MY TWO SETS OF SCORES SEEM RELATED HOW CAN I ANALYZE THESE RELATIONSHIPS?

Dr. (Mrs.) Marie Farrell and Dr. (Miss) Aparna Bhaduri

Dear Reader,

Sometimes we want to know more than what is average, or where a person scores when compared to another person. Sometimes we are interested in the relationship or connection between one thing and another. In this letter, we will look at the issue of relationships, in order to see whether or not our tests or measures are doing what we want them to do.

"What do you mean?"

Suppose you were fortunate enough to develop a test which showed that if babies scored high on the test, they would probably be healthy by one year. Then you waited until the child was one year old. You then had a way of checking on your predictor of health by examining the infant on his first birthday. So you would look to see how closely the scores were related to your criterion of success and the test would tell you how good your test was for that purpose.

So here you are looking at an **index of relationship.** When the number is low, it would give you a coefficient showing a low degree, and when high, would give you a coefficient showing a high degree. This coefficient is called coefficient of correlation. Basically, what correlation tells you is this: "two sets of measures are related if knowledge of an individual's score on one of them reduces the range of possibilities for his position on the other one of them."[1]

"Can you give me an example?"

Sure. Let us say that a measure of mental illness has been devised for your country. The scores on this measure range from 0 to 100 with a measure of family cohesiveness devised with possible scores of 10-50. Now let us say that for a person with a mental health score of 50, he can only score a family cohesive score from 28 to 33. Only one-fifth of the range of the family score can ever be measured for that individual. So knowing about one measure tells us about his ability or characteristic on the other.

If we notice that as one measure rises the other rises, and as one falls the other falls, then we say that a positive **relationship exists.**

"How is this relationship expressed in numbers?"

If a perfect correlation exists, then a 1.00 is scored. If no relationship exists, the score is 0. And, a relation can be a positive or negative one. So, a coefficient correlation provides a scale from 1.00 to .00 telling us the strength of the relation, and the direction expressed as a + or —.

Now, in other parts of this text, you will be learning how to determine if a relationship exists. Tests such as the t-test and Chi-square test will establish the fact that a correlation exists or it does not. A large Chi-square value which attains a predetermined level of significance allows you to say that the two variables are related. But you do not know the extent of the relationship. To determine this, you must do further analysis.

Before going on, let us clarify what is meant by positive and negative correlation, and how one interprets **correlation coefficient.**

A perfect positive correlation, as noted, means that changes in one variable, say change in height, is accompanied by an equivalent change in the **same** direction of another variable, say weight. A positive correlation can be low, .2, .3, or high positive, .8, .9. Remember that the size of the coefficient has nothing to do with its direction. So, you may have a low negative correlation or a high negative correlation. The range from +1.00 to —1.00 is not continuous, however. There are two separate dimensions, one negative, one positive. Each one ranges from .00 to 1.00.

"So a 1.00 correlation is not greater than a —1.00 correlation".

Right. The negative dimension is separate. In our example, as height of the infant increases, weight tends to increase. Suppose that as the height increased, the infant's weight decreased. This would render a negative correlation. Thus a negative correlation means that changes in one variable, in this example height, are accompanied by equivalent changes in the opposite direction in the other variable, weight.

Further, computing correlation coefficient requires two sets of measurements, or the use of **bi-variate data.** The measurements are computed on the same groups of individuals, or matched pairs of individuals. A correlation coefficient cannot be computed on one person alone.

Two correlational methods used frequently are the rank-difference and the product-moment.

Rank-difference Coefficient

Table 1: Ordered Data for Rank Difference and Height and Weight of Rural Infants.

Subject	Rank for Variable Height	Rank for Variable Weight	Defference between Two Ranks (D)	Square of Difference between Ranks (D^2)
A	6	5	1	1
D	4	6	2	4
B	5	4	1	1
C	2	3	1	1
E	3	1	2	4
F	1	2	1	1
				12

In Table 1, the first column identifies the subject, the second column his rank on the first variable, height, the third column his rank on the second variable, weight; the fourth column, the differences, and the fifth the square of the differences. These data are needed for use in a formula which computes ρ* or rho coefficient.

To illustrate the Pearson r, we can begin with a scatter plot where we actually plot our scores.

$$*\rho = \frac{1-7\Sigma D^2}{N(N^2-1)}$$

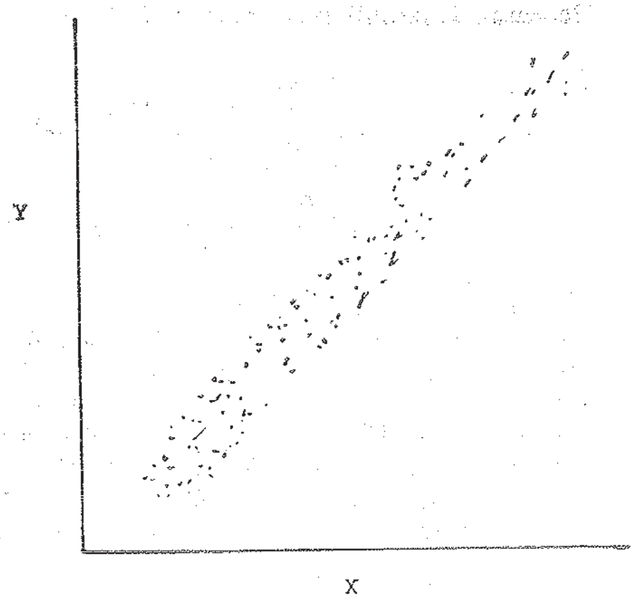

Figure 1—Scatter plot of a High Positive Correlation between Education of Mothers and Completion of Infant Immunization Schedule of Infants.

It is a place on which both the scores of each person can be represented by a single point.

Each mother is noted by a single dot which represents her on two variables X and Y. In this example, a straight regression line is formed. The scatter of scores determine the slope of the line; the smaller the scatter—

"The bigger the correlation".

Right. And the steeper the slope. Here a perfect correlation exists, for there is no scatter.

"Review again what this tells me."

In the formula*, the rank differences are in the numerator of a fraction that is subtracted from 1.00. So the larger the rank differences the smaller rho, until you reach .00 or no correlation.

Pearson "r"

The second measurement of correlation is the Pearson Product moment coefficient of correlation, otherwise known as the Pearson "r".

If the relationship is strong, the deviations are small, there is little scatter. Thus, we can make fairly accurate predictions about Y by knowing about X. Now look at the other possibility. In Figure 2, the relationship is obviously—

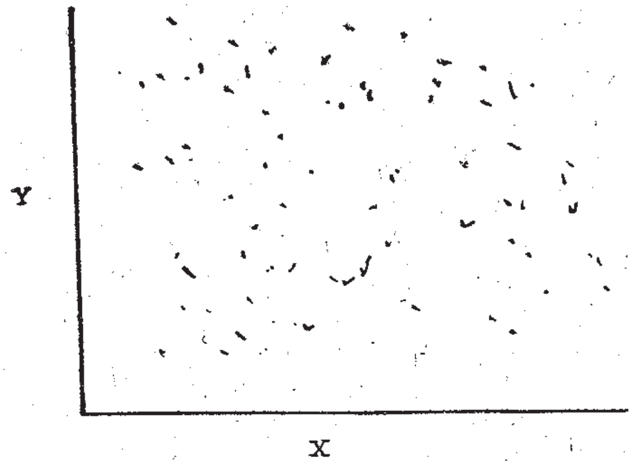

Figure 2—Scatter plot of zero correlation between use of prevention of accidents, slogans and Reduction in Accidents.

"Weak. The dots look almost randomly scattered. The deviations are large and there is a great deal of scatter."

This means that there is no tendency for a low X to be associated with a low Y, or a high X with a high Y. So knowing X does not help us at all in knowing.

The Pearson r is used to describe the degree of relationship between two variables of the kind for which a scatter diagram may be used. By using the formula* and sloving for r, you are representing the situations in the scatter plot.

The Pearson (r) is like rho except that it keeps data not kept by rho—that is the distances that separate subject of adjacent ranks on either of the variables being correlated.[2]

$$*r = \frac{(X-\bar{X})(Y-\bar{Y})}{NSxSy}$$

$\bar{X}$=mean of the X variable
Y=mean of Y variable

Sx=Standard Deviation for X variable

Sy=Standard Deviation for Y variable

N=Total number of pairs of scores

Further, the pearson r assumes continuous data, while rank order correlation (rho) assumes ranked data. In addition, rank order is used when the number of cases is under 30, or for a small sample size. And finally, for both types of tests a correlation is relevant only to the population from which the sample has been drawn.

Summary

In this chapter, we have examined the ideas having to do with indices of relationships. We noted that relationships can range from 0 to 1.0, and can be either positive or negative. We reviewed the two major statistical tests used to compute correlations; these are the rank-order correlation and the Pearson "r".

References

[1]Willemsen, E. E., **Understanding Statistical Reasoning,** San Francisco: W. H. Freeman and Co., 1974, p. 64.

[2]Phillips, J. R., **Statistical Thinking,** San Francisco: W. H. Freeman and Co., 1973, p. 50.

TNAI SOUTH ZONE REQUESTS

The nursing officers and senior nursing personnel of all the hospitals (Govt., quasi govt., and private) in the South Zone, TNAI, comprising Andhra, Kerala and Tamil Nadu are kindly requested to patiently read my appeal, understand the spirit of it, fill in the proforma enclosed and despatch it to their respective State branch Secretary (Mr. P. V. Ramachandran, Tamil Nadu, Mr. K. Papa Rao, A.P. and Miss P. M. Mariamma, Kerala). I beg of my friends, the nursing teachers, administrators and supervisory nurses to kindly rationalise the issue of membership in the TNAI and if convinced to kindly enrol themselves as **LIFE MEMBERS** of TNAI at the earliest possible convenince.

See page 107 and 109 for appeal and proforma.

— T. Stephens,
Secretary,
South Zone, TNAI