SCIENTIFIC SECTION

THE APPLICATION OF STATISTICAL METHODS TO PHARMACEUTICAL RESEARCH. II. CORRELATION COEFFICIENTS.*

BY JAMES C. MUNCH.

The correlation coefficient, r, is a measure of the degree or extent of co-relationship. It measures the proportional change in one variable of a pair when the other variable undergoes successive changes in magnitude. It is an abstract number which ranges in magnitude from 1.00 (perfect relationship) to 0.00 (complete lack of relationship). Between these limits, closer degrees of relationship are denoted by larger values of the coefficient. When one variable increases in value as the other increases, the relation is direct and the sign of the correlation coefficient is positive. When one variable decreases as the other increases, the correlation is inverse and the coefficient has a negative sign (2, 3, 4, 6, 9). The certainty of correlation is not directly proportional to the magnitude of the correlation coefficients. A correlation coefficient of 0.90 is eleven times as certain as a coefficient of 0.30 (6).

The first step in the determination of the correlation coefficient should be the plotting of the experimental results. The most accurate or most dependable value is taken as the abscissa, the other value of the pair as the ordinate. Inspection of the scatter diagram thus obtained indicates at once whether the degree of relationship is high or low, and whether the correlation is linear or not. The correlation coefficient, r, is usable only when linear correlation is indicated. Non-linear correlation may be calculated by appropriate methods; since they are of more limited use they will not be discussed in this paper.

Partial correlation coefficients are calculated to determine the degree of co-relationship between two variables when the effect of one or more additional variables is made constant. They tend to reveal spurious relationships (6, 9, 10).

The application of correlation coefficients reveals the extent of relationship directly and with mathematical certainty. As an example of the application of this type of procedure, the results obtained in the assay of a series of ouabain and strophanthus preparations by the M. L. D. frog and by the Knudsen and Dresbach colorimetric methods (8) are given in Table I. The values for the M. L. D. frog method are given in the first column. The mean of these values is 96.5. The deviation of each value in the first column from the mean is recorded in the second column under the heading "X" and the square of X is given in the third column. The sum of the squares is 80,683. When this sum is divided by "N," the numbers of pairs of observations (28 in this instance), the average value of the squares is found to be 2882. The square root of this value, which is 53.7, is the standard deviation, σ.

The fourth column contains the values for the colorimetric assay of the same samples and the next two columns the deviations and squares of the deviations of "Y." The sum of the squares of the deviations is 93,026, the average 3322 and σ 57.6.

* Scientific Section, A. PH. A., Baltimore meeting, 1930.

TABLE I.—CORRELATION BETWEEN M. L. D. FROG AND COLORIMETRIC ASSAYS OF OUABAIN AND STROPHANTHUS.

	M. L. D. frog.	X.	X^2.	Colorimetric assays.	Y.	Y^2.	XY. Plus.	XY. Minus.
	25	−71	5041	26	−84	7056	5964	..
	29	−67	4489	25	−85	7225	5695	..
	31	−65	4225	33	−77	5929	5055	..
	31	−65	4225	40	−70	4900	4550	..
	50	−46	2116	50	−60	3600	2760	..
	61	−35	1225	78	−32	1024	1120	..
	61	−35	1225	100	−10	100	350	..
	61	−35	1225	140	30	900	...	1050
	64	−32	1024	71	−39	1521	1248	..
	67	−29	841	75	−35	1225	1015	..
	73	−23	529	100	−10	100	230	..
	73	−23	529	100	−10	100	230	..
	80	−16	256	75	−35	1225	560	..
	85	−11	121	100	−10	100	110	..
	90	− 6	36	90	−20	400	120	..
	100	4	16	100	−10	100	...	40
	110	14	196	100	−10	100	...	140
	120	24	576	100	−10	100	...	240
	120	24	576	120	10	100	240	..
	120	24	576	135	25	625	600	..
	120	24	576	135	25	625	600	..
	120	24	576	200	90	8100	2160	..
	130	34	1156	144	34	1156	1156	..
	130	34	1156	155	45	2025	1530	..
	145	41	2401	143	33	1089	1617	..
	155	51	3481	150	40	1600	2360	..
	155	51	3481	215	105	11025	6195	..
	293	197	38809	286	176	30976	34492	..
Sum	2699	..	80683	3086	..	93026	79957	1470
Mean	96.5	..	2882	110.2	..	3322	...	..
σ	....	..	53.7	....	..	57.6	...	..

$$r = \frac{\Sigma XY}{N\sigma_X\sigma_Y} = \frac{79{,}957 - 1470}{28 \times 53.7 \times 57.6} = +0.90$$

$$PE_r = 0.6745 \frac{1 - r^2}{\sqrt{N}} = \pm 0.02$$

The last two columns, headed "XY," are obtained by multiplying each deviation of X by the corresponding deviation of Y, noting the arithmetic signs of the deviations. These are assembled in columns of plus and minus products, giving sums of plus 79,757 and minus 1470, respectively. The formula for the calculation of the correlation coefficient is

$$r = \frac{\Sigma XY}{N\sigma_X\sigma_Y}$$

in which ΣXY represents the sum of the products of the deviations of X and Y from their respective means, N is the number of pairs of observations, and σ_X and σ_Y are the standard deviations of X and Y. Substituting the observed values in this equation

$$r = \frac{79{,}957 - 1470}{28 \times 53.7 \times 57.6} = +0.90.$$

The probable error of r is obtained by substitution in the formula

$$PE_r = 0.6745 \times \frac{1 - r^2}{\sqrt{N}} = 0.6745 \times \frac{1 - (0.90)^2}{\sqrt{28}} = \pm 0.02.$$

The value for r is positive, which indicates that there is a change in the value of X in the same direction with any change in the value of Y. The absolute value, 0.90, is large, indicating the probability that significant correlation is present. The value for r is 0.90/0.02 or 45 times its probable error, which demonstrates that the correlation coefficient is statistically significant (1, 2, 3, 6, 9).

The regression equation, a linear equation representing the change in one variable corresponding to a unit change in the other, is calculated from the equation

$$Y - Y' = r \frac{\sigma_X}{\sigma_Y} (X - X')$$

in which Y' and X' are the means of the observed values for X and Y. Substitution in this instance gives the equation

$$Y - 110.2 = 0.90 \times \frac{53.7}{57.6} X \ (X - 96.5).$$

This simplifies to

$$Y = 0.84\, X + 29.2.$$

In connection with the investigation from which these data were taken, comparisons were also made of the results obtained in the assay of tinctures of digitalis by the M. L. D. frog and the colorimetric methods (8). The correlation coefficient for these determinations was found to be +0.71 ± 0.05. This correlation coefficient is also positive indicating direct relationship. The numerical value is lower, indicating a lower degree of relationship. However, it is 0.71/0.05 or 14 times its probable error, which indicates that it is statistically certain. The regression equation is

$$Y - 194 = 0.71 \times \frac{59.1}{71} \times (X - 147)$$

which simplifies to

$$Y = 0.59\, X + 107.4.$$

In determining the relationship between the results of analysis of oil of chenopodium (7) a series of values for r was obtained. It was desired to determine whether these were real or spurious relationships. For this purpose partial correlation coefficients (6, 7, 9, 10) were calculated by the formula

$$r_{12\cdot 3} = \frac{r_{12} - r_{13}r_{23}}{(1 - r_{13}^2)(1 - r_{23}^2)}$$

in which $r_{12.3}$ represents the correlation between variables 1 and 2 when the effect of variable 3 upon each of them is made constant; r_{12}, r_{23} represent the correlation between the indicated pairs of variables. This same formula may be extended to stabilize the effect of two or more variables in the same way (Table II).

The correlation between ascaridol and specific gravity of oil of chenopodium was found to be 0.89 ± 0.02, which indicates definite relationship. When the rotation or the index of refraction were fixed, values of 0.89 and 0.87 were obtained, showing that these variables had no effect upon the observed relationship. When the alcohol solubility was fixed, the value for r decreased to 0.76, indicating that this variable had an effect upon the observed relationship. The correlation of ascaridol content to alcohol solubility was minus 0.73 and of specific gravity to alcohol solubility was minus 0.74. In other words, about the same degree of relationship was noted between these variables. When the rotation, the index of refraction and the alcohol solubility were all fixed, the correlation between ascaridol content and specific gravity was found to be 0.79. This indicates that the observed correlation is real and not spurious.

TABLE II.—CORRELATION COEFFICIENTS OF OIL OF CHENOPODIUM ASSAYS.

Asc:SpGr	0.89	Asc:Alc/Rot	−0.66
Asc:SpGr/Rot	0.89	Asc:Alc/nD	−0.66
Asc:SpGr/nD	0.87	Asc:Alc/Rot, nD	−0.50
Asc:SpGr/Alc	0.76		
Asc:SpGr/Rot, nD	0.85	SpGr:Rot	−0.60
Asc:SpGr/Rot, Alc	0.80		
Asc:SpGr/nD, Alc	0.77	SpGr:nD	−0.42
Asc:SpGr/Rot, nD, Alc	0.79	SpGr:nD/Rot	−0.69
Asc:Rot	−0.38	SpGr:Alc	−0.74
Asc:Rot/SpGr	0.42	SpGr:Alc/Asc	−0.29
Asc:Rot/nD	−0.59	SpGr:Alc/Rot	−0.77
Asc:Rot/Alc	−0.33	SpGr:Alc/nD	−0.67
		SpGr:Alc/Rot, nD	−0.61
Asc:nD	−0.40		
Asc:nD/SpGr	−0.06		
Asc:nD/Rot	−0.53	Rot:nD	−0.21
Asc:nD/Alc	0.02	Rot:Alc	0.23
Asc:Alc	−0.73	nD:Alc	0.56
Asc:Alc/SpGr	−0.23	nD:Alc/Rot	0.61

Asc represents ascaridol
SpGr represents specific gravity
Rot represents specific rotation
nD represents index of refraction
Alc represents volumes of 70 per cent alcohol for solution.

CONCLUSIONS.

1. The correlation coefficient, r, is a definite statistical measure of the degree of co-relationship.
2. It is only significant of linear correlation.
3. Partial correlation coefficients serve to demonstrate whether an observed relationship is real or is spurious.

BIBLIOGRAPHY.

(1) W. Chauvenet, A treatise on the method of least squares (1868).
(2) C. B. Davenport, Statistical methods, with special reference to biological variation, 3rd Edition (1914).
(3) W. P. Elderton, Frequency curves and correlation.
(4) R. A. Fisher, Statistical methods for research workers, 2nd Edition (1928).
(5) M. Merriman, A textbook on the method of least squares (1913).

(6) James C. Munch, "Bioassays" (1931).

(7) J. C. Munch and W. F. Reindollar, A statistical study of the chemical constants of Oleum Chenopodii, Paper read at Baltimore meeting, A. PH. A., 1930.

(8) L. W. Rowe, The colorimetric assay of strophanthus, JOUR. A. PH. A., 16 (1927), 113–115; The colorimetric assay of digitalis, *Ibid.*, pages 510–516.

(9) K. Pearson, On the criterion that a given system of deviation from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling, *Phil. Mag.*, 5 (1900), 157–175.

(10) H. R. Tolley, The theory of correlation as applied to farm-survey data on fattening baby beef, *U. S. Dept. Agr. Bull.*, 504 (1917).

PROPADIENE.*

BY W. A. LOTT AND W. G. CHRISTIANSEN.

L. K. Riggs predicted that due to the presence of two double bonds in propadiene $CH_2{=}C{=}CH_2$, the gas should have very high anesthetic potency; and that since the molecular weight is low, it should not be highly toxic. His studies indicated that the gas is disappointingly toxic.[1]

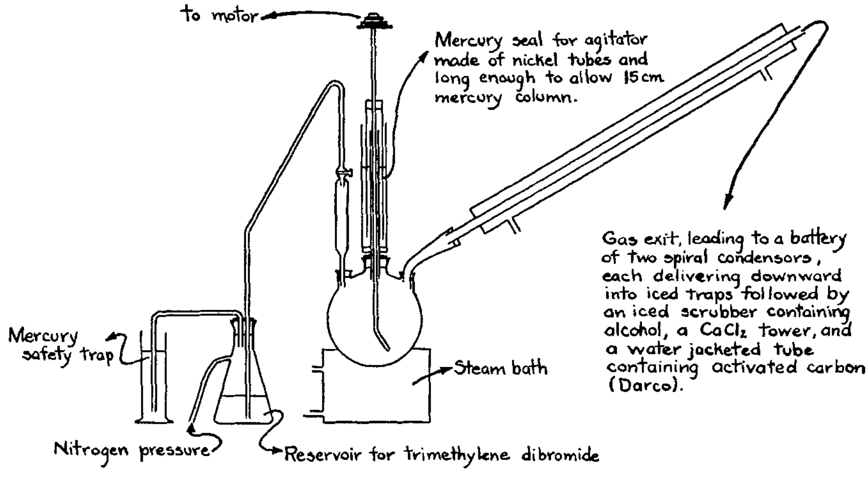

Fig. 1.

It seemed to the present authors that the method by which the material used in the above work was prepared did not adequately provide for the removal of vapors of 2,3-dibrompropene, and that the violent nervous symptoms noted by Dr. Riggs might be avoided by carefully removing these vapors from the propadiene.

In this laboratory we had already found that when propylene is prepared from propylene dibromide, incomplete removal of the propylene dibromide vapors causes the gas to be less satisfactory for anesthetic purposes. Likewise, in the case of cyclopropane we had much better results from that gas which we had rigidly purified from trimethylene dibromide.

---

* Scientific Section, A. PH. A., Baltimore meeting, 1930.

[1] *Proc. Soc. Experimental Biology & Medicine,* 22 (1925), 269, and private communications.