Donald E. Sands
University of Kentucky
Lexington, 40506

Correlation and Covariance

In estimating the uncertainties or errors in calculated results, a frequent implicit assumption is that the variables are independent. That such an assumption may be unwarranted is demonstrated by the following example. At the start of a trip the odometer on a car reads 50000 km. At the end of the trip it reads 50100 km. If the standard deviation on each reading is 1000 km, what is the standard deviation of the distance traveled? The calculation of the distance from the two odometer readings is given by

$$d = t_2 - t_1 = 100 \text{ km} \tag{1}$$

With the assumption of independence, the standard deviation is given by

$$\sigma^2(d) = \sigma^2(t_2) + \sigma^2(t_1) \tag{2}$$

from which $\sigma(d) = 1400$ km. The absurdity of a 1400 km uncertainty in a 100 km trip shows the hazard of using a formula such as eqn. (2) without consideration of its validity. The two odometer readings are certainly not independent, and $\sigma(d)$ had better be determined in some other way—perhaps by assuming a 2% error in all distances measured. This example (which we shall return to in eqn. (22)) illustrates the necessity for being alert to the possibility of non-independent variables. Statistical calculations that neglect correlations often result in incorrect results and erroneous conclusions.

This paper aims to explain the concept of correlation, to describe some of the principal methods of calculating correlation coefficients, and to discuss some applications.

Definitions

When variables are not statistically independent, they are said to be correlated. The covariance of quantities x and y is defined as

$$\text{cov}(x,y) = \langle (x - \bar{x})(y - \bar{y}) \rangle \tag{3}$$

where both the bars and the brackets denote mean values. For the special case of x and y being the same, eqn. (3) produces the variance

$$\text{var}(x,x) = \sigma^2(x) = \text{cov}(x,x) \tag{4}$$

where $\sigma(x)$ is the standard deviation of x. Thus

$$\sigma^2(x) = \langle (x - \bar{x})^2 \rangle \tag{5}$$

The correlation coefficient of x and y, $r(x,y)$, is

$$\text{r}(x,y) = \text{cov}(x,y)/(\sigma(x)\sigma(y)) \tag{6}$$

The quantities defined by eqns. (3), (5), and (6) may be evaluated from lists of corresponding values of x and y. Equation (3) may be reduced to

$$\text{cov}(x,y) = \overline{xy} - \bar{x}\,\bar{y} \tag{7}$$

and eqn. (5) then becomes

$$\sigma^2(x) = \overline{x^2} - \bar{x}^2 \tag{8}$$

Thus, $\sigma^2(x)$ is the difference between the mean of the square of x and the square of the mean of x. In computing $\sigma^2(x)$ from eqn. (5) statistical considerations suggest that the appropriate summation should be divided by the number of degrees of freedom rather than by the number of observations, n. That is

$$\sigma^2(x) = \sum(x - \bar{x})^2/(n - 1) \tag{9}$$

where one has been subtracted from n because one degree of freedom has been used up in computing the mean

$$\bar{x} = \sum x/n \tag{10}$$

To avoid inconsistencies, eqn. (9) should be the basis for calculation rather than eqn. (8), or the result from eqn. (8) should be multiplied by $n/(n - 1)$. In applying eqn. (3), we might be tempted to divide the summations by $n - 2$ since two averages have been computed. However, to obtain consistent and correct correlation coefficients, we should use $n - 1$.

$$\text{cov}(x,y) = \sum(x - \bar{x})(y - \bar{y})/(n - 1) \tag{11}$$

In terms of summations

$$\text{r}(x,y) = \sum(x - \bar{x})(y - \bar{y})/[\sum(x - \bar{x})^2 \sum(y - \bar{y})^2]^{1/2} \tag{12}$$

The Schwarz inequality applied to eqn. (12) shows that

$$1 \geq \text{r}(x,y) \geq -1 \tag{13}$$

A correlation coefficient of +1 implies perfect positive correlation: y increases linearly with x, and if y is plotted versus x all the points fall on a straight line. If $\text{r}(x,y) = -1$, the correlation is perfect but negative: y decreases as x increases, and a plot of y versus x will be linear with a negative slope.

Example

The data in Table 1 will illustrate the use of these equations. We have $\Sigma x = 635$, $\bar{x} = 63.5$, $\Sigma(x - \bar{x})^2 = 2828.5$, $\Sigma y = 684$, $\bar{y} = 68.4$, $\Sigma(y - \bar{y})^2 = 2452.4$, $\Sigma(x - \bar{x})(y - \bar{y}) = 1595$, $n = 10$. For purposes of illustration we have included more significant figures than justified. Now, $\sigma^2(x) = 2828.5/(10 - 1)$, $\sigma(x) =$

Table 1. Scores of Ten Students on Two Examinations

x	y	$x - \bar{x}$	$y - \bar{y}$	y_c
45	47	−18.5	−21.4	58
69	72	5.5	3.6	72
29	51	−34.5	−17.4	49
73	72	9.5	3.6	74
59	73	−4.5	4.6	66
91	94	27.5	25.6	84
72	87	8.5	18.6	73
80	49	16.5	−19.4	78
60	81	−3.5	12.6	66
57	58	−6.5	−10.4	65

Table 2. Correlation Matrices for Examination Grades in Two Freshman Classes

First Semester	1	0.68	0.32	0.58	0.58
	0.68	1	0.17	0.63	0.63
	0.32	0.17	1	0.34	0.36
	0.58	0.63	0.34	1	0.76
	0.58	0.63	0.36	0.76	1
Second Semester	1	0.51	0.50	0.44	0.41
	0.51	1	0.68	0.55	0.59
	0.50	0.68	1	0.57	0.54
	0.44	0.55	0.57	1	0.55
	0.41	0.59	0.54	0.55	1

17.7, $\sigma^2(y) = 2452.4/(10-1)$, $\sigma(y) = 16.5$, $\mathrm{cov}(x,y) = 1595/(10-1) = 177$. Using eqn. (6) for the correlation coefficient, $r(x,y) = 177/(17.7 \times 16.5) = 0.61$; using eqn. (12), $r(x,y) = 1595/(2828.5 \times 2452.4)^{1/2} = 0.61$. This result is a rather high positive correlation. As we would expect, good students generally will do well on most examinations, and poor students rather consistently will be below average.

Table 2 lists the correlation matrices for the examination grades in two semesters of a chemistry class (130–160 students per class). In each semester there were four equally spaced hour examinations plus a comprehensive final examination. In the first semester, examination 3 has rather low correlations with everything else, and its correlation coefficient with examination 2, r(2,3), was only 0.17; this weak correlation may perhaps be attributed to the fact that examination 3 contained a much higher proportion of problems than the others, and problems were anathema to many of the students in this nonscience majors course. Also, some deviation from perfect correlation probably is healthy, in order to ensure that many facets of the students' knowledge and ability get tested during the semester. Another phenomenon that is sometimes revealed by correlation coefficients of examination grades is a fading with time; adjacent examinations may be highly correlated, but the correlations tend to decrease as the separations increase.

Covariance and Independence

If two variables are statistically independent, their covariance and correlation coefficient will be zero. The converse of this may not be true, however. That is, $r(x,y) = 0$ does not imply that x and y are statistically independent. A simple example of such a situation is with the function $y = x^2$, where the x's are symmetrically distributed about the origin (*1*). For this case, the positive x's would contribute to a positive value of $\mathrm{cov}(x,y)$, the negative x's would contribute to a negative value, and over the whole range $\mathrm{cov}(x,y) = 0, r(x,y) = 0$, even though x and y are related by a functional relationship.

Covariance and Causality

Caution must be exercised in drawing inferences about cause and effect from correlation coefficients. A correlation may suggest that possible causative effects should be investigated, as in the case of the relationship between smoking and health, but the mere existence of a nonzero correlation coefficient does not prove that one of the events influences the other. The populations of Miami, Florida and Tulsa, Oklahoma have both increased fantastically in this century, and census figures for 1900, 1950, 1960, and 1970 indicate a correlation coefficient of 0.98. But the increase in one city can hardly be held responsible for the increase in the other. The correlation coefficient between the number of color television sets sold annually and the incidence of poliomyelitis is surely quite negative, but we cannot conclude that TV prevents polio, nor that conquering polio caused TV. Langley (*2*) gives several entertaining examples of the fallacy of inferring cause from correlation.

Continuous Variables

Evaluation of correlation coefficients between continuous variables may be accomplished by replacing the summations by integrals. The average value of the function $f(x)$ over the interval $[a,b]$ is

$$\langle f(x) \rangle = \frac{1}{b-a} \int_a^b f(x)\,dx \tag{14}$$

Equation (14) enables us to evaluate all appropriate averages for calculating

$$r(f(x),g(x)) = (\overline{fg} - \overline{f}\,\overline{g})/[(\overline{f^2} - \overline{f}^2)(\overline{g^2} - \overline{g}^2)]^{1/2} \tag{15}$$

It is instructive to examine the correlation between x and Ax^p, where $p > -1$, over the interval $[0,b]$. The result is

$$r(Ax^p,x) = \sqrt{3(2p+1)}/(p+2) \tag{16}$$

When $p = 1$, the two straight lines have perfect correlation. As p increases, r decreases. This emphasizes that the correlation coefficient, as we have defined it, is a measure of linear correlation; $|r(x,y)| = 1$ implies a straight line relationship between x and y, and the linearity decreases as $|r(x,y)|$ deviates from 1.

Covariance Matrix from Least Squares

A powerful result of statistical theory is that the method of least squares provides a way of estimating the matrix of variances and covariances of the set of parameters being determined. Briefly, applying the method of least squares to fitting the equation $y = f(x)$ requires solving the matrix equation

$$\mathbf{A}\vec{\mathbf{u}} = \vec{\mathbf{v}} \tag{17}$$

for the vector $\vec{\mathbf{u}}$ which gives the parameter shifts. Matrix **A** consists of weighted sums of products of the partial derivatives of $f(x)$ with respect to the parameters, and each element of vector $\vec{\mathbf{v}}$ is the weighted sum of the product of a partial derivative and the corresponding value of $y_o - y_c$, where y_o and y_c are the observed and calculated values of y. Reference (*1*) gives a thorough treatment of the matrix formulation of the method of least squares, and references (*3*) and (*4*) discuss the principles of proper weighting. The matrix of variances and covariances, **V**, is given by

$$\mathbf{V} = \mathbf{A}^{-1} \sum w(y_o - y_c)^2/(n-m) \tag{18}$$

where $\mathbf{A}^{-1}$ is the inverse of **A**, w is a weighting factor, n is the number of observations, and m is the number of parameters being adjusted.

The elements of the correlation matrix, **R**, may be obtained from eqn. (18) by means of eqn. (6). **R** then depends upon **A** and, except for the weights, **A** depends only upon the values of the independent variables (x in $y = f(x)$), so the correlation matrix obtained from least-squares methods is determined principally by the choice of points x at which the measurements y are to be carried out. In fitting an equation of the form

$$y = \alpha + \beta x \tag{19}$$

a value of $r(\alpha,\beta)$ very close to -1 can result if the x's are all positive and span a narrow range, as is the case, for example, if $x = 1/T$. This effect can produce apparent linear relationships between parameters, and such relationships have long been known for the parameters of the Arrhenius equation and related equations (*5*). Highly correlated parameters can lead to refinement problems and indeterminacies in the interacting parameters (*6*).

Transformations

Once the variances and covariances of a set of quantities have been determined, they can be used to calculate the variances and covariances of other quantities. The equation

$$\mathrm{cov}(f_k,f_l) = \sum \frac{\partial f_k}{\partial x^i} \frac{\partial f_l}{\partial x^j} \mathrm{cov}(x^i,x^j) \tag{20}$$

permits calculation of the covariance of functions f_k and f_l from the covariances of x^i and x^j; the summation is over all i and j (*7*). In matrix terminology this is

$$\mathbf{V}' = \mathbf{D}\mathbf{V}\overline{\mathbf{D}} \tag{21}$$

where $\partial f_i/\partial x^j$ is the ijth element of **D**, and $\overline{\mathbf{D}}$ is the transpose of **D**. If we apply eqn. (20) to the problem of eqn. (1), we obtain the corrected form of eqn. (2)

$$\sigma^2(d) = \sigma^2(t_2) + \sigma^2(t_1) - 2\,\mathrm{cov}(t_1,t_2) \tag{22}$$

and eqn. (22) will give a sensible answer for $\sigma(d)$ in this problem with $\sigma(t_2) = \sigma(t_1) = 1000$ if $r(t_2,t_1) = 1 - (2 \times 10^{-6})$.

As another application, suppose we make measurements of two of the angles of a triangle. If $A = 50.0°$, $\sigma(A) = 0.3°$, B

$= 55.0°$, $\sigma(B) = 0.4°$, the third angle is

$$C = 180.0° - A - B \tag{23}$$

and, since A and B are independent

$$\sigma^2(C) = \sigma^2(A) + \sigma^2(B) \tag{24}$$

so $C = 75.0°$, $\sigma(C) = 0.5°$. Now, let us use A, B, and C to calculate $S = A + B + C$ and then $\sigma(S)$. Obviously, $S = 180.0°$, and $\sigma(S)$ must come out to be identically 0. We have

$$\sigma^2(S) = \sigma^2(A) + \sigma^2(B) + \sigma^2(C) + 2\,\mathrm{cov}(A,C) + 2\,\mathrm{cov}(B,C) \tag{25}$$

We can find the covariances by applying eqn. (20) or eqn. (21). If $f_1 = A$ and $f_2 = 180 - A - B$, we obtain $\mathrm{cov}(A,C) = -\sigma^2(A)$ and with our numbers $\mathrm{r}(A,C) = -0.6$. If $f_1 = B$ and $f_2 = 180 - A - B$, the result is $\mathrm{cov}(B,C) = -\sigma^2(B)$, $\mathrm{r}(B,C) = -0.8$. Equation (25) now gives $\sigma^2(S) = 0$, as it must.

Hypothesis Testing

The probability distribution of the correlation coefficient r is rather complex. However, if the number of observations, n, is large (greater than about 50) the statistic

$$Z = \frac{1}{2}\ln\frac{1+r}{1-r} \tag{26}$$

is approximately normally distributed with mean 0 and variance $1/(n-3)$. For example, suppose we have $r = 0.17$ with $n \approx 130$, (one of the entries in Table 2). Let us test the hypothesis that r differs from 0 only by chance; that is, how likely is it that examinations 2 and 3 are really independent? From eqn. (26), $Z = 0.172$, and the variance is 0.0079. Since $\sigma = \sqrt{0.0079} = 0.089$, we are asking how likely it is that the observed Z be 1.93 standard deviations from 0. This probability is

$$p = 1 - \frac{1}{\sqrt{2\pi}\sigma}\int_{-1.93\sigma}^{1.93\sigma} e^{-x^2/2\sigma^2}\mathrm{d}x \tag{27}$$

Appropriate tables of the error function show that $p = 0.05$. That is, there is about a 5% probability that Z be this far from 0 by chance alone. Note that this is a two-sided probability; in about 2.5% of the cases where r is really 0 the experimental value of Z will exceed 0.172, and in another 2.5% Z will be below −0.172. (Langley (2) presents a lucid discussion of the distinction between one-sided and two-sided probabilities).

Problems with fewer than about 50 observations may be treated by applying the Student t test with $n - 2$ degrees of freedom to

$$t = r[(n-2)/(1-r^2)]^{1/2} \tag{28}$$

Further details on these tests may be found in one of many excellent textbooks on statistics (*1, 2, 8, 9*).

Regression and Prediction

The results of fitting sets of corresponding values of x and y to linear relationships by the method of least-squares may be expressed by the linear *regression* equations

$$y_c = \bar{y} + (x - \bar{x})r\sigma(y)/\sigma(x) \tag{29}$$

$$x_c = \bar{x} + (y - \bar{y})r\sigma(x)/\sigma(y) \tag{30}$$

where r is the correlation coefficient between x and y. The quantities y_c and x_c that may be calculated from these equations may be interpreted as predicted values: for each observed x eqn. (29) provides a prediction of y, and for each observed y eqn. (30) predicts x. Equations (29) and (30) correspond to the same straight line if and only if x and y are perfectly correlated, in which case $r = +1$ or $r = -1$ and all of the points lie on the straight line. Let us now define

$$\sigma^2(y_c) = \langle(y - y_c)^2\rangle \tag{31}$$

That is, $\sigma^2(y_c)$ measures the dispersion or scatter of the y values about y_c, in contrast with $\sigma^2(y)$, which measures the scatter about $\bar{y}$. If y_c from eqn. (29) is substituted into eqn. (31), evaluation of the averages leads to

$$\sigma^2(y_c) = (1 - r^2)\sigma^2(y) \tag{32}$$

Similarly,

$$\sigma^2(x_c) = (1 - r^2)\sigma^2(x) \tag{33}$$

In these equations, r^2 represents the fractional decrease in variance resulting from using the predicted values rather than the means. Comparisons of σ's rather than variances may be carried out with the *coefficient of predictive efficiency* (*9*)

$$E \approx 100(1 - \sqrt{1 - r^2}) \tag{34}$$

The utility of these equations may be illustrated by returning to the data of Table 1, where the two examinations x and y have respective means 63.5 and 68.4, variances 314 and 272, and correlation coefficient 0.61. After having given examination x, how well can we predict how our students will do on examination y? The means and variances are quantities that are subject to some control, either by carefully designing the examinations or by curving the results to fit some desired distribution. The predicted score for an average student on examination y is 68.4, and the laws of probability tell us that in a large random sample 68.3% of the observations will be within σ of the mean. Therefore, we should expect 68.3% of our students to score between 52 and 85 on the second examination. When we look at individual students, for whom we have tabulated first examination scores, better predictions than $\bar{y}$ can be obtained from eqn. (29), which gives

$$y_c = 68.4 + 0.564(x - 63.5) \tag{35}$$

The values of y_c generated by eqn. (35) are given in the last column of Table 1. According to eqn. (32), $\sigma^2(y_c) = 172$, $\sigma(y_c) = 13.1$, and y_c will be within 13.1 of y in 68.3% of the cases in a large normally distributed sample. The improvement in going from $\sigma(y)$ to $\sigma(y_c)$ is given by eqn. (34) as 20.4%.

Bivariate Distribution

The joint distribution of two normally distributed correlated variables is

$$\phi(x,y) = \frac{1}{2\pi\sqrt{1-r^2}\sigma(x)\sigma(y)}\exp -\frac{1}{2(1-r^2)}[(x-\bar{x})^2/\sigma^2(x) + (y-\bar{y})^2/\sigma^2(y) - 2(x-\bar{x})(y-\bar{y})r/\sigma(x)\sigma(y)] \tag{36}$$

The probability that x is between x and $x + \mathrm{d}x$ and that simultaneously y is between y and $y + \mathrm{d}y$ is given by $\phi(x,y)\mathrm{d}x\,\mathrm{d}y$. When r is 0 this is just the product of the independent probabilities, but no such factorization is possible in the case of correlated variables. Equations analogous to eqn. (36) for more than two variables are most conveniently expressed in matrix notation (*1*).

Integrations of eqn. (36) over intervals corresponding to equal probability ranges for uncorrelated variables lead to tables that provide practical insight into the meaning of correlation. If, for example, we use integration limits of $[-\infty, -0.6745]$, $[-0.6745, 0]$, $[0, 0.6745]$, and $[0.6745, \infty]$, each of these integrations of a single normal distribution would give 0.25 since these intervals divide the area under the normal distribution curve into four equal parts. These limits applied to eqn. (36), however, lead to tables such as those in Table 3. For purposes of illustration, let us examine the table labeled $r = 0.4$. Suppose we have a method of prediction that enables us to sort items into four groups. Our predictor may, for example, be an examination administered to a group of students, and we thereby divide the students into four equally populated groups or quartiles. We now examine the efficacy of this examination by observing the performance of these students on some other measure of success. If there were no correlations between these two measures, all the entries in the tables would be 0.25. If there were perfect correlation, each student would be in the same quartile on the subsequent test as on the orig-

Table 3. Predictions and Results for Various Correlation Coefficients

$r = 0.10$	Prediction 1	2	3	4
Result 1	0.291	0.259	0.239	0.211
2	0.259	0.254	0.248	0.239
3	0.239	0.248	0.254	0.259
4	0.211	0.239	0.259	0.291

$r = 0.20$	Prediction 1	2	3	4
Result 1	0.335	0.267	0.225	0.173
2	0.267	0.259	0.248	0.225
3	0.225	0.248	0.259	0.267
4	0.173	0.225	0.267	0.335

$r = 0.40$	Prediction 1	2	3	4
Result 1	0.429	0.277	0.191	0.103
2	0.277	0.280	0.253	0.191
3	0.191	0.253	0.280	0.277
4	0.103	0.191	0.277	0.429

$r = 0.50$	Prediction 1	2	3	4
Result 1	0.481	0.278	0.168	0.072
2	0.278	0.296	0.258	0.168
3	0.168	0.258	0.296	0.278
4	0.072	0.168	0.278	0.481

$r = 0.60$	Prediction 1	2	3	4
Result 1	0.538	0.276	0.141	0.044
2	0.276	0.319	0.264	0.141
3	0.141	0.264	0.319	0.276
4	0.044	0.141	0.276	0.538

$r = 0.80$	Prediction 1	2	3	4
Result 1	0.676	0.252	0.066	0.006
2	0.252	0.411	0.272	0.066
3	0.066	0.272	0.411	0.252
4	0.006	0.066	0.252	0.676

$r = 0.90$	Prediction 1	2	3	4
Result 1	0.764	0.208	0.019	0.001
2	0.208	0.524	0.248	0.019
3	0.019	0.248	0.524	0.208
4	0.001	0.019	0.208	0.764

$r = 0.95$	Prediction 1	2	3	4
Result 1	0.839	0.158	0.003	0.000
2	0.158	0.643	0.196	0.003
3	0.003	0.196	0.643	0.158
4	0.000	0.003	0.158	0.839

inal and the tables would consist of 1.0 in each diagonal position and 0.0 elsewhere. For our $r = 0.4$ case, of the students who were in the upper quartile on the first test, 42.9% were again in the upper quartile on the second test, 27.7% in the next quartile (a total of 70.6% in the upper half), 19.1% in the third quartile down, and 10.3% at the bottom. Of those placed by the predictor in the next highest quartile, 27.7% subsequently ended up at the top, another 28.0% ended up in the quartile predicted, 25.3% fell down one quartile, and 19.1% were at the bottom. Of the students placed in the upper half by the predictor, 63% were in the upper half of the later measure. These comparisons also apply at the other end of the scale, and 63% of those starting out in the lower half remained in the lower half. That 37% switched halves should provide a warning against complacency for those who started high and a note of encouragement for those who started low.

Educational applications of these tables have been discussed by Thorndike and Hagen (*10*).

Correlations between Subsets

In order to illustrate that $r(x,y) = 0$ does not imply statistical independence, we considered the function $y = x^2$ with x symmetrically distributed about 0. If, on the other hand, we restrict x to positive values, the correlation coefficient according to eqn. (16) is $\sqrt{15}/4$. This demonstrates a hazard in using and interpreting correlation coefficients when only a portion of the data is available. The correlations between subsets are not necessarily even close to those between the complete sets of data.

As another example of this effect, consider the following hypothetical situation. Suppose that Middling University rates prospective graduate students by assigning $G = +1$ for good grades, $G = -1$ for poor grades, $M = +1$ for high motivation, and $M = -1$ for low motivation. If a student eventually completes the program and obtains an advanced degree, he is assigned $S = +1$; if he fails he gets $S = -1$. Let us assume that 90% of the (+,+) students (those who are high on both grades and motivation) will succeed, 50% of the (+,−) or (−,+) students will succeed, and 10% of the (−,−) students will succeed. The correlation coefficient, $r(S,G)$, for the entire set of data is easily worked out and is found to be 0.40. If, however, all the (+,+) students go to Harvard and the (−,−) students do not go to graduate school at all, the Middling University (+,−) and (−,+) sample would give $r(S,G) = r(S,M) = 0$, and the Middling faculty and administration might conclude that neither grades nor motivation has anything to do with potential success in graduate school. Harvard, with nothing but (+,+) students, would also observe no correlation.

The reliability of these statistical prognostications can be enhanced by using a combination of predictors. If, in the example above, P is a linear combination of G and M, chosen so as to maximize $r(P,S)$, it turns out that

$$r^2(P,S) = \frac{r^2(G,S) + r^2(M,S) - 2r(G,M)r(G,S)r(M,S)}{1 - r^2(G,M)} \quad (37)$$

and $r(P,S)$ for the complete data set may be as high as 0.57 (if $r(G,M) = 0$). The (+,−) and (−,+) subset will give $r(P,S) = 0$, again failing to detect correlation.

These examples are intended to show the need for caution in drawing inferences from data. Predictions involving any human activity are especially subject to great uncertainty, and much research is still needed in these areas. Some of the problems of establishing reliable admission standards have been discussed by Willingham and by Dawes (*11*). Efforts to evaluate teaching quality have been the subject of lively controversy (*12*).

Mean of Correlated Variables

The weighted mean of a set of correlated variables is given by

$$\bar{x} = \sum w_i x_i / \sum w_i \quad (38)$$

The "best" value of $\bar{x}$ is that which has the minimum variance, and the minimum $\sigma^2(\bar{x})$ is obtained in the case of uncorrelated variables by choosing the weights as

$$w_i = 1/\sigma_i^2 \quad (39)$$

where σ_i^2 is the variance of x_i. If the variables are correlated, minimization of $\sigma^2(\bar{x})$ leads to the matrix equation

$$\vec{w} = \mathbf{V}^{-1}\vec{v} \quad (40)$$

where $\vec{w}$ is the column vector of the weights, $\mathbf{V}^{-1}$ is the reciprocal of the covariance matrix of the x's, and each element of $\mathbf{v}$ is 1 (*7*). If there are just two variables, eqn. (*40*) gives

$$w_i = 1/(\sigma^2(x_i) - \mathrm{cov}(x_1,x_2)) \quad (41)$$

As another example, consider determining the mean of a large number of measurements of a quantity. Let us assume that these measurements all have variance V, and that the covariances between the measurements are all equal to c. That is, all diagonal elements of matrix $\mathbf{V}$ are equal to V, and all off-diagonal elements are c. The weights are all equal in this case, and the variance of $\bar{x}$ is

$$\sigma^2(\bar{x}) = [V + (n-1)c]/n \quad (42)$$

For the case of uncorrelated observations this reduces to the usual formula that shows how the variance improves with increasing n. With correlated observations, however, $\sigma^2(\bar{x})$ does not approach zero; the limit of $\sigma^2(\bar{x})$ as n approaches infinity is c.

Entropy of Correlation

The information theory definition of entropy (*13*) is

$$S = -k\sum p \ln p \tag{43}$$

where k is a constant that determines the units and p is a probability. For a continuous probability distribution, the summation is replaced by integration

$$s = -k\int p \ln p \mathrm{d}x \tag{44}$$

If, in particular, p is the normal distribution function, eqn. (44) yields

$$S = \frac{1}{2} k \ln (2\pi e \sigma^2) \tag{45}$$

While this so-called Shannon entropy is not necessarily identical with the usual thermodynamic or Gibbs entropy (*14*), the correspondences are sufficient to be useful; it is enlightening to think of entropy as related to the dispersion or spread of a probability distribution, which is a measure of lack of information. With more than one variable, the additive property of entropy applies, and if the variables are independent and normally distributed

$$S = \frac{1}{2} k\sum \ln (2\pi e \sigma_i^2) \tag{46}$$

If the variables are correlated, a set of new uncorrelated variables may be obtained by diagonalizing the covariance matrix, and the entropy will be given by eqn. (46) applied to the new variances $\sigma_i'^2$. When the $\sigma_i'^2$ are expressed in terms of the original covariance matrix, the result is

$$S = \frac{1}{2} k\sum \ln (2\pi e \sigma_i^2) + \frac{1}{2} k \ln |R| \tag{47}$$

where $|R|$ is the determinant of the correlation matrix. The term $\frac{1}{2}k \ln |R|$ represents an entropy of correlation. Since $|R| < 1$, entropy of correlation is negative; the correlations between variables provide an increase in information and a concomitant decrease of entropy (*15*).

Literature Cited

(1) Hamilton, W. C., "Statistics in Physical Science," The Ronald Press Co., New York, **1964.**
(2) Langley, R., "Practical Statistics Simply Explained," Dover Publications, Inc., New York, **1971.**
(3) Sands, D. E., J. CHEM. EDUC., **51,** 473 (1974).
(4) Christian, S. D., Lane, E. H., and Garland, F., J. CHEM. EDUC. **51,** 475 (1974).
(5) Fairclough, R. A., and Hinshelwood, C. N., *J. Chem. Soc.*, **538,** 1573 (1937); Waring, C. E., and Becher, P., *J. Chem. Phys.*, **15,** 488 (1947); Misra, B. N., and Varshni, Y. P., *J. Chem. Eng. Data*, **6,** 194 (1961).
(6) Geller, S., *Acta Crystallogr.*, **14,** 1026 (1961).
(7) Sands, D. E., *Acta Crystallogr.*, **21,** 868 (1966).
(8) Mendenhall, W., and Scheaffer, R. L., "Mathematical Statistics with Applications," Duxbury Press, North Scituate, Mass., **1973;** Spiegel, M. R., "Statistics," Schaum's Outline Series, McGraw-Hill Book Co., New York, **1961.**
(9) Senter, R. J., "Analysis of Data," Scott, Foresman and Co., Glenview, Illinois, **1969.**
(10) Thorndike, R. L., and Hagen, E., "Measurement and Evaluation in Psychology and Education," 3rd Ed., John Wiley and Sons, New York, **1969.**
(11) Willingham, W. W., *Science*, **183,** 273 (1974); Dawes, R. M., *Science*, **187,** 721 (1975).
(12) Rodin, M., and Rodin, R., *Science*, **177,** 1164 (1972); Gessner, P. K., *Science*, **180,** 566 (1973); Frey, P. W., *Science*, **182,** 83 (1973); Rodin, M., *Science*, **187,** 555 (1975); Frey, P. W., *Science*, **187,** 557 (1975); Gessner, P. K., *Science*, **187,** 558 (1975).
(13) Landsberg, P. T., "Thermodynamics," Interscience Publishers, New York, **1961.**
(14) Jaynes, E. T., *Amer. J. Phys.*, **33,** 391 (1965); Lindblad, G., *J. Stat. Phys.*, **11,** 231 (1974); Liboff, R. L., *J. Stat. Phys.*, **11,** 343 (1974).
(15) Kullback, S., "Information Theory and Statistics," Dover Publications, New York, **1968,** p. 199.