

GENOTYPE \times ENVIRONMENT INTERACTIONS

I. REGRESSION OF INTERACTION ON OVERALL EFFECT OF THE ENVIRONMENT

KENNETH MATHER and P. D. S. CALIGARI

Department of Genetics, University of Birmingham, Birmingham B15 2TT

Received 1.ix.73

SUMMARY

The conditions are discussed under which the regression ($b_{D/S}$) over a range of environments of the difference between two genotypes, X and Y, on their sum will be linear. It is shown that if the slope of this regression falls outside the range -1 to 1 , the two genotypes must be responding in opposite directions to the changes in the environment. The relations of $b_{D/S}$ to $b_{Y/X}$ and $b_{X/Y}$, the regressions of X and Y on one another, are derived and the effects of error variation are considered. The three regressions are derivable from one another in principle, and the most useful values will be obtained by estimating from the data the one which is least subject to distortion by error variation and deriving the others from it. The best starting point will commonly, but not always, be $b_{D/S}$.

The treatment is extended to the analysis of multi-line experiments, due to Perkins and Jinks, in which each line is regressed on the mean of them all. It is shown how estimates of such regressions, not open to the statistical objections of regressing a variate onto another of which it itself is a part, can be obtained by starting with the regression of each line on the mean of the rest. The value of $b_{D/S}$ for any pair of the constituent lines can be derived directly from the multi-line analysis, and indeed once the analysis is available any group of lines can be readily compared with any other such group.

1. INTRODUCTION

It was observed by Yates and Cochran (1938) that the magnitude of the genotype \times environment interaction in the determination of the phenotypes shown by a number of genetically different lines or varieties raised in a range of different environments, could be related to the overall effects of the environments. Their treatment was wholly statistical, but a similar relation between the genotype \times environment interaction was also observed with two inbred lines of *Nicotiana rustica* and their F_1 by Bucio Alanis (1966) and Bucio Alanis and Hill (1966) whose analyses were carried out and results expressed in terms of the parameters which biometrical genetics has taught us to use. In particular they showed that with the two inbred lines the relationship of the interaction, g (see Mather and Jones, 1958) to the overall effect of the environment, e , could be found from the regression of the difference between the mean phenotypes of two lines in any given environment on the corresponding sum of these mean phenotypes. Furthermore they found this regression to be linear. The departure of the mean phenotype of the F_1 from the midparent similarly showed a linear relation to the e . This approach was generalised by Perkins and Jinks (1968*a, b*) who considered the case of more than two lines and showed that the interaction properties of any line could be inferred from the regression of that line's phenotype

on the mean of all the lines under observation. They also established, however, that the regressions are commonly not linear, even in *Nicotiana rustica*, the departures from linearity showing evidence of being themselves related to the genotypes of the lines under investigation.

These findings raise many questions, some of which will require further examination of the genetical aspects of these interactions with the environment. We will, however, confine ourselves here to looking into the general conditions for linearity of the regression line and certain aspects of the information that can be derived from the slopes of these lines.

2. THE TWO-LINE CASE

(i) Linearity

Let us consider two genotypic lines, X and Y, raised in a range of environments which may differ in any number of factors such as temperature, humidity, availability of nutrients, crowding of habitat and so on, all prospectively affecting the phenotypes of X and Y. Let us further suppose that these environmental factors may be measured by means independent of the development of X and Y in them, and the measurements combined into an overall metric characterising the environment that the factors combine to produce. Then in the environment characterised by the value z_1 we observe line X to express the character under consideration to the value x_1 , and line Y to the value y_1 . In environment z_2 we similarly observe x_2 and y_2 , and so on. Then we can plot x and y against z , and assuming

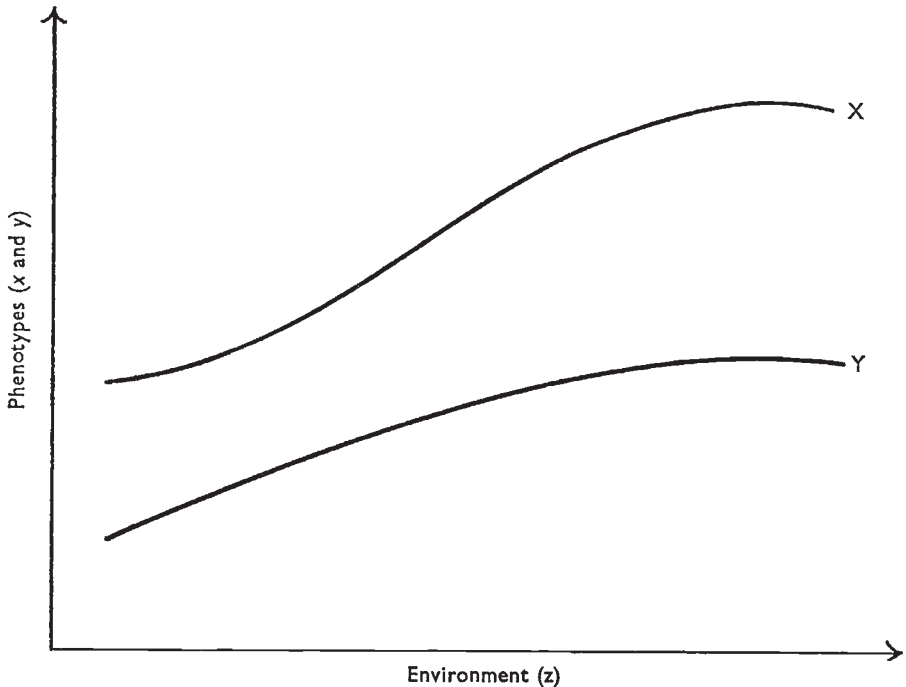


FIG. 1.—The phenotypes, x and y , of two genetically different lines, X and Y, varying with the environment as measured by an independent metric, z .

the three to be continuous variates, we can obtain two lines representing the phenotypes and the changes that X and Y show over this range of environments (fig. 1).

Following Bucio Alanis and Hill, we measure the overall effect of the environment, e , by $x+y$. The genotype x environment interaction, g , is one of two components whose sum is measured by $x-y$. The second component is d , the overall effect of the genotypic difference between X and Y, but since this is by definition constant over environments we may neglect it in considering, as we are concerned to do, the slope of the regression line. Thus leaving error variation out of account for the present, the slope of the regression of interaction on overall effect, that is of g on e , is the rate of change of $x-y$ on $x+y$.

Now

$$\frac{d(x-y)}{d(x+y)} = \frac{d(x-y)}{dx} \bigg/ \frac{d(x+y)}{dx} = \left(1 - \frac{dy}{dx}\right) \bigg/ \left(1 + \frac{dy}{dx}\right).$$

The regression of g on e as measured in this way is independent of z and hence of the way that the various environmental factors were combined in producing the environment as measured by z . At the same time, if the regression of $x-y$ on $x+y$ is to be simply linear, $\frac{d(x-y)}{d(x+y)}$ must be invariable

with x and y . This in turn requires that $\frac{dy}{dx}$ does not vary with x . Further-

more since $\frac{dy}{dx} = \frac{dy}{dz} \bigg/ \frac{dx}{dz}$, linear regression of g on e as measured in this way

requires that x and y are related in a basically similar way to the various environmental factors that jointly determine z . Or to put it more precisely, if $f(x, z)$ and $f(y, z)$ are the functions relating x and y to z $f(y, z) = kf(x, z)$ where k is independent of x, y and z . k may be of any size or even negative (which would imply that the two lines responded to change in the environment by change of phenotype in opposite directions); but it is independent of the environment itself. We may note that k as so defined must be $\frac{dy}{dx}$ which

is the regression of the phenotype of line Y on that of line X.

Writing $b_{D/S}$ for the regression of $x-y$ ($= D$, the difference) on $x+y$ ($= S$, the sum) and $b_{Y/X}$ for the regression of Y on X, we can rewrite the relation as

$$b_{D/S} = (1 - b_{Y/X}) / (1 + b_{Y/X})$$

and $b_{D/S} > 1$ only if $b_{Y/X}$ is negative.

Equally, using the same definition of $b_{D/S}$, the regression of $y-x$ on $y+x$ will be

$$-b_{D/S} = (1 - b_{X/Y}) / (1 + b_{X/Y}) = (b_{Y/X} - 1) / (b_{Y/X} + 1).$$

Where, as at present, we are neglecting error variation, $b_{X/Y} = 1/b_{Y/X}$ and the values of $b_{D/S}$ corresponding to equal values of $b_{Y/X}$ and $b_{X/Y}$ will be equal in magnitude but opposite in sign.

The relations of $b_{D/S}$ to $b_{Y/X}$ and $b_{X/Y}$ are shown graphically in fig. 2. It will be observed that when $b_{D/S}$ lies between -1 and 1 , $b_{Y/X}$ and $b_{X/Y}$ are positive. Thus for $-1 < b_{D/S} < 1$ the two genotypes X and Y are

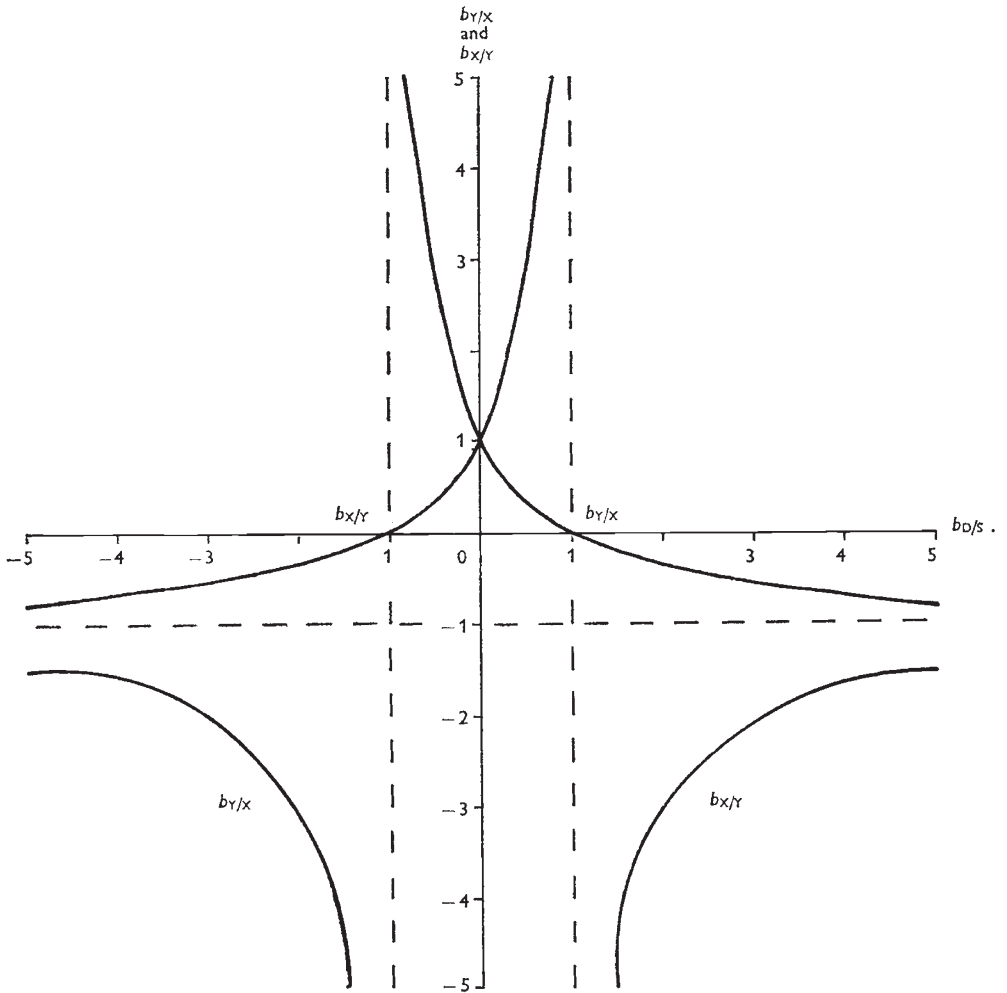


FIG. 2.—The relations of $b_{Y/X}$ and $b_{X/Y}$ with $b_{D/S}$ in the two line case.

responding in the same direction to the change in the environment, though at different rates where $b_{D/S} \neq 0$. When $b_{D/S} = 0$, $b_{Y/X} = b_{X/Y}$ and the two genotypes are responding at the same rate: in other words $g = 0$ and there is no interaction. Where $b_{D/S}$ lies outside the range -1 to $+1$, $b_{Y/X}$ and $b_{X/Y}$ must be negative, and the genotypes X and Y are responding in opposite directions to change in the environment.

These relations hold precisely only where $b_{D/S}$, $b_{Y/X}$ and $b_{X/Y}$ are, so to speak, the ideal regression coefficients which would be obtained by the normal process of estimation only in the absence of error variation. We must now turn to examine the effects of error variation.

(ii) *The multiplicity of regression estimates*

We have been considering the properties of $b_{D/S}$, $b_{Y/X}$ and $b_{X/Y}$ all of which can be estimated from the data, and any one of which, once known, can be used to generate expectations for the others. All these estimates will,

however, be affected by the error variation of the observations of the phenotypes, x and y , displayed by the two genotypic lines (see Hardwick and Wood, 1972). We will assume that, as can be secured by adequate experimental design, the error variations of x and y can be measured and are independent.

The estimates of $b_{Y/X}$ and $b_{X/Y}$ will, of course, be found as $b_{Y/X} = W_{y,x}/V_x$ and $b_{X/Y} = W_{y,x}/V_y$ where x and y are taken as deviations from their respective means. Assuming that the regressions are simply linear, and in the absence of error variation, these two regression coefficients will be reciprocals of each other; but error variation in x will raise the value of V_x while leaving $W_{y,x}$ unaltered and so will reduce the value of $b_{Y/X}$. Similarly error variation in y will reduce the value of $b_{X/Y}$. If x and y are subject to equal error variation, the reduction will be proportionately less in $b_{Y/X}$ where genotype X reacts more to change in the environment than does genotype Y, and the reduction will similarly be less in $b_{X/Y}$ where Y reacts more to environmental change than does X. Now if we derive $b_{D/S}$ from $b_{Y/X}$ we find:

$$b_{D/S} = \frac{1 - b_{Y/X}}{1 + b_{Y/X}} = \frac{1 - W_{y,x}/V_x}{1 + W_{y,x}/V_x} = \frac{V_x - W_{y,x}}{V_x + W_{y,x}} = \frac{W_x(x-y)}{W_x(x+y)}$$

where a variance is regarded as the covariance with itself of the variate in question (though we must remember that error variation inflates variances, but not covariances in an adequately designed experiment).

Similarly starting with $b_{X/Y}$, we find

$$b_{D/S} = \frac{V_y - W_{y,x}}{V_y + W_{y,x}} = \frac{W_y(y-x)}{W_y(y+x)}$$

Since for any given $W_{y,x}$, the greater the value of V_x the nearer the value of $(V_x - W_{y,x})/(V_x + W_{y,x})$ will be to 1, the effect of error variation in x and y on the estimate of $b_{D/S}$, as derived from $b_{Y/X}$ or $b_{X/Y}$, will be to reduce its departure from 1, the reduction being greater where the error variation is proportionately greater.

If, however, we estimate $b_{D/S}$ directly from $x-y$ and $x+y$ we have:

$$b_{D/S} = \frac{W_{(x-y)(x+y)}}{V_{(x+y)}} = \frac{V_x - V_y}{V_x + 2W_{y,x} + V_y}$$

Clearly $b_{D/S}$ will be 0 where V_x and V_y are equal. More particularly however, since error variation in x and y will tend to cancel out in the numerator but to reinforce in the denominator, its effect will be to reduce the estimate of $b_{D/S}$, to bring it in fact nearer to 0 rather than towards 1, as is the effect of such variation when $b_{D/S}$ is found from $b_{Y/X}$ or $b_{X/Y}$.

Having estimated $b_{D/S}$ directly from $x-y$ and $x+y$, we can of course find from it:

$$\begin{aligned} b_{Y/X} &= \frac{1 - b_{D/S}}{1 + b_{D/S}} = \frac{1 - (W_{(x-y)(x+y)}/V_{(x+y)})}{1 + (W_{(x-y)(x+y)}/V_{(x+y)})} \\ &= \frac{V_{(x+y)} - W_{(x-y)(x+y)}}{V_{(x+y)} + W_{(x-y)(x+y)}} \\ &= \frac{W_y(x+y)}{W_x(x+y)} \end{aligned}$$

and similarly $b_{X/Y} = \frac{W_x(x+y)}{W_y(x+y)}$

Although when given one of the three regression coefficients the others can be derived from it, the estimates so obtained will evidently vary according to the particular regression coefficient with which we start. Thus, for example, starting with $b_{Y/X} = W_{y, x}/V_x$, we can find $b_{X/Y} = V_x/W_{y, x}$ as its reciprocal, and $b_{D/S} = W_{x(x-y)}/W_{x(x+y)}$ as shown above. Starting with $b_{X/Y}, b_{D/S}$ becomes $W_{y(x-y)}/W_{y(x+y)}$ and if found directly it is $W_{(x-y)(x+y)}/V_{(x+y)}$. The three sets of estimates, found when starting with the three regressions are set out in table 1.

TABLE 1
The interrelations of the three regression coefficients as derived from each other (see in the text)

Starting with	Structure of		
	$b_{Y/X}$	$b_{X/Y}$	$b_{D/S}$
$b_{Y/X}$	$W_{y, x}/V_x$	$V_x/W_{y, x}$	$W_{x(x-y)}/W_{x(x+y)}$
$b_{X/Y}$	$V_y/W_{y, x}$	$W_{y, x}/V_y$	$W_{y(x-y)}/W_{y(x+y)}$
$b_{D/S}$	$W_{y(x+y)}/W_{x(x+y)}$	$W_{x(x+y)}/W_{y(x+y)}$	$W_{(x-y)(x+y)}/V_{(x+y)}$

The various estimates of each regression differ from one another in two ways. First, they differ in the impact that error variation has on them. Thus $b_{X/Y}$ found as the reciprocal of $b_{Y/X}$ will be inflated by error variation in x just as $b_{Y/X}$ is itself reduced, while $b_{X/Y}$ found directly will be reduced by error variation in y , just as $b_{Y/X}$ found as its reciprocal will be inflated. Similarly $b_{D/S}$ found directly will be reduced by error variation in $x+y$, but if found from $b_{Y/X}$ or $b_{X/Y}$ the reduction will be in its departure not from 0 but from 1. When $b_{Y/X}$ and $b_{X/Y}$ are found from $b_{D/S}$, one will be inflated by error variation and the other reduced according to the relative values of V_x and V_y .

Secondly, if we regard a variance as the covariance of a variate with itself (subject always, of course, to the differing effects of error variation on covariances and variances), $b_{Y/X}$ can be regarded as a weighted mean of y/x using x as the weight, and $b_{X/Y}$ and $b_{D/S}$ found from it are similar weighted means of x/y and $(x-y)/(x+y)$ with x as the weight. When we start with $b_{X/Y}$ we obtain similarly weighted means with y as the weight, and when we start with $b_{D/S}$ the weight is $x+y$. In the absence of error variation the use of the various weights will make no difference, since with a linear relation such as we are discussing between x and y , the two variates will bear a constant relation to one another and weighting by x, y and $x+y$ will come to the same thing in the end. With error variation present, however, the various weights will not be equivalent, as they will be differently affected by the error variation. Often we might expect $x+y$ to be subject to proportionately less disturbance from the error variation, in which case it is better to make $b_{D/S}$ the basis of the analysis, but, as we shall see, this is by no means always the case.

Where estimates of the error variation in x and y are available the variances can be adjusted by their deduction, and with linearity the regressions once again become equivalent to one another. Where, on the other hand, the regressions are not linear, the deviations from linearity will have effects similar to error variation on the estimates of the regression coefficients and the deduction of the error variation itself will not suffice to make these regressions equivalent to one another.

(iii) *Examples*

An example of the effect of non-linearity of the regression is afforded by sternopleural chaeta number in two inbred lines, Samarkand and Wellington, of *Drosophila melanogaster*, raised in six different environments, being all the combinations of two temperatures (25° and 18° C.) with three types of culture container (third pint milk-bottles, 3" × 1" tubes with yeasted medium and similar tubes with unyeasted medium). The actual variate was the mean chaeta number of 10 males and 10 females from each culture.

The observations were made in duplicate on each of two occasions. Since there was no significant difference between occasions there are thus 18 degrees of freedom each for the estimation of error variation in x and y , the chaeta numbers of Samarkand and Wellington respectively. The estimates of error variation in x and y proved not to differ significantly and so were pooled to give a joint estimate of 0.37083 based on 36 degrees of freedom. The values of V_x , V_y , $W_{y,x}$, $V_{(x+y)}$, $V_{(x-y)}$ and $W_{(x-y)(x+y)}$ are shown in table 2, the last three statistics having been divided by 2 to make the error variation content of the two variances equal to that of V_x and V_y . Samarkand is denoted as line X and Wellington as line Y.

TABLE 2
*The Samarkand (X) and Wellington (Y) inbred lines of Drosophila melanogaster.
Sternopleural chaetae in six environments*

V_x	1.12242	V_y	8.68542	$W_{y,x}$	-1.21008	Error variance	0.37083
$V_{(x+y)}$	3.69383	$V_{(x-y)}$	6.11400	$W_{(x-y)(x+y)}$	-3.78150		
				Regression coefficients			
Derived from				$b_{Y/X}$	$b_{X/Y}$	$b_{D/S}$	
				<i>-1.078</i>	<i>-0.928</i>	26.608	
$b_{Y/X}$				<i>-1.610</i>	<i>-0.621</i>	4.279	
				<i>-7.178</i>	<i>-0.139</i>	1.324	
$b_{X/Y}$				<i>-6.871</i>	<i>-0.146</i>	1.341	
				<i>-85.267</i>	<i>-0.012</i>	1.024	
$b_{D/S}$				<i>-15.495</i>	<i>-0.065</i>	1.138	
$\sqrt{V_D/V_S}$				—	—	1.287	
						1.315	
Perfect fit				-6.406	-0.156	1.370	

The figures in italics are the values of the coefficients estimated directly from the data and those in roman are the values of the remaining coefficients derived from them.

The upper figure in each case is where no correction has been made for error variation. The lower figure is that obtained after subtracting the value of the error variation from each variance.

Estimates of $b_{Y/X}$, $b_{X/Y}$ and $b_{D/S}$ were each calculated directly from the data and are shown in italics in table 2. Each is also accompanied by the values of the other two regressions derived from it. Thus $b_{Y/X}$ calculated from the data is $W_{y,x}/V_x = -1.21008/1.12242 = -1.0781$, from which we derive $b_{X/Y} = 1/b_{Y/X} = -0.9276$ and $b_{D/S} = (1 - b_{Y/X})/(1 + b_{Y/X}) = 26.6084$. Needless to say, the three values found for each regression, one directly from the data and others by derivation from the other two regressions, do not agree: indeed some of the disagreements are large. We can, however,

subtract the estimated error variance (0.37083) from all the variances and repeat the calculations. The results, also shown in table 2, are much more consistent, some of the improvements being very marked; but even so inconsistencies remain. The reason is, of course, that the regressions are not simply linear, and the residual variation round the straight regression line mimics the error variation in its effects but is not removed by subtraction of the error variation. Thus regression of y on x accounts for only 1.30459 of the value of 8.68542 for V_y , leaving a residuum of 7.38083 against which the error variance at 0.37083 is small. However, $b_{Y/X}$ is extreme among the three regressions in the effect of residual variation on it, and if instead of 0.37083 we subtract 0.93352 from each of the variances all the residual variation, due to both error and non-linearity, is removed.

If after subtracting 0.93352 from the variances, the covariances remaining of course unaltered, we recalculate the regressions it makes no difference which regression we calculate directly, deriving the others from it: the results are all the same at $b_{Y/X} = -6.4059$, $b_{X/Y} = -0.1561$ and $b_{D/S} = -1.3699$, as shown in the bottom row of table 2. We can now see that of the earlier calculations, obtaining $b_{X/Y}$ from the data and deriving the others from it gave the best results in that not only were the regressions so found least sensitive to the effects of error variation but also least upset by residual variation stemming from non-linearity of the regression. Indeed when we look at $b_{D/S}$, the value found from $b_{X/Y}$ before any correction was 1.32, which became 1.34 when the correction had been made for error variation, by comparison with 1.37 for the perfect fit. The reason for this superiority of $b_{X/Y}$ is not far to seek. Its estimation utilises V_y as the denominator and this is the largest of the variances, all of which, however, have the same component of error variation. Of all the variances, therefore, V_y contains the lowest proportion of error variation and hence is the least distorted by it, with the consequence that the estimate of $b_{X/Y}$ is least reduced by error variation, and also by residual variation simulating error variation. The values found for $b_{Y/X}$ and $b_{D/S}$ by derivation from $b_{X/Y}$ will thus be more useful for analytical and predictive purposes than those found by direct estimation.

Thus $b_{X/Y}$, and not $b_{D/S}$, is the best starting point for consideration of the genotype \times environment interaction shown by the Samarkand and Wellington lines of *Drosophila melanogaster*, even though it may be useful to recast the value found for it into the form of $b_{D/S}$. This will not, however, always be the case. The data given by Bucio Alanis (1966) for plant height in the two lines P_1 and P_5 of *Nicotiana rustica* are insufficient for us to estimate the true error variation from them. We can, however, find the amounts of residual variation by which V_{P_1} , V_{P_5} and $V_{(P_1+P_5)}$ must be reduced to give the perfect fit values for the three regressions. This turns out to be 3.37 in respect of P_1 and P_5 , and it will be twice that value for P_1+P_5 since in this case the variance of the sum has not been divided by 2 to keep the error component constant as was done with the *Drosophila* results. It may be noted that this figure of 3.37 compares with 6.76, 3.82 and 3.83 which Mather and Vines (1952) observed as the corresponding error variances in the years 1946 to 1948. It is thus extremely likely that the whole of the 3.37 is accounted for by error variation, and that there is no residual variation due to non-linearity of the regressions which are thus straight lines.

The observed values of V_{P_1} , V_{P_5} etc. are given in table 3, together with the regressions derived from these observations and also the perfect fit values of the regressions obtained after deducting the appropriate items for residual variation. This time $b_{D/S}$ appears to be least affected by the residual variation with b_{P_1/P_5} running it a close second. Again this is not surprising since $V_{(P_1+P_5)}$ is just over twice as large as V_{P_5} and over eight times as large as V_{P_1} , so that even though its component of residual variation is twice that of V_{P_5} and V_{P_1} , this constitutes a somewhat lower proportion of the total in $V_{(P_1+P_5)}$ than in V_{P_5} and a very much lower proportion than in V_{P_1} . So $b_{D/S}$ will be disturbed less by residual variation than will b_{P_1/P_5} and very much less than b_{P_5/P_1} . $b_{D/S}$ will thus provide the best starting point for the consideration of interaction, though b_{P_1/P_5} will not be greatly inferior.

Before we leave the comparison of these various estimates of the regression, it should be noted that a fourth estimate is possible, since $\sqrt{V_{(x-y)}/V_{(x+y)}} = b_{D/S}$ (Bucio Alanis, Perkins and Jinks, 1969). In this case, by contrast with the three earlier estimates, the numerator is a variance as well as the denominator, and as such it too will be inflated by error variation or residual variation round the simple regression line. The values for $b_{D/S}$ arrived at in this way are given for *Drosophila* in table 2 and *Nicotiana* in table 3. In both cases they are greater than $b_{D/S}$ found directly as

TABLE 3
Plant height of *Nicotiana rustica* lines 1 and 5 grown in 16 environments
(Bucio Alanis, 1966)

V_{P_1}	16.05	V_{P_5}	63.10	W_{P_1/P_5}	27.52
$V_{(P_1+P_5)}$	133.56	$V_{(P_1-P_5)}$	23.48	$W_{(P_1-P_5)/(P_1+P_5)}$	47.05
		Regression coefficients			
Derived from		⏟			
		b_{P_1/P_5}	b_{P_5/P_1}	$b_{D/S}$	
b_{P_1/P_5}		<i>0.436</i>	2.293	0.393	
b_{P_5/P_1}		0.583	<i>1.715</i>	0.263	
$b_{D/S}$		0.479	2.088	<i>0.352</i>	
$\sqrt{V_D/V_S}$		—	—	0.419	
Perfect fit		0.461	2.170	0.369	

The figures in italics are the values of the coefficients estimated directly from the data, and those in roman are the values of the remaining coefficients derived from them.

$W_{(x-y)(x+y)}/V_{(x+y)}$, as would be expected since the denominator is the same in both cases but the one has error variation in the numerator and the other does not. In *Nicotiana* $\sqrt{V_{(x-y)}/V_{(x+y)}}$ would appear to be a less good estimator of $b_{D/S}$ than $W_{(x-y)(x+y)}/V_{(x+y)}$; but in *Drosophila* it appears to be better and is certainly less sensitive to the subtraction of error variation. Even so, it is more disturbed than is $b_{X/Y}$ and the value of $b_{D/S}$ derived from it. So in general there is no indication that the variance ratio offers a better approach than the regression calculated in the normal way, provided the best of the regressions is taken as the starting point. In general error variation should tend to push the variance ratio, and hence the estimate of $b_{D/S}$ found from it, nearer towards 1.

3. THE MULTI-LINE CASE

(i) *General properties*

The treatment of two lines discussed by Bucio Alanis has been generalised for any number of lines by Perkins and Jinks (1968*a, b*). Taking n genetically different lines, $X_1 \dots X_n$ giving phenotypes $x_1 \dots x_n$ they use the mean \bar{x} in each environment as a measure of the effects of that environment, in parallel with Bucio Alanis' use of the sum of the phenotypes of his two lines. The equivalent of the difference between two lines is the difference between the phenotype of any one of the n lines, say x_1 , and the mean. This reflects the overall genetic departure, d , of the line in question from the mean as well as the genotype \times environment interaction, but since d is constant β_1 , the regression of $x_1 - \bar{x}$ on \bar{x} , is a measure of the rate change in the interaction, g , with change in the environment, e .

Thus

$$\beta_1 = \frac{d(x_1 - \bar{x})}{d\bar{x}}.$$

It is often, however, more convenient to follow Perkins and Jinks and find the regression of x_1 on \bar{x} *i.e.* the rate of change of $g + e$ on e . This is, of course,

$$1 + \beta_1 = \frac{dx_1}{d\bar{x}} = n \frac{dx_1}{dS(x)}$$

where $\sum_1^n S(x)$ is the sum of all the x 's in a given environment. $\sum_1^n S(x)$ is obviously not independent of x_1 , but noting that $\sum_1^n S(x) = x_1 + \sum_2^n S(x)$ we can write

$$\begin{aligned} 1 + \beta_1 &= n \frac{dx_1}{dS(x)} = n \left/ \frac{dS(x)}{dx_1} \right. \\ &= n \left/ \left(\frac{dx_1}{dx_1} + \frac{dS(x)}{dx_1} \right) \right. \\ &= n \left/ \left(1 + \frac{dS(x)}{dx_1} \right) \right. = n \left/ \left(1 + \frac{dx_2}{dx_1} + \frac{dx_3}{dx_1} \dots \frac{dx_n}{dx_1} \right) \right. \end{aligned}$$

This will be a straight line only if $\frac{dS(x)}{dx_1}$ is constant, that is only if $f(x_2 + x_3 \dots x_n, z) = kf(x_1, z)$ where k is independent of x and z . That

$\frac{dS(x)}{dx_1}$ is constant does not, of course, imply that $\frac{dx_2}{dx_1}, \frac{dx_3}{dx_1}$ etc. are constant since they could individually contain terms in x which balanced out on summing over $x_2 \dots x_n$. If, however, a number of the lines $X_1 \dots X_n$ all gave rectilinear regressions when compared with the sums, or means, of their fellows in this way, it would be most unlikely for their relations one with another not to be such that their relative rates of change were constant

(that is for their relations to be such as to give $\frac{dx_2}{dx_1}, \frac{dx_3}{dx_1}, \dots, \frac{dx_n}{dx_1}, \dots$, etc. all constant) though of course not necessarily with the same relative rates of change.

Now $\frac{dS(x)}{dx_1}$ measures the rate of change of the summed phenotypes of all the remaining lines, $X_2 \dots X_n$, on that of X_1 . It reduces to $\frac{dy}{dx}$ in the two line case and by analogy with $b_{Y/X}$ the regression to which it gives rise may be denoted as $b_{(X_2 \dots X_n)/X_1}$.

Then

$$1 + \beta_1 = n / (1 + b_{(X_2 \dots X_n)/X_1})$$

and

$$b_{(X_2 \dots X_n)/X_1} = \frac{n}{1 + \beta_1} - 1.$$

Hence if $1 + \beta_1$ is negative or is greater than n , the number of lines, $b_{(X_2 \dots X_n)/X_1}$ must be negative and X_1 is responding to change in the environment in the opposite direction to the sum, or mean, of the rest of the lines. If $1 + \beta_1$ lies outside the range 0 to n , β_1 itself must lie outside the range -1 to $n-1$, which in the two line case reduces to relation already found since β_1 then becomes $b_{D/S}$ and $b_{Y/X}$ must be negative if it lies outside the range -1 to 1. The relation of β_1 and $b_{(X_2+X_3)/X_1}$ is shown in fig. 3 for the three-line case.

One further point should be noted about this relationship. The calculation of $1 + \beta_1$, as the regression of x_1 on \bar{x} is open to criticism since \bar{x} must include $\frac{x_1}{n}$ and so is not independent of x_1 (Freeman and Perkins, 1971; but see also Freeman, 1973). We now see, however that,

$$1 + \beta_1 = \frac{nb_{x_1/(X_2 \dots X_n)}}{1 + b_{(X_1/X_2 \dots X_n)}}$$

and is obtainable therefore from the regression of x_1 on $S(x_2 \dots x_n)$ or, by derivation, from the regression of x_1 on the mean of the remaining $n-1$ x 's. The statistical difficulty can therefore easily be avoided without need to resort to the inclusion in the experiment of additional individuals to supply an independent estimate of the environment, as these authors recommended. We shall see, in the next section, a comparison of values so derived for a set of $1 + \beta$'s with those obtained directly by regression on \bar{x} .

In the multi-line analysis we implicitly compare each line in turn, through the relevant $1 + \beta$, with the mean of the remaining $n-1$ lines. We can, however, if we so wish compare each line with every other in pairs by the two-line analysis of Bucio Alanis and Hill. This may, of course, be undertaken directly in each case by finding the differences and sums, environment by environment, and then estimating the regression $b_{D/S}$ as we did in the previous section. There would be $\frac{n}{2}(n-1)$ such calculations, and if n is at all large considerable labour could be involved. The values of the $b_{D/S}$'s are, however, obtainable much less laboriously from the $1 + \beta$'s.

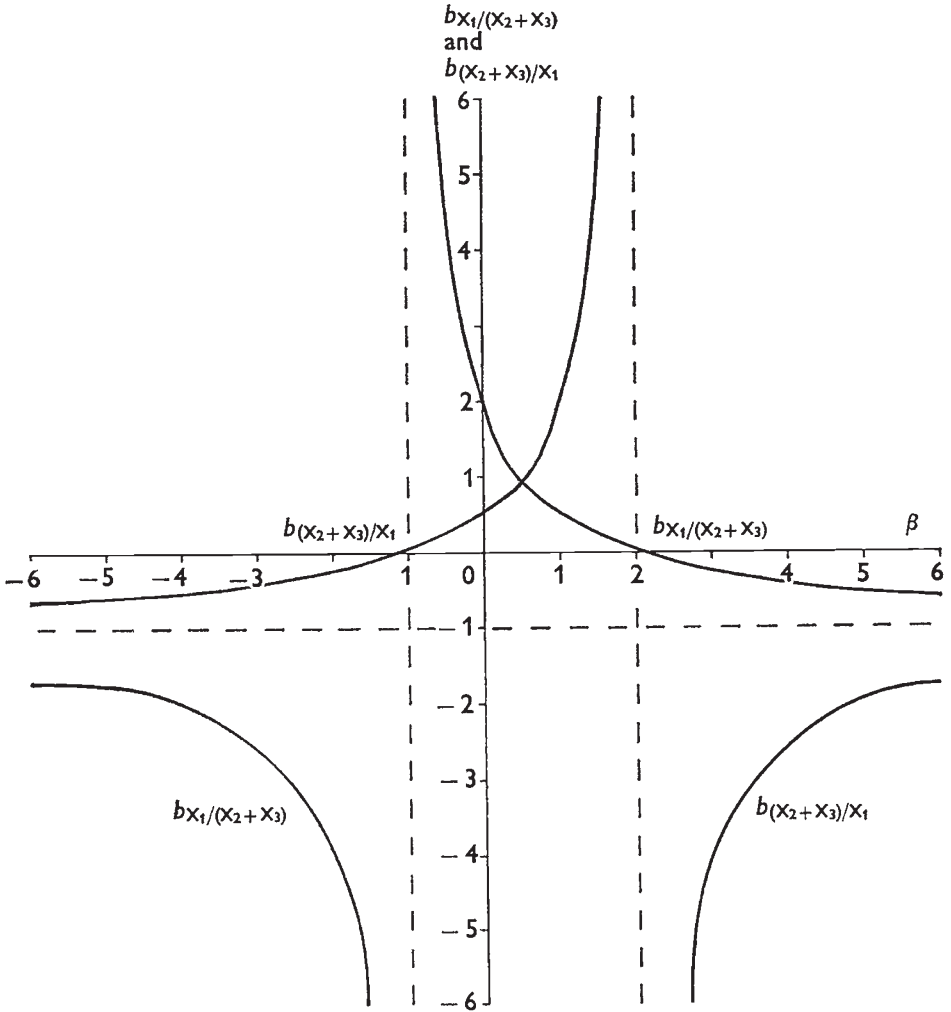


FIG. 3.—The relations of $b_{X_1/(X_2+X_3)}$ and $b_{(X_2+X_3)/X_1}$ with β in the three line case.

Considering lines X_1 and X_2 , the regression of the one on the other, is

$$b_{X_2/X_1} = \frac{dx_2}{dx_1} = \frac{dx_2}{d\bar{x}} \bigg/ \frac{dx_1}{d\bar{x}} = \frac{1 + \beta_2}{1 + \beta_1}$$

and

$$b_{D/S} = \frac{1 - b_{X_2/X_1}}{1 + b_{X_2/X_1}} = \left(1 - \frac{1 + \beta_2}{1 + \beta_1}\right) \bigg/ \left(1 + \frac{1 + \beta_2}{1 + \beta_1}\right) = \frac{(1 + \beta_1) - (1 + \beta_2)}{(1 + \beta_1) + (1 + \beta_2)}$$

Indeed, by simple extension, we can compare any group of lines with any other group, with which it has no line in common, whether the two groups together do or do not include all n lines. Let us consider two groups, which we will denote as A and B, comprising lines $1 \dots j$ and $k \dots l$ respectively.

Then, where β_A is the mean of $\beta_1 \dots \beta_j$ and β_B is the mean of $\beta_k \dots \beta_l$

$$1 + \beta_A = \frac{1}{j} [(1 + \beta_1) + \dots (1 + \beta_j)] = \frac{1}{j} \left[\frac{dx_1}{d\bar{x}} + \dots \frac{dx_j}{d\bar{x}} \right] = \frac{d\bar{x}_{(1\dots j)}}{d\bar{x}}$$

and similarly

$$1 + \beta_B = \frac{d\bar{x}_{(k\dots l)}}{d\bar{x}}$$

Then,

$$b_{\bar{X}_B/\bar{X}_A} = \frac{d\bar{x}_{(k\dots l)}}{d\bar{x}_{(1\dots j)}} = \frac{d\bar{x}_{(k\dots l)}/d\bar{x}}{d\bar{x}_{(1\dots j)}/d\bar{x}} = \frac{1 + \beta_B}{1 + \beta_A}$$

and

$$b_{(\bar{X}_A - \bar{X}_B)/(\bar{X}_A + \bar{X}_B)} = \frac{(1 + \beta_A) - (1 + \beta_B)}{(1 + \beta_A) + (1 + \beta_B)}$$

Since the properties in interaction of any group of lines may be compared with those of any other group by this means, the way is open for a great variety of detailed analyses, whether hierarchial or orthogonal (where this is possible and meaningful) of the n lines in the multi-line case.

(ii) *An example*

Data for illustrating multi-line analysis are afforded by the behaviour of five inbred lines of *Drosophila melanogaster*, Wellington (W), Samarkand (S), Edinburgh (E), Texas 19 (T19) and Texas 20 (T20) raised in third-pint milk bottles at three temperatures, 25°, 21.5° and 18° C. The character recorded was again number of sternopleural chaetae, and the variate was the mean chaeta number of 10 females and 10 males. Also, the observations were made in duplicate so allowing an estimate of error variation to be obtained for 15 degrees of freedom from the comparison of the $5 \times 3 = 15$ pairs of duplicate observations.

TABLE 4

Sternopleural chaeta numbers in five inbred lines of Drosophila melanogaster raised at three temperatures. Each entry is the average of 20 males and 20 females taken equally from two replicates

Line	Temperature (°C.)			Mean	1 + β	1 + β'
	25°	21.5°	18°			
W	18.275	18.975	19.225	18.8250	0.8534	0.8533
S	21.050	20.825	20.975	20.9500	-0.1082	-0.1136
E	23.575	23.350	23.950	23.6250	0.2163	0.1684
T19	16.350	18.450	18.550	17.7833	2.0913	2.0007
T20	16.825	18.475	18.975	18.0916	1.9471	1.9431
Means						
Overall	19.2150	20.0150	20.3350	19.8550		
Excluding W	19.4500	20.2750	20.6125	20.1125		

The results are set out in table 4. The entries in this table are the means of the two duplicate observations and as such were found to be subject to an error variance of 0.064725. The table also gives the mean for each of the five lines over all three temperatures and the mean for each temperature over the five lines. It will be noted that the lines differ in their overall chaeta numbers, and that the temperatures also differ, chaeta number rising

in general as temperature falls though 21.5° is much nearer to 18° than it is to 25° in its effect on chaeta number.

We can then find the value of $1 + \beta$ for each line as the regression of the values for that line at the three temperatures on the corresponding means of all five lines. These five values of $1 + \beta$, one for each line, are shown down the right hand margin of the table. As we have seen these estimates of $1 + \beta$ are open to criticism in that the line means are not independent of the overall means on which they are regressed. As already noted, however, we may estimate $1 + \beta$ from the regression of each line on the mean of the remaining four. To take Wellington as an example, the mean of the remaining four lines is shown for each temperature along the bottom of the table. The regression of W on the mean of the other four lines is then found as the ratio the sum of cross products

$$\frac{[(18.2750 \times 19.4500) + (18.9750 \times 20.2750) + (19.2250 \times 20.6125)]}{-\frac{1}{3}(56.4750 \times 60.3375)}$$

bears to the sum of squares of the means of the four remaining lines,

$$(19.4500^2 + 20.2750^2 + 20.6125^2) - \frac{1}{3}(60.3375)^2$$

and turns out to be 0.823067.

Now where for brevity we denote by R the remaining four lines as a group

$$b_{W/\bar{R}} = 4b_{W/S(R)}$$

and since $n = 5$, from the result on page 53

$$1 + \beta'_W = \frac{5b_{W/S(R)}}{1 + b_{W/S(R)}} = \frac{5 \cdot \frac{1}{4}b_{W/\bar{R}}}{1 + \frac{1}{4}b_{W/\bar{R}}} = \frac{5b_{W/\bar{R}}}{4 + b_{W/\bar{R}}}$$

where $1 + \beta'$ is used to distinguish this estimate from $1 + \beta$ as found earlier. Thus for Wellington

$$1 + \beta'_W = \frac{5 \times 0.823067}{4 + 0.823067} = 0.853261.$$

The five values of $1 + \beta'$, so found, one for each line, are entered on the right of table 4 where they are convenient for comparison with the five corresponding values found earlier for $1 + \beta$. The differences between the two sets of estimates are small, in some cases very small. Evidently the lack of independence of the line mean and the overall mean has made but little trouble in the direct estimate of $1 + \beta$. In any case, however, the values of $1 + \beta'$ which are not open to this objection are available for use in their place, except in the analyses of variance of the lines where they cannot be employed to calculate the regression and remainder sums of squares directly in the customary fashion. If we carry out such analyses using $1 + \beta$, we obtain the results shown in table 5 from which it can be seen that, when the mean squares are compared with the error variance of 0.064725, the regressions for W , T19 and T20 are significant, that there is heterogeneity of the regressions and that no remainder mean square is significant. Evidently the lines respond differently to changes in temperature but all these changes are adequately described by a simple linear relation with the overall effect of the environment, and most likely with one another.

Whether judged by $1 + \beta$ or $1 + \beta'$, T19 and T20 respond to the environmental change in much the same way. W responds in the same direction

but to a somewhat lesser extent. E responds but little to environmental change as does S also, but whereas such change as E shows is in the same direction as T19, T20 and W, the change in S appears, insofar as it is real, to be in the opposite direction, whether judged by $1 + \beta_S$ or $1 + \beta'_S$ since both are negative.

TABLE 5

Regression analysis of variance of the experiment with the five inbred lines of Drosophila melanogaster

Mean squares for	d.f.	Line				
		W	S	E	T19	T20
Regression	1	0.4847	0.0078	0.0311	2.9112	2.5235
Remainder	1	0.0003	0.0185	0.1526	0.1755	0.0082
		d.f.	S.S.	M.S.	P	
		4	5.2927	1.323	< 0.001	
		5	0.3551	0.071	> 0.20	
		15	—	0.0647	—	

It is of interest to compare the difference in response to change in the environment between W and S in this experiment with that observed for the same lines in the other experiment discussed in an earlier section. We may proceed in several ways to find $b_{D/S}$ for these two lines from the present data. We may derive it as

$$b_{D/S} = \frac{(1 + \beta_W) - (1 + \beta_S)}{(1 + \beta_W) + (1 + \beta_S)} = \frac{0.8534 - (-0.1082)}{0.8534 + (-0.1082)} = 1.290$$

or, for comparison, from $1 + \beta'_W$ and $1 + \beta'_S$ as

$$b_{D/S} = \frac{0.8533 - (-0.1136)}{0.8533 + (-0.1136)} = 1.307.$$

We may also, of course, proceed by abstracting the data for W and S and treating them by the methods we have earlier discussed for the two-line case. We then find $b_{D/S} = 1.195$ by direct estimation from the differences and sums of W and S. First finding the regression of S on W and then deriving $b_{D/S}$ from it gives 1.303 and the same procedure but regressing W on S gives 2.400, again showing how the high proportion of error variation in the variance of Samarkand distorts the estimate. Finally $b_{D/S} = \sqrt{\overline{V_D}/\overline{V_S}} = 1.290$. These various values agree well with $b_{D/S}$ as found in the earlier experiment, and help to convince us that, despite the relative insensitivity of S to change of temperature, it must in fact be altering in the opposite direction to W, *i.e.* whereas W's chaeta number rises with falling temperatures that of S falls though at a much lower rate.

All the estimates of $b_{D/S}$ for W and S, whether derived from the multi-line analysis or the two-line treatment, agree well apart from that obtained from the regression of W on S, which as we saw in the earlier section is distorted to a much greater extent than the rest by error variation. Similar sets of estimates of $b_{D/S}$ for all the 10 pairs of lines are set out in table 6. All but one of the pairs, S and E to which we return below, give sets of estimates showing good internal agreement. The only proviso we must make is, of course, that where an estimate is derived from the regression of one line

on the other distortion must be expected if the line used as the independent variate has a low variance and hence a high proportion of error variation. In particular the estimates of $b_{D/S}$ derived from the multi-line analysis are fully consonant with those from the two-line treatment, as indeed the discussion in the previous section would lead us to expect—and once the values of $1 + \beta$ are available they can be obtained with much less labour. It will be seen too that the estimate of $b_{D/S}$ derived from the multi-line analysis is in general affected but little by the choice of $1 + \beta$ or $1 + \beta'$ from which to derive it.

TABLE 6

*Values of $b_{D/S}$ in the experiment with five inbred lines of *Drosophila melanogaster**

	W/S	W/E	W/T19	W/T20	S/E
(a)	1.290	0.596	0.420	0.391	3.000
(b)	1.307	0.670	0.402	0.389	5.145
(c)	1.195	0.334	0.436	0.391	0.553
(d)	2.400	0.225	0.442	0.391	0.661
(e)	1.303	0.613	0.422	0.391	0.176
(f)	1.290	0.696	0.445	0.391	0.688
	S/T19	S/T20	E/T19	E/T20	T19/T20
(a)	1.109	1.118	0.813	0.800	0.036
(b)	1.120	1.124	0.844	0.840	0.015
(c)	1.134	1.112	0.818	0.732	0.050
(d)	1.144	1.128	0.914	0.824	0.041
(e)	1.290	1.416	0.144	0.143	0.058
(f)	1.144	1.127	0.919	0.833	0.105

The sources of the estimates are:

- Multi line analysis* (a) from $1 + \beta$
 (b) from $1 + \beta'$
Two line treatment (c) by direct calculation
 (d) from $b_{A/B}$
 (e) $b_{B/A}$
 (f) $\sqrt{V_D/V_S}$

where A/B are the two lines being compared

The comparison of S and E requires special mention, however. Here all the estimates of $b_{D/S}$ obtained by a two-line treatment are less than 1, and all lie between 0.55 and 0.69 except for that obtained from the regression of E on S which at 0.18 would appear to show the distortion commonly encountered when the variance of Samarkand is directly involved in the estimation. Yet the multi-line analysis yields an estimate of $b_{D/S}$ in excess of 1, whether based on $1 + \beta$ or $1 + \beta'$. Thus the multi-line analysis suggests that E and S are responding to change in temperature in opposite directions, while the two-line treatment suggests they are changing in the same direction. Now the two-line estimates all involve the covariance of E and S, and this is positive. At the same time it is small, 0.0375. Thus should this small positive covariance be merely the outcome of sampling variation, all the two-line estimates will reflect it by being less than 1. The multi-line analysis suggests that this may well be the case. S is consistent in showing $b_{D/S} > 1$, *i.e.* change in opposite direction, when compared with W, T19 and T20. E is consistent in showing $b_{D/S} < 1$ *i.e.* change in the same direction, when compared with these same three lines, and W, T19 and T20 are consistent among themselves in giving every indication that they all change in the

same direction. Thus the two-line treatment of S and E gives a result inconsistent with all the other comparisons among the five lines. The multi-line analysis yields a $b_{D/S}$ which makes the comparison of S and E consistent with all the other comparisons among the five lines, again because the covariances of S and E with the mean of all five lines appear with opposite signs, that of S being negative: indeed S is the only line to produce such a negative covariance with the overall mean. Thus the discrepancy stems from the covariance of S and E being small and positive (0.0375) while that of S and the five-line mean is small and negative (-0.0720). Whichever analysis is used, the sign of the small covariance that it turns on is critical, and from this stems the difficulty and uncertainty in interpreting the behaviour of a line, which like S is relatively insensitive to the environmental factor, here temperatures, at issue. Other data, like that quoted in the earlier section, combine to point to S responding to change in temperature in the opposite direction to W, and so support the outcome of the present multi-line analysis: but however this may be, it is clear that S is relatively very much less sensitive to changes of temperature than are the other lines, and this is indicated more clearly by the low value of its variance over environments, than by any of the regression coefficients.

Acknowledgments.—We gratefully acknowledge financial assistance from the Agricultural Research Council and the Leverhulme Trust, and we thank Professor J. L. Jinks, F.R.S for helpful comments.

4. REFERENCES

- BUCIO ALANIS, L. 1966. Environmental and genotype-environmental components of variability: I. Inbred lines. *Heredity*, 21, 387-397.
- BUCIO ALANIS, L., AND HILL, J. 1966. Environmental and genotype-environmental components of variability. II. Heterozygotes. *Heredity*, 21, 399-405.
- BUCIO ALANIS, L., PERKINS, J. M., AND JINKS, J. L. 1969. Environmental and genotype-environmental components of variability. V. Segregating generations. *Heredity*, 24, 115-127.
- FREEMAN, G. H. 1973. Statistical methods for the analysis of genotype-environment interactions. *Heredity*, 31, 339-354.
- FREEMAN, G. H., AND PERKINS, J. M. 1971. Environmental and genotype-environmental components of variability. VIII. Relations between genotypes grown in different environments and measures of these environments. *Heredity*, 27, 15-23.
- HARDWICK, R. C., AND WOOD J. T. 1972. Regression methods for studying genotype-environment interactions. *Heredity* 28, 209-222.
- MATHER, K., AND JONES, R. M. 1958. Interaction of genotype and environment in continuous variation. I. Description. *Biometrics*, 14, 343-359.
- MATHER, K., AND VINES, A. 1952. The inheritance of height and flowering time in a cross of *Nicotiana rustica*. *Quantitative Inheritance* (ed. E. C. R. Reeve and C. H. Waddington), pp. 49-80, H.M.S.O., London.
- PERKINS, J. M., AND JINKS, J. L. 1968a. Environmental and genotype-environmental components of variability. III. Multiple lines and crosses. *Heredity*, 23, 339-356.
- PERKINS, J. M., AND JINKS, J. L. 1968b. Environmental and genotype-environmental components of variability. IV. Non-linear interactions for multiple inbred lines. *Heredity*, 23, 525-535.
- YATES, F., AND COCHRAN, W. G. 1938. The analysis of groups of experiments. *J. agric. Sci.*, 28, 556-580.