NOTES

A COEFFICIENT OF CORRELATION BETWEEN SCHOLARSHIP AND SALARIES

INTRODUCTION

Some might doubt that it is correct to apply a coefficient of correlation to show the relationship between scholarship and salaries. This coefficient can be trusted to give at least a rough approximation, which is all that is necessary in the inexact science of vocation. It is fictitious accuracy to be too finical in the application of formulas. Therefore, a coefficient of correlation between scholarship and salaries is a valuable part of human knowledge.

Would it be worth while to find this coefficient if it is based upon the experience of the American Telegraph and Telephone Company? Since the employment practices of this company are not representative of the employment practices of business at large, one might doubt the validity of drawing general conclusions from such specialized data. The coefficient for business at large is probably less than the coefficient for the Bell System; the value of this knowledge is enhanced if we know the latter coefficient. Since this company is very large, a coefficient between scholarship and salaries would be valuable, even if this coefficient applies only to the Bell System and to other companies having approximately the same employment practices.

An article[1] by Mr. Walter S. Gifford, President of the Bell System, contains a discussion of some of the relationships between scholarship and salaries. President Gifford, however, did not determine in the case of the Bell System a coefficient of correlation between scholarship and salaries.

The purpose of this article is not a new contribution to statistical method, but is an application of the method[2] of finding the coefficient of correlation when the two variables have not been quantitatively measured. This method will be applied to the chart on page 672 of President Gifford's article, in order to determine for the Bell System the coefficient of correlation between scholarship and salaries.

FINDING THE COEFFICIENT OF CORRELATION

An explanation of the chart. It is based on the experience of 2,144 Bell System employees over five years out of college. First, assume these employees

[1] It is entitled "Does Business Want Scholars?" and was printed in the May 1928 issue of Harper's Magazine.

[2] It can be found in Elderton's "Frequency Curves and Correlation."

are grouped according to their grades in college. In the high scholarship group put those who graduated in the highest third of their classes. The middle and low scholarship groups are formed in like manner. Secondly, suppose the same employees are divided into three equal groups according to their salaries. Then, the salary of any one of the employees would be high, middle, or low.

Assume a hypothetical group of 300 employees who are college graduates. Suppose that the scholarship of 100 of them was high, that the scholarship of 100 of them was middle, and that the scholarship of the others was low. Also assume that the salary experience of these 300 employees is the same as that of the 2,144 employees of the Bell System.

The 300 employees can be grouped according to the following table.

TABLE NO. 1

Salary	Scholarship			Totals
	Low	Middle	High	
High	22	24	48	94
Middle	31	39	27	97
Low	47	37	25	109
Totals	100	100	100	300

This table can be combined as follows.

TABLE NO. 2

Salary	Scholarship	
	Low & Middle	High
High	c	d
Middle & Low	a	b

Then, $c = 46$, $a = 154$, $d = 48$, and $b = 52$. Assume $N = 300$.

Assume x is a function of grades received in college. Suppose y is a function of salaries received. Assume that the frequencies x and y both follow the normal curve of error whose standard deviation is equal to one. Also assume that the average of x and the average of y are both equal to zero. It is a matter of common knowledge that salaries are not arranged in a symmetrical fashion; y is not a linear function of salaries.

In the formulas which follow, r is the symbol for the coefficient of correlation. These formulas are applied to Table No. 2. We have

$$\frac{1}{\sqrt{2\pi}} \int_0^h e^{-\frac{1}{2}x^2}\, dx = \frac{(a + c) - (b + d)}{2N} = .167, \text{ and } h = .4316.$$

Also

$$\frac{1}{\sqrt{2\pi}} \int_0^k e^{-\frac{1}{2}y^2}\, dy = \frac{(a + b) - (c + d)}{2N} = .187, \text{ and } k = .4874.$$

Then,

$$H = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}h^2} = .3635, \text{ and } K = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}k^2} = .3543.$$

All the quantities except r in the following approximate equation are known:

$$\frac{ad - bc}{N^2HK} = r + \frac{r^2}{2} hk + \frac{r^3}{6} (h^2 - 1)(k^2 - 1) + \frac{r^4}{24} h(h^2 - 3)k(k^2 - 3) + \frac{r^5}{125} (h^4 - 6h^2 + 3)(k^4 - 6k^2 + 3).$$

Therefore,

$$.0261r^5 + .0681r^4 + .1034r^3 + .1052r^2 + r - .4314 = 0.$$

Then, r is approximately equal to .4051. Consequently, for practical purposes we can assume that $r = .4$.

28 Boody Street
Brunswick, Maine

John L. Roberts

NOTE ON THE DERIVATION OF THE MULTIPLE CORRELATION COEFFICIENT

Consider N observed values of each of n variables. These $n \cdot N$ values may be tabulated in a double-entry table as follows:

$$\begin{matrix} X_{11} & X_{12} & X_{13} & \cdots & X_{1N} \\ X_{21} & X_{22} & X_{23} & \cdots & X_{2N} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ X_{n1} & X_{n2} & X_{n3} & \cdots & X_{nN} \end{matrix}$$

where X_{ik} is the k^{th} value of the i^{th} variable.

Using the i^{th} variable as the dependent variable, the general linear relationship between the n variables may be expressed by

$$x_i = {}_ia_1 x_1 + {}_ia_2 x_2 + \cdots + {}_ia_{i-1} x_{i-1} + {}_ia_{i+1} x_{i+1} + \cdots + {}_ia_n x_n \qquad (1)$$

where

${}_ia_j$ is the general parameter which is to be determined empirically;

$x_j = X_j - M_j$;

M_j is the arithmetic mean of the j^{th} variable.