

THE USE OF REGRESSIONS TO INCREASE HERITABILITY

By J. M. RENDEL*

[Manuscript received March 18, 1954]

Summary

The difficulties of applying correction factors to eliminate non-genetic variance in selective breeding work are discussed. The use of a second character as an index of environmental change is explored and related to the general formula for selection indices. Tibia length and body weight in mice are used to demonstrate points made in the theoretical discussions of the first part of the paper.

I. INTRODUCTION

When selection for a trait is practised in order to increase its expression in a population the expected rate of genetic change, ΔG , is proportional to the selection differential \bar{i} , the heritability of the trait in the population h^2 , and the genetic variance σ_G^2 , such that

$$\Delta G = \bar{i}h\sigma_G \dots \dots \dots (1)$$

Whenever the genetic situation cannot be analysed into the effects of individual gene loci the breeder attempts to maximize expression (1) by manipulating those factors which he can control. \bar{i} , the difference in standard deviations between the mean of the population and the mean of those individuals which become parents, is determined by the fraction of the population saved as parents, and so by the rate of reproduction of the species; σ_G^2 is not readily changed except by crossing, a procedure which may undo the effects of selection; so most breeding programmes aim to maximize h^2 without reducing the other terms or increasing the generation length. Family selection and progeny testing can both be regarded as increasing h^2 by obtaining a measure of the parents selected that is a more accurate estimate of their genotype than a simple phenotypic score. The commonest method, however, is to reduce the phenotypic variation to a minimum by calculating environmental trends and correcting individual scores for the calculated effects of definable environmental effects. As $h^2 = \sigma_G^2 / (\sigma_G^2 + \sigma_E^2)$ in its simplest form, anything which reduces σ_E^2 increases h^2 .

II. CORRECTION FACTORS

Before a correction can be applied it is first necessary to classify the individual as having been subjected to some environment with known phenotypic effect. It is simple to score an individual as belonging to a litter of one, two, three, or more litter mates, or as being in her first, second, third, or later lacta-

* Animal Genetics Section, C.S.I.R.O., University of Sydney.

tion, or as having had an interval of so many days between the first and second lactation. It is not always so easy to say that an individual has suffered from some infection, and difficult, if not impossible, to subdivide the infected individuals into classes of those which were more and those which were less severely infected. The number of possible trends which would have to be calculated to give a complete or nearly complete description of environmental effects is obviously large and would require much laborious computation to work out. But even when the computations have been made and where the environmental classes appear to be unequivocal there are difficulties in applying corrections which may not at first appear.

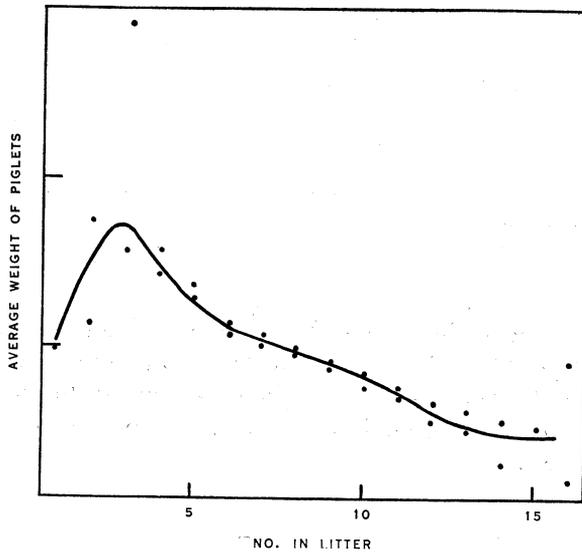


Fig. 1.—Effect of litter size on weight of pig at 3 wk (after Korkmann).

Let us take as an example the effect of number in the litter on 3-wk weight of piglets. It can be seen from Korkmann's (1947) data, Figure 1, that in general the larger the number in the litter the smaller the piglet. But as the number becomes very small a new trend makes its appearance and as the number decreases the weight of piglets also decreases. It can be shown for mice, where the trend of decrease in size with increase in number is slight (Table 1; Fig. 2), that as the number in the litter decreases, the variance increases by an amount which is greater than would be expected to result from the smaller sample size. It appears as though a new classification is required at this end of the curve. The old trend which affects litters of larger number continues to operate in so far as the biggest mice come from the smallest litters, but there is another sort of small litter in which the progeny are themselves much smaller than average. At this point a single correction on the basis of litter number is no longer applicable to all litters of a given number. Some litters are affected one way,

and some another. It appears as though there were at least two kinds of litter, those which are small in number but otherwise normal, and those which are small in number because of some factor which affects the growth of mice adversely. If the same correction is applied to both sorts of litter on the grounds

TABLE 1
VARIANCE OF MEAN LITTER WEIGHT IN MICE CLASSIFIED ACCORDING TO NUMBERS IN LITTER

Number in the Litter	2	3	4	5	6	7	8	9	10
Mean wt. (g)	12.4 (12.4)	13.8 (12.1)	14.3 (13.1)	12.8 (11.4)	14.1 (11.6)	12.9 (11.4)	12.7 (9.8)	12.3	13.9
Variance of mean	10.15 (11.54)	6.41 (7.84)	4.48 (6.12)	2.08 (3.10)	1.24 (4.31)	3.42 (3.73)	1.41 (2.45)	2.25	0.85
Number of litters	4 (6)	6 (16)	7 (11)	10 (22)	8 (15)	9 (9)	7 (4)	4	2

Variance of weights of individuals of same litter = 1.68.

Figures in brackets come from crosses between pure lines and outbred mice, the remaining figures are from mice in this experiment.

that they both belong to the class of litter which contains so many mice, some will be corrected insufficiently for a trend due to an environment to which they were subjected, and others will be corrected for a trend due to an environment to which they were not subjected, in a direction opposite to the one by which they were in fact affected.

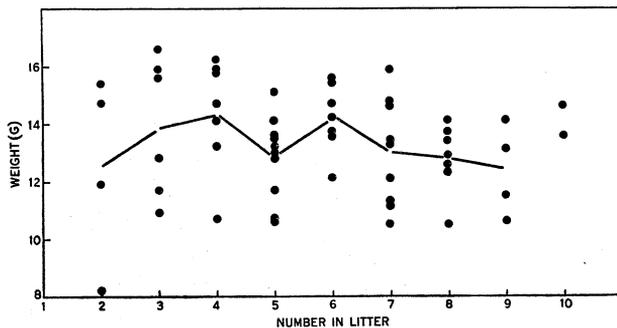


Fig. 2.—Effect of number in litter on weight of mice at 4 wk.

An index is required which will not only register the degree to which each individual has been exposed to a particular environment, but also integrate the effect of all possible environments into one figure. Hazel (1943) has suggested that the measure of one character can be used not only as an index of the genotype of the individual with respect to another character, but also of the environmental effects to which it has been exposed. Suppose the correlation between

two characters is entirely environmental, then deviation from the mean population score in one is a measure of the extent to which the second has been affected by the environment and the regression of one character on the other can be used to correct for all those environmentally produced deviations which give rise to the correlation between them. The extent to which environmental variance in one character is removed by its regression on a second will depend on the closeness of the correlation between them.

III. RELATIONSHIP BETWEEN SELECTION INDEX AND CORRECTION OF ONE CHARACTER FOR VARIATION IN ANOTHER

If two characters are measured on the same animal or plant the breeder must decide how best to use the information at his disposal when selecting parents for the next generation. In making his choice the weight he puts on the two scores will depend on the heritability of the characters and the economic value of an advance in them. The solution of the problem of how to weight the two traits to the best advantage was first put forward by Fairfield Smith (1937), using Fisher's discriminant functions, and later arrived at by Hazel (1943), using Sewall Wright's path coefficients. The problem and its solution are set out again here for convenience.

Suppose selection is being practised for two characters 1 and 2. The genotype of an individual may be written $G_1 + G_2$ and its phenotype $X_1 + X_2$. The economic value of an individual (H) depends on the money value a_1, a_2 of each character:

$$H = a_1G_1 + a_2G_2.$$

The best index on which to select, I , depends on the heritability of each character and its economic value, which must be taken into account when weighting X_1 and X_2 :

$$I = S_1X_1 + S_2X_2.$$

The best index will be one for which H and I are maximally correlated. The problem is to find the weighting factors S_1 and S_2 . The correlation of I with H can be written

$$R_{HI} = \frac{\text{cov } HI}{\sqrt{\text{var } H \times \text{var } I}} \dots \dots \dots (2)$$

The covariances and variances of H and I are simply expressed in terms of $a, S, G,$ and X . The assumption is made that there is no correlation between environment and genotype so that, for example, the expression in the covariance term $S_1a_1X_1G_1$ is equivalent to $S_1a_1 \text{ var } G_1$ and the term $a_1S_2G_1X_2$ is equivalent to $a_1S_2 \text{ cov } G_1G_2$. If equation (2) is written out in full and converted to logarithms R_{HI} can be differentiated with respect to S_1 and S_2 and by equating these derivatives to zero, values for S_1 and S_2 found which maximize the correlation R_{HI} . As S_1 and S_2 are weighting factors, S_1 can be taken as unity and S_2 expressed in terms of S_1 . S_2 then becomes

$$\frac{\text{cov } X_1 X_2 (a_1 \text{ var } G_1 + a_2 \text{ cov } G_1 G_2) - \text{var } X_1 (a_1 \text{ cov } G_1 G_2 + a_2 \text{ var } G_2)}{\text{cov } X_1 X_2 (a_1 \text{ cov } G_1 G_2 + a_2 \text{ var } G_2) - \text{var } X_2 (a_1 \text{ var } G_1 + a_2 \text{ cov } G_1 G_2)}$$

When all that is required of the second character, 2, is that it should indicate the environment to which 1 was exposed, it can be chosen so as to have no economic value itself, so that a_2 becomes zero and all terms containing a_2 disappear. If there is no genetic correlation between 1 and 2 all terms containing $\text{cov } G_1 G_2$ also disappear, leaving $-\text{cov } X_1 X_2 / \text{var } X_2$. Thus when the second character is of no economic importance and when there is no genetic correlation between 1 and 2, or if it is desired not to increase 2, the regression of 1 on 2 is the best correction to use to make the most of information available.

$$I = X_1 - b_{1.2} X_2,$$

where $b_{1.2}$ is the regression of X_1 on X_2 .

As it is usually difficult to get an accurate estimate of genetic correlation it is of interest to note what happens if the regression is used as a correction and there is some genetic as well as environmental correlation between the two characters. Reframing (1), we have before correction

$$\Delta G = \bar{i} h^2 \sigma_p.$$

Correction removes a portion of the phenotypic variance due to regression equal to $r_{p_1 p_2}^2 \sigma_{p_1}^2$, where $r_{p_1 p_2}$ is the correlation between the two phenotypes. Correction also removes $r_{G_1 G_2}^2 \sigma_{G_1}^2$ of the genetic variance of the first trait where $r_{G_1 G_2}$ is the genetic correlation between them. After correction we find

$$\begin{aligned} \Delta G \text{ corrected} &= \frac{\bar{i} \sigma_G^2 (1 - r_{G_1 G_2}^2) \sigma_p (\sqrt{1 - r_{p_1 p_2}^2})}{\sigma_p^2 (1 - r_{p_1 p_2}^2)} \\ &= \frac{\Delta G (1 - r_{G_1 G_2}^2)}{\sqrt{1 - r_{p_1 p_2}^2}} \end{aligned}$$

This is greater than ΔG so long as the genetic correlation is not large compared to the phenotypic one. Values of $r_{p_1 p_2}$ below which correction decreases ΔG as shown in Figure 5 for different values of $r_{G_1 G_2}$. It will be seen that in the absence of genetic correlation between two traits, and if one of them has no economic value, the selection index commonly used in plant and animal breeding leads to the conclusion that the best use of the second trait is as an indicator of the environment to which the first was subjected. That is, selection is made with the regression line as the base line and in this way h^2 is increased since σ_E^2 is reduced with respect to the character of economic value.

IV. EXPERIMENTAL MATERIAL

The measurements on which the method of correction discussed here has been tried out, came from a selection experiment using mice in which 15 males were mated each to four half sibs and the litters used to discriminate between the sires. The young mice were weighed at 4 wk, they were then killed and the tibia of one leg prepared after cooking in a pressure cooker for 5 min. The

bone was projected onto a screen and the shadow measured. The length of the shadow, given in centimetres, is used in this analysis. Magnification was about $11\frac{1}{4}$ times. A total of 327 mice from 57 litters were finally measured. Mean weight was 13.25 g, and tibia length (shadow) 15.21 cm, so that actual tibias were about 1.35 cm long.

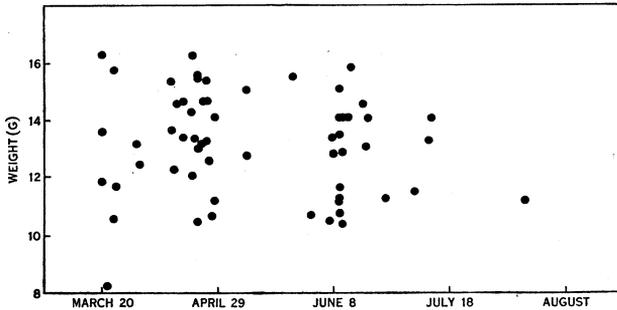


Fig. 3.—Effect of season on body weight of mice at 4 wk.

V. ANALYSIS

Two possible environmental trends were first investigated to see what could be done to decrease environmental effects by direct corrections. These are shown in Figures 2 and 3. The effect of litter number on individual weight (Fig. 2) is not very great and it is clear that the regression, which is about 0.12

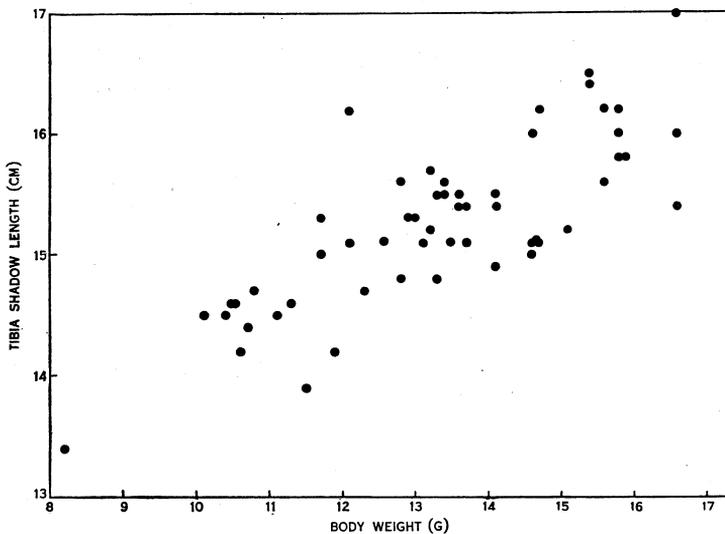


Fig. 4.—Correlation of tibia shadow length and body weight in mice at 4 wk.

g decrease in weight for each increase of one in litter number, is not going to remove much of the variance of weight. A seasonal trend of some sort seems to be indicated in Figure 3 but the data do not group themselves closely round

any line, and the variance removed by correcting for season would be very small indeed in this instance. The effects of these two environments on tibia length are similar and have not been shown.

TABLE 2
SUMS OF SQUARES AND CROSS PRODUCTS OF BODY WEIGHT AND TIBIA LENGTH AT 4 WK

Source	D.f.	Sums of Squares, Body Weight	Sum of Cross Products	Sums of Squares, Tibia Length
Between sires	14	287·9050	57·3800	23·8480
Between litters of the same sire	42	577·7697	184·0134	91·5735
Between individuals of the same litter	270	454·8000	74·2344	33·6846

The plot of tibia length against weight (Fig. 4) shows a much closer clustering round the regression line. The correlation is about 0·7 and the regression line will account for some 50 per cent. of the total variance. This is in marked contrast to the effect of season and litter number and, provided the correlation is largely an environmental one, it will remove a great deal more environmentally caused variance than the other two put together. It is clear that a considerable search through possible environmental effects would have to be made before as much variance could be accounted for as is removed by the regression line of weight on tibia length and vice versa. This search has not been made here because factors which might have influenced body weight and tibia length, such as temperature and disease, were not recorded in sufficient detail.

TABLE 3
REGRESSIONS OF BODY WEIGHT ON TIBIA LENGTH AND TIBIA LENGTH ON BODY WEIGHT

Source	Regression	
	Body Weight on Tibia Length	Tibia Length on Body Weight
Between sires	2·405	0·199
Between litters of the same sire	2·010	0·319
Between individuals of the same litter	2·204	0·163

The sums of squared deviations and cross products of tibia length and body weight are set out in Table 2. The regressions derived from them are set out in Table 3. The regression of body weight on tibia length is the same, no matter what the source of variation. The change in body weight accom-

panying a unit change in tibia length is the same whether the comparison is between individuals of a litter, litters of a sire, or different sire progenies. This indicates either that the causes of the differences in tibia length are all of the same kind (e.g. environmental), no matter what the source of variance, or that changes in tibia length, whether their cause is genetic or environmental, are always accompanied by the same change in weight. The analysis of the components of variance in Table 4 shows that the former is the true explanation. There is no detectable genetic component of variance of tibia length so that an overwhelming portion of the variance will be environmental no matter what the source.

TABLE 4
ANALYSIS OF VARIANCE OF BODY WEIGHT AND TIBIA LENGTH
BODY WEIGHT

Source	Uncorrected			Corrected		
	D.f.	Mean Sq.	Component	D.f.	Mean Sq.	Component
Between sires	14	20.5646	0.3137	13	11.5343	0.2977
Between litters of same sire	42	13.7564	2.1094	41	5.0732	0.6897
Between individuals of same litter	270	1.6844	1.6844	269	1.1260	1.1260
Heritability			0.3055	0.5634		

TIBIA LENGTH

Source	Uncorrected			Corrected		
	D.f.	Mean Sq.	Component	D.f.	Mean Sq.	Component
Between sires	14	1.7034	—	13	0.9548	0.0069
Between litters of the same sire	42	2.1803	0.3592	41	0.8041	0.1265
Between individuals of same litter	270	0.1248	0.1248	269	0.0802	0.0802
Heritability				0.1292		

The regression of tibia length on weight is different according to the source of variation. The steepest regression is of tibia length on weight of litter. Now the genetic variance will be divided so that one-half is between individuals of a litter, one-quarter between sires, and one-quarter between dams. In this experiment between dams is the same as between litters. As the genetic variance between dam and sire should be equal, obviously the differences between

litters are mostly environmental and the environmental component is much larger between dams than it is between sires or between individuals of a litter.

Where the genetic component is increased in importance relative to the non-genetic component, as it is between sires and between individuals of a litter, the regression is less, suggesting that if there is a genetic correlation between the two characters it is considerably smaller than the environmental one. The component of covariance due to differences between sires is -0.0003 which, despite the absence of a component of variance in tibia length due to sire, suggests that there is no genetic correlation.

The analysis of variance shows that heritability for weight is about 0.3, assuming that all the sire component is genetic and that it includes one-quarter of the total genetic variance. There is no component from which heritability of tibia length can be measured.

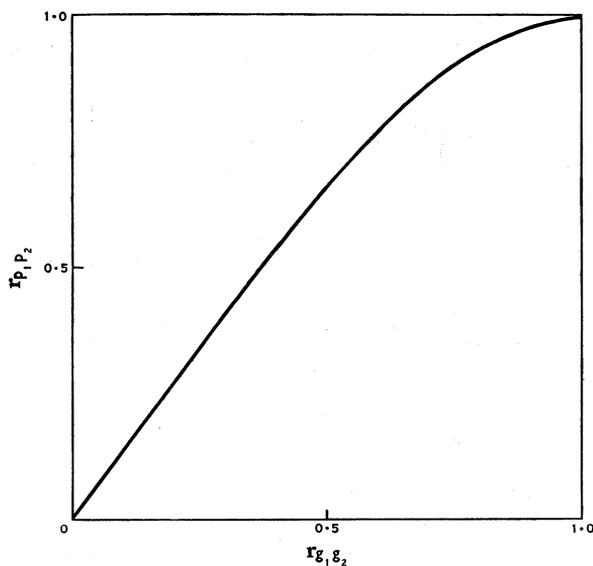


Fig. 5.—Values of r_p for different values of r_G at which ΔG is unchanged when the correction $I = X_1 - bX_2$ is applied.

After the figures were corrected for the regression of weight on tibia length and tibia length on weight the analysis was repeated. It is shown in Table 4 under the heading "corrected." In every case the regression calculated for the class corrected was used. Correction reduces the variance of both characters by about one-half, the greatest reduction being of the variance between litters. This is particularly noticeable of tibia length. The component of variance due to differences between individuals of a litter is slightly lower after correction than before, the litter component is reduced to a third and the sire component practically unchanged for weight, but now just positive for tibia length. Heritability is increased to 0.5634 for weight and about 0.1 (still a very unreliable figure) for tibia length.

The mean scores of the progeny of the 15 sires have been computed on the corrected and uncorrected basis to show what difference in practice correction would make, supposing the best five sires were selected in each case. It will be seen that the order of merit is quite considerably altered in both characters (Table 5).

TABLE 5
MEAN WEIGHT OF SIRE PROGENIES ARRANGED IN ORDER OF MERIT,
CORRECTED AND UNCORRECTED

Weight			Tibia Length			
Uncorrected		Corrected	Uncorrected		Corrected	
(a)	15.15	(d) 14.33 (4)	(b)	15.84	(p)	15.60 (5)
(b)	14.56	(c) 14.02 (3)	(a)	15.80	(b)	15.58 (1)
(c)	14.24	(f) 13.93 (6)	(g)	15.47	(g)	15.48 (3)
(d)	14.23	(j) 13.75 (9)	(e)	15.43	(a)	15.42 (2)
(e)	14.02	(a) 13.73 (1)	(p)	15.33	(o)	15.34 (9)
(f)	13.79	(e) 13.49 (5)	(c)	15.30	(e)	15.28 (4)
(g)	13.19	(h) 13.36 (8)	(d)	15.17	(m)	15.26 (12)
(h)	13.02	(b) 13.04 (2)	(f)	15.15	(l)	15.21 (10)
(j)	12.98	(k) 13.04 (10)	(o)	15.11	(k)	15.19 (11)
(k)	12.75	(l) 12.94 (11)	(l)	15.10	(h)	15.12 (13)
(l)	12.68	(n) 12.92 (13)	(k)	15.09	(j)	15.11 (14)
(m)	12.40	(m) 12.69 (12)	(m)	15.09	(c)	15.10 (6)
(n)	12.15	(g) 12.56 (7)	(h)	15.07	(f)	15.04 (8)
(o)	12.09	(o) 12.33 (14)	(j)	14.89	(d)	14.97 (7)
(p)	11.90	(p) 11.61 (15)	(n)	14.89	(n)	14.94 (15)

Letters in brackets identify the sires.

Figures in brackets indicate the ranking of the sire before correcting.

VI. DISCUSSION

The foregoing exercise shows that where a suitable character is available it is far simpler and more effective to discount environmental effects by correcting one character for its regression on another than by searching out the separate environmental trends. A suitable character is one which has no economic value, is strongly correlated phenotypically with the desired character, but has little or no genetic correlation with it. It is worth noting that even if there is a genetic correlation there are two general situations in which the correction is still worth applying; first, when the phenotypic correlation is sufficiently greater than the genetic one, and second, where the second character is one which it is desired to keep reasonably constant. For example, in Denmark, selection for milk yield and fat yield has been accompanied by an increase in body size, to a point where some farmers are faced with rebuilding their cow sheds, or changing their breed. It is believed that the Danish Red by Jersey cross has gained in popularity for this reason. Under such circumstances one might be willing to sacrifice increase in yield which accompanies increase in size to save capital expenditure. But where a genetic corre-

lation is known to exist and has been measured, the full index should be used to determine S_2 .

There is one possible danger in using this method of correction as a basis of selection, which is that under normal methods animals sensitive to bad environments will be selected against, but after correcting for environmental trends, although there will probably be some selection against sensitivity to the environment, it will not be nearly so strong.

VII. ACKNOWLEDGMENTS

I should like to thank Miss H. N. Turner and Dr. F. H. W. Morley for their helpful suggestions during the preparation of this paper, and Miss Turner's unit of the Mathematical Statistics Section, C.S.I.R.O., for assistance with the computations.

VIII. REFERENCES

- FAIRFIELD SMITH, H. (1937).—A discriminant function for plant selection. *Ann. Eugen.*, *Lond.* **7**: 240-50.
- HAZEL, L. N. (1943).—The genetic basis of constructing selection indexes. *Genetics* **28**: 476-90.
- KORKMANN, N. (1947).—Causes of variation in the size and weight of litters from sows. *Acta Agric. Suec.* (2) **3**: 253-310.