CORRELATION AND CAUSATION

By Sewall Wright

Senior Animal Husbandman in Animal Genetics, Bureau of Animal Industry, United States Department of Agriculture

PART I. METHOD OF PATH COEFFICIENTS

INTRODUCTION

The ideal method of science is the study of the direct influence of one condition on another in experiments in which all other possible causes of variation are eliminated. Unfortunately, causes of variation often seem to be beyond control. In the biological sciences, especially, one often has to deal with a group of characteristics or conditions which are correlated because of a complex of interacting, uncontrollable, and often obscure causes. The degree of correlation between two variables can be calculated by well-known methods, but when it is found it gives merely the resultant of all connecting paths of influence.

The present paper is an attempt to present a method of measuring the direct influence along each separate path in such a system and thus of finding the degree to which variation of a given effect is determined by each particular cause. The method depends on the combination of knowledge of the degrees of correlation among the variables in a system with such knowledge as may be possessed of the causal relations. In cases in which the causal relations are uncertain the method can be used to find the logical consequences of any particular hypothesis in regard to them.

CORRELATION

Relations between variables which can be measured quantitatively are usually expressed in terms of Galton's (4)[1] coefficient of correlation, $r_{XY} = \dfrac{\Sigma X'Y'}{n\sigma_X\sigma_Y}$ (the ratio of the average product of deviations of X and Y to the product of their standard deviations), or of Pearson's (7) correlation ratio, $\eta_{X \cdot Y} = \dfrac{\sigma\left(\frac{M}{Y\,X}\right)}{\sigma_X}$ (the ratio of the standard deviation of the mean values of X for each value of Y to the total standard deviation of X), the standard deviation being the square root of the mean square deviation.

Use of the coefficient of correlation (r) assumes that there is a linear relation between the two variables—that is, that a given change in one variable always involves a certain constant change in the corresponding average value of the other. The value of the coefficient can never exceed

---

[1] Reference is made by number (italic) to "Literature cited," p. 585.

$+1$ or -1. For many purposes it is enough to look on it as giving an arbitrary scale between $+1$ for perfect positive correlation, o for no correlation, and -1 for perfect negative correlation.

The correlation ratio (η) equals the coefficient of correlation if the relation between the variables is exactly linear. It does not, however, depend on the assumption of such a relation, and it is always larger than r when the relations are not exactly linear. It can only take values between o and $+1$, and it can be looked upon as giving an arbitrary scale between o for no correlation and 1 for perfect correlation.

The numerical value of the coefficient of correlation (r) takes on added significance in connection with the idea of regression. It gives the average deviation of either variable from its mean value corresponding to a given deviation of the other variable, provided that the standard deviation is the unit of measurement in both cases. The regression in terms of the actual units can, of course, be obtained by multiplying by the ratio of the standard deviations. Thus, for the deviation of X corresponding to a unit deviation of Y, we have $reg_{X \cdot Y} = r_{XY}\dfrac{\sigma_X}{\sigma_Y}$. This formula may be deduced from the theory of least squares as the best linear expression for X in terms of Y. The formula for what Galton later called the coefficient of correlation was, in fact, first presented in this connection by Bravais (*1*) in 1846. Any such interpretation is of course impossible with the correlation ratio.

The numerical values of both coefficients, however, have significance in another way. Their squares (η^2, or r^2 if regression is linear) measure the portion of the variability of one of the variables which is determined by the other and which disappears in data in which the second is constant. Thus if $_Y\sigma^2_X$ is the mean square deviation of X for constant Y, Pearson has shown that:

$$_Y\sigma^2_X = \sigma^2_X(1 - \eta^2_{X \cdot Y})$$

or $_Y\sigma^2_X = \sigma^2_X(1 - r^2_{XY})$ if regression is linear.

It often happens that it is desirable to consider simultaneously the relations in a system of more than two variables. For such cases, involving only linear relations between the various pairs of variables, Pearson (*6*) has devised the coefficient of multiple correlation.

$$R_{X(ABC \cdots N)} = \sqrt{1 - \frac{\Delta}{\Delta_{XX}}}$$

in which

$$\Delta = \begin{vmatrix} 1 & r_{XA} & r_{XB} & \cdot & \cdot & r_{XN} \\ r_{AX} & 1 & r_{AB} & \cdot & \cdot & r_{AN} \\ r_{BX} & r_{BA} & 1 & \cdot & \cdot & r_{BN} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ r_{NX} & r_{NA} & r_{NB} & \cdot & \cdot & 1 \end{vmatrix}$$

and Δ_{xx} is the minor made by deleting row X and column X. $R^2_{x(ABC\cdots N)}$ measures the degree of determination of X by the whole set of other factors, and $1 - R^2_{x(ABC\cdots N)} = \dfrac{\Delta}{\Delta_{xx}}$ is the maximum possible squared correlation between X and a factor independent of those considered. This formula for multiple correlation leads to one for multiple regression. Letting X', A', B', etc., be the deviations of variables X, A, B, etc., from their mean values, Pearson has shown that the most probable value of X' for known values of the other variables is given by the formula

$$\frac{X'}{\sigma_X} = \frac{\Delta_{XA}}{\Delta_{XX}}\frac{A'}{\sigma_A} + \frac{\Delta_{XB}}{\Delta_{XX}}\frac{B'}{\sigma_B} \quad . \quad . \quad . \quad . \quad \frac{\Delta_{XN}}{\Delta_{XX}}\frac{N'}{\sigma_N}$$

$$\sigma_{X'} = {}_{N\cdots BA}\sigma_X = \sigma_X\sqrt{\frac{\Delta}{\Delta_{XX}}}.$$

Analogous but more complex formulae have recently been published by Isserlis (5) for the multiple correlation ratio for use in cases in which the regressions are not necessarily linear.

CAUSATION

In all the preceding results no account is taken of the nature of the relationship between the variables. The calculations thus neglect a very important part of the knowledge which we often possess. There are usually *a priori* or experimental grounds for believing that certain factors are direct causes of variation in others or that other pairs are related as effects of a common cause. In many cases, again, there is an obvious mathematical relationship between variables, as between a sum and its components or between a product and its factors. A correlation between the length and volume of a body is an example of this kind. Just because it involves no assumptions in regard to the nature of the relationship, a coefficient of correlation may be looked upon as a fact pertaining to the description of a particular population only to be questioned on the grounds of inaccuracy in computation. But it would often be desirable to use a method of analysis by which the knowledge that we have in regard to causal relations may be combined with the knowledge of the degree of relationship furnished by the coefficients of correlation.

The problem can best be presented by using a concrete example. In a population of guinea pigs it will be found that the birth weights, early gains, sizes of litters, and gestation periods are all more or less closely correlated with each other. The influence of heredity, environmental conditions, health of dam, etc., are also easily shown. In a rough way, at least, it is easy to see why these variables are correlated with each other. These relations can be represented conveniently in a diagram like that in figure 1, in which the paths of influence are shown by arrows.

The variety and complexity of the relations which may be back of a correlation are well illustrated in this case. Thus, the weight at weaning (33 days of age) should be correlated with the birth weight and with the gain between birth and weaning simply because it is their sum. The relations of birth weight with gestation period and the prenatal rate of growth are also essentially mathematical rather than causal. Birth weight is necessarily fully determined by the character of the prenatal growth curve and the time at which this is interrupted by birth.

In the relation between gestation period and size of litter we come to a case in which there is no necessary mathematical relationship. We naturally attempt to account for the high negative correlation by the hypothesis that a large number in a litter in some way causes early
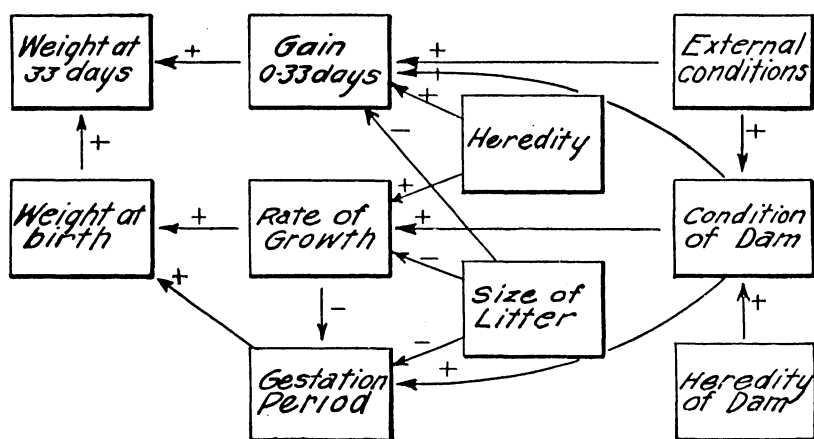

FIG. 1.—Diagram illustrating the interrelations among the factors which determine the weight of guinea pigs at birth and at weaning (33 days).

parturition. Similarly, a large number in a litter might be expected to be a cause of slow growth in the foetuses.

Birth weight and gain after birth are highly correlated. Here neither variable can be spoken of as the cause of variation in the other, and the relation is not mathematical. They are evidently influenced by common causes, among which heredity, size of litter, and conditions which affect the health of the dam up to the time of birth at once come to mind.

Most of the variables are connected with each other through more than one path. Thus, weight at birth is correlated with weight at weaning both as a component of a sum and as the effect of common causes.

There may be a conflict of the paths. Thus, a large number in a litter has a fairly direct tendency to shorten the gestation period, but this is probably balanced in part by its tendency to reduce the rate of growth of the foetuses, slow growth permitting a longer gestation period. Large litters tend to reduce gestation period and rate of growth before and after birth. But large litters are themselves most apt to come when

external conditions are favorable, which also favors long gestation periods and vigorous growth.

The coefficient of correlation is a resultant of all paths connecting the two variables. It would be valuable in many cases to be able to determine the relative importance of each particular path. The usual method in such cases is to calculate the partial correlation between two variables for a third constant, using Pearson's well-known formula

$$c^{r}_{AB} = \frac{r_{AB} - r_{AC}r_{BC}}{\sqrt{(1 - r^2_{AC})(1 - r^2_{BC})}}$$

for correlation between A and B for constant C. Such partial correlations, however, must be interpreted with caution. It is true that by making constant a connecting link between two variables, whether it is a common cause or the cause of one and effect of the other, we eliminate the path in question. This elimination of connecting paths in which the constant factor is a link is not, however, the only way in which correlation is affected. If an effect of a number of causes is made constant, spurious negative correlations appear among the causes and their other effects. Thus, if weight at 33 days is made constant, the correlation between birth weight and gain necessarily becomes -1. We are simply picking out a population in which any deficiencies in birth weight happen to be exactly balanced by excess in gain after birth. This is an extreme case, but where the relations of cause and effect are at all complex it is evident that the correlation between two variables may be changed in more than one way by making a third variable constant, making the interpretation doubtful.

Where there is a network of causes and effects, the interrelations could be grasped best if a coefficient could be assigned to each path in the diagram designed to measure the direct influence along it. The following is an attempt to provide such a coefficient, which may be called a path coefficient.

<center>DEFINITIONS</center>

We will start with the assumption that the direct influence along a given path can be measured by the standard deviation remaining in the effect after all other possible paths of influence are eliminated, while variation of the causes back of the given path is kept as great as ever, regardless of their relations to the other variables which have been made constant. Let X be the dependent variable or effect and A the independent variable or cause. The expression $\sigma_{X \cdot A}$ will be used for the standard deviation of X, which is found under the foregoing conditions, and may be read as the standard deviation of X due to A. In a system in which variation of X is completely determined by A, B, and C we have $\sigma_{X \cdot A} = \sigma_{ACB}\sigma_X$ representing the constant factors, B and C, and also the variation of A itself (σ_A) by subscripts to the left. The path

coefficient for the path from A to X will be defined as the ratio of the standard deviation of X due to A to the total standard deviation of X.

$$p_{\mathrm{x \cdot a}} = \frac{\sigma_{\mathrm{x \cdot a}}}{\sigma_{\mathrm{x}}}.$$

Just as the regression of X on A is expressed by $r_{\mathrm{xy}} \frac{\sigma_{\mathrm{x}}}{\sigma_{\mathrm{a}}}$ the deviation of X directly caused by a unit deviation of A is given by the formula

$$p_{\mathrm{x \cdot a}} \frac{\sigma_{\mathrm{x}}}{\sigma_{\mathrm{a}}} = \frac{\sigma_{\mathrm{x \cdot a}}}{\sigma_{\mathrm{a}}}.$$

Another coefficient which it will be convenient to use, the coefficient of determination of X by A, $d_{\mathrm{x \cdot a}}$, measures the fraction of complete determination for which factor A is directly responsible in the given system of factors. This definition implies that the sum of such coefficients must equal unity if all causes are accounted for.

SYSTEMS OF INDEPENDENT CAUSES

The degree of determination of one variable by another is most easily found where the variables are connected by a mathematical relationship. The simplest mathematical relationship is that between a sum and its components. For the standard deviation of a sum the following relation is well known:

$$\sigma^2_{\mathrm{A+B}} = \frac{\Sigma (A' + B')^2}{n} = \sigma^2_{\mathrm{A}} + \sigma^2_{\mathrm{B}} + 2\sigma_{\mathrm{A}}\sigma_{\mathrm{B}}r_{\mathrm{AB}}.$$

If A and B are independent of each other, $r_{\mathrm{AB}} = 0$, and we have

$$\sigma^2_{\mathrm{A+B}} = \sigma^2_{\mathrm{A}} + \sigma^2_{\mathrm{B}}.$$

The degree to which variation of the sum is determined by that of each component is obvious.

$$d_{\mathrm{x \cdot a}} = \frac{\sigma^2_{\mathrm{A}}}{\sigma^2_{\mathrm{x}}} \text{ and } d_{\mathrm{x \cdot b}} = \frac{\sigma^2_{\mathrm{B}}}{\sigma^2_{\mathrm{x}}}, \text{ where } X = A + B,$$

giving $d_{\mathrm{x \cdot a}} + d_{\mathrm{x \cdot b}} = 1$, as required by definition.
For the standard deviation of X due to A we have in this case, $\sigma_{\mathrm{x \cdot a}} = \sigma_{\mathrm{a}}$.
Thus, $p_{\mathrm{x \cdot a}} = \frac{\sigma_{\mathrm{x \cdot a}}}{\sigma_{\mathrm{x}}} = \frac{\sigma_{\mathrm{a}}}{\sigma_{\mathrm{x}}}$ by definition.
Again, $r_{\mathrm{xa}} = \frac{\Sigma (A' + B')A'}{n\sigma_{\mathrm{x}}\sigma_{\mathrm{a}}} = \frac{\Sigma A'^2}{n\sigma_{\mathrm{x}}\sigma_{\mathrm{a}}} = \frac{\sigma_{\mathrm{a}}}{\sigma_{\mathrm{x}}}.$
Summing up, $p_{\mathrm{x \cdot a}} = \sqrt{d_{\mathrm{x \cdot a}}} = r_{\mathrm{xa}}$.

It can easily be shown that the same formulae hold in case we are dealing with the sum of multiples of a number of independent factors instead of with their own sum.

We can pass at once from this case to cases in which variation of X is caused in the physical or physiological sense by variation in several causes

provided that these causes are independent of each other, have linear relations to the dependent variable X, and that the deviations which they determine are additive. They are independent of each other if there is no correlation between their variations. A cause has a linear relation to the effect and is combined additively with the other factors if a given amount of change in it always determines the same change in the effect, regardless of its own absolute value or that of the other causes. The conclusion is that, under these conditions, the path coefficient equals the coefficient of correlation between cause and effect, and the degree of determination equals the square of either of the preceding coefficients.

CHAINS OF CAUSES

If we know the extent to which a variable X is determined by a certain cause M, which is independent of other causes, combines with them additively, and acts on X in a linear manner, and if we know the extent to which M is determined by a more remote cause A, the degree of determination of X by A must be the product of the component degrees of determination.

Let $X = M + N$, and $M = A + B$

$$d_{\text{X·M}} = \frac{\sigma^2_\text{M}}{\sigma^2_\text{X}}, \; d_{\text{M·A}} = \frac{\sigma^2_\text{A}}{\sigma^2_\text{M}}, \text{ and } d_{\text{X·A}} = \frac{\sigma^2_\text{A}}{\sigma^2_\text{X}}.$$

Thus $d_{\text{X·A}} = d_{\text{X·M}} d_{\text{M·A}}$
and $p_{\text{X·A}} = p_{\text{X·M}} p_{\text{M·A}}$.

NONADDITIVE FACTORS

In cases in which a factor does not act additively with the other factors in determining the variations in the dependent variable, its influence on the latter can not be completely expressed apart from the other factors, at least in terms of the ordinary measures of variability. This can be made clearer by an illustration. Multiplying factors are among the most important of those which do not combine by addition.

Let $X = AB$ and assume that $r_{\text{AB}} = o$

$$\sigma^2_\text{X} = M^2_\text{B}\sigma^2_\text{A} + M^2_\text{A}\sigma^2_\text{B} + \frac{\Sigma A'^2 B'^2}{n}$$

where A' and B' are deviations of A and B from their mean values M_A and M_B. Putting B constant, we have $\sigma^2_{\text{X·A}} = M^2_\text{B}\sigma^2_\text{A}$; and similarly putting A constant, we have $\sigma^2_{\text{X·B}} = M^2_\text{A}\sigma^2_\text{B}$. There remains a portion of σ^2_X which is due to A and B jointly and which can not be separated into parts due to each alone. If we write $d_{\text{X·A}} = \dfrac{M^2_\text{B}\sigma^2_\text{A}}{\sigma^2_\text{X}}$ as the degree of determination of X by variation of A alone, and $d_{\text{X·B}} = \dfrac{M^2_\text{A}\sigma^2_\text{B}}{\sigma^2_\text{X}}$ as the corresponding degree of determination of X by variation of B alone, we must recognize an additional term $d_{\text{X·}\overline{\text{AB}}} = \dfrac{\Sigma A'^2 B'^2}{n\sigma^2_\text{X}}$, in order that the sum of the

coefficients of determination may equal unity. Regression is linear and $r^2_{XA} = \eta^2_{X \cdot A} = \dfrac{M^2_{B} \sigma^2_{A}}{\sigma^2_{x}}$. Thus $d_{X \cdot A} = r^2_{XA}$ as in the case of independent additive factors. The term $\dfrac{\Sigma A'^2 B'^2}{n \sigma^2_{x}}$ is small unless the amounts of variation in A and B are large in comparison with the mean values. In many cases it is safe to deal with path coefficients and degrees of determination in the case of multiplying factors just as in the case of additive factors.

As a concrete illustration of these points take two independent variables, for each of which the values 1, 2, and 3 occur in the frequencies 1, 2, and 1, respectively. Below is the correlation table between one of these factors and their product.

Product (X).

		1	2	3	4	5	6	7	8	9	
	1......	1	2	1	....	....	....	....	....	....	4
Factor (A).	2......	....	2	....	4	....	2	....	....	....	8
	3......	....	....	1	....	....	2	....	....	1	4
		1	4	2	4	0	4	0	0	1	16

$M_A = 2 \quad \sigma_A = \sqrt{1/2} \quad r_{AX} = \sqrt{8/17} \quad d_{X \cdot A} = 8/17$

$M_X = 4 \quad \sigma_X = \sqrt{17/4} \quad \dfrac{\Sigma A'^2 B'^2}{n \sigma^2_x} = 1/17 \quad d_{X \cdot B} = 8/17$

$d_{X \cdot \overline{AB}} = \dfrac{1/17}{1}$

In this case the amounts of variation in the factors are relatively large compared with their mean values, making the distribution surface markedly heteroscedastic, yet the degree of determination by either factor comes out only slightly less than one-half.

NONLINEAR RELATIONS

Pearson's definition of the correlation ratio, $\eta_{X \cdot A} = \dfrac{\sigma(_A M_x)}{\sigma_X}$, has already been given. The variations of the mean value of X for different values of A are the variations which can be attributed to the direct influence of A, assuming that A is cause, X effect, and that other causes are combined with A additively. Thus $\sigma_{X \cdot A} = \sigma(_A M_x)$ and we have at once $p_{X \cdot A} = \eta_{X \cdot A}$.

Again, as the total variation of X is composed of the variation of its mean values for different values of A, plus the variation about these mean values, we have $\sigma^2_X = \sigma^2(_A M_x) + _A \sigma^2_x$, giving $_A \sigma^2_x = \sigma^2_X (1 - \eta^2_{X \cdot A})$, as already noted.

Thus $\eta^2_{X \cdot A}$ measures the portion of σ^2_x lost by making A constant, so that as before $d_{X \cdot A} = \eta^2_{X \cdot A} = p^2_{X \cdot A}$.

Unfortunately we can not deal with chains of factors which involve nonlinear relations by mere multiplication of the path coefficients of the component links.   In the present paper, unless otherwise stated, it will be assumed that all correlations are essentially linear.

EFFECTS OF COMMON CAUSES

Suppose that two variables, X and Y, are affected by a number of causes in common, (B, C, D). Let A represent causes affecting X alone and E causes affecting Y alone (fig. 2).

Let $p_{X \cdot A} = a$ $p_{Y \cdot A} = o$
        $p_{X \cdot B} = b$ $p_{Y \cdot B} = b'$
        $p_{X \cdot C} = c$ $p_{Y \cdot C} = c'$
        $p_{X \cdot D} = d$ $p_{Y \cdot D} = d'$
        $p_{X \cdot E} = o$ $p_{Y \cdot E} = e'$
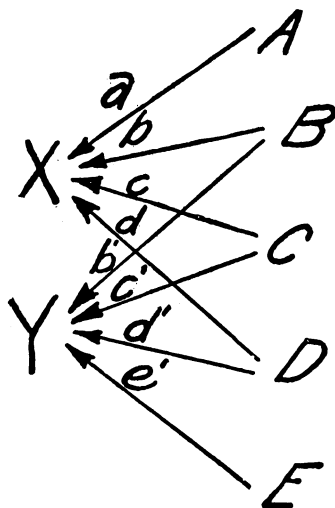
B, C, and D are assumed to be independent of each other—that is, $r_{BC} = o$, etc.

Hence $p_{X \cdot B} = r_{XB}$, etc.

FIG. 2.—Diagram showing relations between two variables, X and Y, whose values are determined in part by common causes, B, C, and D, which are independent of each other.

$$_{B}r_{XY} = \frac{r_{XY} - bb'}{\sqrt{(1 - b^2)(1 - b'^2)}}$$

$$_{CB}r_{XY} = \frac{_{B}r_{XY} - _{B}r_{XC}\,_{B}r_{YC}}{\sqrt{(1 - _{B}r^2_{XC})(1 - _{B}r^2_{YC})}} = \frac{r_{XY} - bb' - cc'}{\sqrt{(1 - b^2 - c^2)(1 - b'^2 - c'^2)}}.$$

When all common causes have been made constant, $_{DCB}r_{XY} = o$
$$r_{XY} = bb' + cc' + dd' = \Sigma p_{X \cdot B}\, p_{Y \cdot B}.$$

Thus, in those cases in which the causes are independent of each other, the correlation between two variables equals the sum of the products of the pairs of path coefficients which connect the two variables with each common cause.   An illustration of the use of this principle was given in an earlier paper (8) in analyzing the nature of size factors in rabbits.

It may be deduced from the foregoing formula that two variables may even be completely determined by the same factors and yet be uncorrelated with each other. Let variation of X be completely determined by factors B and C, the path coefficients being b and c, respectively. Let Y be completely determined by the same factors, the path coefficients being b' and c' (fig. 3). Then $r_{XY} = bb' + cc'$. The condition
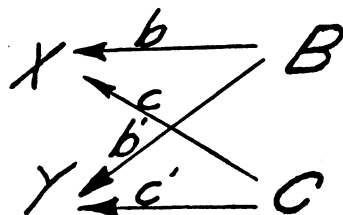

FIG. 3.—Diagram showing relations between two variables, X and Y, whose values are completely determined by common causes, B and C, which are independent of each other.

under which r_{XY} may equal zero is evidently that $bb' = -cc'$. An example may be found in the absence of correlation between the sum and difference of pairs of numbers picked at random from a table.

In many cases a small actual correlation between variables will be found on analysis to be the resultant of a balancing of very much more important but opposed paths of influence leading from common causes.

<div style="text-align:center">SYSTEMS OF CORRELATED CAUSES</div>

The discussion up to this point has dealt wholly with causes which act independently of each other. It is necessary to consider the effects of correlation among the causes.

Let us consider the sum of two correlated variables (fig. 4).

$$\text{Let } X = M + N$$
$$\sigma^2_X = \sigma^2_M + \sigma^2_N + 2\sigma_M\sigma_N r_{MN}.$$

We have defined $\sigma_{X \cdot M}$ as the standard deviation of X when factors other than M are constant, but M varies as much as before. The latter qualification is important in the present case, since the making of N constant tends to reduce the variation of M, reducing σ_M to $\sigma_M\sqrt{1 - r^2_{MN}}$.
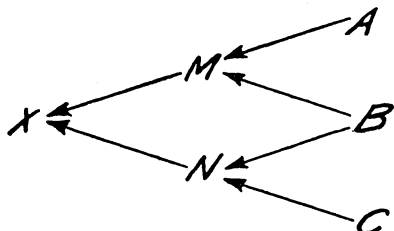
The definition of $\sigma_{X \cdot M}$ implies that not only is N made constant but that there is such a readjustment among the more remote causes, A, B, and C, that σ_M is unchanged. Under the definition it is evident that in this case $\sigma_{X \cdot M} = \sigma_M$ and $\sigma_{X \cdot N} = \sigma_N$. Thus $p_{X \cdot M} = \frac{\sigma_M}{\sigma_X}$ and $p_{X \cdot N} = \frac{\sigma_N}{\sigma_X}$.

In attempting to find the degrees of determination of X by M and N we meet a difficulty somewhat similar to that met in the case of non-additive factors. The squared standard deviation is made up in part of elements due wholly to M and N, respectively, but in part to a portion which can not be divided between them. The term $2\sigma_M\sigma_N r_{MN}$ is due solely to the fact that the variations of X, which M and N determine, tend to be in the same direction and so have greater effect than if variations M and N were combined at random. It seems best to define $d_{X \cdot M}$ as the degree of determination of X due to M alone. Thus $d_{X \cdot M} = \frac{\sigma^2_M}{\sigma^2_X}$, $d_{X \cdot N} = \frac{\sigma^2_N}{\sigma^2_X}$. The remaining term may be considered as determination by M and N jointly and may be written $d_{X \cdot \overline{MN}} = 2p_{X \cdot M}p_{X \cdot N}r_{MN}$.

These rules can be extended at once to the sums of more than two variables, to sums of multiples of variables, and hence, as before, to

linear relations of cause and effect in which the influence of the causes is combined additively. It is also easy to show that the formulae apply approximately for multiplying factors.

Summing up, $p_{\mathrm{X \cdot M}} = \sqrt{d_{\mathrm{X \cdot M}}} = \dfrac{\sigma_{\mathrm{X \cdot M}}}{\sigma_{\mathrm{X}}}$

$$\Sigma d_{\mathrm{X \cdot M}} + 2 \Sigma p_{\mathrm{X \cdot M}} p_{\mathrm{X \cdot N}} r_{\mathrm{MN}} = 1.$$

The next problem is to find the degree of determination of X by a factor such as B, which is connected with X by more than one path (fig. 5).

Assume that A, B, C, and D are independent and completely determine X. $d_{\mathrm{X \cdot A}} + d_{\mathrm{X \cdot B}} + d_{\mathrm{X \cdot C}} + d_{\mathrm{X \cdot D}} = 1$. But also $d_{\mathrm{X \cdot M}} + d_{\mathrm{X \cdot N}} + 2 p_{\mathrm{X \cdot M}} p_{\mathrm{X \cdot N}} r_{\mathrm{MN}} + d_{\mathrm{X \cdot D}} = 1$.

$d_{\mathrm{X \cdot B}} = d_{\mathrm{X \cdot M}} - d_{\mathrm{X \cdot A}} + d_{\mathrm{X \cdot N}} - d_{\mathrm{X \cdot C}} + 2 p_{\mathrm{X \cdot M}} p_{\mathrm{X \cdot N}} p_{\mathrm{M \cdot B}} p_{\mathrm{N \cdot B}}$, remembering that $r_{\mathrm{MN}} = p_{\mathrm{M \cdot B}} p_{\mathrm{N \cdot B}}$.
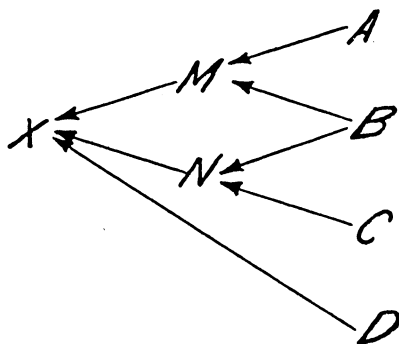

FIG. 5.—A system in which the value of X is affected by a factor, B, along two different paths, BMX and BNX.

Since $d_{\mathrm{M \cdot A}} + d_{\mathrm{M \cdot B}} = 1$, etc., we have $d_{\mathrm{X \cdot M}} = d_{\mathrm{X \cdot M}} d_{\mathrm{M \cdot A}} + d_{\mathrm{X \cdot M}} d_{\mathrm{M \cdot B}} = d_{\mathrm{X \cdot A}} + d_{\mathrm{X \cdot M}} d_{\mathrm{M \cdot B}}$, and $d_{\mathrm{X \cdot N}} = d_{\mathrm{X \cdot C}} + d_{\mathrm{X \cdot N}} d_{\mathrm{N \cdot B}}$.

Therefore $d_{\mathrm{X \cdot B}} = d_{\mathrm{X \cdot M}} d_{\mathrm{M \cdot B}} + d_{\mathrm{X \cdot N}} d_{\mathrm{N \cdot B}} + 2 p_{\mathrm{X \cdot M}} p_{\mathrm{X \cdot N}} p_{\mathrm{M \cdot B}} p_{\mathrm{N \cdot B}}$

$$= p^2_{\mathrm{X \cdot M}} p^2_{\mathrm{M \cdot B}} + p^2_{\mathrm{X \cdot N}} p^2_{\mathrm{N \cdot B}} + 2 p_{\mathrm{X \cdot M}} p_{\mathrm{X \cdot N}} p_{\mathrm{M \cdot B}} p_{\mathrm{N \cdot B}}$$

$$= (p_{\mathrm{X \cdot M}} p_{\mathrm{M \cdot B}} + p_{\mathrm{X \cdot N}} p_{\mathrm{N \cdot B}})^2$$

$$p_{\mathrm{X \cdot B}} = p_{\mathrm{X \cdot M}} p_{\mathrm{M \cdot B}} + p_{\mathrm{X \cdot N}} p_{\mathrm{N \cdot B}}.$$

These results are easily extended to cases in which B acts on X through any number of causes. If a path coefficient is assigned to each component path, the combined path coefficient for all paths connecting an effect with a remote cause equals the sum of the products of the path coefficients along all the paths. Since B is independent of A, C, and D, $r_{\mathrm{X \cdot B}} = p_{\mathrm{X \cdot B}} = p_{\mathrm{X \cdot M}} p_{\mathrm{M \cdot B}} + p_{\mathrm{X \cdot N}} p_{\mathrm{N \cdot B}}$.

GENERAL FORMULA

We are now in a position to express the correlation between any two variables in terms of path coefficients. Let X and Y be two variables which are affected by correlated causes M and N. Represent the various path coefficients by small letters as in the diagram. Let A, B, and C be hypothetical remote causes which are independent of each other (fig. 6).

$$r_{\mathrm{XY}} = p_{\mathrm{X \cdot A}} p_{\mathrm{Y \cdot A}} + p_{\mathrm{X \cdot B}} p_{\mathrm{Y \cdot B}} + p_{\cdot \mathrm{C}} p_{\mathrm{Y \cdot C}}$$

$$= mam'a + (mb + nb')(m'b + n'b') + ncn'c$$

$$= mm' + mbb'n' + nn' + nb'bm'.$$

Thus, the correlation between two variables is equal to the sum of the products of the chains of path coefficients along all of the paths by which they are connected.

If we know only the effects, X and Y, and correlated causes, such as M and N, it will be well to substitute r_{MN} for bb' in the foregoing formula.

$$r_{XY} = p_{X \cdot M} p_{Y \cdot M} + p_{X \cdot M} r_{MN} p_{Y \cdot N} + p_{X \cdot N} p_{Y \cdot N} + p_{X \cdot N} r_{MN} p_{Y \cdot M}.$$

We have reached a general formula expressing correlation in terms of path coefficients. This is not the order in which knowledge of the coefficients must be obtained, but, nevertheless, by means of simultaneous equations the values of the path coefficients in a system can often be calculated from the known correlations. Additional equations are furnished by the principle that the sum of the degrees of determination must
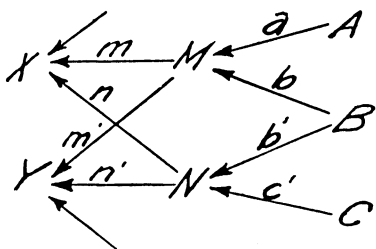

FIG. 6. Diagram showing relations between two variables, X and Y, whose values are determined in part by common causes, M and N, which are correlated with each other.
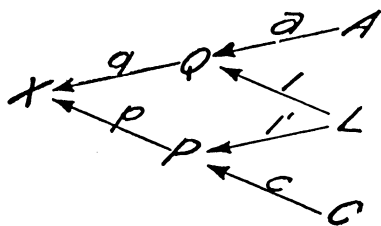
FIG. 7.—Simplified diagram of factors which determine birth weight in guinea pigs.

equal unity. The fundamental equations can be written in general form as follows:

$$d_{X \cdot A} = p^2_{X \cdot A}$$

$$d_{X \cdot \overline{AB}} = 2 p_{X \cdot A} p_{X \cdot B} r_{AB}$$

$$\Sigma d_{X \cdot A} + \Sigma d_{X \cdot \overline{AB}} = 1$$

$$r_{XY} = \Sigma p_{X \cdot A} p_{Y \cdot A}.$$

APPLICATION TO BIRTH WEIGHT OF GUINEA PIGS

As a simple example, we may consider the factors which determine birth weight in guinea pigs (fig. 7).

Let X be birth weight; Q, prenatal growth curve; P, gestation period; L, size of litter; A, hereditary and environmental factors which determine Q, apart from size of litter; C, factors determining gestation period apart from size of litter.

For the sake of simplicity, it will be assumed that the interval between litters (if less than 75 days) accurately measures the gestation period

and that the variables are connected only by the paths shown above. In a certain stock of guinea pigs the following correlations were found:

Birth weight with interval, $r_{XP} = +0.5547$.

Birth weight with litter, $r_{XL} = -0.6578$.

Interval with litter, $r_{PL} = -0.4444$.

We are able to form three equations of type $r_{XY} = \Sigma p_{X \cdot A} p_{Y \cdot A}$ and three of type $\Sigma p^2_{X \cdot A} + 2\Sigma p_{X \cdot A} p_{X \cdot B} r_{AB} = 1$. These six equations will enable us to calculate six unknown quantities. The six path coefficients in the diagram in figure 7 can thus be calculated from the information given here, but no others.

The equations are as follows:

(1) $r_{XP} = +0.5547 = p + qll'$.

(2) $r_{XL} = -0.6578 = ql + pl'$.

(3) $r_{PL} = -0.4444 = l'$.

(4) $q^2 + p^2 + 2qpll' = 1$.

(5) $a^2 + l^2 = 1$.

(6) $l'^2 + c^2 = 1$.

From (3),	$p_{P \cdot L} = l' = -0.4444$	$d_{P \cdot L} = l'^2$	$= 0.1975$
From (6),	$p_{P \cdot C} = c = 0.8958$	$d_{P \cdot C} = c^2$	$= .8025$
			1.0000

From (1) and (2),	$p_{X \cdot P} = p = 0.3269$	$d_{X \cdot P} = p^2$	$= 0.1069$
	$ql = -0.5125$	$d_{X \cdot Q} = q^2$	$= .7442$
From (4),	$p_{X \cdot Q} = q = 0.8627$	$d_{X \cdot PQ} = 2pqll'$	$= .1489$
			1.0000

	$p_{Q \cdot L} = l = -0.5941$	$d_{Q \cdot L} = l^2$	$= 0.3530$
	$p_{Q \cdot A} = a = 0.8044$	$d_{Q \cdot A} = a^2$	$= .6470$
			1.0000

	$d_{X \cdot Q \cdot L} = q^2 l^2$	$= 0.2627$
	$d_{X \cdot P \cdot L} = p^2 l'^2$	$= .0211$
	$d_{X \cdot PQ \cdot L} = 2pqll'$	$= .1489$
	$d_{X \cdot L} = (ql + pl')^2$	$= .4327$
	$d_{X \cdot A} = q^2 a^2$	$= .4815$
	$d_{X \cdot C} = p^2 c^2$	$= .0858$
		1.0000

Assuming that the diagrams in figures 7, 8, and 9 accurately represent the causal relations, it appears that birth weight is determined to a very much greater extent by v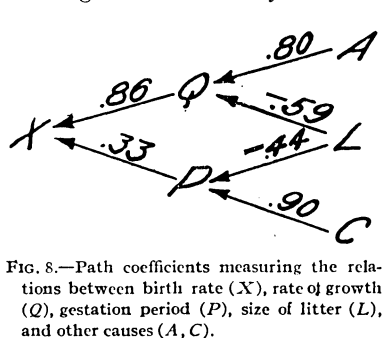ariations in the rate of growth of the foetuses than by variations in the length of the gestation period ($d_{\text{X}\cdot\text{Q}}=0.74$, $d_{\text{X}\cdot\text{P}}=0.11$). Size of litter has much more effect on birth weight by reducing the rate of growth of the foetuses than by causing early parturition ($d_{\text{X}\cdot\text{Q}\cdot\text{L}}=0.26$, $d_{\text{X}\cdot\text{P}\cdot\text{L}}=0.02$). The difference in birth weight caused by a difference of a day in gestation period can be calculated from the path coefficient and the standard deviations

FIG. 8.—Path coefficients measuring the relations between birth rate (*X*), rate of growth (*Q*), gestation period (*P*), size of litter (*L*), and other causes (*A,C*).

by the formula for path regression, $p.^{\cdot}\ reg_{\text{X}\cdot\text{P}}=p_{\text{X}\cdot\text{P}}\dfrac{\sigma_{\text{X}}}{\sigma_{\text{P}}}$. The result, 3.34

gm. per day, should measure the average rate of growth just preceding parturition. The actual regression, 5.66 gm. per day of delay in parturition, is larger because a long gestation period means not merely a longer time for growth but also, in general, a smaller litter and hence more rapid growth.

On introducing other data the analysis can be carried much farther. There are other paths of influence which should be recognized, positive paths connecting *A*, *C*, and *L*, representing the favorable effects of good health in the dam on rate of growth, gestation period, and size of litter, and a negative path from *Q* to *P* to represent the tendency of rapid growth to induce early parturition. The relations between the observed interval between litters and the actual gestation period should also be considered. The results presented here are thus intended merely to furnish a simple illustration of the method. A more complete analysis

FIG. 9.—Coefficients of determination. Symbols as in figure 7.

of the relations among the factors which affect birth weight and later growth will be presented in a later paper.

DETERMINATION IN TERMS OF CORRELATION
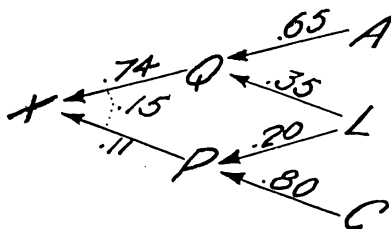
Having obtained a formula for correlation in terms of determination, the question arises whether the converse is possible. For a special class of cases such a formula is easily obtained.

For a single cause and effect the required formula is merely $d_{x \cdot A} = r^2_{XA}$ (fig. 10).
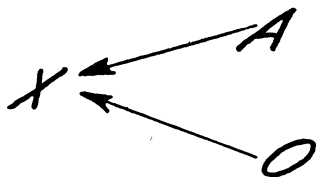

FIG. 10.—Effect and one known cause.

The degree of determination by residual factors; that is, $d_{x \cdot o}$, is thus $1 - r^2_{XA}$.

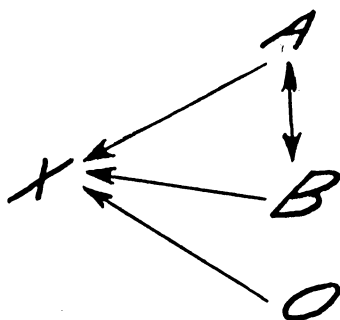If two causes are known, and the degree of correlation between them, we have (fig. 11)—

FIG. 11.—Effect and two correlated known causes.

$$_B r^2_{XA} + _B r^2_{XO} = 1$$

$$\frac{(r_{XA} - r_{XB}r_{AB})^2}{(1 - r^2_{XB})(1 - r^2_{AB})} = 1 - \frac{r^2_{XO}}{1 - r^2_{XB}}$$

$$r^2_{XO} = d_{x \cdot o} = \frac{1 - r^2_{XA} - r^2_{XB} - r^2_{AB} + 2r_{XA}r_{XB}r_{AB}}{1 - r^2_{AB}}.$$
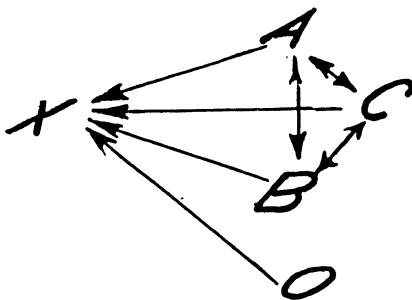

FIG. 12.—Effect and three correlated known causes.

If three causes and their correlations are known (fig. 12), we have $_{CB} r^2_{XA} + _{CB} r^2_{XO} = 1$, from which

$$r^2_{XO} = d_{x \cdot o} = \frac{1 - \Sigma r^2_{XA} + 2\Sigma r_{XA}r_{AB}r_{BX} - 2\Sigma r_{XA}r_{AB}r_{BC}r_{CX} + \Sigma r^2_{XA}r^2_{BC}}{1 - r^2_{AB} - r^2_{AC} - r^2_{BC} + 2r_{AC}r_{CB}r_{BA}}.$$

In this expression Σr^2_{XA} means the sum of squares of the six known correlations. $\Sigma r_{XA} r_{AB} r_{BX}$ means the sum of the products of the groups of three correlations, corresponding to the sides of triangles. There are four of these triangles, XAC, XAB, XCB, ABC. $\Sigma r_{XA} r_{AB} r_{BC} r_{CX}$ means the sum of the three products of the groups of correlations which are arranged in closed quadrilaterals, and $\Sigma r^2_{XA} r^2_{BC}$ means the sum of the product of squared correlations in pairs which involve no common variable ($r^2_{XA} r^2_{BC}$, $r^2_{XC} r^2_{AB}$, $r^2_{XB} r^2_{AC}$) (fig. 13).

The formula for four known causes is easily found by a continuation of the methods used to find the others if attention is paid to the sym-
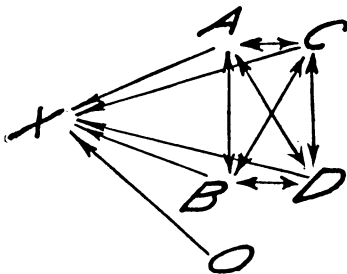

FIG. 13.—Effect and four correlated known causes.

metry of the expressions. Since, however, this formula, as well as that just given for the case of three causes, is somewhat cumbersome, it will be convenient to use a more condensed notation. $\phi(XABC \ldots)$ may be used for a function involving all possible correlations among the variables $(XABC \ldots)$. In the definitions Σr^2 means the sum of the squares of all correlations; $\Sigma r^2 r^2$, the sum of the product of all pairs of squared correlations which involve no variables in common; Σrrr, $\Sigma rrrr$, and $\Sigma rrrrr$ are the sums of the products of all groups of correlations which, represented by paths, form closed figures, triangles, quadrilaterals, and pentagons, respectively. $\Sigma r^2 rrr$ is the sum of the products made by multiplying each triangle of correlations in the sense above by the second power of those correlations which do not involve any of the variables in the triangle. The number of terms of each kind is given above the brace, where it is more than one.

$\phi(AB) \quad = 1 - r^2$ (2 terms).

$$\phi(ABC) \quad = 1 - \overset{3}{\overbrace{\Sigma r^2}} + 2\Sigma rrr \text{ (5 terms).}$$

$$\phi(ABCD) \quad = 1 - \overset{6}{\overbrace{\Sigma r^2}} + \overset{4}{\overbrace{2\Sigma rrr}} - \overset{3}{\overbrace{2\Sigma rrrr}} + \overset{3}{\overbrace{\Sigma r^2 r^2}} \text{ (17 terms).}$$

$$\phi(ABCDE) = 1 - \overset{10}{\overbrace{\Sigma r^2}} + \overset{10}{\overbrace{2\Sigma rrr}} - \overset{15}{\overbrace{2\Sigma rrrr}} + \overset{12}{\overbrace{2\Sigma rrrrr}} + \overset{15}{\overbrace{\Sigma r^2 r^2}} - \overset{10}{\overbrace{2\Sigma r^2 rrr}} \text{ (73 terms).}$$

The formulae for degree of determination by residual factors may be written as follows:

$d_{x \cdot o} = \phi(XA)$ in system XA.

$d_{x \cdot o} = \dfrac{\phi(XAB)}{\phi(AB)}$ in system XAB.

$d_{x \cdot o} = \dfrac{\phi(XABC)}{\phi(ABC)}$ in system $XABC$.

$d_{x \cdot o} = \dfrac{\phi(XABCD)}{\phi(ABCD)}$ in system $XABCD$

The degree of determination by the known causes is now easily calculated. When all causes of variation in X are constant except A, variation of X is measured by $_{o \cdots cB}\sigma_x$ and variation of A is measured by $_{o \cdots cB}\sigma_A$, writing the constant factors as subscripts to the left. Assuming that the relation between A and X is linear, the deviation of X determined by a unit deviation of A should be constant, whatever the amount of variation in A. Thus:

$$p_{x \cdot A} \frac{\sigma_x}{\sigma_A} = \frac{\sigma_{x \cdot A}}{\sigma_A} = \frac{_{o \cdots cB}\sigma_x}{_{o \cdots cB}\sigma_A}.$$

In the case of the residual factor O, assumed to be independent of the known factors A, B, C, etc., $_{\cdots cBA}\sigma_o = \sigma_o$, and we have $\sigma_{x \cdot o} = {}_{\cdots cBA}\sigma_x$

$$d_{x \cdot o} = \frac{\phi(XABC...)}{\phi(ABC...)} = \frac{\sigma^2_{x \cdot o}}{\sigma^2_x} = \frac{_{\cdots cBA}\sigma^2_x}{\sigma^2_x}.$$

Thus:

$$_{\cdots cBA}\sigma^2_x = \frac{\phi(XABC...)}{\phi(ABC...)} \sigma^2_x.$$

This should be the general formula for the squared standard deviation with a number of constant factors.
Hence:

$$\frac{\sigma^2_{x \cdot A}}{\sigma^2_A} = \frac{\phi(XBC...O)}{\phi(BC...O)} \sigma^2_x \bigg/ \frac{\phi(ABC...O)}{\phi(BC...O)} \sigma^2_A$$

$$\sigma^2_{x \cdot A} = \frac{\phi(XBC...O)}{\phi(ABC...O)} \sigma^2_x$$

$$p_{x \cdot A} = \sqrt{\frac{\phi(XBC...O)}{\phi(ABC...O)}}$$

$$d_{x \cdot A} = \frac{\phi(XBC...O)}{\phi(ABC...O)} = \frac{\phi(XBC...) - d_{x \cdot o}\phi(BC...)}{\phi(ABC...)}.$$

The general formula for partial correlation can easily be expressed in the present terminology.

$$_{\text{DCBA}}\sigma^2_x = {_{\text{DCB}}}\sigma^2_x(1 - {_{\text{DCB}}}r^2_{xA})$$

$$_{\text{DCB}}r^2_{xA} = 1 - \frac{_{\text{DCBA}}\sigma^2_x}{_{\text{DCB}}\sigma^2_x} = 1 - \frac{\phi(XABCD)\phi(BCD)}{\phi(ABCD)\phi(XBCD)}.$$

In some cases it may be of interest to find the degree of determination when a number of factors not in the direct path between cause and effect are assumed constant.

$$_{\text{UTS}}d_{x\cdot A} = \frac{_{\text{UTS}}\sigma^2_{x\cdot A}}{_{\text{UTS}}\sigma^2_x} = \frac{\left(_{\text{o}\cdots\text{UTS}\cdots\text{CB}}\sigma^2_x\right)\left(_{\text{UTS}}\sigma^2_A\right)}{\left(_{\text{o}\cdots\text{UTS}\cdots\text{CB}}\sigma^2_A\right)\left(_{\text{UTS}}\sigma^2_x\right)}$$

$$= \frac{\phi(XBC...STU...O)\phi(ASTU)}{\phi(ABC...STU)\phi(XSTU)}.$$

RELATION TO MULTIPLE CORRELATION

The expressions defined as $\phi(XABC...)$, etc., suggest the expansion of determinants. It is in fact easy to show that $\phi(XABC...N) = \Delta$.

Where
$$\Delta = \begin{vmatrix} 1 & r_{XA} & r_{XB} & \cdot & \cdot & r_{XN} \\ r_{AX} & 1 & r_{AB} & \cdot & \cdot & r_{AN} \\ r_{BX} & r_{BA} & 1 & \cdot & \cdot & r_{BN} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ r_{NX} & r_{NA} & r_{NB} & \cdot & \cdot & 1 \end{vmatrix}$$

The formula for Pearson's coefficient of multiple correlation has already been given, $R_{x(\text{ABCO})} = \sqrt{1 - \dfrac{\Delta}{\Delta_{xx}}}$ where Δ_{xx} is the minor made by deleting row X, column X.

Evidently in this class of cases the coefficient of determination degenerates into a function of the coefficient of multiple correlation. For the degree of determination by residual factors we have

$$d_{x\cdot o} = \frac{\phi(XABC...)}{\phi(ABC...)} = \frac{\Delta}{\Delta_{xx}} = 1 - R^2_{x(\text{ABC}\cdots)}$$

in agreement with Pearson's results.

For the degree of determination by a known factor we have

$$d_{x\cdot A} = \frac{\phi(XBC...O)}{\phi(ABC...O)} = \frac{\phi(XBC...) - d_{x\cdot o}\phi(BC...)}{\phi(ABC...)} = \frac{\Delta_{AA}\Delta_{xx} - \Delta\Delta_{AAxx}}{\Delta^2_{xx}}$$

$$= \frac{\Delta^2_{xA}}{\Delta^2_{xx}}.$$

$$p_{x\cdot A} = \frac{\Delta_{xA}}{\Delta_{xx}}.$$

The last formula brings out the close relation between the path coefficients and multiple regression. As already noted, the most probable deviation of X for known deviations of A, B, C, etc., is given by the formula

$$\frac{X'}{\sigma_X} = \frac{\Delta_{XA}A'}{\Delta_{XX}\sigma_A} + \frac{\Delta_{XB}B'}{\Delta_{XX}\sigma_B} \cdots = p_{X \cdot A}\frac{A'}{\sigma_A} + p_{X \cdot B}\frac{B'}{\sigma_B} \cdots$$

As already stated, Pearson's coefficients of multiple correlation and regression were not devised especially for the analysis of causal relations. The formula for multiple regression, for example, gives the most probable value of one of the variates for given values of the others regardless of causal relations. In cases in which all the correlations are known in a system including an effect and a number of causes the method can be used to find the path coefficients and the degrees of determination of the effect by each cause in the sense used in this paper. Such cases in which the direct methods can be used are, however, relatively uncommon. Where the system of paths of influence is at all complex, involving perhaps hypothetical factors, the causal relations can be analyzed only by the indirect method of expressing the known correlations in terms of the unknown path coefficients, making the sums of the degrees of determination unity and solving the simultaneous equations.

PART II. APPLICATION TO THE TRANSPIRATION OF PLANTS

A large body of experimental data on the factors which affect the rate of transpiration in plants has been published by Briggs and Shantz (2). These data are well adapted for use in illustrating the methods of analyzing causal relations presented in part I of this paper.

The experiments which are used in this paper were conducted at Akron, Colo., in 1914. A variety of crop plants were grown in sealed pots. The total transpiration was measured each day. Among the environmental factors studied were the total solar radiation during the day, the wind velocity, the air temperature (in the shade), the rate of evaporation from a shallow tank, and the wet-bulb depression (sheltered from sun but not wind). The correlations between the daily transpiration of each kind of plant and the integrated values of the environmental factors were published by Briggs and Shantz. In order to avoid the effect of seasonal change in the plants, the logarithms of the ratios of the transpiration on succeeding days were correlated with similar figures for the various factors. The correlations between the various environmental factors for the 100 days from June 18 to September 25, 1914, have been calculated by the writer from the data presented by Briggs and Shantz. This period covers all the crop periods but is longer than most of them. None of the correlations appeared to depart much from linearity.

The daily averages, the standard deviations, and the correlations are given in Table I.

TABLE I.—*Daily averages, standard deviations, and correlations from experiments on transpiration in crop plants made by Briggs and Shantz at Akron, Colo., 1914*

CORRELATIONS

	Wind.	Radiation.	Temperature.	Wet-bulb depression.	Evaporation.
Wind...................	....	−0.01 ±0.07	−0.02 ±0.07	0.28 ±0.06	0.38 ±0.06
Radiation...............	−0.01 ±0.07	....	.47 ±.05	.48 ±.05	.68 ±.04
Temperature.............	− .02 ± .07	.47 ± .05	....	.59 ± .05	.56 ± .05
Wet-bulb depression.......	.28 ± .06	.48 ± .05	.59 ± .05	....	.83 ± .02
Evaporation..............	.38 ± .06	.68 ± .04	.56 ± .05	.83 ± .02	....
Small grains[a].........	.22 ± .04	.65 ± .03	.71 ± .02	.88 ± .01	.87 ± .02
Rye....................	.19 ± .10	.65 ± .06	.73 ± .05	.94 ± .01	.91 ± .02
Sorghum, millet[b]......	.218± .041	.570± .030	.653± .026	.788± .018	.713± .021
Sudan grass (in inclosure).	.52 ± .07	.55 ± .06	.84 ± .03	.83 ± .03	.93 ± .01
Sudan grass (in open)......	.32 ± .08	.52 ± .07	.81 ± .03	.85 ± .03	.82 ± .03
Dent corn...............	.28 ± .08	[e] .52 ± .06	.71 ± .04	.81 ± .03	.79 ± .03
Algerian corn...........	.33 ± .09	.62 ± .06	.79 ± .04	.88 ± .02	.85 ± .03
Cowpea, lupine[c].......	.335± .057	.570± .042	.675± .035	.785± .025	.775± .025
Alfalfa[d]	.290± .035	.430± .030	.495± .029	.700±. 019	.705± .019
Amaranthus.............	.04 ± .10	.40 ± .09	.45 ± .08	.60 ± .07	.56 ± .06

	Mean.	σ
Evaporation (shallow tank) (kilograms per square meter)...........	9.70	2.76
Integrated radiation (calories per square centimeter)..............	753	134
Air temperature, integrated mean (degrees Centigrade).............	20.10	3.48
Integrated wet-bulb depression (hour degrees, Centigrade)..........	143	58
Wind velocity (miles per hour)................................	5.54	2.24

[a] Averages of six similar correlations involving Kubanka and Galgalos wheat, Swedish Select and Burt oats, Hannchen barley, and spring rye. The last, having on the whole the largest correlations, is also given separately.
[b] Averages of four correlations, Minnesota Amber and Dakota Amber sorghum and Kursh and Siberian Millet. These correlations were all very similar.
[e] Average of the similar correlations for cowpeas and lupine.
[d] Average of four tests with alfalfa.
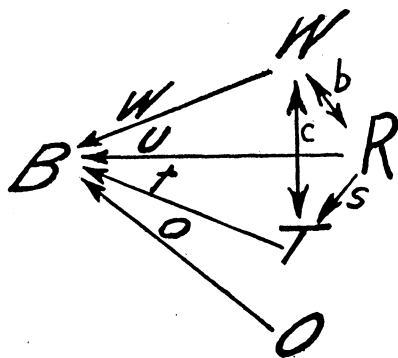[e] Published as + 0.80, which seems too large. Recalculation gives + 0.52.

FIG. 14.—Relations between wet-bulb depression (*W*), wind velocity (*W*), radiation (*R*), and temperature (*T*) as assumed for direct analysis.

It will be interesting first to compare the direct and indirect methods of calculating path coefficients and coefficients of determination. Let us consider the relations of wet-bulb depression (B) to temperature (T), wind velocity (W), and radiation (R). Since the direct methods are only applicable in systems in which each variable is connected with every other variable, the diagram of relations is as shown in figure 14. Outstanding factors, independent of W, R, and T are represented by O.

INDIRECT METHOD

Six equations can be formed, expressing the six known correlations in terms of the unknown path coefficients. A seventh equation represents the complete determination of B by W, R, T, and O.

(1) $r_{BW} = \quad 0.28 = w + t(c + bs) + ub.$

(2) $r_{BR} = \quad .48 = wb + ts + u.$

(3) $r_{BT} = \quad .59 = w(c + bs) + t + us.$

(4) $r_{WR} = - \quad .01 = b.$

(5) $r_{WT} = - \quad .02 = c + bs.$

(6) $r_{RT} = \quad .47 = s.$

(7) $o^2 + w^2 + t^2 + u^2 + 2wt(c + bs) + 2wub + 2uts = 1.$

The values of b and s are given directly from equations (4) and (6), and the value of c ($= -0.0153$) can then be obtained from (5). The solution of (1), (2), and (3) gives $w = 0.2921$, $t = 0.4735$, and $u = 0.2604$. Finally, from (7) we obtain $o^2 = 0.5138$ as the degree of determination by outstanding factors.

$$d_{B \cdot O} = o^2 \qquad = \quad 0.5138$$
$$d_{B \cdot W} = w^2 \qquad = \quad .0853 \qquad p_{B \cdot W} = w = 0.2921$$
$$d_{B \cdot T} = t^2 \qquad = \quad .2242 \qquad p_{B \cdot T} = t = \quad .4735$$
$$d_{B \cdot R} = u^2 \qquad = \quad .0678 \qquad p_{B \cdot R} = u = \quad .2604$$
$$d_{B \cdot \overline{WT}} = 2wt(c + bs) = - \quad .0055$$
$$d_{B \cdot \overline{WR}} = 2wub \qquad = - \quad .0015$$
$$d_{B \cdot \overline{RT}} = 2uts \qquad = \quad \underline{.1159}$$
$$1.0000$$

DIRECT METHODS

According to the formulae given in part I we have—

$$d_{B \cdot O} = \frac{\phi(BWRT)}{\phi(WRT)}$$

$$d_{B \cdot W} = \frac{\phi(BRT) - d_{B \cdot O}\phi(RT)}{\phi(WRT)}$$

$$d_{B \cdot R} = \frac{\phi(BWT) - d_{B \cdot O}\phi(WT)}{\phi(WRT)}$$

$$d_{B \cdot T} = \frac{\phi(BRW) - d_{B \cdot O}\phi(RW)}{\phi(WRT)}$$

where

$$\phi(BWRT) = 1 - r^2_{BW} + 2r_{BW}r_{WR}r_{RB} - 2r_{BW}r_{WR}r_{RT}r_{TB} + r^2_{BW}r^2_{RT}$$
$$- r^2_{BR} + 2r_{BW}r_{WT}r_{TB} - 2r_{BW}r_{WT}r_{TR}r_{RB} + r^2_{BR}r^2_{WT}$$
$$- r^2_{BT} + 2r_{BR}r_{RT}r_{TB} - 2r_{BR}r_{RW}r_{WT}r_{TB} + r^2_{BT}r^2_{WR}$$
$$- r^2_{WR} + 2r_{WR}r_{RT}r_{TW}$$
$$- r^2_{WT}$$
$$- r^2_{RT}$$

$$\phi(WRT) = 1 - r^2_{WR} - r^2_{WT} - r^2_{RT} + 2r_{WR}r_{RT}r_{TW}$$

$\phi(BWR)$, etc., are analogous to $\phi(WRT)$

$$\phi(RT) = 1 - r^2_{TR} \quad \phi(WT), \text{ etc., are analogous to } \phi(RT).$$

By substitution of the correlations in these formulae the following results are obtained:

$$\phi(BWRT) = 0.4002$$

$\phi(BWR)$ = .6884	$\phi(BW)$ = 0.9216	$\phi(WR)$ = 0.9999
$\phi(BWT)$ = .5665	$\phi(BR)$ = .7696	$\phi(WT)$ = .9996
$\phi(BRT)$ = .4668	$\phi(BT)$ = .6519	$\phi(RT)$ = .7791
$\phi(WRT)$ = .7788		

These give values of the coefficients of determination identical with those given by the indirect method.

This method, as was shown in part I, is essentially the same as Pearson's method of calculating multiple regression.

$$\text{Let } \Delta = \begin{vmatrix} 1 & r_{BR} & r_{BT} & r_{BW} \\ r_{RB} & 1 & r_{RT} & r_{RW} \\ r_{TB} & r_{TR} & 1 & r_{TW} \\ r_{WB} & r_{WR} & r_{WT} & 1 \end{vmatrix} = \begin{vmatrix} 1 & 0.48 & 0.59 & 0.28 \\ .48 & 1 & .47 & -.01 \\ .59 & .47 & 1 & -.02 \\ .28 & -.01 & -.02 & 1 \end{vmatrix} = 0.4002$$

Let $\Delta_{BB} = \Delta$ with column B, row B, deleted.

$$\Delta_{BB} = 0.7788, \Delta_{BR} = 0.2028, \Delta_{BT} = 0.3687, \Delta_{BW} = 0.2275$$

$$p_{B \cdot W} = \frac{\Delta_{BW}}{\Delta_{BB}} = 0.2921 \qquad\qquad d_{B \cdot 0} = \frac{\Delta}{\Delta_{BB}} = 0.5139$$

$$p_{B \cdot R} = \frac{\Delta_{BR}}{\Delta_{BB}} = 0.2604$$

$$p_{B \cdot T} = \frac{\Delta_{BT}}{\Delta_{BB}} = 0.4735.$$

These values are identical with those obtained by the preceding methods.

It will be seen that the first method, while apparently less direct than the others, is really less laborious. The solution of three simultaneous equations requires merely the evaluation of a determinant of the third order instead of one of the fourth order, as in the last method. The expression $\phi(BWRT)$ in the second method is, of course, merely an expansion of the same determinant of the fourth order as that used in the last. The indirect method, moreover, gives more insight into the processes followed than the others in which there is a substitution in what appear to be arbitrary formulae. In line with this last point, the indirect method is more flexible in that it can be used to test out the consequences of any assumed relation among the factors.

ANALYSIS OF CAUSAL RELATIONS

In attempting to interpret the present results in terms of causation, we see at once that the scheme of relations chosen is not a very satisfactory one. The wet-bulb depression was measured under shelter. Consequently the coefficient of determination, $d_{B \cdot R} = 0.0678$, can not measure

the degree of direct determination by radiation, but determination by some factor other than wind or temperature with which radiation is correlated.

One should not attempt to apply in general a causal interpretation to solutions by the direct methods. In these cases, determination can usually be used only in the sense in which it can be said that knowledge of the effect determines the probable value of the cause. This is the sense in which Pearson's formula for multiple regression must be interpreted. If W', T', and R' are given deviations of wind, temperature, and radiation from their mean values, the most probable value of the wet-bulb depression, B', is given by the following formula:

$$\frac{B'}{\sigma_B} = \frac{W'}{\sigma_W}p_{B\cdot W} + \frac{R'}{\sigma_R}p_{B\cdot R} + \frac{T'}{\sigma_T}p_{B\cdot T}.$$

This formula can only be used for conditions which are similar to those for which the values of the path coefficients were calculated. If path coefficients were calculated in a system which truly represented the causal relations, the formula would give the value of the wet-bulb depression under any set of conditions in so far as it is determined by the factors considered.

The causal factors which actually determine wet-bulb depression are temperature, absolute humidity (H), and wind velocity (fig. 15). Radiation can be introduced into the scheme
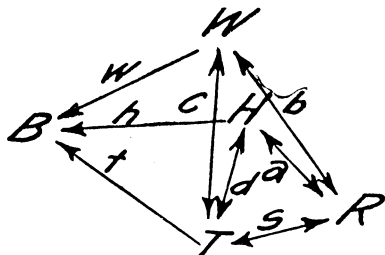

FIG. 15.—Relations between factors of figure 14 and absolute humidity (H) expressing causal relations better than figure 14 but adapted only to indirect analysis.

as a factor correlated with these causal factors. Wind velocity is correlated to such a very slight extent with temperature and radiation that its correlation with absolute humidity can probably be neglected without serious error. The relations between radiation, temperature, and absolute humidity are undoubtedly very complex. Radiation has a direct positive influence on temperature. Both radiation and temperature have positive effects on absolute humidity by increasing evaporation. Correlation between absolute humidity and temperature would be expected, because with reduced temperature the saturation point is reached at a lower absolute humidity and the excess moisture is precipitated. Increase in humidity, on the other hand, tends to reduce the radiation which reaches the earth, and directly or indirectly this has a negative influence on all three of the correlations.

There are not enough data to estimate the importance of all of these paths of influence. Even if we represent the complex of paths connecting H, R, and T merely by three correlations, the diagram has eight paths to solve. The six correlations between B, W, R, and T and the statement

in regard to complete determination of B by W, H, and T furnish only seven equations.

Fortunately, data are given in another paper by Briggs and Shantz (*3*) from which an eighth equation can be derived. In this paper the average value of each of the measured factors is given for each hour of the day. The cycle of changes in wet-bulb depression follows very closely the changes in temperature. In fact, there should be very little, if any, regular hourly cycle of changes in absolute humidity, so that the wet-bulb depression should be wholly determined by the temperature changes except for some influence of wind velocity.

Let $p_{B \cdot T} = t$ be the path coefficient which measures the relative influence of temperature on wet-bulb depression in the variations from day to day. Let $p_{B \cdot H} = h$, $p_{B \cdot W} = w$, and let σ_T, σ_H, σ_W, and σ_B be the standard deviations of the daily differences in the various factors and in wet-bulb depression. Let $T' - T''$, etc., be the actual differences in temperature, etc., at certain times. The difference to be expected in wet-bulb depression, $B' - B''$, is as follows:

$$\frac{B' - B''}{\sigma_B} = \frac{T' - T''}{\sigma_T} t + \frac{W' - W''}{\sigma_W} w + \frac{H' - H''}{\sigma_H} h.$$

While t, w, and h are assumed to measure the relative influence of temperature, wind, and humidity in the variations from day to day, the foregoing formula should apply under any conditions, if t, w, and h were calculated from a system which represented truly causal relations. The expression $\frac{\sigma_B}{\sigma_T} t$ is shown in part I to give the change in wet-bulb depression (B) directly caused by a unit change in temperature. The relative importance of the various factors in determining the variations from hour to hour is very different from that from day to day, but the change in wet-bulb depression caused by unit changes in temperature, wind velocity, or absolute humidity should always be the same so long as the relations are substantially linear.

The greatest difference, in temperature within an average day in the data was between 5 a. m. and 3 p. m. This is given as 32.7° F., or 18.167° C. The difference in wet-bulb depression between these hours was 21.8° F., or 12.111° C. The difference in average wind velocity was 2.5 miles per hour. The standard deviations of the daily variations have already been given. $\sigma_T = 3.48$ day degrees C., $\sigma_B = 58$ hour degrees C. integrated for 24 hours. This means 2.4167 degrees C. $\sigma_W = 2.24$ miles per hour. We will assume that there is no difference in absolute humidity ($H' - H'' = 0$). Substituting those values in the formula for wet-bulb depression, we get

$$\frac{12.111}{2.4167} = \frac{18.167}{3.48} t + \frac{2.50}{2.24} w$$

$$5.0114 = 5.2204 t + 1.1161 w.$$

We now have eight equations from which to find eight unknown path coefficients.

(1) $r_{BW} = 0.28 = w + tc$.

(2) $r_{BR} = .48 = ts + bw + ah$.

(3) $r_{BT} = .59 = t + dh + wc$.

(4) $r_{WR} = -.01 = b$.

(5) $r_{WT} = -.02 = c$.

(6) $r_{RT} = .47 = s$.

(7) $w^2 + h^2 + t^2 + 2wtc + 2htd = 1$.

(8) $5.0114 = 5.2204t + 1.1161w$.

Equations (4), (5), and (6) give b, c, and s directly. Solution of (1) and (8) gives $t = 0.8963$, $w = 0.2979$.

From (2)　$ah = 0.0617$

From (7)　$h^2 = .6570$, $h = -0.8105$, $a = -0.0761$

From (3)　$dh = -.3003$, $d = .3706$

$r_{BH} = h + td = -0.4784$.

The coefficients of determination, the path coefficients, and the correlations are thus as follows:

$d_{B \cdot T} = 0.8034$　　$p_{B \cdot T} = 0.8963$　　$r_{BT} = 0.5900$

$d_{B \cdot H} = .6570$　　$p_{B \cdot H} = -.8105$　　$r_{BH} = -.4784$

$d_{B \cdot W} = .0888$　　$p_{B \cdot W} = .2979$　　$r_{BW} = .2800$

$d_{B \cdot \overline{HT}} = -.5384$

$d_{B \cdot \overline{WT}} = -.0107$　　　　　　　　　　$r_{HR} = -.0761$

　　　　$\overline{1.0001}$　　　　　　　　　　$r_{HT} = .3706$

　　　　　　　　　　　　　　　　　$r_{RT} = .4700$.

It turns out that the differences between different days in wet-bulb depressions are due to a somewhat greater extent to differences in temperature (0.80) than to absolute humidity (0.66). The variation in wet-bulb depression would be much greater were it not that these factors vary together but act on wet-bulb depression in opposite directions and so tend to balance each other ($d_{B \cdot \overline{HT}} = -0.54$). Temperature shows a rather strong positive correlation with absolute humidity (0.37) as well as with radiation (0.47), but the various paths of influence between radiation and absolute humidity almost balance each other ($r_{HR} = -0.08$).

These results can now be used in finding the relative importance of the various factors which determine evaporation or transpiration. In figure 16, X may represent either evaporation or the transpiration of any plant. Radiation must be considered as a direct causal factor in these cases.

The following four equations can be made with which to solve the path coefficients from W, H, R, and T to X:

$$r_{XW} = w' \quad + t'c \quad + u'b$$
$$r_{XT} = w'c \quad + t' \quad + u's \quad + h'd$$
$$r_{XR} = w'b \quad + t's \quad + u' \quad + h'a$$
$$r_{XB} = w'r_{BW} + t'r_{BT} + u'r_{BR} + h'r_{BH}.$$

Substituting the values already found for a, b, c, d, w, h, t, and r_{BH}, we have

$$r_{XW} = + 1.00w' - 0.02t' - 0.01u'$$
$$r_{XT} = - .02w' + 1.00t' + .47u' + 0.3706h'$$
$$r_{XR} = - .01w' + .47t' + 1.00u' - .0761h'$$
$$r_{XB} = + .28w' + .59t' + .48u' - .4784h'.$$

The solution is as follows:

$$w' = p_{X \cdot W} = + 0.9971r_{XW} + 0.0143r_{XT} - 0.0022r_{XR} + 0.0114r_{XB}$$
$$t' = p_{X \cdot T} = - .2207r_{XW} + .8943r_{XT} - .8175r_{XR} + .8228r_{XB}$$
$$u' = p_{X \cdot R} = + .1488r_{XW} - .3633r_{XT} + 1.4155r_{XR} - .5067r_{XB}$$
$$h' = p_{X \cdot H} = + .4607r_{XW} + .7468r_{XT} + .4107r_{XR} - 1.5772r_{XB}.$$

It is merely necessary to substitute the values of the correlations of evaporation or transpiration with wind velocity, temperature, radia-
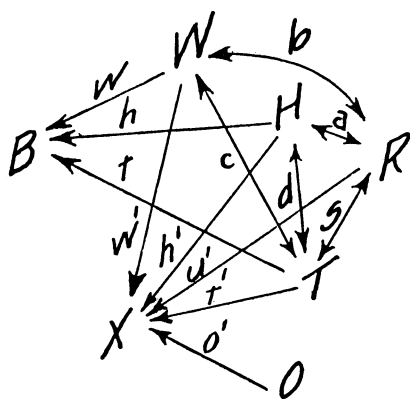

FIG. 16.—Relations between evaporations or transpiration (X) and the system shown in figure 15.

tion, and wet-bulb depression, as given in Table I, to find the four path coefficients in each case. The results are given in Table II. These have all been checked by substitution in the fourth equation ($r_{XB} = +$ $0.28w' + 0.59t' + 0.48u' - 0.4784h'$). The correlation between evaporation and the transpiration of any plant can be deduced from the formula $r_{XE} = w'r_{EW} + t'r_{ET} + u'r_{ER} + h'r_{EH}.$ The correlations of evaporation with wind velocity, temperature, and radiation have been given in Table I as 0.38, 0.56, and 0.68, and that with humidity can be calculated by the formula $r_{EH} = p_{E \cdot H} + a p_{E \cdot R} + d p_{E \cdot T} = -0.2651.$ Thus $r_{XE} = 0.38w' + 0.56t' + 0.68u' - 0.2651h'.$ The calculated results in column 6 of Table II are compared with actual correlations between evaporation and transpiration in column 7. The correlation of evaporation with itself comes out 0.839 by this formula. There should, however, be an additional term ($p_{X \cdot O}r_{EO}$) in the formula to allow for correlation through other factors (O) than W, T, R, and H. From Table III we find that evaporation is determined

to a considerable extent ($d_{E \cdot O} = 0.161$) by outstanding factors. The additional term in this case would have this value and when added to 0.839 gives 1, as it should. With one exception, the calculated correlation between transpiration and evaporation is a little smaller than the actual correlation. This means either that there is some additional factor which should be allowed for or else that the path coefficients with W, T, R, and H are not given quite their due weight, owing perhaps to lack of complete linearity in the correlations.

TABLE II.—*Table of calculated path coefficients*

	Wind.	Tempera-ture.	Radia-tion.	Absolute humidity.	Correlation with evaporation.	
					Calcu-lated.	Actual.
	$p_{X \cdot W}$	$p_{X \cdot T}$	$p_{X \cdot R}$	$p_{X \cdot H}$	r_{XE}	r_{XE}
Wet-bulb depression...................	0. 298	0. 896	0	—0. 811	0. 830	0. 83
Evaporation (shallow tank).............	. 395	. 544	. 395	— . 437	(. 839)	1. 00
Transpiration:						
Small grains......................	. 238	. 779	. 249	— . 489	. 826	. 87
Rye.............................	. 209	. 853	. 207	— . 583	. 852	. 91
Sorghum and millet................	. 234	. 718	. 203	— . 421	. 741	. 713
Sudan grass (inclosure).............	. 539	. 870	. 130	— . 216	. 838	. 93
Sudan grass (open).................	. 339	. 928	. 059	— . 375	. 788	. 82
Dent corn.........................	. 297	. 815	. 109	— . 405	. 751	. 79
Algerian corn......................	. 349	. 851	. 194	— . 391	. 844	. 85
Cowpea and lupine.................	. 351	. 710	. 214	— . 346	. 768	. 775
Alfalfa............................	. 303	. 603	. 117	— . 424	. 645	. 705
Amaranthus.......................	. 052	. 560	. 105	— . 428	. 518	. 560
Average transpiration.................	. 279	. 733	. 181	— . 420	. 751	. 781

TABLE III.—*Coefficients of determination*

	Wind.	Tem-pera-ture.	Radi-ation.	Abso-lute hu-mid-ity.	Joint determination.					Residual.
	$d_{X \cdot W}$	$d_{X \cdot T}$	$d_{X \cdot R}$	$d_{X \cdot H}$	$d_{X \cdot WT}$	$d_{X \cdot WR}$	$d_{X \cdot TR}$	$d_{X \cdot TH}$	$d_{X \cdot RH}$	$d_{X \cdot O}$
Wet-bulb depression.......	0. 089	0. 803	0	0. 657	—0. 011	0	0	—0. 538	0	0
Evaporation...............	. 156	. 296	0. 156	. 191	— . 009	— . 003	0. 202	— . 176	+0. 026	0. 161
Transpiration:										
Small grain...........	. 057	. 607	. 062	. 240	— . 007	— . 001	. 182	— . 283	+ . 019	. 125
Rye...................	. 044	. 728	. 043	. 340	— . 007	— . 001	. 166	— . 369	+ . 018	. 018
Sorghum and millet...	. 055	. 516	. 041	. 177	— . 007	— . 001	. 137	— . 224	+ . 013	. 293
Sudan (inclosure).....	. 290	. 757	. 017	. 047	— . 019	— . 001	. 106	— . 140	+ . 004	(— . 062)
Sudan (open).........	. 115	. 801	. 003	. 141	— . 013	— . 000	. 051	— . 258	+ . 003	. 096
Dent corn.............	. 088	. 664	. 012	. 164	— . 010	— . 001	. 084	— . 244	+ . 007	. 237
Algerian corn.........	. 122	. 724	. 038	. 153	— . 012	— . 001	. 155	— . 247	+ . 012	. 057
Cowpea and lupine.....	. 123	. 504	. 046	. 120	— . 010	— . 002	. 143	— . 182	+ . 011	. 247
Alfalfa...............	. 092	. 364	. 014	. 180	— . 007	— . 001	. 067	— . 190	+ . 008	. 474
Amaranthus...........	. 003	. 314	. 011	. 183	— . 001	— . 000	. 055	— . 178	+ . 007	. 607
Average transpiration......	. 078	. 537	. 033	. 176	— . 008	— . 001	. 124	— . 228	+ . 012	. 277

The coefficients of determination are given in Table III. The difference between their sum and unity is given in the last column as $d_{X \cdot O}$, the determination by outstanding factors. As suggested above, the assumption that all the fundamental correlations are linear may involve

some error which would tend to underweight the coefficients of determination between transpiration and the known factors and so overweight the apparent degree of determination by outstanding factors. In certain cases, however, the residue is so small, in one case actually coming out negative, that it is probable that this is not an important source of error. The residual determination is greatest for the crops which were cut twice during the season—namely alfalfa and amaranthus. There were considerable periods following each cutting during which the absolute value of the transpiration was small.

Wind velocity has about the same relative value as a factor in determining transpiration as it has in determining wet-bulb depression. Its relative importance is a little greater for determining evaporation from the shallow tank.

Temperature is somewhat more important than absolute humidity in determining the variations in wet-bulb depression and rate of evaporation from day to day. It is very much the most important factor in determining the rate of transpiration in all the plants.

Radiation is an important factor in evaporation, coming out equal to wind velocity and only slightly less important than absolute humidity. In the plants, on the other hand, it is almost a negligible factor.

Comparing transpiration in the average plant with evaporation in the sun from a shallow tank, we find that the former is influenced relatively much more by temperature, to about the same degree by absolute humidity, somewhat less by wind velocity, and very much less by radiation. The four factors are much more nearly equal in importance in the case of evaporation ($d_{E \cdot T} = 0.30$, $d_{E \cdot H} = 0.19$, $d_{E \cdot W} = 0.16$, $d_{E \cdot R} = 0.16$) than in the case of transpiration ($d_{X \cdot T} = 0.55$, $d_{X \cdot H} = 0.18$, $d_{X \cdot W} = 0.09$, $d_{X \cdot R} = 0.04$). In comparing the importance of these factors it should be added that radiation has an importance somewhat in excess of its direct influence, in that its variations are correlated with those of temperature. Humidity has reduced importance, since, though correlated with temperature, it affects evaporation and transpiration in the opposite direction.

OTHER APPLICATIONS

The method of analysis presented here can readily be applied to the problem of the relative importance of heredity and environment. An application of this kind to the case of the piebald pattern of guinea pigs has already been published (9), and one to the resistance of the same animal to tuberculosis is in press.[1] The method can be applied also to such a problem as the determination of the effects of various systems of mating, such as inbreeding, line breeding, and assortative mating on the genetic composition of an originally random-bred stock.[2]

---

[1] WRIGHT, SEWALL, and LEWIS, PAUL A. FACTORS IN THE RESISTANCE OF GUINEA PIGS TO TUBERCULOSIS WITH SPECIAL REGARD TO INBREEDING AND HEREDITY. *In* Amer. Nat., v. 55. 1921. In press.

[2] WRIGHT, SEWALL. SYSTEMS OF MATING, I TO V. *In* Genetics, v. 6. 1921. In press.

LITERATURE CITED

(1) BRAVAIS, A.

    1846. ANALYSE MATHÉMATIQUE. SUR LES PROBABILITÉS DES ERREURS DE SITUATION D'UN POINT. *In* Mem. Acad. Roy. Sci. Inst. France, Sci. Math. et Phys., t. 9, p. 255–332.

(2) BRIGGS, Lyman J., and SHANTZ, H. L.

    1916. DAILY TRANSPIRATION DURING THE NORMAL GROWTH PERIOD AND ITS CORRELATION WITH THE WEATHER. *In* Jour. Agr. Research, v. 7, no. 4, p. 155–212, 18 fig., 6 pl.

(3) ——— ———

    1916. HOURLY TRANSPIRATION RATE ON CLEAR DAYS AS DETERMINED BY CYCLIC ENVIRONMENTAL FACTORS. *In* Jour. Agr. Research, v. 5, no. 14, p. 583–649, 22 fig., pl. 53–55.

(4) GALTON, Francis.

    1888. CO-RELATIONS AND THEIR MEASUREMENT, CHIEFLY FROM ANTHROPOMETRIC DATA. *In* Proc. Roy Soc. London, v. 45, no. 274, p. 135–145.

(5) ISSERLIS, L.

    1914–15. ON THE PARTIAL CORRELATION RATIO. I–II. *In* Biometrika, v. 10, pt. 2/3, p. 391–411, 1914; v. 11, pt. 1/2. p. 50–66, 1915.

(6) PEARSON, Karl.

    1897. MATHEMATICAL CONTRIBUTIONS TO THE THEORY OF EVOLUTION.—III. REGRESSION, HEREDITY, AND PANMIXIA. *In* Phil. Trans. Roy. Soc. London, s. A., v. 187, 1896, p. 253–318.

(7) ———

    1905. MATHEMATICAL CONTRIBUTIONS TO THE THEORY OF EVOLUTION.—XIV. ON THE GENERAL THEORY OF SKEW CORRELATION AND NON-LINEAR REGRESSION. Drapers' Co. Research Mem. Biom. Ser. 2, 54 p. 3 pl.

(8) WRIGHT, Sewall.

    1918. ON THE NATURE OF SIZE FACTORS. *In* Genetics, v. 3, no. 4, p. 367–374.

(9) ———

    1920. THE RELATIVE IMPORTANCE OF HEREDITY AND ENVIRONMENT IN DETERMINING THE PIEBALD PATTERN OF GUINEA PIGS. *In* Proc. Nat. Acad. Sci., v. 6, no. 6, p. 320–332. 6 fig.