150

THE COMPARISON OF VARIABILITY IN POPULATIONS HAVING UNEQUAL MEANS. AN EXAMPLE OF THE ANALYSIS OF COVARIANCE WITH MULTIPLE DEPENDENT AND INDEPENDENT VARIATES

By B. DAY (*U.S. Forest Service*) and R. A. FISHER, Sc.D., F.R.S.

I. Introduction

Natural populations of *Plantago maritima* in a number of Scottish localities have recently formed the subject of a remarkably thorough ecological investigation by J. W. Gregor. He also collected samples of seed from many populations and grew them for two years under comparable conditions in an experimental garden. In this way characteristic values of means and standard deviations were obtained for a large number of measurable features of 100 plants from each of the different populations. The methods of culture and measurement are given by Gregor *et al.* (1936) and the results are to be presented in detail by Gregor. One of the many points arising in the consideration of data of this kind is that of obtaining a method of comparing the variability of populations differing in their mean values. The present paper is solely concerned with this question. The original data were kindly placed at the disposal of the authors, in advance of publication, by Dr Gregor. Comparison of variabilities undisturbed by differences in the population means are needed for several distinct purposes. Such differences between different species represent the capacity of the species to maintain a store of genetic variance on the maintenance of which their evolutionary progress depends. This evolutionary point was recognized early by Darwin, and one of the authors has already used two bodies of data involving interspecific comparisons for the purpose of verifying, respectively, in moths and in birds, Darwin's law connecting variability with abundance. In comparisons within species variability is also an important index of the extent to which different ecological situations tolerate variation in genotype, and to which observable populations owe their origin to the mingling of material from various sources. It is, above all, important that complex inferences of these kinds should be freed from the confusion that must follow if mere consequences of differences in the average values are mistaken for genuine differences in variability.

II. Ecological classification by regions and types

The twenty-nine populations available are classified ecologically into three regions: Inland, Coastal and Island. The stations of the Coastal region are subdivided into four types of coastal habitat, designated (*a*), (*b*), (*c*) and (*d*), and characterized as follows: (*a*) waterlogged mud, (*b*) typical salt marsh, (*c*) drained mud, (*d*) coastal meadow above highest tide mark.

The Inland populations are usually concentrated in upland districts, far removed from their Coastal relatives. Dr Gregor suggests from the distributional evidence that *P. maritima* spread post-Glacially from the Coastal districts to the higher mountains, subsequently being displaced from the intermediate stations, leaving the mountains as "permanent" loci from which present-day migrations radiate. The downward migration is primarily into small areas on the borders of streams and paths, i.e. into recently formed and more or less isolated and unoccupied habitats.

In contrast to the Inland populations, the Island population represents an isolated unit, whose constituent local populations are more or less continuous geographically. The five Island populations discussed come three from Lewes and two from North Uist. Ecologically the sites are somewhat strongly differentiated. From Lewes there is one population from coastal waterlogged mud, a second from partially populated sea-cliff, and a third from a meadow above highest tide mark. From North Uist both populations come from pastures on the shores of a sea loch.

Dr Gregor's ecological classification may be set out as follows:

Table I. *Ecological classification of twenty-nine populations*

	In types	In regions
Inland	8	8
Coastal (*a*)	5	
(*b*)	5	
(*c*)	3	
(*d*)	3	
Total		16
Island	5	5
Total		29

The classification carries with it the subdivision of the 28 degrees of freedom, or independent comparisons, among twenty-nine populations, into three parts. We have first the comparisons among regions, supplying 2 degrees of freedom. Within the Coastal region we have the comparison among the four types, giving 3 degrees of freedom between types within regions, while within types, taking the Inland or Island regions as each comprising only one type, there remain 23 degrees of freedom. These latter could, of course, be subdivided if desired, seven being within the Inland type, four within Coastal (*a*), and so on. All analyses of the material will therefore be of the form

	Degrees of freedom
Among regions	2
Among types within regions	3
Within types	23
Total	28

As a preliminary example we may take the variation in leaf length shown by the twenty-nine populations (Table XVI). Since the standard deviations are already available, it will be convenient to use the logarithm of these (to the base 10) to three decimal places. The logarithm of the standard deviation is, of course, half the logarithm of the corresponding variance, so that the analyses will be equivalent. For simplicity the three-figure logarithms are multiplied by 1000, so as to yield whole numbers. The process of analysis is, of course, much expedited by the use of machines. It may, however, be as well to illustrate the analysis of a hierarchical classification of this kind, as it would be carried out without mechanical aids.

Table II

	Total values for			Squares	Subtotals	Total × Mean	
	Populations	Types	Regions			Types	Regions
Inland	616			7,056			
	609			8,281			
	770			4,900			
	754			2,916			
	629			5,041			
	589			12,321			
	734			1,156			
	576	5,277	5,277	15,376	57,047	13,041·12	13,041·12
Coastal (*a*)	722			484			
	732			1,024			
	732			1,024			
	695			25			
	740	3,621		1,600	4,157	2,928·20	
(*b*)	770			4,900			
	762			3,844			
	713			169			
	731			961			
	683	3,659		289	10,163	5,056·20	
(*c*)	649			2,601			
	754			2,916			
	663	2,066		1,369	6,886	385·33	
(*d*)	754			2,916			
	614			7,396			
	682	2,050	11,396	324	10,636	833·33	2,401·00
Island	585			13,225			
	702			4			
	652			2,304			
	702			4			
	652	3,293	3,293	2,304	17,841	8,569·80	8,569·80
			19,966		−3,846·76	−3,846·76	−3,846·76
					102,883·24	26,967·22	20,165·16
Degrees of freedom					28	5	2

Table II shows the three-figure values of 1000 log (S.D.) in a parallel column with their squares, or rather the squares of their deviations from the working mean taken at 700.

Parallel with the main left-hand column are subordinate columns showing subtotals by types, and by regions. The main column and the two subordinate columns would all alike give the same grand total, 19,966, for the twenty-nine populations. In the main column for squares are shown the squares of the deviations of the population values from the working mean 700. A subordinate column, which is not strictly necessary, unless the 23 degrees of freedom within types are to be subdivided, shows the subtotals of these values.

The two remaining columns on the right give the products of total × mean, still using deviations from 700, for each type and region. Thus, for the Inland region the deviation from 5600 is -323, and the entry is $323^2 \div 8 = 13{,}041{\cdot}12$, which is entered both as a typical and as a regional value.

The correction to the general mean is carried out by a similar calculation on the grand total. 29×700 is 20,300; the deviation is -334, and the correction to the general mean, $334^2 \div 29$, is deducted from all three of the right-hand columns. After this deduction, the totals of these columns now represent the sums of squares required in an analysis of variance table, corresponding respectively to the twenty-eight comparisons among all populations, the five comparisons among the six ecological types distinguished, and the two comparisons among the three regions. Evidently, each total includes those which follow it. The separate contributions are shown in the analysis of variance below.

Table III. *Analysis of variance of variability figures for leaf length in twenty-nine ecologically classified populations*

	Degrees of freedom	Sums of squares	Mean square
Regions	2	20,165·16	10,082·58
Types within Coastal region	3	6,802·06	2,267·35
Total types	5	26,967·22	
Populations within types	23	75,916·02	3,300·70
Total populations	28	102,883·24	

It is evident that populations from different types of the Coastal region are not differentiated in respect of variability in leaf length. The mean square for regions is great enough to suggest differentiation, though, so far as the present analysis goes, it falls short of the 5 per cent level of significance.

Table IIIa. *Test of significance of differentiation among regions in variability of leaf length*

	Degrees of freedom	Mean square	$\frac{1}{2}\log_e$
Regions	2	10,083	1·1554
Within types	23	3,301	0·5971
z			0·5583
5% value			0·6151

III. Allowance for differences in average size between different populations

For populations having the same mean value in any measurement, comparisons of variability may be made directly, using the standard deviations or variances. When, however, the populations to be compared differ also in their means, such comparisons may not provide what is wanted; for, great differences in average size would naturally be expected to be accompanied by similar differences in the variances, and the comparison of these, without allowance for differences in the means, might supply little information beyond a repetition of what would be learnt from an analysis of the means themselves.

In early biometrical work it was a common practice to divide the standard deviation by the mean, and obtain the so-called "coefficient of variation". If logarithms are used, this is equivalent to analysing the differences, obtained from each population, between the log (S.D.) and the log (mean). If, in fact, the nature of the populations was such that, other things being equal, an increase in the mean carried with it a proportional increase in the standard deviation, this method of allowance would be correct. There is, however, no logical necessity for this to be so. The use of the coefficient of variation rests on a postulate about natural variation, which may or may not be true. If it is not true, the coefficient of variation is a mere convention. When tested empirically it has been found to be untrue in, for example, the variation of the egg size of wild birds. It is, therefore, preferable in all cases to derive the allowance to be made from the data themselves, by means of the analysis of the covariance of the means and standard deviations in the populations observed.

In Section II we have analysed the differences in variability of leaf lengths in portions appropriate to the ecological classification. An exactly similar analysis may be applied to the mean leaf lengths (or rather their logarithms) of the twenty-nine populations. Further, using the product of the two variates, we may obtain a third analysis in the same form. We shall speak of the log (mean) of each population as the independent variates, x_L, and the log (S.D.) as the dependent variate, y_L, since it is the variation of the latter which is required, after allowance has been made for variation in the mean.

The analysis of covariance, sums of squares and products, for x_L and y_L, is shown in Table IV.

Table IV. *Analysis of covariance for logarithmic measures of means and standard deviations of leaf length*

	Degrees of freedom	x_L^2	$x_L y_L$	y_L^2
Populations within types	23	108,577·37	14,147·96	75,916·02
Types within Coastal region	3	161,372·37	−28,738·72	6,802·06
Populations within regions	26	269,949·74	−14,590·76	82,718·08
Regions	2	62,559·50	25,620·00	20,165·16
All populations	28	332,509·24	11,029·24	102,883·24

All the entries in the two columns for squares must necessarily be positive. In the column for products positive values will occur when groups characterized by high mean values also have generally higher variances; when the reverse is the case the sign will be negative. It is seen that within types there is generally a small positive association, but that this is more than counterbalanced by a negative association among the types of the Coastal region, where types (*a*) and (*b*) have distinctly lower means, and slightly higher standard deviations than (*c*) or (*d*). Between regions, however, the association is again positive, giving a small positive balance.

For the populations as a whole, unit increase in x_L is followed by an average increase of 11,029·24 ÷ 332,509·24, or 0·0332 in y_L. This is the regression of y_L on x_L. The allowance made by using the coefficient of variation is thus about 30 times too large, when only this measurement is considered. For comparisons within regions the regression is −0·0540, so that the use of the coefficient of variation introduces an allowance in the wrong direction. Finally, for comparisons within types the regression is 0·1303, about an eighth of that supposed in the use of the coefficient of variation. It is clear that such coefficients would be useless in this material.

In order to obtain an analysis of y_L, after an appropriate allowance for x_L has been made, we may, in the three lines for populations, with 23, 26 and 28 degrees of freedom respectively make a deduction from y_L^2 representing the one degree of freedom for the dependence of y_L on x_L. This deduction is found by multiplying the regression by the product entry, or, alternatively, by squaring the entry under $x_L y_L$, and dividing by the entry in x_L^2. The deduction is thus always positive, and the sum of squares left after deducting it is ascribed to one degree of freedom less than it had before. Table V shows the process of adjustment.

Table V. *Analysis of variance of y_L, making allowance for the variation of x_L*

	Degrees of freedom	y_L^2	Deduction	y_L^2 eliminating x_L	Degrees of freedom	Mean square
Populations within types	23	75,916·02	1,843·52	74,072·50	22	3,367
Types within Coastal region	3	6,802·06		7,856·95	3	2,619
Populations within regions	26	82,718·08	788·63	81,929·45	25	
Regions	2	20,165·16		20,587·95	2	10,294
All populations	28	102,883·24	365·84	102,517·40	27	

As might be judged from the low regressions seen in the analysis of covariance, the elimination of variation in x_L has made scarcely any difference. Actually all three mean squares have been raised slightly in value. The ratio of that for regions to that within types is practically unchanged. So far as concerns variability in leaf length only, the influence of mean leaf length might have been ignored.

Table VA. *Test of significance of differentiation among regions in variability in leaf length allowing for differentiation in mean leaf length*

	Degrees of freedom	Mean square	$\frac{1}{2}\log_e$
Regions	2	10,294	1·1655
Within types	22	3,367	0·6070
z			0·5585
5% value			0·6182

IV. The multiplicity of dependent variates

We have used the analysis of covariance to show how to eliminate the influence of one variate on another, while retaining the structure of the ecological classification. It is apparent, however, from Table IV, that, if we are interested in the variability of more than one measurement of the same organ, such as the length, breadth and thickness of the leaf, we should need three analyses of sums of squares, and three of sums of products, in order to specify the situation adequately. The simultaneous variation of a system of multiple variates is of special importance, for, while the organ itself is a physical reality, any one measurement of it, such as its length, is only an abstraction, and it is usually only a set of measurements which can specify the real situation at all adequately. The simultaneous analysis of variability in length, breadth and thickness of leaf, as measured by the logarithms of the standard deviations, is set out below in relation to the ecological classification of the twenty-nine populations.

Table VI. *Analysis of covariance of variability in length, breadth and thickness of leaf, for twenty-nine ecologically classified populations*

	Within types	Types of Coastal region	Within regions	Regions	All populations
Degrees of freedom...	23	3	26	2	28
y_L^2	75,916	6,802	82,718	20,165	102,883
$y_L y_B$	22,046	−12,176	9,870	38,715	48,585
$y_L y_T$	33,120	22,000	55,120	38,200	93,320
y_B^2	69,780	32,477	102,257	75,318	177,574
$y_B y_T$	−9,427	−61,766	−71,193	72,472	1,279
y_T^2	154,881	134,516	289,397	73,126	362,523

A great deal may be learnt from the inspection of such a table. There is more difference in the variabilities in breadth than in length, both among types in the same region, and among regions; within types they are nearly equally uniform. There is more difference in the variabilities in thickness than in breadth, except among regions. Between one region and another the variabilities in thickness and in breadth differ about equally. The variabilities in length and breadth are positively associated within types, while among the types of the Coastal region high variability in length goes on the contrary with low variability in breadth. Among regions, again, the association is strong and positive. Without the ecological classification, and the corresponding segregation of the degrees of freedom, these different associations would be hopelessly confused.

Variabilities, again, in length and thickness are positively but slightly associated in comparisons within types, more strongly among types within the Coastal region, and most strongly among regions. In contrast with these, variabilities in breadth and thickness are negatively, but weakly, associated within types, more strongly and still negatively associated in comparisons between the types of the Coastal region, whereas between regions they are almost identical. The ecological associations of variability, and especially of the different variabilities of different measurements, appear to be extremely complex. We must, however, remember that the figures of Table VI are based on actual logarithmic standard deviations, and that no allowance has been made in them for the rather large differences which the different populations show in their means for some of these measurements.

V. The multiplicity of independent variates

In order to adjust such material as is shown in Table VI for the three independent variates provided by the means of the length, breadth and thickness of the leaves, we shall need first six analyses of the three variances and the three covariances of the means, which can be set out in a manner similar to Table VI and, next, nine analyses of the covariances of each dependent with each independent variate in turn.

Table VII. *Six analyses of variance and covariance for the independent variates; mean lengths, breadths and thicknesses of leaves*

	Within types	Types of Coastal region	Within regions	Regions	All populations
Degrees of freedom...	23	3	26	2	28
x_L^2	108,577	161,372	269,949	62,560	332,508
$x_L x_B$	85,623	227,442	313,065	75,910	388,975
$x_L x_T$	20,384	3,214	23,598	18,698	42,296
x_B^2	126,988	324,779	451,767	115,673	567,440
$x_B x_T$	−3,674	774	−2,900	53,423	50,523
x_T^2	44,709	7,934	52,643	45,677	98,320

It appears from Table VII that in all comparisons the mean breadths of different populations, when measured logarithmically, differ more than the mean lengths. The difference is striking between regions, and between types of the Coastal region, but less so within types. Differences between the mean lengths are greater than those between mean thicknesses to an enormous extent among the types of the Coastal region, and only moderately between regions. In all comparisons differences in mean length and breadth are positively associated, but mean thickness is almost independent of mean breadth within types and within regions, being but slightly associated with mean length in both these comparisons; whereas between regions breadth and thickness are strongly associated. With these rather complex relationships among the independent variates, it becomes particularly important to eliminate the influence of mean length, breadth and thickness before judging of the relationships among the variabilities.

Table VIII. *Nine analyses of covariance between the three dependent variates* y_L, y_B, y_T, *and the three independent variates* x_L, x_B, x_T

	Within types	Types of Coastal region	Within regions	Regions	All populations
Degrees of freedom...	23	3	26	2	28
$x_L y_L$	14,148	−28,739	−14,591	25,620	11,029
$x_B y_L$	−15,821	−42,680	−58,501	46,185	−12,316
$x_T y_L$	3,979	−230	3,749	27,349	31,098
$x_L y_B$	37,041	68,660	105,701	54,631	160,332
$x_B y_B$	51,665	98,847	150,512	91,796	242,308
$x_T y_B$	−11,020	−3,667	−14,687	49,598	34,911
$x_L y_T$	−1,552	−114,930	−116,482	43,750	−72,732
$x_B y_T$	−7,833	−172,458	−180,291	84,743	−95,547
$x_T y_T$	46,764	18,020	64,784	54,368	119,152

From Table VIII it appears at once that though the covariance of x_L and y_L is generally small, yet some of the values for $x_B y_B$ and $x_T y_T$ are very considerable; so that the elimination of the effects of mean breadth and mean thickness may have important effects. Some pairs of cross-covariances also, such as $x_L y_B$ and $x_B y_T$ have large values, so that we can scarcely expect to get comparable values for the variability of any measurement unless the two other measurements of the leaf are also allowed for.

To allow simultaneously for the effects within types of x_L, x_B and x_T on any other variate, such as y_L, we need the simultaneous regressions given by the equations

$$\begin{aligned} 108{,}577b_1 + 85{,}623b_2 + 20{,}384b_3 &= 14{,}148, \\ 85{,}623b_1 + 126{,}988b_2 - 3{,}674b_3 &= -15{,}821, \\ 20{,}384b_1 - 3{,}674b_2 + 44{,}709b_3 &= 3{,}979. \end{aligned}$$

The coefficients of the regressions on the left are all taken from the first column of Table VII. The right-hand numbers are drawn from the corresponding column of Table VIII, selecting the three values involving y_L. It is usually advantageous to solve such equations by inverting the matrix of coefficients, i.e. by replacing the numbers on the right by 1, 0, 0; 0, 1, 0; 0, 0, 1 and solving these three sets of equations. The solutions form the inverted matrix. For the 23 degrees of freedom within types we have:

Table IX. *Matrix of multipliers for 23 degrees of freedom within types, expressed in millionths*

25·5578	−17·6117	−13·1000
−17·6117	20·0296	9·6758
−13·1000	9·6758	29·1347

If the rows of the matrix be multiplied respectively by the sums of products of y_L, i.e. 14,148, −15,821, 3,979, and added, we obtain the simultaneous regressions of y_L on x_L, x_B and x_T. Doing the same with the sums of products of y_B and y_T, we obtain their simultaneous regressions on the same three variates. We have thus the three regression formulae, in which Y_L, Y_B, Y_T stand for expectations based on these mean measurements

$$Y_L = 0{\cdot}588{,}099x_L - 0{\cdot}527{,}555x_B - 0{\cdot}222{,}498x_T,$$
$$Y_B = 0{\cdot}181{,}139x_L + 0{\cdot}275{,}852x_B - 0{\cdot}306{,}391x_T,$$
$$Y_T = -0{\cdot}514{,}335x_L + 0{\cdot}322{,}932x_B + 1{\cdot}306{,}994x_T.$$

Hence we see that, when allowance is made also for mean breadth and mean thickness, the standard deviation in length is considerably affected by the mean length, contrary to what appears in the preliminary analysis. where the mean breadths and thicknesses were ignored. Associated variation in mean breadth and mean thickness had thus obscured the correction for mean length. For the rest, the values above show clearly how far from adequate would be an analysis based on the coefficients of variation, which would be equivalent to putting unity for the values in the leading diagonal, and zero for all other regressions.

The regression equations supply the means of adjusting the observed standard deviations of length, breadth and thickness so as to obtain comparable values for populations having different mean measurements. It is, however, of more general interest to use them to adjust the analysis of variance and covariance shown in Table VI.

If we multiply the three regressions of y_L by the corresponding sums of products of y_L with x_L, x_B and x_T, we obtain 15,781, which is the portion of the sum of squares of y_L accounted for by its covariance with the independent variates. Similarly, we may evaluate the parallel contribution to the sum of squares of y_B and y_T. Moreover, if we use the regressions of y_L as multipliers of the sums of the products of y_B, or vice versa, we obtain the corresponding ingredient of the analysis of the covariance of y_L and y_B. The contributions of linear regression within types on the mean measurements and the residual variation unaccounted for by this regression are shown in Table X.

In all three characters the residual variation is seen to be substantially diminished, so that, not only the sums of squares, but the mean squares are reduced. For breadth and thickness the reduction is significant (Table XB). It is noticeable, also, that the covariances are in all cases increased, so that the small negative association of y_B and y_T is reversed while y_L is significantly associated with y_B and y_T (Table XA).

Table X. *Portion of variance and covariance within types calculable from mean values*

	Within types, total	Due to regression on mean values	Remainder, allowing for mean values
Degrees of freedom...	23	3	20
y_L^2	75,916	15,781	60,135
$y_L y_B$	22,046	−3,020	25,066
$y_L y_T$	33,120	−7,185	40,305
y_B^2	69,780	24,338	45,442
$y_B y_T$	−9,427	−16,770	7,343
y_T^2	154,881	59,389	95,492

Table XA. *Correlations between variabilities within types*

	r Before adjustment	r After adjustment
y_L and y_B	+0·3029	+0·4795
y_L and y_T	+0·3054	+0·5319
y_B and y_T	−0·0907	+0·1115
5% value	0·4060	0·4329

Table XB. *Significance of reductions in variability*

Degrees of freedom	y_L^2		y_B^2		y_T^2	
	Mean squares	$\frac{1}{2}\log_e$	Mean squares	$\frac{1}{2}\log_e$	Mean squares	$\frac{1}{2}\log_e$
3	5,260	0·8300	8113	1·0468	19,796	1·4927
20	3,007	0·5504	2272	0·4104	4,775	0·7817
		0·2796		0·6364		0·7110
5% value		0·5654		0·5654		0·5654

As in the table for simple regression (Table IV), we may now proceed in a parallel fashion to determine the regressions for the 26 degrees of freedom, within regions. The matrix of coefficients of the simultaneous equation is:

269,949	313,065	23,598
313,065	451,767	−2,900
23,598	−2,900	52,643

Inverting this we find (in millionths) the multiplier-matrix:

24·6664	−17·1704	−12·0032
−17·1704	14·1667	8·4775
−12·0032	8·4775	24·8437

giving the simultaneous correction equations:

$$Y_L = \quad 0{\cdot}599{,}584x_L - 0{\cdot}546{,}455x_B - 0{\cdot}227{,}673x_T,$$
$$Y_B = \quad 0{\cdot}199{,}196x_L + 0{\cdot}192{,}830x_B - 0{\cdot}357{,}653x_T,$$
$$Y_T = -0{\cdot}555{,}151x_L - 0{\cdot}004{,}875x_B + 1{\cdot}479{,}214x_T.$$

These supply corrections appropriate for comparisons within regions. Compared with the corrections within types, the only coefficients largely altered are those of x_B and x_T in the formula for Y_T.

We may now separate the contribution of the regression from the remainder in the third column of Table VI, as follows:

Table XI. *Portion of variance and covariance within regions calculable from mean values*

	Within regions, total	Due to regression on mean values	Remainder, allowing for mean values
Degrees of freedom...	26	3	23
y_L^2	82,718	22,366	60,352
$y_L y_B$	9,870	−15,528	25,398
$y_L y_T$	55,120	13,931	41,189
y_B^2	102,257	55,331	46,926
$y_B y_T$	−71,193	−81,138	9,945
y_T^2	289,397	161,373	128,024

As in Table X we see that all the variabilities are positively associated, when allowance is made for the mean values. Indeed the residuals of Table XI differ little from those of Table X, save for the increase in y_T^2. The reduction in the mean squares is near the 5 % point for y_L and beyond 1 in 1000 for y_B and y_T (Table XIA).

Table XIA. *Significance of reduction of mean squares*

Degrees of freedom	y_L		y_B		y_T	
	Mean squares	$\frac{1}{2}\log_e$	Mean squares	$\frac{1}{2}\log_e$	Mean squares	$\frac{1}{2}\log_e$
3	7,455	1·0044	18,444	1·4574	53,791	1·9926
23	2,624	0·4824	2,040	0·3564	5,566	0·8583
z		0·5220		1·1010		1·1343
5 % value		0·5540				
0·1 % value				1·0186		1·0186

For adjusting the last column of Table VI we take from Table VII the matrix of coefficients:

332,508	388,975	42,296
388,975	567,440	50,523
42,296	50,523	98,320

the inverse of which is:

$$\begin{matrix} 15\cdot32704 & -10\cdot39509 & -1\cdot25191 \\ -10\cdot39509 & 8\cdot89694 & -0\cdot09992 \\ -1\cdot25191 & -0\cdot09992 & 10\cdot76077 \end{matrix}$$

giving the simultaneous correction formulae:

$$Y_L = 0\cdot258{,}138x_L - 0\cdot227{,}331x_B + 0\cdot322{,}062x_T,$$
$$Y_B = -0\cdot105{,}115x_L + 0\cdot485{,}655x_B + 0\cdot150{,}741x_T,$$
$$Y_T = -0\cdot270{,}712x_L - 0\cdot105{,}928x_B + 1\cdot382{,}762x_T.$$

These are the corrections appropriate for comparison between regions; they differ rather largely from those within regions. The portions of the general variance and covariance accounted for by them are shown in Table XII.

Table XII. *Portions of general variance and covariance calculable from mean values*

	All populations, total	Due to regression on mean values	Remainder, allowing for mean values
Degrees of freedom...	28	3	25
y_L^2	102,883	15,662	87,221
$y_L y_B$	48,585	−2,453	51,038
$y_L y_T$	93,320	41,320	52,000
y_B^2	177,574	106,088	71,486
$y_B y_T$	1,279	−20,797	22,076
y_T^2	362,523	194,569	167,954

The reduction of the mean square of y_L is here definitely insignificant, while the mean squares of y_B and y_T are very significantly reduced (Table XIIA).

Table XIIA. *Significance of reduction of mean squares*

Degrees of freedom	y_L		y_B		y_T	
	Mean squares	$\frac{1}{2}\log_e$	Mean squares	$\frac{1}{2}\log_e$	Mean squares	$\frac{1}{2}\log_e$
3	5,221	0·8263	35,363	1·7828	64,856	2·0861
25	3,489	0·6248	2,859	0·5252	6,718	0·9524
z		0·2015		1·2576		1·1337
5% point		0·5478				
0·1% point				1·0041		1·0041

Tables X, XI and XII now enable us to make a full analysis of the simultaneous variation of the three standard deviations, when allowance is made for variations in the means.

Table XIII. *Analysis of covariance of variability in length, breadth and thickness of leaf for twenty-nine ecologically classified populations, eliminating linear regressions on the three means. The values for types and regions are found by subtraction*

	Within types	Types of Coastal region	Within regions	Regions	All populations
Degrees of freedom...	20	3	23	2	25
y_L^2	60,135	217	60,352	26,869	87,221
$y_L y_B$	25,066	332	25,398	25,640	51,038
$y_L y_T$	40,305	884	41,189	10,811	52,000
y_B^2	45,442	1,484	46,926	24,560	71,486
$y_B y_T$	7,343	2,602	9,945	12,131	22,076
y_T^2	95,492	32,532	128,024	39,930	167,954

Table XIIIA. *Significance of differences among the types of the Coastal region*

	Degrees of freedom	y_T		Degrees of freedom	x_L		x_B	
		Mean squares	$\frac{1}{2}\log_e$		Mean squares	$\frac{1}{2}\log_e$	Mean squares	$\frac{1}{2}\log_e$
Between types	3	10,844	1·1918	3	53,791	1·9926	108,260	2·3423
Within types	20	4,775	0·7817	23	4,721	0·7760	5,521	0·8543
z			0·4101			1·2166		1·4880
5% value			0·5654					
0·1% value						1·0186		1·0186

The information of Table XIII about variability is now independent of the information of Table VII about mean values.

VI. Discussion

From Table VII it appears that the different types of the Coastal region are differentiated in mean length and breadth of leaf, these two variates being somewhat highly associated. Their mean thickness is practically constant for all types. From Table XIII it appears further that the variability of both length and breadth of leaf is practically constant in these types, but that the variability in thickness approaches significance (Table XIIIA), when allowance is made for the mean measurements. At this stage it is useful to look at the values given by the individual types (Table XIV).

Table XIV. *Variates showing differences among types of the Coastal region; y_T' is the mean of y_T adjusted by the observed regressions within types*

Type	x_L	x_B	y_T'
Waterlogged mud (*a*)	278	704	295
Typical salt marsh (*b*)	370	805	254
Drained mud (*c*)	509	1006	197
Meadow (*d*)	525	1063	83

The same linear sequence is maintained for all three variates. The ecological conditions which have favoured a large leaved population have also favoured one with low variance in leaf thickness. Among populations growing in the same type of habitat, on the other hand, the regressions show that variability in thickness is greatly favoured by a high average thickness, somewhat favoured by low average length, and practically uninfluenced by average breadth. The differentiation of the types in variability for thickness is, therefore, in no sense, an indirect physiological consequence of the differences in leaf length and breadth, which has been fully eliminated in the comparison.

The three regions show significant differences both in the means and in the variabilities. Table XV shows the averages of the means x_L, x_B, x_T, and of the standard deviations y'_L, y'_B, y'_T adjusted for regression within regions.

Table XV. *Differentiation of regions; y corrected for regression within regions*

	x_L	x_B	x_T	y'_L	y'_B	y'_T
Inland	368	761	958	621	205	207
Coastal	396	860	1050	731	307	193
Island	268	702	1010	659	235	89

The major differences in mean values are that the Inland region has thin leaves; the Coastal region has leaves broad, on the average, but very variable in length and breadth between types; the Island region has short leaves. The Inland populations are least variable in length and breadth, but most variable (on a logarithmic scale) in thickness. The broad-leaved Coastal populations are most variable in length and breadth, and nearly so in thickness. The (short leaved) Island populations are distinctly the least variable in thickness. As in the previous comparison, the contrasts in variability have been completely freed from any necessary association with the mean values.

Table XVA. *Significance of differences in variability between regions*

	Degrees of freedom	y'_L		y'_B		y'_T	
		Mean squares	$\frac{1}{2}\log_e$	Mean squares	$\frac{1}{2}\log_e$	Mean squares	$\frac{1}{2}\log_e$
Between regions	2	13,434	1·2989	12,280	1·2540	19,865	1·4945
Within regions	23	2,624	0·4824	2,040	0·3564	5,566	0·8583
z			0·8165		0·8976		0·6362
5% point							0·6151
1% point			0·8670		0·8670		

Table XVI. *Primary data used for analysis*

		Log standard deviations			Log means		
		Y_L	Y_B	Y_T	X_L	X_B	X_T
Coastal (*a*)		722	262	255	199	617	1,009
		732	248	532	303	771	1,093
		732	215	255	220	590	1,072
		695	218	230	344	736	1,072
		740	362	230	322	806	986
$N=5$	Total	3,621	1,305	1,502	1,388	3,520	5,232
Coastal (*b*)		770	301	255	367	736	1,097
		762	386	301	380	811	1,075
		713	228	230	344	751	1,056
		731	330	230	367	870	1,033
		683	270	230	394	856	991
$N=5$	Total	3,659	1,515	1,246	1,852	4,024	5,252
Coastal (*c*)		649	290	380	430	972	1,079
		754	326	255	538	961	1,123
		663	389	79	558	1,086	1,068
$N=3$	Total	2,066	1,005	714	1,526	3,019	3,270
Coastal (*d*)		754	380	79	550	1,059	1,025
		614	431	0	508	1,069	1,000
		682	354	41	517	1,061	1,029
$N=3$	Total	2,050	1,165	120	1,575	3,189	3,054
$N=16$	Coastal total	11,396	4,990	3,582	6,341	13,752	16,808
Inland		616	204	146	356	696	1,000
		609	223	114	364	756	1,017
		770	258	176	294	696	913
		754	190	79	330	688	919
		629	274	79	400	820	903
		589	134	41	324	731	913
		734	253	230	442	838	1,025
		576	207	79	436	865	977
$N=8$	Inland total	5,277	1,743	944	2,946	6,090	7,667
Island		585	104	114	61	588	991
		702	267	114	260	694	954
		652	193	176	309	728	1,072
		702	164	176	365	670	1,041
		652	262	79	346	831	991
$N=5$	Island total	3,293	990	659	1,341	3,511	5,049
$N=29$	Grand total	19,966	7,723	5,185	10,628	23,353	29,524

REFERENCE

Gregor, Davey & Lang (1936). "Experimental taxonomy. I. Experimental garden technique in relation to the recognition of the small taxonomic units." *New Phytol.* **35**, 323–50.