THE TEACHER'S CORNER

Classroom Use of Rank-Order Data

by THOMAS R. KNAPP
University of Rochester

In a recent article in *The American Statistician*, Edwards [1] discussed a method for constructing two series of numbers X and Y which yield predetermined means, standard deviations, correlation coefficients, etc. which are "nice to work with" integers or decimal fractions. Such numbers are very useful for the teacher of statistics who wants to go through an example on the blackboard and does not want to get involved with complicated calculations and roundoff problems. The present article is also concerned with the use of "faked" data, for the very special case where each variable consists of rankings from 1 to N (ties excluded).

It is well known that for $N = 2$ and for two variables X and Y there are only two possible values of the Spearman rank-difference correlation coefficient obtaining between the two variables, namely -1 and $+1$. For $N = 3$ the possible values are -1, $-.50$, $+.50$, and $+1$, there being one pairing of the variables that yields -1, two that yield $-.50$, two that yield $+.50$, and one that yields $+1$, a total of $6 = 3!$ pairings. For $N = 4$ there are $4! = 24$ pairings and 11 possible values of the rank-correlation coefficient (I use the notion r_s) having the following frequency distribution:

r_S	f
—1.00	1
— .80	3
— .60	1
— .40	4
— .20	2
.00	2
.20	2
.40	4
.60	1
.80	3
1.00	1
	24

In general there are $N!$ different arrangements of the digits 1 through N which when paired with the digits in their natural order $(1, 2, 3, \ldots, N)$ yield various values of the rank-correlation coefficient, the frequency of each having been determined by Olds[2] for $N \leqq 7$ in his work on the sampling distribution of r_S[1].

But what the teacher of statistics needs is not only the number of ways one can produce a given value of r_S but the actual permutations of the digits which produce it, so that he can write the two series X and Y on the board. With the aid of the IBM 650 computer, I have generated, for $N \leqq 7$, all the possible permutations of the digits 1 through N and the values of r_S that are produced when each of these permutations is paired with the natural sequence $1, 2, 3, \ldots, N$.[2] It is not the purpose of this article merely to list these permutations, since for $N = 7$, for example, the list would contain $7! = 5040$ different sequences, which took up 87 pages when listed on the IBM 407 tabulator. What I do want to do is suggest a couple of the uses to which the classroom teacher can put such information.

I. Consider the case of $N = 5$. Table 1 contains all of the permutations of the digits 1 through 5 and the rank-correlation of each with the permutation (1,2,3,4,5). There are six arrangements of the digits 1,2,3,4, and 5 which when correlated with the series (1,2,3,4,5,) produce a rank-correlation coefficient identically equal to zero. They are:

a. (5,1,2,3,4)
b. (4,3,2,1,5)
c. (4,1,3,5,2)
d. (2,5,3,1,4)
e. (2,3,4,5,1)
f. (1,5,4,3,2)

When teaching *Pearson product-moment correlation coefficients* and trying to help students visualize various degrees of relationships, the teacher could use these same series of digits[3] and make six different scatter-plots, each of which represents a correlation coefficient of .00. These scatter-plots are illustrated in Figure 1.

This sort of demonstration can also obviously be used for the *three* arrangements which produce $r_S = -.80$, the *seven* arrangements which produce $r_S = .60$, etc.

II. Consider the case of $N = 7$ and the following set of data:

INDIVIDUAL	X_1	X_2	X_3
A	1	2	6
B	2	3	1
C	3	5	5
D	4	7	2
E	5	6	7
F	6	4	3
G	7	1	4

The 3 x 3 correlation matrix consisting of the intercorrelations (again product-moment) of these variables is the *identity matrix* of order 3 (since $r_{12} = r_{13} = r_{23} = 0$) and provides an interesting classroom example when discussing multiple-regression or factor analysis. The numbers 1, 2, 3, 4, 5, 6, 7 are especially convenient since their mean is an integer, 4, and their variance is a perfect square, also 4, which results in an integral

[1] Olds actually obtained the frequency distribution of the various values of Σd^2, but since $r_s = 1 - \frac{6 \Sigma d^2}{N^3 - N}$, one correspondence between values of Σd^2 and values of r_S for a fixed value of N.

[2] I might say that I was gratified to find that the frequency of occurrence of each r_S checked perfectly with the frequencies given by Olds.

[3] This is prefectly legitimate since the rank-correlation coefficient is identically equal to the product-moment correlation coefficient obtaining between the ranks.

standard deviation, 2. A perfect-square variance is particularly difficult to fake "on the spot" in the classroom situation.

I cherish my "book" of digit-permutations and would be most happy to share these data with anyone who may be interested. And if someone has the courage to carry the process out for $N = 8$ and 9 I would be beside myself with joy at the prospect of receiving a copy of the results.

BIBLIOGRAPHY

1. Edwards, Bruce, "Constructing Simple Correlation Problems with Predetermined Answers", *The American Statistician*, Vol. 13, No. 5, (December, 1959), 25-27.
2. Olds, E. G., "Distributions of Sums of Squares of Rank Differences for Small Numbers of Individuals", *The Annals of Mathematical Statistics*, Vol. VIII, (1937), 133-148.

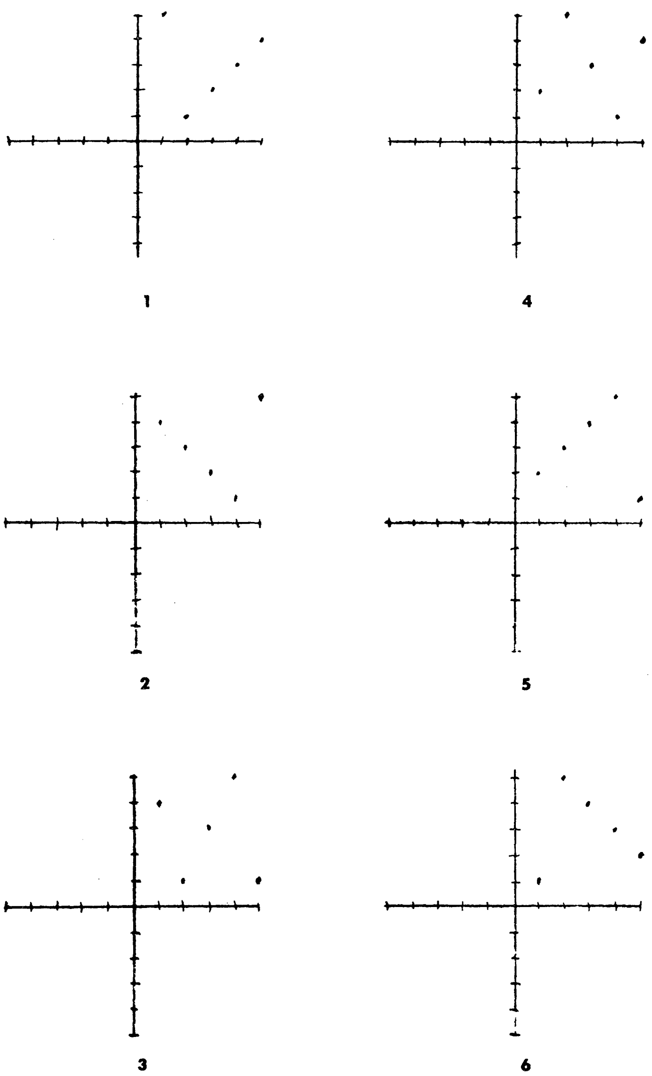

FIGURE 1

Table 1

$X = 12345$				$N = 5$	
Y	r_S	Y	r_S	Y	r_S
54321	—1.00	52143	—.20	42135	.30
		43152		41325	
54312	— .90	41532		32415	
54231		35214		31524	
53421		32541		24315	
45321		25413		24153	
				15324	
53412	— .80	52134	—.10	15243	.30
45312		51324		14352	
45231		51243		13542	
		41523			
54213	— .70	35124		41235	.40
54132		34152		23415	
53241		32451		15234	
52431		24513		13452	
43521		24351			
35421		23541		32154	.50
				31425	
54123	— .60	51234	.00	24135	
52341		43215		21543	
45213		41352		14253	
45132		25314		13524	
43512		23451			
35412		15432		32145	.60
34521				31254	
		43125	.10	23154	
53142	— .50	42315		21534	
52413		42153		21453	
45123		34215		14325	
42531		32514		12543	
35241		31542			
34512		25143		31245	.70
		15423		23145	
53214	— .40	15342		14235	
51432		14532		13425	
43251				12534	
25431		41253	.20	12453	
		34125			
53124	— .30	31452		21435	.80
52314		25134		21354	
51423		23514		13254	
51342		14523			
42513				21345	.90
42351				13245	
35142				12435	
34251				12354	
25341					
24531				12345	1.00