

# AN ANALYSIS OF SHORT-TERM SELECTION EXPERIMENTS\*

R. H. RICHARDSON and K. KOJIMA

*University of Texas, Austin*

and

H. L. LUCAS

*N.C. State University, Raleigh*

Received 8.i.68

## 1. INTRODUCTION

### 1.1. *Basic orientation*

The change in the mean of a trait during a few initial cycles of a directional selection imposed on a population is among the most reliable criteria for estimating the exploitable amount of genetic variation in a given genetic population. Estimates obtained by this procedure include not only the "additive" genetic variance but also some of the variance attributable to non-allelic gene interactions. This is an advantageous rather than a disturbing aspect of the method, for one is interested in all of the utilisable genetic variation whether it be "additive" or not. In some studies "two-way selection" may be employed. That is, starting from a reference or "base" population, selection on the right-hand tail of the phenotypic distribution to yield a "high" line and selection on the left-hand tail to yield a "low" line can be done simultaneously.

To estimate rate of genetic "gain", the meaningful quantity we are concerned with is the rate at which the mean of a selected line diverges from the mean of the parent population. Under two-way selection, an index of total genetic variance is provided by the rate of divergence between the high and low lines. It is the detection and the evaluation of these quantities that are of primary concern in this paper.

### 1.2. *Correction for environmental effects*

In evaluating the effects of selection through successive generations, one must also reckon with trends and fluctuations in environment. Effective evaluation of the environmental effects may be accomplished, for example, by carrying the base (unselected) population through the successive cycles simultaneously with the selected line(s) and in the same environment. "Correction" for environmental effects in the selected line can be accomplished by subtracting the base line means from the selection line means, generation by generation, before analysing the data.

### 1.3. *Analytical complications introduced by correction for environmental effects*

In two-way selection studies, the correction process just mentioned involves

\* The research reported here was supported (in part) by NSF Grant GB 7252 and Public Health Grants GM 11546 and GM 15769. Computer time was made available by Public Health Service Grant Fr-00011 (NC. State University) and by USPHS training grant (2T1-GM-337-06 to R. P. Wagner, *et al.*, Univ. of Texas).

a statistical complication. Both the high and the low lines are corrected by subtracting the performance of the base (control) line. This introduces correlations between the high- and low-line data which should be taken into account when analysing experimental results. This complication does not arise in one-way selection experiments involving only one line and a control line. If in the same experiment, however, one or more control lines are used to provide environmental correction, and two or more selection lines are corrected by one or more of the control lines, the complication will be present. One can, of course, analyse such corrected data by a weighted analysis which takes into account the correlations between observations (Richardson, 1965). The main point of this paper, however, is to elucidate a method of statistical analysis which effects the desired corrections but which is done on the raw data, thereby avoiding the correlations introduced by the subtraction procedure.

#### 1.4. *Associated matters*

1.4.1. *Control populations.* It is important to choose the control lines wisely. Kojima and Kelleher (1963) have discussed several desirable characteristics of control populations. However, if there exist interactions between genetic make-up primarily associated with the trait under study and the environment, the appropriate analysis depends heavily on the choice of the control population. It is assumed that the selection lines and the control lines will respond in similar fashion to environmental fluctuations encountered. Unless this assumption is realised, the correction for environmental trends and fluctuations likely will be biased. It is conceivable that as selection changes the genetic constitution of a population, the same control may not be appropriate for the entire experiment. However, since this possibility is not likely to become a problem for short-term selection experiments, it will be ignored.

1.4.2. *Choice of control populations in general.* If either one- or two-way selection studies are being done starting from two or more base populations which are quite similar genetically, then only one appropriately selected control line will be needed to provide environmental corrections for all. If, however, the two or more base populations from which the selection lines are derived differ considerably in genetic make-up, then the lines should be grouped on the basis of genetic similarity and a proper control line should be carried for each group. In practice, the magnitude of genetic differences between the two or more base populations and the economics of the experimental situation will dictate how finely the base populations should be classified and, hence, the number of different control lines used.

1.4.3. *Asymmetry in two-way selection studies.* Selection on the high side may result in a rate of divergence from the base or control population that is different in absolute magnitude from that on the low side. Problems involved have been discussed by Falconer (1953). These include the scale of measurement, inbreeding depression during the experiment, and genetic interactions which may be uncovered by the changing genetic composition of the selection lines. Interpretation of the regressions fitted to the selection data depend heavily on a consideration of these possible factors, but in the short-term experiment they offer no difficulties to the regression model for analysis.

1.4.4. *Nature of selection trends through time.* Selection through successive

generations results in a time-trend of responses. The magnitude of responses usually diminishes, yielding a non-linear trend, with the total response to selection approaching an asymptote. That is, the response to selection, barring mutation and sampling and environmental vagaries, approaches an upper (lower) asymptote. Even though the effect of selection is non-linear through time, the effects of selection can be practically approximated by a linear trend over the first several generations (cycles) of the selection process. The linearity assumption is made throughout this paper.

1.4.5. *Replication of tests.* Replications in time and/or location provide a better estimate of the error involved than do single experiments in assessing the progress to be expected in the selection process and in the estimation of genetic variance for the one or more base populations under study. Selection experiments involve genetic-environmental interactions. Further, they are stochastic processes and performance in a given generation is dependent on the genetic samples retained in previous generations. If desirable alleles are lost at one stage as a result of sampling vagaries, the subsequent effects of selection are restricted. In one replication some desirable genes might be lost early, and in another replication not at all. Thus the progress realised over a given number of generations can vary from one replication to another. If the population retained at each step is small, the progress can be subject to high sampling variance. The analysis of several replications permits a proper estimate of experimental error. If there is only one replication, the estimate of the error component of the regression line for genetic sampling is not estimable, and consequently the estimate of error fails to include it.

Generally, the analytical approach to replicated experiments depends on whether they are conducted together or separately. When separately, an additional analysis performed on the estimates of regression coefficients obtained by the analysis to be presented may be done to evaluate differences among rates of selection progress of the various replications. If measured together, extra factors for replications and trend by replication interactions can be included in the model with identifying subscripts and a composite analysis performed.

1.4.6. *Final orientative comments.* The statistical analysis to be presented will not encompass the problem of replicating tests nor take into account that selection experiments are stochastic processes. Extensions to encompass such complications will, however, be briefly discussed. The method will be described for experiments involving one or more control lines with one or more selection lines associated with each control. It is an improvement over those most frequently used in that it explicitly considers more of the characteristics of selection data. The numerical example is a relatively simple one, namely, one with two-way selection on two base populations and a single control line.

## 2. METHOD OF ANALYSIS

### 2.1. Theory

2.1.1. *Identification of observations.* Let  $y_{hijk}$  be the observed performance for the  $k$ th generation ( $k = 0, 1, \dots, g$ ) of the  $j$ th line ( $j = 0, 1, \dots, l_{hi}$ ) of the  $i$ th base population ( $i = 0, 1, \dots, p_h$ ) in the  $h$ th population class ( $h = 1, 2, \dots, c$ ). A class is that group of lines, including control, sharing a common environmental response. Because some  $hijk$  possibilities may not be present in a given set of data, we note the identification matters given in table 1 for

observations in a given population class. All of the possible observations identified in table 1 are not necessarily available. Further, the number of

TABLE 1  
*Details of identification of observations for a particular class*

Item	Population $i =$	Line $j =$	Generation $k =$	Observations identified
1	0	0	0	A control line different from base population(s) before selection has begun on base population(s).
2	0	0	1, 2, ..., $g$	A control line different from base population(s) after selection has begun.
3	0	1, 2, ..., $l_0$	0	Same as item 1 for several control lines different from base population(s).
4	0	1, 2, ..., $l_i$	1, 2, ..., $g$	Same as item 2 for several control lines different from base population(s).
5	1, 2, ..., $p$	0	0	Base population(s) before selection has begun.
6	1, 2, ..., $p$	0	1, 2, ..., $g$	Base population(s) (unselected) carried as control during selection of other lines.
7	1, 2, ..., $p$	1, 2, ..., $l_i$	0	Not possible.
8	1, 2, ..., $p$	1, 2, ..., $l_i$	1, 2, ..., $g$	Selection lines after selection has begun.

base populations is permitted to vary between classes and the number of selection lines to vary among base populations. It is assumed, however, that all lines are observed for the same number of generations and that, at a given generation, all lines in a given population class are observed under the same environment.

2.1.2. *The model.* Write

$$y_{hijk} = \phi_{h..k} + \alpha_{hi..} + z_{hijk} \beta_{hij.} + \epsilon_{hijk} \quad (1)$$

where  $\phi_{h..k}$  = environmental effects occurring at the  $k$ th generation for the  $h$ th class,

$\alpha_{hi..}$  = the expected performance of the  $i$ th base population in the  $h$ th class,

$z_{hijk}$  = a characterisation of the selection cycle; this could be the cumulative selection differential corresponding to  $y_{hijk}$  or it could simply be the cycle number,  $k$ ,

$\beta_{hij.}$  = the slope of the (assumed linear) relation to generation resulting from selection—by definition  $\beta_{h00.} = 0$ ,

$\epsilon_{hijk}$  = random fluctuations assumed to be independent with zero mean and variance,  $\sigma_{hijk}^2$ .

A dot in these notations indicates that the quantity is defined as an average over that particular subscript range. As noted, selection experiments are stochastic processes. The genetic spectrum represented in a given generation is subject to sampling fluctuation which conditions the spectrum represented in the next generation. Strictly, the  $\epsilon_{hijk}$  are not independent from generation to generation in a given line, even though assumed to be so here.

2.1.3. *The model in matrix form.* Note that the model involves no parameters common to two or more population classes. The analysis can be done separately for each population class, and should be done separately if

the variance differs between classes. If there is reason to make comparisons between lines in different population classes, this can be accomplished after the individual analyses are done. It is convenient, therefore, to proceed as though only one population class (and one control line) is involved, and the subscript,  $h$ , will be dropped.

The model in matrix form may be written

$$y = X\beta + \varepsilon \quad (2)$$

where  $y$  = column vector of observed means,

$X$  = matrix of coefficients,

$\beta$  = column vector of effects:  $\phi$ 's,  $\alpha$ 's, and  $\beta$ 's,

$\varepsilon$  = column vector of random errors.

The construction of the  $X$ -matrix may be easily achieved by ordering and partitioning the  $y$ -vector in a convenient way. If the  $y$ -vector is partitioned according to the  $y$ -subscripts as in table 2, then it is a simple matter to place zeros, ones, or  $z$ 's (usually the cumulative selection differentials) along each row of the  $X$ -matrix: ones in columns for effects present in an observation, or  $z$ 's in the case of the  $\beta$ 's, and zeros in all other columns. Where a partition is missing in the  $y$ -vector there will be no corresponding row partition in the  $X$ -matrix. An example of forming the  $X$ -matrix is given in table 2.

This analysis is done on means. All means might not have the same number of observations. Even if, as shall be assumed here, the variance among individuals is the same, say  $\sigma^2$ , for all  $ijk$ , the variance of the means can differ. That is, the variance of  $y_{ijk}$  will be  $\sigma^2/n_{ijk}$ , where  $n_{ijk}$  is the number of observations in  $y_{ijk}$ . It is especially likely, for example, that the initial (prior to selection) observation of a two-way selection experiment (*i.e.* the mean for the base population) will involve more observations than the later means. Thus, a weighting matrix,  $W$ , is needed to weight each mean proportionally to its relative information. Under the present assumptions,  $W$  is a diagonal matrix with the diagonal elements being the numbers of observations in the means. The matrix is diagonal because the  $\varepsilon$ 's are assumed to be independent.

2.1.4. *Computational formulas.* The analysis is easily handled by the usual procedures for weighted multiple regression. The column vector of parameter estimates,  $b$ , is obtained as

$$b = (X'WX)^{-1}X'Wy, \quad (3)$$

where priming indicates the transpose of a matrix or vector. The variance-covariance matrix of  $b$  is given by

$$V(b) = (X'WX)^{-1}Q_1. \quad (4)$$

The quantity,  $Q_1$ , is taken from table 3, which outlines the preliminary analysis of variance.

Interest will lie in given comparisons, each of which can be written as

$$c = v'b \quad (5)$$

where  $v$  is an appropriate column vector of coefficients. The variance of a comparison is

$$V(c) = v'(X'WX)^{-1}vQ_1 \quad (6)$$

and the sum of squares for a comparison is

$$SS \text{ (comparison)} = \mathbf{b}'\mathbf{v}[\mathbf{v}'(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{v}]^{-1}\mathbf{v}'\mathbf{b}. \tag{7}$$

Some examples of  $\mathbf{v}$  are given in table 4. The examples of comparisons given are not exhaustive, nor are they independent, but the experimenter may select a set appropriate to his interest.

TABLE 2

*Arrangement of the X-matrix relative to that of the y-vector (for rows) and model parameters (for columns)*

Sub- scripts of $y$	Coefficients of												Description of partition	
	$\phi_{..0}$	$\phi_{..1}$	$\phi_{..2}$	...	$\phi_{..g}$	$\alpha_{100}$	$\alpha_{200}$	...	$\alpha_{p00}$	$\beta_{11}$	$\beta_{12}$	...		$\beta_{p1g}$
000	1	0	0	...	0	0	0	...	0	0	0	...	0	Control population different from parents of selection lines.
001	0	1	0	...	0	0	0	...	0	0	0	...	0	
002	0	0	1	...	0	0	0	...	0	0	0	...	0	
.	.	.	.	.	.	.	.	.	.	.	.	.	.	
.	.	.	.	.	.	.	.	.	.	.	.	.	.	
00g	0	0	0	...	1	0	0	...	0	0	0	...	0	
100	1	0	0	...	0	1	0	...	0	0	0	...	0	Initial observation of population 1.
101	0	1	0	...	0	1	0	...	0	0	0	...	0	Control population if it is unselected population 1.
102	0	0	1	...	0	1	0	...	0	0	0	...	0	
.	.	.	.	.	.	.	.	.	.	.	.	.	.	
.	.	.	.	.	.	.	.	.	.	.	.	.	.	
.	.	.	.	.	.	.	.	.	.	.	.	.	.	
10g	0	0	0	...	1	1	0	...	0	0	0	...	0	
110	1	0	0	...	0	1	0	...	0	0	0	...	0	Selection line 1 of population 1; observation $y_{110}$ is deleted unless it is different from $y_{100}$ .
111	0	1	0	...	0	1	0	...	0	$z_{111}$	0	...	0	
112	0	0	1	...	0	1	0	...	0	$z_{112}$	0	...	0	
.	.	.	.	.	.	.	.	.	.	.	.	.	.	
.	.	.	.	.	.	.	.	.	.	.	.	.	.	
11g	0	0	0	...	1	1	0	...	0	$z_{11g}$	0	...	0	
121	0	1	0	...	0	1	0	...	0	0	$z_{121}$	...	0	Selection line 2 of population 1.
122	0	0	1	...	0	1	0	...	0	0	$z_{122}$	...	0	
.	.	.	.	.	.	.	.	.	.	.	.	.	.	
.	.	.	.	.	.	.	.	.	.	.	.	.	.	
.	.	.	.	.	.	.	.	.	.	.	.	.	.	
12g	0	0	0	...	1	1	0	...	0	0	$z_{12g}$	...	0	
.	.	.	.	.	.	.	.	.	.	.	.	.	.	
.	.	.	.	.	.	.	.	.	.	.	.	.	.	
.	.	.	.	.	.	.	.	.	.	.	.	.	.	
200	1	0	0	...	0	0	1	...	0	0	0	...	0	Initial observation of population 2.
201	0	1	0	...	0	0	1	...	0	0	0	...	0	Control population if it is unselected population 2.
202	0	0	1	...	0	0	1	...	0	0	0	...	0	
.	.	.	.	.	.	.	.	.	.	.	.	.	.	
.	.	.	.	.	.	.	.	.	.	.	.	.	.	
.	.	.	.	.	.	.	.	.	.	.	.	.	.	
20g	0	0	0	...	1	0	1	...	0	0	0	...	0	
etc.														etc., similar to above.

TABLE 3

Computing formulas for sums of squares

Variance source	D.F.*	Sums of squares	Mean squares
Total	$n$	$S_T^2 = y'Wy$	—
Due to $\beta$ (all parameters)	$n_p$	$S_p^2 = b'X'Wy$	$Q_2 = S_p^2/n_p$
Residual (error)	$n - n_p$	$S_R^2 = S_T^2 - S_p^2$	$Q_1 = S_R^2/(n - n_p)$

\*  $n$  = number of elements of  $y$ ;  $n_p$  = number of columns of  $X$ .

TABLE 4

Examples of row vectors ( $v$ ) for subdividing the sum of squares due to all parameters ( $\beta$ )

Null hypothesis	Vector elements corresponding to											
	$\phi_{..0}$	$\phi_{..1}$	...	$\alpha_{100}$	$\alpha_{200}$	...	$\beta_{11}$	$\beta_{12}$	...	$\beta_{21}$	$\beta_{22}$	...
$\alpha_{100} = 0$	0	0	.0.	1	0	.0.	0	0	.0.	0	0	.0.
$\alpha_{200} = 0$	0	0	.0.	0	1	.0.	0	0	.0.	0	0	.0.
$\alpha_{100} - \alpha_{200} = 0$	0	0	.0.	1	-1	.0.	0	0	.0.	0	0	.0.
$\beta_{11} = 0$	0	0	.0.	0	0	.0.	1	0	.0.	0	0	.0.
$\beta_{12} = 0$	0	0	.0.	0	0	.0.	0	1	.0.	0	0	.0.
$\beta_{11} - \beta_{12} = 0$	0	0	.0.	0	0	.0.	1	-1	.0.	0	0	.0.
$\beta_{11} + \beta_{12} = 0$	0	0	.0.	0	0	.0.	1	1	.0.	0	0	.0.
$\beta_{11} - \beta_{21} = 0$	0	0	.0.	0	0	.0.	1	0	.0.	-1	0	.0.
$(\beta_{11} - \beta_{12}) - (\beta_{21} - \beta_{22}) = 0$	0	0	.0.	0	0	.0.	1	-1	.0.	-1	1	.0.
$(\beta_{11} + \beta_{12}) - (\beta_{21} + \beta_{22}) = 0$	0	0	.0.	0	0	.0.	1	1	.0.	-1	-1	.0.

2.2. Numerical example

For purposes of illustration, a relatively simple case is considered, using *Drosophila* fecundity selection data previously given by Richardson (1965) and summarised by Richardson and Kojima (1965). Each cycle consisted of the high and low selection lines of two different populations (called *Base* and *RRS*) with each line having about 85 families observed daily for four days at each generation. The control population consisted of about 72 families observed daily for four days, and which were taken from a constant set of hybrid flies obtained by crossing a set of inbred lines selected to have similar homeostasis and phenotypic variation to the experimental populations. The average performances of these five groups for days (table 5) formed the elements of the  $y$ -vector. Cumulative selection differentials were used for the  $z$  portions (table 6) of the  $X$ -matrix. Figure 1 is a graph of the performances of table 5 plotted against the cumulative selection differentials of table 6 for the four selection lines, and fig. 3, top, is a graph of performance versus generations for the control. The initial observations for the two populations, *Base* and *RRS*, were averages of about 170 families observed daily for four days with the same control as for later generations.

Note that from table 5, we can partition the  $y$ -vector into subvectors, corresponding to selection lines as follows:

$$y = \begin{bmatrix} y_{00} \\ y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \end{bmatrix}$$



TABLE 5  
Mean egg production

Generation <i>k</i> =	Control <i>y</i> <sub>00<i>k</i></sub> =	Base High <i>y</i> <sub>11<i>k</i></sub> =	Base Low <i>y</i> <sub>12<i>k</i></sub> =	RRS High <i>y</i> <sub>21<i>k</i></sub> =	RRS Low <i>y</i> <sub>22<i>k</i></sub> =
0	31.95	34.27	—	40.62	—
1	36.63	50.47	46.25	47.69	49.04
2	43.61	51.47	43.15	48.70	48.76
3	40.73	47.59	41.42	48.59	44.24
4	35.13	38.26	32.79	43.82	36.68
5	40.28	46.34	36.39	46.15	46.98
6	35.62	39.52	27.98	44.34	32.43
7	43.20	51.20	39.22	54.41	47.54
8	40.71	46.86	31.25	45.81	40.84

TABLE 6  
Cumulative selection differentials, *z*<sub>*ijk*</sub>, incorporated in the columns of the *X*-matrix corresponding to β<sub>*ij*</sub>.

Generation <i>k</i> =	Base High <i>ij</i> = 11	Base Low <i>ij</i> = 12	RRS High <i>ij</i> = 21	RRS Low <i>ij</i> = 22
1	5.5	5.7	5.4	5.6
2	15.2	13.4	9.7	12.4
3	23.5	20.1	15.0	20.6
4	27.0	28.7	20.2	27.7
5	34.7	35.8	25.5	34.5
6	44.5	44.9	31.4	40.0
7	52.2	54.1	36.8	45.6
8	58.7	62.9	43.0	51.6

The weighting matrix, *W*, is of 43-square dimension, and may be partitioned to correspond to the various lines as follows:

- for Control, *W*<sub>1</sub> = diag. (1, 1, 1, 1, 1, 1, 1, 1, 1)
- for Base High, *W*<sub>2</sub> = diag. (2, 1, 1, 1, 1, 1, 1, 1, 1)
- for Base Low, *W*<sub>3</sub> = diag. (1, 1, 1, 1, 1, 1, 1, 1, 1)
- for RRS High, *W*<sub>4</sub> = diag. (2, 1, 1, 1, 1, 1, 1, 1, 1)
- for RRS Low, *W*<sub>5</sub> = diag. (1, 1, 1, 1, 1, 1, 1, 1, 1)

where

$$W = \begin{bmatrix} W_1 & 0 & 0 & 0 & 0 \\ 0 & W_2 & 0 & 0 & 0 \\ 0 & 0 & W_3 & 0 & 0 \\ 0 & 0 & 0 & W_4 & 0 \\ 0 & 0 & 0 & 0 & W_5 \end{bmatrix} \tag{9}$$

The diagonal elements of 1's and 2's give the same weight as actual *n*<sub>*ijk*</sub>, where all *n*<sub>*ijk*</sub> except two were equal, and these two were twice as large as the others.

For this example, the *X'**WX* matrix may be partially given for reference, as below

$$X'WX = \begin{bmatrix} 5.0000 & 0 & \dots & 0 & 0 \\ 0 & 5.0000 & \dots & 5.4000 & 5.6000 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 5.4000 & \dots & 5595.7400 & 0 \\ 0 & 5.6000 & \dots & 0 & 8908.9400 \end{bmatrix} \tag{10}$$



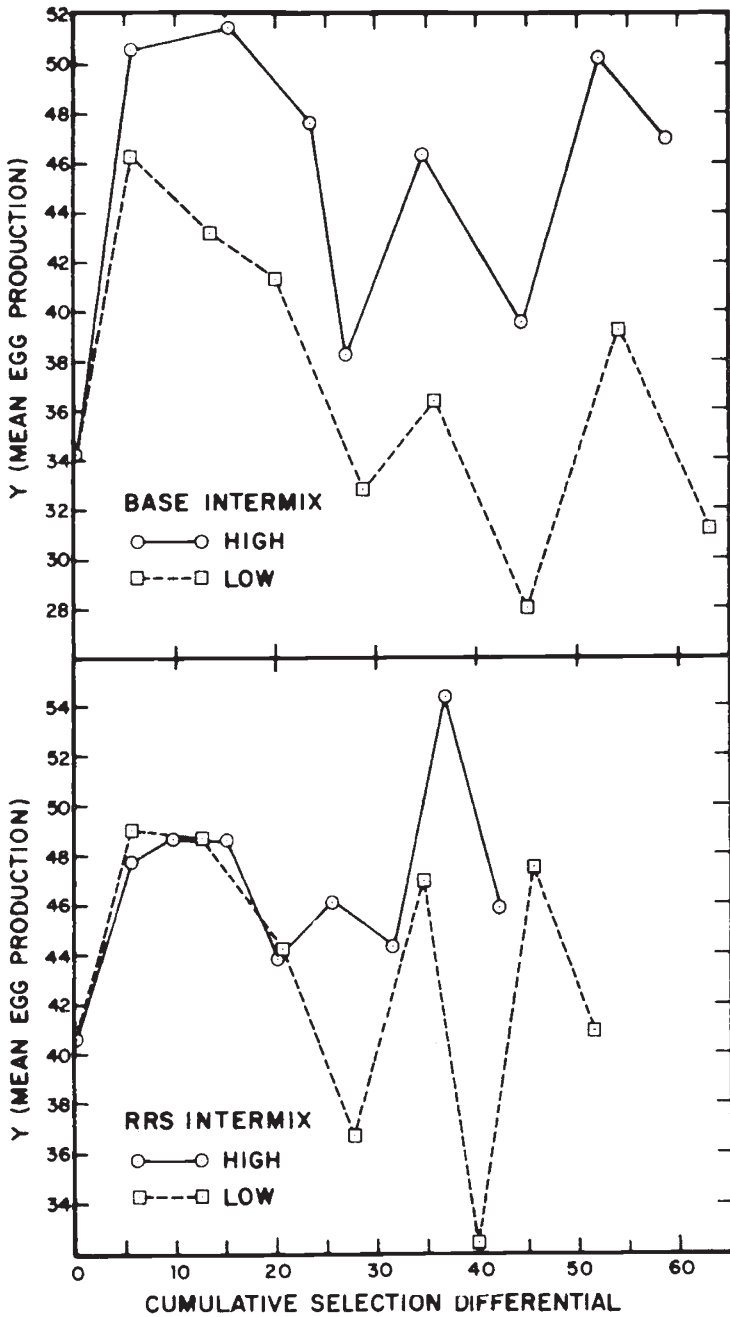


FIG. 1.—Average egg production per female per day *versus* cumulative selection differentials.

and the inverse

$$(X'WX)^{-1} = \begin{bmatrix} 0.4872 & 0.2448 & \dots & 0.0121 & 0.0096 \\ 0.2448 & 0.4114 & \dots & 0.0095 & 0.0075 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0.0121 & 0.0095 & \dots & 0.0010 & 0.0006 \\ 0.0096 & 0.0075 & \dots & 0.0006 & 0.0006 \end{bmatrix} \quad (11)$$

The vector of parameter estimates is

$$\mathbf{b}' = (30.43, 40.60, 42.36, 40.38, 33.90, 40.41, 33.83, 45.64, 40.30, 5.97, 8.83, 0.0049, -0.2476, -0.0343, -0.1721) \quad (12)$$

Note that the first nine values (30.43 to 40.30) are estimates of the  $\phi_{..k}$ , the environmental components. The "adjusted" means shown in fig. 2 are the elements of the  $\mathbf{y}$ -vector (equation 8) minus the  $\phi_{..k}$ . The next two elements of (12) are intercepts for Base Intermix and RRS Intermix, respectively, and the last four values are regression coefficients for Base High and Low and for RRS High and Low, respectively. The regression lines also are given in fig. 2. Figure 3, bottom, is a graph of adjusted performance of the control. The extensive "smoothing" of the curves in fig. 2 obtained by removal of much of the environmental variation graphically illustrates the assistance in detection of the responses to selection.

The variance-covariance matrix for the intercepts and regression coefficients is

$$V = \begin{bmatrix} 2.5983 & 1.7685 & -0.0618 & -0.0595 & -0.0585 & -0.0463 \\ & 2.7287 & -0.0423 & -0.0410 & -0.0912 & -0.0725 \\ & & 0.0024 & 0.0019 & 0.0020 & 0.0016 \\ & \text{(symmetrical)} & & 0.0022 & 0.0020 & 0.0016 \\ & & & & 0.0048 & 0.0031 \\ & & & & & 0.0030 \end{bmatrix} \quad (13)$$

An analysis of variance is given in table 7. Only a few of the many possible individual contrasts are shown. Here the most interest revolves around the estimates of slopes and the contrasts among them.

TABLE 7

*Analysis of variance for mean egg production*

Source	D.F.	SS	MS
Total	43	81,440.2054	
All Parameters	15	81,301.9326	
Intercepts $\times$ Populations	1	4.5550	4.56
Base Intermix Divergence	1	357.5788	357.58
RRS Intermix Divergence	1	58.4811	58.48
Base Intermix Symmetry	1	35.1063	35.11
RRS Intermix Symmetry	1	14.9440	14.94
Divergence $\times$ Population	2	26.4391	13.22
Symmetry $\times$ Population	2	0.8086	0.40
Residual (Error)	28	138.2728	4.94

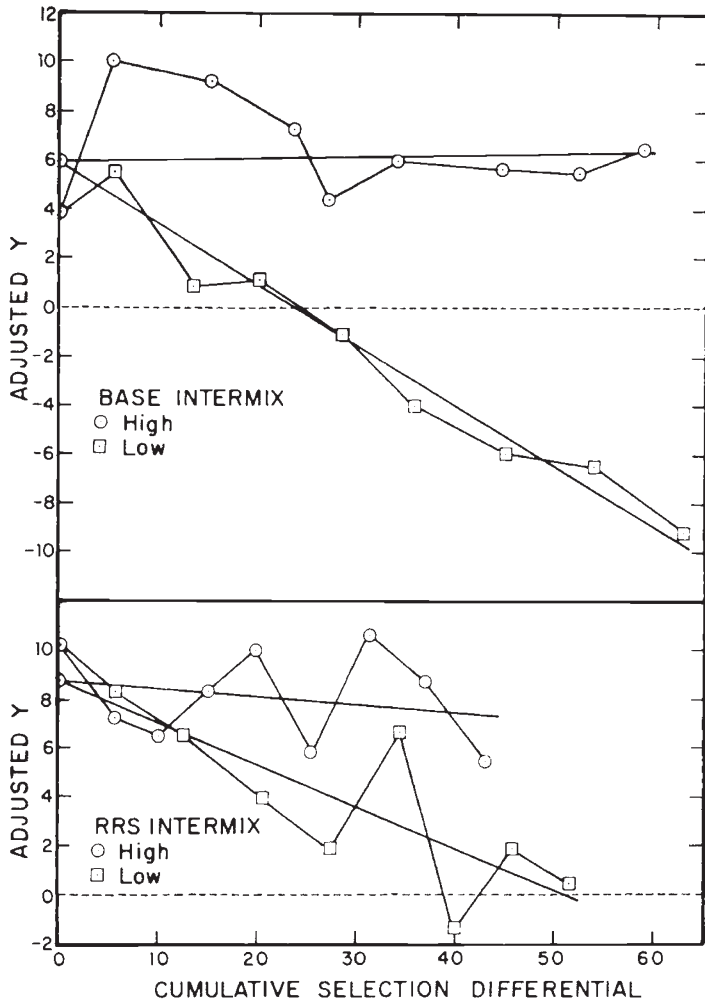


FIG. 2.—Adjusted average egg production per female per day *versus* cumulative selection differentials; scale same as FIG. 1, but with origin shifted.

For obtaining estimates of genetic parameters, one must consider the type of selection involved. For example, the slopes estimate realised heritabilities. If the selection system was one involving selection of full-sib families, then either the regression coefficient of a high or of a low selection line involves one-half the additive genetic variance, while the sum of the absolute values of the slopes for a high and low line from the same base population involves all the additive genetic variance. If selection is only in one direction, the genetic situation which results in asymmetry would not be detected, and estimates of realised heritability might be biased in either direction.

### 3. DISCUSSION

#### 3.1. *Generality of approach*

The intrinsic ideas of the analysis developed are basically the same for

short- and for long-term experiments. Certainly for most long-term experiments, the assumption of a linear trend would not adequately approximate the situation, and appropriate modifications of the model would have to be made. In fact, the data for Base Intermix (especially Base High) suggest that the linear trend might be inadequate for some short-term experiments, regardless of whether the model used is linear or non-linear. Comments regarding the stochastic nature of the process still hold, and the desirability of replicated experiments still exists.

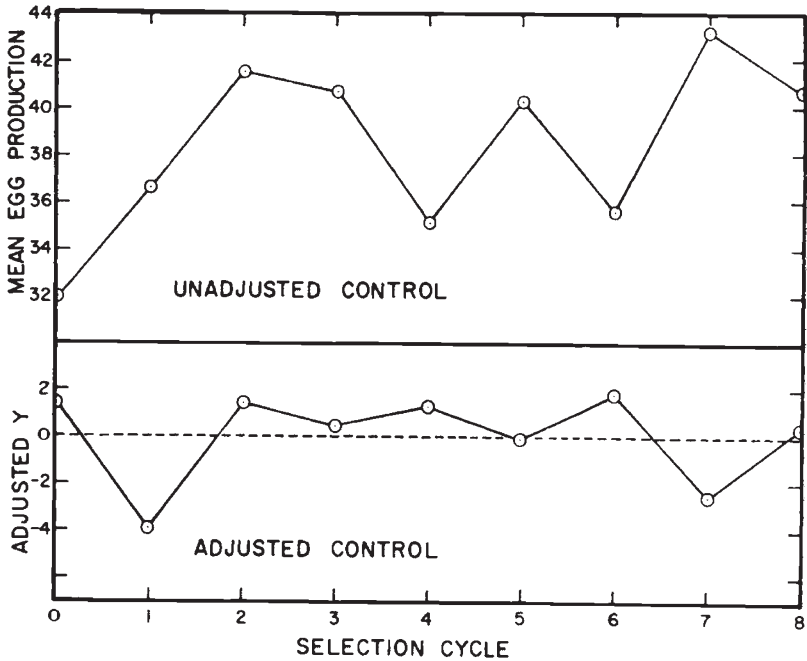


FIG. 3.—Comparison of control average egg production per female per day with the adjusted average egg production (origin shifted).

A possible way of characterising a non-linear trend is to introduce quadratic, cubic, etc. terms in addition to the linear ones. Although the polynomial approach could be descriptive of the data over some ranges, from the viewpoint of genetic analysis such a description is sterile. It is therefore appropriate to consider non-linear models for a trend which have a basis in genetic theory.

### 3.2. Some possible non-linear alternatives

3.2.1. *Models which disregard linkage and epistasis.* For short-term experiments, the linear model assumed in the foregoing theory and example ordinarily will be adequate. However, for long-term experiments, models such as that of James (1965), namely

$$\eta = \mu + \kappa\rho^x \quad (0 < \rho < 1; \kappa = \mu - \eta_0) \quad (14)$$

or the more specialised one of Frahm and Kojima (1966), *i.e.*

$$\eta = \mu(1 - e^{-\beta x}) \quad (\beta > 0; e^{-} = \rho) \quad (15)$$

might be used. These models replace the intercept and slope terms of the models, illustrated in the present development. For statistical analysis, iterative weighted least squares techniques would be required for estimating  $\mu$ ,  $\kappa$ ,  $\rho$  of (14) or  $\mu$ ,  $\beta$  of (15).

3.2.2. *Models which consider genetic recombination and epistasis.* Responses have been observed which are not hyperboloid in form, as required by models (14) and (15). For example, a "stair-stepped" response was observed by Mather and Harrison (1949) which probably was a result of recombination of tightly linked loci releasing concealed genetic variation after a considerable delay.

Similar effects could occur in short-term experiments, thus requiring replacement of the straight line by some other functional form. This possibility was pointed out for Base Intermix, and is supported by subsequent observations (Richardson, 1966 and unpublished). Models for substitution in place of the straight line cannot be suggested. Any such used, however, should be developed from genetic premises.

#### 4. SUMMARY

1. An improved regression approach for the analysis of selection experiments is described. The specific formulation presented in detail is particularly useful for studies designed to measure genetic variation that is immediately available for "short run" response.

2. Extensions of the analysis for non-linear response and replicated tests are discussed.

3. The major statistical improvement in the analysis lies in the estimation of environmental effects from one generation to the next using all the information available in the data. Consequently, estimates of genetic effects are more efficient.

4. The method enables the estimation of a point of common origin (intercept) for several selection lines arising from the same base population. As a result, the experiments may use *a priori* knowledge of the origin of selection lines.

5. Genetic progress is estimated separately for each line. Thus, with replicated selection lines, it is possible to estimate stochastic aspects of a genetic nature in addition to environmental fluctuations affecting estimates of response.

#### REFERENCES

- FALCONER, D. S. 1953. Asymmetrical response in selection experiments. Symp. on Genetics of Population Structure. *Internat. Union of Biol. Sci., Naples*, Series B, No. 15, pp. 16-41.
- FRAHM, R. R., AND KOJIMA, K. 1966. Comparison of selection responses on body weight under divergent larval density conditions in *Drosophila pseudoobscura*. *Genetics*, 54, 625-637.
- JAMES, J. W. 1965. Response curves in selection experiments. *Heredity*, 20, 57-63.
- KOJIMA, K., AND KELLEHER, T. 1963. Selection studies of quantitative traits with laboratory animals. In *Statistical Genetics and Plant Breeding* (W. D. Hanson and H. F. Robinson, editors). National Academy of Science—National Research Council Pub. 982, pp. 295-422.
- MATHER, K., AND HARRISON, B. J. 1949. The manifold effect of selection. *Heredity*, 3, 1-52, 131-162.

- RICHARDSON, R. H. 1965. A study of genetic variation of *Drosophila* fecundity by analysis of selection responses. Ph. D. thesis, N.C. State University, Raleigh.
- RICHARDSON, R. H. 1966. Effects of epistasis and linkage disequilibrium on average daily egg production in *Drosophila pseudoobscura*. (Abstract) *Genetics*, 54, 356-357.
- RICHARDSON, R. H., AND KOJIMA, K. 1965. The kinds of genetic variability in relation to selection responses in *Drosophila* fecundity. *Genetics*, 52, 583-598.
- ROBERTSON, A. 1960. A theory of limits in artificial selection. *Proc. Roy. Soc., B*, 153, 234-249.