Coefficients Of Regression

Leonard P. Ayres

To cite this article: Leonard P. Ayres (1920) Coefficients Of Regression, The Journal of Educational Research, 1:5, 398-402, DOI: [10.1080/00220671.1920.10879066](https://doi.org/10.1080/00220671.1920.10879066)

To link to this article: <http://dx.doi.org/10.1080/00220671.1920.10879066>

Published online: 15 Dec 2014.

Submit your article to this journal [↗](#)

View related articles [↗](#)

COEFFICIENTS OF REGRESSION

LEONARD P. AYRES

Russell Sage Foundation

In the March and April numbers of the JOURNAL OF EDUCATIONAL RESEARCH there were published articles by the writer on shorter methods for computing the coefficient of correlation. The object of the present article is to describe a short method for computing the coefficients of regression and to comment briefly on the practical utility of that measure of relationship.

The coefficient of regression is a measure which tells how much change takes place in one of two series of paired values for each change of one unit in the other series. Suppose we have a set of figures giving the ages and grades of the children in a school system. Since the plan of the system is that children shall be promoted one grade each year, it is sure that there will be a high correlation between the two series.

We know, however, that in nearly all school systems many children fail to go forward at the rate of a grade each year and that the average progress is in marked degree less than the ideal. The coefficient of regression is the measure which will tell us what the average rate of progress is. It will tell what the change is in grades for each change in years or what part of a grade the average child tends to advance for each additional year of age. This is what is meant in this particular case by the statement that the coefficient of regression shows the change in one series for a unit of change in the other.

The formula for the coefficient of regression, as it has ordinarily been stated, is not easy to use or understand, but its difficulties are mostly easy to avoid. Its use will be illustrated by finding it for two correlations, one for a short list of paired values and the other for a table of distribution. The formula for finding the coefficient of correlation is the one explained by the writer in the two previous articles published in this magazine. It is as follows:

$$r = \frac{\begin{array}{c} \text{Sum of products of subject and relative items} \\ \text{minus} \\ \text{av. of subject items} \times \text{tot. of relative items} \end{array}}{\sqrt{\begin{array}{c} \text{Sum of squares of sub. items} \\ \text{minus} \\ \text{av.} \times \text{total of these items} \end{array} \times \begin{array}{c} \text{Sum of squares of rel. items} \\ \text{minus} \\ \text{av.} \times \text{total of these items} \end{array}}}$$

Let us apply this formula to a short list of paired values representing the ages and grades of ten school children. Two of them began school four years ago, two three years ago, two two years ago, two last year, and two this year. They all entered at the age of six. One of each pair went forward one grade each year while the other spent two years in each grade. Their ages and grades, together with the computation of the coefficient of correlation, are shown in the following tabulation:

	Grades (Sub.)	Ages (Rel.)	Sub. ²	Rel. ²	S × R
	1	6	1	36	6
	1	6	1	36	6
	1	7	1	49	7
	2	7	4	49	14
	2	8	4	64	16
	2	9	4	81	18
	3	8	9	64	24
	3	10	9	100	30
	4	9	16	81	36
	5	10	25	100	50
Total	24	80	74	660	207
Average	2.4	8			

$$80 \times 2.4 = 192$$

$$24 \times 2.4 = 57.6$$

$$80 \times 8.0 = 640$$

$$207 - 192 = 15$$

$$74 - 57.6 = 17.4$$

$$660 - 640 = 20$$

$$r = \frac{15}{\sqrt{17.4 \times 20}} = \frac{15}{\sqrt{348}} = \frac{15}{18.65} = 0.80$$

In this example the coefficient of correlation is equal to the fraction $\frac{15}{\sqrt{17.4 \times 20}}$. The numerator (15) is derived from the

products of the items of the two series. The denominator is equal to the square root of the product of two numbers, which are 17.4 and 20. The first of these two numbers is derived from the subject series, giving the data for grades, and the second from the relative series, giving the data for ages.

The coefficients of regression are derived from this fraction and are $\frac{15}{17.4}$ and $\frac{15}{20}$. The first has as its denominator the number 17.4, which was derived from the data as to grades. The coefficient of regression $\left[\frac{15}{17.4} \right]$ tells the amount of change in ages taking place on the average in this group for a change of one in the grades. This information does not in this particular case give us information of much value.

The other coefficient of regression is the fraction $\frac{15}{20}$ in which the denominator was derived from the data on ages. This coefficient tells us how much the children in this group go forward in the grades on the average for each advancing year of age. It tells us that they go forward $\frac{15}{20}$ or $\frac{3}{4}$ of a grade for each year of age. This is in accord with the original conditions of the problem, for it was stated that half the children made 100 percent of normal progress and the other half 50 percent, or an average of 75 percent of the full rate.

This is information of an important type that cannot be secured by the simpler processes of working directly with the averages of the series and trying to compute the relationship between the time the children have been in school and the progress they have made. In many sorts of results from tests of classroom work the coefficients of regression can be used to great advantage.

The method described for finding these coefficients is far simpler than the one usually used. The formula commonly given (rho standing for the coefficient of regression) is $\rho = r \frac{\sigma_1}{\sigma_2}$.

The steps by which the method here described was derived from the standard formula may be readily followed through by anyone familiar with the ordinary notations for the coefficient of correlation. They are as follows:

$$\rho = r \frac{\sigma_1}{\sigma_2}, \quad \text{or}$$

$$\rho = \frac{\sum xy}{n \sigma_1 \sigma_2} \times \frac{\sigma_1}{\sigma_2}, \quad \text{or}$$

$$\rho = \frac{\sum xy}{n \sqrt{\frac{\sum x^2}{n}} \sqrt{\frac{\sum y^2}{n}}} \times \frac{\sqrt{\frac{\sum x^2}{n}}}{\sqrt{\frac{\sum y^2}{n}}}, \quad \text{or}$$

$$\rho = \frac{\sum xy}{\sum y^2}$$

The application of the method described may be again illustrated by using as an example an age and grade table. The one chosen is a supposititious distribution for a system where 160 children enter each year at the age of six. They are promoted annually at a 75 percent rate and none drop out. The data for the table, together with the computations of the coefficients of correlation and regression, are shown in the accompanying tabulation.

S Grades	R—AGES								T	ST	SST
	6	7	8	9	10	11	12	13			
1	160	40	10	2					212	212	212
2		120	60	23	8	2			213	426	852
3			90	67	34	15	6	2	214	642	1,926
4				68	67	42	22	9	208	832	3,328
5					51	63	47	29	190	950	4,750
6						38	56	49	143	858	5,148
7							29	49	78	546	3,822
8								22	22	176	1,408
T	160	160	160	160	160	160	160	160	1,280	4,642	21,446
										av. of sub. items	
RT	960	1,120	1,280	1,440	1,600	1,760	1,920	2,080	12,160	$\frac{4,642}{1,280} = 3.627$	
RRT	5,760	7,840	10,240	12,960	16,000	19,360	23,040	27,040	122,240		
ΣSf	160	280	400	521	641	760	880	1,000		av. of rel. items	
R(ΣSf)	960	1,960	3,200	4,689	6,410	8,360	10,560	13,000	49,139	$\frac{12,160}{1,280} = 9.5$	

$$4,642 \times 9.5 = 44,099$$

$$4,642 \times 3.627 = 16,836$$

$$12,160 \times 9.5 = 115,520$$

$$49,139 - 44,099 = 5,040$$

$$21,446 - 16,836 = 4,610$$

$$122,240 - 115,520 = 6,720$$

$$r = \frac{5,040}{\sqrt{4,610 \times 6,720}} = \frac{5,040}{\sqrt{30,979,200}} = \frac{5,040}{5,566} = 0.91$$

$$\rho_1 = 1.09 \quad \rho_2 = 0.75$$

In this case the coefficient of correlation is 0.91. Such coefficients based on age and grade tables are always high and there is surprisingly small range between those for relatively efficient school systems, having little retardation and those for far less effective systems showing high percentages of retardation.

The coefficients of regression are of greater interest and more value. In this case the coefficient of regression based on the grades is 1.09. This indicates that in this school system the children tend to average 1.09 increase in years of age for each advance of one grade.

The other coefficient in this case gives us more valuable information. It is 0.75 and is based on the ages. It tells us that the children in this system tend to go forward on the average three-fourths of a grade for each advance of one year in age. This is in agreement with the conditions laid down at the outset which were that these children were regularly promoted at a 75 percent rate.

Coefficients of regression can be made to yield information of great practical value. They may well be used in studying such questions as: regularity of attendance and promotion rates, deficiency in weight and hemoglobin counts, progress through the grades and expenditures for supervision, pages of material read in the first three grades and ability in silent reading, time spent on spelling drills and scores in spelling ability, years of schooling and earnings up to the age of fifty, and many others of the most varied character.