 5 & 5 & 0 & 0 \end{bmatrix}$$

6. Standard Deviation

By substituting the computed values for the parameters into the equations of expected values for the observation, one has a *predicted value* to compare to the actual observation. The difference, d, between the observed and predicted value is called the *deviation* and is used to determine an estimate, s, of the standard deviation, σ, of the process

$$s = \sqrt{\frac{\Sigma d_i^2}{n-k+m}}$$

where n is the number of measurements, k is the number of parameters and m is the number of restraints.

Ordinarily one has available a sequence of values of the standard deviation say s_1, s_2, s_3, . . . , s_n based on ν_1, ν_2, ν_3, . . . , ν_n degrees of freedom. One forms the estimate of σ by combining these in quadrature

$$\hat{\sigma} = \sqrt{\frac{\nu_1 s_1^2 + \nu_2 s_2^2 + \ldots + \nu_n s_n^2}{\nu_1 + \nu_2 + \ldots + \nu_n}}$$

with degrees of freedom $N = \Sigma\nu$. In assigning a standard deviation to the parameters or linear combinations of them, the value $\hat{\sigma}$ is used rather than the value of s from a single experiment.

The variance of the sums of two parameter values is given by adding the corresponding diagonal terms (variances) in the inverse of the matrix of normal equations and the appropriate off diagonal terms (covariances) and multiplying by σ^2. For the case of the intercomparison of 5 standards given at the end of Section 5:

14

$$\text{s.d. } (\hat{A}+\hat{B}) = \sqrt{\sigma_A^2 + \sigma_B^2 + 2\sigma_{AB}} = \frac{\sigma}{\sqrt{25}}\sqrt{[4+4+2(-1)]} = \frac{\sigma\sqrt{6}}{5}$$

For the variance of the difference, the covariance terms enter negatively so that for the same example

$$\text{s.d.}(\hat{A}-\hat{B}) = \sqrt{\sigma_A^2 + \sigma_B^2 - 2\sigma_{AB}} = \frac{\sigma}{\sqrt{25}}\sqrt{[4+4-2(-1)]} = \frac{\sigma\sqrt{10}}{5}$$

For other linear combinations, formula 1.10-M of the Appendix would be used.

For the linear function example, the predicted value of y for x_0 is $\hat{y}_0 = \hat{\alpha} + \hat{\beta}x_0$ which has a variance of

$$[1 \quad x_0]\begin{bmatrix} C_{11} & C_{12} \\ C_{12} & C_{22} \end{bmatrix}\begin{bmatrix} 1 \\ x_0 \end{bmatrix}\sigma^2 = (C_{11} + x_0^2 C_{22} + 2x_0 C_{12})\sigma^2$$

where the terms C_{11}, C_{12}, C_{22} are the elements of $(X'X)^{-1}$ given in Section 5 for the case of y as a linear function of x.

7. Correlated Measurements

In the previous section it was assumed that the observations were uncorrelated, i.e., that $V(y_i) = \sigma^2$, $\text{cov}(y_i, y_j) = 0$ or in matrix form $V = \text{Var}(y) = \sigma^2 I$ where I is the identity matrix. Section 4 of the Appendix discusses the general case where one knows the matrix, V, of variances and covariances for the observations.

Quite often a transformation of variables can be achieved to obtain variables that are uncorrelated. A simple example is provided by the case of cummulative errors, i.e., in the case where

$$y_1 = \mu_1 + \varepsilon_1$$

$$y_2 = \mu_2 + \varepsilon_1 + \varepsilon_2$$

$$y_3 = \mu_3 + \varepsilon_1 + \varepsilon_2 + \varepsilon_3$$

The variance covariance matrix of the y's assuming $E(\varepsilon_i) = 0$, $\text{Var}(\varepsilon) = \sigma^2$, $\text{cov}(\varepsilon_i \varepsilon_j) = 0$ is given by

$$V = \sigma^2 \begin{bmatrix} 1 & 1 & 1 & .. & 1 \\ 1 & 2 & 2 & & 2 \\ 1 & 2 & 3 & & 3 \\ \cdot & & & & \\ \cdot & & & & \\ 1 & 2 & 3 & .. & n \end{bmatrix}$$

If one transforms to variables x_i where

$$x_1 = y_1 \qquad = \mu_1 + \varepsilon_1$$

$$x_2 = y_2 - y_1 \quad = \mu_2 - \mu_1 + \varepsilon_2$$

$$x_3 = y_3 - y_2 \quad = \mu_3 - \mu_2 + \varepsilon_3$$

$$\cdot$$

$$\cdot$$

$$x_n = y_n - y_{n-1} = \mu_n - \mu_{n-1} + \varepsilon_n$$

The expected values and variances become

$$E(X) = \begin{bmatrix} \mu_1 \\ \mu_2 - \mu_1 \\ \cdot \\ \cdot \\ \mu_n - \mu_{n-1} \end{bmatrix} \qquad V(X) = \begin{bmatrix} \sigma^2 & 0 & . & . & 0 \\ 0 & \sigma^2 & & & 0 \\ \cdot & & & & \\ \cdot & & & & \\ 0 & 0 & & & \sigma^2 \end{bmatrix}$$

In matrix form $X = Ty$ where $T = \begin{bmatrix} 1 & 0 & 0 & . & . & 0 & 0 \\ -1 & 1 & 0 & & & 0 & 0 \\ 0 & -1 & 1 & & & 0 & 0 \\ \cdot & & & & & & \\ \cdot & & & & & & \\ 0 & 0 & 0 & & & -1 & 1 \end{bmatrix}$

16

and if one computes $\text{Var}(Ty) = TVT'$, one gets

$$\text{Var}(Ty) = \begin{bmatrix} 1 & 0 & 0 & . & . \\ -1 & 1 & 0 & & \\ 0 & -1 & 1 & & \\ . & & & & \\ . & & & & \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & . & . & . & 1 \\ 1 & 2 & 2 & & & & 2 \\ 1 & 2 & 3 & & & & 3 \\ . & & & & & & \\ . & & & & & & \end{bmatrix} \begin{bmatrix} 1 & -1 & 0 & . & . \\ 0 & 1 & -1 & & \\ 0 & 0 & 1 & & \\ . & & & & \\ . & & & & \end{bmatrix} \quad \sigma^2 = \sigma^2 I$$

REFERENCES

(1) Cameron, J. M., and Hailes, G. E., "Designs for the Calibration of Small Groups of Standards in the Presence of Instrumental Drift," NBS Technical Note 844, Aug. 1974.

(2) Eisenhart, C., "Realistic Evaluation of the Precision and Accuracy of Instrument Calibration Systems," NBS J. Research, 67C (1963), 161-187.

(3) Eisenhart, C., "The Meaning of 'Least' in Least Squares," J. Wash. Acad. Sci., 54 (1964), 24-33.

(4) Goldman, A. J., and Zelen, M., "Weak Generalized Inverses and Minimum Variances Linear Unbiased Estimation," NBS J. Research 68B (1964) 151-172.

(5) Zelen, M., "Linear Estimation and Related Topics," Chapter 17 of Survey of Numerical Analysis, edited by J. Todd, McGraw Hill, New York, 1957, 558-584.

APPENDIX: FORMULAS FROM STATISTICS

1. Background and Notation

(a) Expected Value

The expected value, μ, of a random variable, y, will be written

$$E(y) = \mu$$

The mean μ may represent a linear function of some basic parameters $\beta_1, \beta_2, \ldots \beta_k$ with known coefficients $x_1, x_2, \ldots, x_k$

$$E(y) = \mu = x_1\beta_1 + x_2\beta_2 + \ldots + x_k\beta_k$$

The expected value of n observed Values $y_1, y_2, \ldots, y_n$ can then be written

$$E(y_1) = x_{11}\beta_1 + x_{12}\beta_2 + \ldots + x_{1k}\beta_k \qquad (1.1)$$

$$E(y_2) = x_{21}\beta_1 + x_{22}\beta_2 + \ldots + x_{2k}\beta_k$$

$$\cdot$$
$$\cdot$$
$$\cdot$$

$$E(y_n) = x_{n1}\beta_1 + x_{n2}\beta_2 + \ldots + x_{nk}\beta_k$$

This may be written in matrix notation as

$$\begin{bmatrix} E(y_1) \\ E(y_2) \\ \cdot \\ \cdot \\ \cdot \\ E(y_n) \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ & & & \\ & & & \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \\ \beta_k \end{bmatrix} \qquad (1.1\text{-M})$$

or as $E(y) = X\beta$

where the vectors y and β and the matrix, X, are easily identified.

(b) Variance, Covariance

The variance, σ_i^2, of a random variable, y_i, is defined as

$$\sigma_i^2 = E\{(y_i - \mu_i)^2\} = E(y_i^2) - 2\mu_i E(y_i) + \mu_i^2 = E(y_i^2) - \mu_i^2 \qquad (1.2)$$

and the covariance σ_{ij} of the variables y_i and y_j by

$$\sigma_{ij} = E\{(y_i - \mu_i)(y_j - \mu_j)\} = E(y_i y_j) - \mu_i \mu_j \qquad (1.3)$$

The variance of cy where c is some constant is

$$\text{Var}(cy) = E\{(cy - c\mu)^2\} = c^2 \sigma^2 \qquad (1.4)$$

The variance of a sum of two variables

$$\text{Var}(y_1 + y_2) = E\{[y_1 + y_2 - (\mu_1 + \mu_2)]^2\} = E\{[(y_1 - \mu_1) + (y_2 - \mu_2)]^2\}$$

$$= E(y_1 - \mu_1)^2 + E(y_2 - \mu_2)^2 + 2E\{(y_1 - \mu_1)(y_2 - \mu_2)\}$$

$$= \sigma_1^2 + \sigma_2^2 + 2\sigma_{12} \qquad (1.5)$$

which we may write as

$$\sigma_1^2 + \sigma_2^2 + 2\sigma_{12} = [1 \quad 1] \begin{bmatrix} \sigma_1^2 + \sigma_{12} \\ \sigma_{12} + \sigma_2^2 \end{bmatrix} = [1 \quad 1] \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad (1.5\text{-M})$$

For independent random variables $\sigma_{ij} = 0$ and

$$\text{Var}(\Sigma y_i) = \Sigma \sigma_i^2 \qquad (1.6)$$

EXAMPLE:

$$\text{Var}(ay_1 + by_2 + cy_3) = E\{[(ay_1 - a\mu_1) + (by_2 - b\mu_2) + (cy_3 - c\mu_3)]^2\}$$

$$= a^2 \sigma_1^2 + b^2 \sigma_2^2 + c^2 \sigma_3^2 + 2ab\sigma_{12} + 2ac\sigma_{13} + 2bc\sigma_{23} \qquad (1.7)$$

which may be written as

$$[a \quad b \quad c] \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} \qquad (1.7\text{-M})$$

20

(c) Linear Function of Random Variables

A linear function

$$L = a_1 y_1 + a_2 y_2 + \ldots + a_n y_n \tag{1.8}$$

has expected value

$$E(L) = a_1 E(y_1) + a_2 E(y_2) + \ldots + a_n E(y_n) \tag{1.9}$$

or in matrix notation

$$E(L) = (a_1 \; a_2 \; \ldots \; a_n) \begin{bmatrix} E(y_1) \\ E(y_2) \\ E(y_n) \end{bmatrix} = a'\mu \tag{1.9-M}$$

The variance is given, by analogy with (1.7) by

$$V(L) = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_n \end{bmatrix} \tag{1.10-M}$$

which reduces to the usual formula

$$V(\Sigma a_i y_i) = \Sigma a_i^2 \sigma_i^2 \tag{1.11}$$

if $\sigma_{ij} = 0$.

For two linear functions L_1 and L_2 the covariance term is given by

$$E\{[a_1(y_1-\mu_1) + \ldots\ldots + a_n(y_n-\mu_n)][b_1(y_1-\mu_1) + \ldots\ldots + b_n(y_n-\mu_n)]\}$$

$$= a_1 b_1 E(y_1-\mu_1)^2 + a_2 b_2 E(y_2-\mu_2)^2 + \ldots + a_n b_n E(y_n-\mu_n)^2$$

$$+ (a_1 b_2 + a_2 b_1)E(y_1-\mu_1)(y_2-\mu_2) + (a_1 b_3 + a_3 b_1)E(y_1-\mu_1)(y_3-\mu_3)$$

$$+ (a_2 b_3 + a_3 b_2)E(y_2-\mu_2)(y_3-\mu_3) + \ldots$$

This reduces to the usual formulas:

$$\text{If } \sigma_{ij} = 0 \qquad \text{then } \text{Cov}(L_1, L_2) = \Sigma a_i b_i \sigma_i^2 \qquad (1.12)$$

$$\text{If } \sigma_i = \sigma \qquad \text{then } \text{Cov}(L_1, L_2) = \sigma^2 \Sigma a_i b_i$$

For the case of $L_1 = a_1 y_1 + a_2 y_2 + a_3 y_3$ and $L_2 = b_1 y_1 + b_2 y_2 + b_3 y_3$, the covariance can be written:

$$(a_1\ a_2\ a_3) \begin{bmatrix} b_1\sigma_1^2 + b_2\sigma_{12} + b_3\sigma_{13} \\ b_2\sigma_2^2 + b_1\sigma_{12} + b_3\sigma_{23} \\ b_3\sigma_3^2 + b_1\sigma_{13} + b_2\sigma_{23} \end{bmatrix} = (a_1\ a_2\ a_3) \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \qquad (1.12\text{-M})$$

giving the general formula for the variance and covariance of two linear functions

$$\begin{bmatrix} a_1 & a_2 & \cdots & a_n \\ b_1 & b_2 & \cdots & b_n \end{bmatrix} \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ \sigma_{1n} & \sigma_{2n} & \cdots & \sigma_n^2 \end{bmatrix} \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ a_n & b_n \end{bmatrix} \qquad (1.13\text{-M})$$

or in general for p such function, i.e., for a pxn matrix A

$$\text{Var}(AY) = AVA' \qquad (1.14\text{-M})$$

(d) <u>Quadratic Forms in Random Variables</u>

We have from (1.2)

$$E(y^2) = \sigma^2 + \mu^2 \qquad (1.15)$$

We wish to extend this to include the case of a more general quadratic expression in the y's, consider for example

$$E[(ay_1 + by_2)^2] = Ea^2 y_1^2 + Eb^2 y_2 + 2ab E(y_1 y_2)$$

$$= a^2\sigma_1^2 + a^2\mu_1^2 + b^2\sigma_2^2 + b^2\mu_2^2 + 2ab\mu_1\mu_2 + 2ab\sigma_{12}$$

22

which may be displayed as a matrix product as follows:

$$[y_1 \; y_2] \begin{bmatrix} a^2 & ab \\ ab & b^2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = [\mu_1 \; \mu_2] \begin{bmatrix} a^2 & ab \\ ab & b^2 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + [a \; b] \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$$

This example illustrates the general formula:

$$E[(a_1 y_1 + \ldots + a_n y_n)^2] = E \left\{ [y_1 \; y_2 \cdot \cdot \; y_n] \begin{bmatrix} a_1^2 & a_1 a_2 & \cdot \cdot & a_1 a_n \\ a_2 a_1 & a_2^2 & \cdot \cdot & a_2 a_n \\ \cdot & & & \\ \cdot & & & \\ a_n a_1 & a_n a_2 & \cdot \cdot & a_n^2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_n \end{bmatrix} \right\}$$

$$= [\mu_1 \cdot \cdot \; \mu_n] \begin{bmatrix} a_1^2 & \cdot \cdot \cdot & a_{1n} \\ \cdot & & \\ \cdot & & \\ a_{1n} & \cdot \cdot \cdot & a_n^2 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \cdot \\ \cdot \\ \mu_n \end{bmatrix} + [a_1 \cdot \cdot \; a_n] \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdot \cdot & \sigma_{1n} \\ \cdot & & \\ \cdot & & \\ \sigma_{1n} & \sigma_{2n} & \cdot & \sigma_n^2 \end{bmatrix} \begin{bmatrix} a_1 \\ \cdot \\ \cdot \\ a_n \end{bmatrix}$$

or

$$E\{y'Ay\} = \mu'A\mu + a'Va \qquad\qquad (1.16\text{-}M)$$

where $A = \begin{bmatrix} a_1^2 & a_1 a_2 & \cdot \cdot \cdot \\ a_1 a_2 & a_2^2 & \cdot \cdot \cdot \\ \cdot & & \\ \cdot & & \end{bmatrix}$ and $V = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdot \cdot \cdot \\ \sigma_{12} & \sigma_2^2 & \cdot \cdot \cdot \\ \cdot & & \\ \cdot & & \end{bmatrix}$

The last term can be replaced by the trace of AV so that we have

$$E(Y'AY) = \mu'A\mu + \text{Trace}(AV) \qquad\qquad (1.17\text{-}M)$$

For an excellent treatment of these statistical topics one should consult Zelen [5].

23

2. The Gauss Theorem on Least Squares (Independent, Equal Variance, Full Rank)

Let the n observations $y_1, y_2, \ldots, y_n$ have expected values

$$E(y_1) = x_{11}\beta_1 + x_{12}\beta_2 \cdots x_{1k}\beta_k$$

$$E(y_2) = x_{21}\beta_1 + x_{22}\beta_2 \cdots x_{2k}\beta_k$$

$$\cdot$$
$$\cdot$$
$$\cdot$$

(2.1)

$$E(y_n) = x_{n1}\beta_1 + x_{n2}\beta_2 \cdots x_{nk}\beta_k$$

and be statistically independent with common variance, σ^2. These two conditions can be expressed in matrix form as follows:

$$E(y) = \begin{bmatrix} E\ y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & & x_{1k} \\ x_{21} & x_{22} & & x_{2k} \\ \cdot & \cdot & & \\ \cdot & \cdot & & \\ \cdot & \cdot & & \\ x_{n1} & x_{n2} & & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \cdot \\ \cdot \\ \beta_k \end{bmatrix} = X\beta$$

(2.1-M)

$$v(y) = \begin{bmatrix} \sigma^2 & 0 & \ldots & 0 \\ 0 & \sigma^2 & \ldots & 0 \\ 0 & 0 & \ldots & \sigma^2 \end{bmatrix} = \sigma^2 I$$

The Gauss theorem states that the minimum variance unbiased linear estimator of any linear function, L, of the parameters, $\beta_1\ \beta_2 \ldots \beta_k$, say

$$L = a_1\beta_1 + a_2\beta_2 + \ldots a_k\beta_k$$

is given by substituting the values of β_i which minimize

$$Q = \Sigma[y_i - (x_{i1}\beta_1 + \ldots + x_{ik}\beta_k)]^2 \qquad (2.2)$$

considered as a function of the β_i. These values, $\hat{\beta}_1$, $\hat{\beta}_2$ . . . $\hat{\beta}_k$ are the solutions to the k equations, called the *normal equations*.

$$\Sigma x_{i1}^2 \hat{\beta}_1 + \Sigma x_{i1}x_{i2}\hat{\beta}_2 + \ldots + \Sigma x_{i1}x_{ik}\hat{\beta}_k = \Sigma x_{i1}y_i$$

$$\Sigma x_{i2}x_{i1}\hat{\beta}_1 + \Sigma x_{i2}^2\hat{\beta}_2 + \ldots + \Sigma x_{i2}x_{ik}\hat{\beta}_k = \Sigma x_{i2}y_i$$

.

. . $\qquad\qquad\qquad\qquad\qquad (2.3)$

.

$$\Sigma x_{ik}x_{i1}\hat{\beta}_1 + \Sigma x_{ik}x_{i2}\hat{\beta}_2 + \ldots + \Sigma x_{ik}^2\hat{\beta}_k = \Sigma x_{ik}y_i$$

or in matrix form

$$(X'X)\hat{\beta} = X'y \qquad (2.3\text{-}M)$$

The solution to these equations can be written as

$$\hat{\beta} = (X'X)^{-1}X'y \qquad (2.4\text{-}M)$$

because X was assumed to be of rank k. The matrix $(X'X)^{-1}$ is the *inverse of the matrix of normal equations* and plays an important role in least squares analysis. Let its elements be denoted by c_{ij} so that

$$(X'X)^{-1} = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1k} \\ c_{21} & c_{22} & \cdots & c_{2k} \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ c_{k1} & c_{k2} & \cdots & c_{kk} \end{bmatrix} \qquad (2.5\text{-}M)$$

The standard deviation, σ, is estimated from the deviations d_i, where

$$d_i = y_i - (x_{i1}\hat{\beta}_1 + x_{i2}\hat{\beta}_2 + \ldots + x_{ik}\hat{\beta}_k) \qquad (2.6)$$

25

by the quantity, s,

$$s = \sqrt{\frac{\Sigma d^2}{n-k}} \qquad (2.7)$$

and is said to have n-k *degrees of freedom.*

The standard deviation of the values for the coefficients $\hat{\beta}_i$ are given by

$$\text{s.d.}(\hat{\beta}_i) = \sigma\sqrt{c_{ii}} \qquad (2.8)$$

and for a linear function $L = a_1\hat{\beta}_1 + a_2\hat{\beta}_2 \ldots a_k\hat{\beta}_k$ is [see equation (1.10-M)]

$$\text{s.d.}(L) = \sigma\left\{[a_1 \ldots a_k]\begin{bmatrix} c_{11} & \cdots & c_{1n} \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ c_{n1} & \cdots & c_{nn} \end{bmatrix}\begin{bmatrix} a_1 \\ \cdot \\ \cdot \\ \cdot \\ a_k \end{bmatrix}\right\}^{1/2} \qquad (2.9-M)$$

3. The Gauss Theorem on Least Squares (Independent, Equal Variance, With Restraints)

If the parameters, β_i, are required to satisfy the m linear equations

$$\left\{\begin{array}{l} \psi_1 = b_{11}\beta_1 + b_{12}\beta_2 \cdots b_{1k}\beta_k = K_1 \\[3mm] \qquad \cdot \\[2mm] \qquad \cdot \\[2mm] \qquad \cdot \\[3mm] \psi_m = b_{m1}\beta_1 + b_{m2}\beta_2 \cdots b_{mk}\beta_k = K_m \end{array}\right. \qquad (3.1)$$

or in matrix form

$$B'\beta = K \qquad (3.1\text{-}M)$$

then using the method of Lagrangian multipliers, it turns out that the minimum variance unbiased linear estimators are given by minimizing

$$F = Q + 2\lambda_1 (\psi_1 - K_1) + 2\lambda_2(\psi_2 - K_2) + \ldots + 2\lambda_m(\psi_m - K_m) \qquad (3.2)$$

considered as a function of the β's and λ's. ($2\lambda_i$ is chosen rather than just λ_i so that in setting $\partial F/\partial\beta_i = 0$, a common factor of 2 can be divided out.)

This leads to the normal equations

$$\Sigma x_1^2\beta_1 + \ldots + \Sigma x_1 x_k\beta_k + b_{11}\lambda_1 + \ldots + b_{m1}\lambda_m = \Sigma x_1 y$$

$$\cdot$$
$$\cdot$$
$$\cdot$$

$$\Sigma x_k x_1\beta_1 + \ldots + \Sigma x_k^2\beta_k + b_{1k}\lambda_1 + \ldots + b_{mk}\lambda_m = \Sigma x_k y$$

$$b_{11}\beta_1 + \ldots + \beta_{1k}\beta_k \qquad\qquad\qquad = K_1$$

$$\cdot$$
$$\cdot$$
$$\cdot$$

$$b_{m1}\beta_1 + \ldots + \beta_{mk}\beta_k \qquad\qquad\qquad = K_m$$

$$(3.3)$$

or in matrix form

$$\begin{bmatrix} X'X & B \\ B' & 0 \end{bmatrix} \begin{bmatrix} \beta \\ \lambda \end{bmatrix} = \begin{bmatrix} X'y \\ K \end{bmatrix} \qquad\qquad (3.3\text{-}M)$$

and the solution is given by

$$\begin{bmatrix} \hat{\beta} \\ \hat{\lambda} \end{bmatrix} = \begin{bmatrix} X'X & B \\ B' & 0 \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ K \end{bmatrix} \qquad\qquad (3.4\text{-}M)$$

If X'X was already of full rank, then B must be of rank m for the inverse to exist. If X'X is of rank (k-m) and B' consists of m rows, then the indicated inverse will exist if B is orthogonal to X'X, i.e. that (X'X)B = 0, and B is of rank m. Also if B is a combination of such an orthogonal set of restraints, denoted by H, and the vectors of X'X, then the inverse exists if the mxm matrix B'H is of rank m, i.e., the determinant $|B'H| \neq 0$.

EXAMPLE: If the differences A-B, B-C, C-D, D-E, E-A are measured, then the 5 measurements y_1, y_2, y_3, y_4, y_5 (assumed independent with equal variance) can be represented as

$$E(y) = \begin{bmatrix} A-B \\ B-C \\ C-D \\ D-E \\ -A \qquad +E \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \\ -1 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} A \\ B \\ C \\ D \\ E \end{bmatrix} = X\beta$$

$$X'X = \begin{bmatrix} 2 & -1 & 0 & 0 & -1 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ -1 & 0 & 0 & -1 & 2 \end{bmatrix} \qquad \text{rank of } X'X \text{ is } 4$$

The restraint $A+B+C+D+E = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} A \\ B \\ C \\ D \\ E \end{bmatrix} = H' \begin{bmatrix} A \\ B \\ C \\ D \\ E \end{bmatrix} = K$

is orthogonal to X'X because $H'(X'X) = (1\ 1\ 1\ 1\ 1)\ (X'X) = [0\ 0\ 0\ 0\ 0]$. If the given restraint were $A + B = K_0$, then $B' = (1\ 1\ 0\ 0\ 0)$ and $|B'H| = 2 \neq 0$ so that the restraint is sufficient to produce a solution.

The standard deviation estimate is changed from that given in formula (2.7) to become

$$s = \sqrt{\frac{\Sigma d^2}{n-k+m}} \qquad \text{degrees of freedom} = (n-k+m) \qquad (3.5)$$

where m is the number of restraints.

Formulas (2.8) and (2.9) still apply for the standard deviation of the parameter values and of linear combinations of them.

4. The Gauss Theorem on Least Squares (General Case)

If the observed values $y_1\ y_2\ \ldots\ y_n$ have variances σ_i^2 and covariances σ_{ij} so that

$$\text{Var}\ (y) = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdot\ \cdot & \sigma_{1n} \\ \sigma_{12} & \sigma_2^2 & & \sigma_{2n} \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ \sigma_{1n} & \sigma_{2n} & & \sigma_n^2 \end{bmatrix} = V \tag{4.1-M}$$

and the parameters are subject to the m restraints

$$\left\{ \begin{array}{l} b_{11}\beta_1 + \ldots + b_{1k}\beta_k = K_1 \\ \cdot \\ \cdot \\ \cdot \\ b_{m1}\beta_1 + \ldots + b_{mk}\beta_k = K_m \end{array} \right. \tag{4.2}$$

or in matrix form

$$B'\beta = K \tag{4.2-M}$$

Then the least squares estimators for β are given by

$$\begin{bmatrix} \hat{\beta} \\ \hat{\lambda} \end{bmatrix} = \begin{bmatrix} X'V^{-1}X & B \\ B' & 0 \end{bmatrix}^{-1} \begin{bmatrix} X'V^{-1}y \\ K \end{bmatrix}$$

where as before $\hat{\lambda}' = [\hat{\lambda}_1\ \hat{\lambda}_2\ \ldots\ \hat{\lambda}_m]$ is a vector of Lagrangian multipliers entering into the minimization process.

For a discussion of this general case, the reader is referred to the Goldman-Zelen article [4].