Records

Lindquist EF (1940). Statistical analysis in educational research. Boston: Houghton Mifflin.

Title pages

STATISTICAL ANALYSIS IN EDUCATIONAL RESEARCH

E. F. LINDQUIST
Professor of Education
THE STATE UNIVERSITY OF IOWA

HOUGHTON MIFFLIN COMPANY
BOSTON · NEW YORK · CHICAGO · DALLAS
ATLANTA · SAN FRANCISCO
The Riverside Press Cambridge

Key passages

6. THE PROBLEM OF SAMPLING IN EDUCATIONAL RESEARCH

We have already noted that many of the populations with which we are concerned in educational research are such that it is highly impracticable to draw truly random samples from them. Suppose, for example, that we wished to draw a sample of 500 pupils from all pupils in the eighth grades of Iowa public schools. As was noted before in a similar illustration, if we were to select a *random* sample of pupils from this population we would have to give every eighth-grade pupil in every school an equal chance to be selected. If we were to do this, we might find that the 500 pupils finally selected would be widely scattered over the whole state in several hundred different schools, and would therefore, for all practical purposes, be *inaccessible* for measurement, observation, or experimentation. What we would actually do, therefore, would be to secure the co-operation of, say, 10 schools, which together could provide 500 eighth-grade pupils. We might be able to select the *schools* at random from all schools, but usually even this would be impracticable. In general, the best we could do would be to prepare a list of schools which we know in advance might be willing to co-operate in our investigation, and then select 10 schools at random from this list. If then we have no reason to suppose that the schools in our list differ systematically from the other schools in the state with reference to the characteristic(s) we are investigating, we might be justified in considering our sample of 10 schools as *equivalent* to a random selection from *all schools* in the state. Even so, we could *not* consider our 500 pupils as equivalent to a random selection from all *pupils* in the state.

The reasons for this is that the pupils in different schools show large systematic differences in almost any trait that may be the subject of a research investigation. The pupils in one school may have had the advantage of a long succession of superior teachers in the preceding grades, while those in another may have had consistently incompetent teachers under poor supervision. The pupils in one school may come from a high-class residential section of the community, a section made up of professional and successful business men, while those in another may have come from an impoverished and underprivileged section made up largely of illiterate day laborers of recent foreign extraction. Large differences, particularly in educational achievement, are frequently found even between schools that apparently have much the same external advantages. It is almost a commonplace, to those familiar with the results of wide-scale testing programs, that differences in mean achievement from school to school, regardless of the size of school, are of almost the same order of magnitude as differences in individual pupil achievement in a single school. The student is perhaps already familiar with much of the almost overwhelming mass of evidence on this point, but it might be worth while to consider here one representative bit of this evidence. The State University of Iowa annually conducts a state-wide end-of-the-year achievement testing program which involves the administration of objective tests of school achievement to over 50,000 pupils in several hundred high schools. In the 1935 program, an objective test of achievement in English correctness was administered to all ninth-grade pupils in 274 schools. For the 24 largest schools, each of which tested over 100 ninth-grade pupils, the total distribution of pupil scores is given in Table 1 at the left, while the distribution of mean scores in these schools is given at the right. Had the pupils in each school constituted a random sample from the pupils in all these schools, we should, according to sampling theory, expect the standard deviation of the distribution of *means* to be less than $31.54/\sqrt{100} = 3.15$, 31.54 being the standard deviation of the total pupil distribution, and 100 the minimum number of cases in any sample. Actually, however, we see that the stand-

TABLE 1

DISTRIBUTIONS OF INDIVIDUAL AND MEAN SCORES OF NINTH-GRADE PUPILS ON THE 1935 IOWA EVERY-PUPIL TEST IN ENGLISH CORRECTNESS FOR 24 SCHOOLS

(Each school tested over 100 ninth-grade pupils)

Pupil Scores		School Means	
Scores	Frequency	Means	Frequency
170–180	3	91.0–94.49	1
160–169	10	87.5–90.99	
150–159	15	84.0–87.49	
140–149	30	80.5–83.99	
130–139	43	77.0–80.49	
120–129	67	73.5–76.99	1
110–119	84	70.0–73.49	1
100–109	112	66.5–69.99	
90– 99	146	63.0–66.49	3
80– 89	188	59.5–62.99	2
70– 79	266	56.0–59.49	2
60– 69	335	52.5–55.99	2
50– 59	394	49.0–52.49	1
40– 49	553	45.5–48.99	4
30– 39	554	42.0–45.49	5
20– 29	508	38.5–41.99	1
10– 19	263	35.0–38.49	
0– 9	75	31.5–34.99	
		28.0–31.49	1
N	3646	N	24
M	54.08	M	54.58
S.D.	31.54	S.D.	13.29

ard deviation of school means is 13.29, or more than 4 times as large as would be expected on the hypothesis of random sampling.

In consideration of these much-larger-than-chance differences between schools, let us consider further our illustrative sample of 500 pupils of Iowa eighth-grade pupils. It is very obvious that had all of these pupils come from a *single* school, the sample would represent a very poor basis for generalization about the population, particularly in contrast to a truly random sample of equal size. In the random sample, hundreds of different schools would be represented, in the sample just considered only one is represented — and that might be one of the schools in which the level of achievement is very high, or it might be one in which the level is very low. It should be equally evident that a sample in which only 10 schools are represented is neither as good as nor equivalent to a random sample of 500 cases from the population at large. With so few schools involved, the danger is appreciable that we might by chance have selected 10 good schools, or 10 poor ones, or that most of the schools used are good schools or poor ones. If then, we were to use the standard-error-of-the-mean formula with this sample, substituting 500 for n, we would very seriously exaggerate the reliability of the mean. In spite of the fact that it contains 500 pupils, this sample must be considered as a very "small" sample — a sample of only *ten* schools — and in order to evaluate any estimate derived from it we must have a sampling theory appropriate for small samples.

In general, then, many of the samples employed in educational research consist of a small number of intact groups (such as *classes* in the same or different schools, groups of pupils in separate buildings in the same system, or groups of pupils from different communities or geographical regions), or of a small number of subsamples selected from different "strata" in the population (in the case of controlled samples). In all such cases, the "size" of the sample is dependent, not upon the number of individual observations, but upon the number of intact groups or subsamples of which the total sample is constituted. In other words, the *unit* of sampling in educational research is often the class, the school, or the community, rather than the pupil. It is for this reason that the need is so great in educational research for a special small sample theory, and that this text is in so large part devoted to an exposition of this theory.

7. THE TECHNIQUE OF RANDOM SELECTION

It is a fact of extreme practical significance that all mathematical sampling theory is based finally on the assumption of *random selection*, and that any application of this theory is valid only to the degree that the samples employed have been so selected. We may note at once, however, that random selection does not always mean simple random sampling. We have seen, for instance, that controlled samples or matched samples are not simple random samples, but that they may involve random selection, and that it is therefore sometimes possible to deduce the sampling distributions for estimates obtained from such samples. Simple random sampling is often impracticable in educational research, but it is nearly always possible to plan our investigations and experiments so as to provide for random selection, and thus to utilize sampling theory in interpreting our results. Since those interpretations will be valid only to the degree that the selection was actually random, it is obviously important that the student be provided with a technique that will *insure* random selection, in so far as that is possible.

In many of the instances in which random selection is necessary in educational research, the selection is made from a relatively small number of cases. This is particularly true in experimental work. For example, in each of the schools involved in a methods experiment, we may wish to divide the seventh-grade pupils at random into two or more equal groups to be taught by different methods, or we may wish to assign the *classes* (as already organized) at random to the different methods. Again, we may divide the available pupils into levels of intelligence, and *within* each level assign the pupils at random to the experimental treatments.

One method of making random selections in situations of this kind may be described as the "lottery" method. For example, if we wished to split a group of 30 pupils at random into 3 groups of 10 each, we could prepare 30 cards or slips of paper on each of which is written the name of one pupil, shuffle or mix these cards very thoroughly, and then "deal" or draw blindly 3 sets of 10 cards each. This is a troublesome procedure, however, and introduces the danger of bias through improper mixing or drawing of the cards.

A more certain and more convenient procedure is to make use of a table of "random numbers." For the convenience of the student, a part of one such table is reproduced in the Appendix (Table 18). The manner in which the original table was constructed is described on page 18 of *Statistical Tables for Biological, Agricultural and Medical Research*, by R. A. Fisher and F. Yates

(Oliver & Boyd, London, 1938). It is sufficient to say here that the digits in this table were so selected that any digit from 0 to 9 had an equal chance to appear in any given position in the table. The manner in which this table may be used should be made clear by the following illustrations.

Illustration No. 1: *To assign 5 classes at random one to each of 5 experimental treatments.*

Number the classes and the treatments separately from 1 to 5 in any order whatever. Select any point at haphazard in the table of random numbers. Reading in any direction from this point (right to left, bottom to top, diagonally, etc.) read the first five *unlike* two digit numbers (skipping any that may previously have been read) from the table. Assign the first of these numbers to class 1, the second to class 2, etc. The class with the highest random number will then be assigned to treatment 1, that with the second highest to treatment 2, etc.

Suppose, for example, that the first number selected haphazardly is that in the 14th row and the 4th double column on the first page of Table 18. Reading to the right from this point, the first five unlike two-digit numbers are 19, 95, 50, 92 and 26. The second class would therefore be assigned to treatment 1, the fourth to treatment 2, the third to treatment 3, etc.

The "starting point" in the table should be determined before looking at any number in the table. In the preceding case, for instance, the decision to begin with "the 14th row in the 4th column on the first page of the table" should be made before looking in the table. Otherwise one might, without being fully conscious of the fact, begin with a large number, and thus in effect *deliberately* insure that class 1 will receive treatment 1, or otherwise bias the selection. Furthermore, once having selected the starting-point and direction, no peculiarity in the numbers read should be permitted to cause one to discard the results and start anew at another point.

Illustration No. 2: *To select 20 pupils at random from 62 available pupils.*

Number the pupils from 00 to 61 in any order whatever. Turn to the table, and from any point and in any direction read the first 20 two-digit numbers that are less than 62, skipping any number previously read. For example, beginning in the 11th row of the 5th column on the first page of Table 18 and reading downward, the first 20 unlike numbers below 62 are 46, 12, 13, 35, 43, 53, 61, 24, 59, 06, 20, 38, 47, 14, 11, 00, 60, 23, 19, and 53. The pupils who had previously been assigned these numbers would then be the 20 required. If these numbers are checked off in the original list of numbered pupils as they are read from the table, there will be no difficulty in avoiding duplications.

If the selection is made from more than 99 cases, we must read three- or four-digit numbers from the table, as the case may be. These may be secured by combining columns in the table. Suppose, for instance, we wish to select 15 schools at random from the 418 high schools in a given state. We would first number all schools from 000 to 417 in any order whatever. We would then combine, say, the 7th double column (on the first page of Table 18) with the first half of the next column to the right, to secure a "column" of three-digit numbers. Reading upwards from the bottom of this "column," the first 15 unlike numbers less than 418 are 044, 416, 377, 358, 061, 057, 389, 325, 091, 373, 299, 278, 271, 332, and 395. The schools previously assigned these numbers would then constitute our sample of 15. If the total number from which the selection is to be made is a number like 160, for example, considerable "hunting" would be required in the table to find three-digit numbers less than this value. A more convenient procedure in such cases is explained by the following example.

Suppose that we wished to select 5 cases at random from 120 available cases, numbered from 0 to 119 in any order whatever. We would first observe that 120 is contained in 999 eight times, or that $8 \times 120 = 960$ is the largest multiple of 120 which is contained in 999. We would then select random numbers less than

960 from a three-digit column of the table, and divide each by 8, dropping any remainder. The first five unlike *quotients* would then be the numbers of the cases selected. Suppose, for instance, that we begin at a point in the table in which the following three-digit numbers appear in the order given below.

Random Numbers	Quotients
562	70
815	101
982	
322	40
057	7
815	101
723	90

The cases numbered 70, 101, 40, 7, and 90 would then constitute our random sample of 5.

Illustration No. 3: *To Split a Group at Random into a Number of Equal Groups*

The problem of splitting a group at random into a number of equal groups would simply involve an extension of the procedure just described. Suppose we wish to select 3 random groups of 20 each from 62 available pupils. We would first draw one sample of 20 in the manner already illustrated. We would then continue reading from the table until we had 20 more numbers (not previously read) that are less than 62. The pupils with these numbers would constitute the second group. We would then continue as before to get *two* more numbers (not previously read) less than 62. The pupils with these numbers would be discarded, the remaining 20 pupils would constitute the third group. If, again, the numbers in the original list were checked off as read, there would be no difficulty in avoiding duplications.

Other methods of employing tables of random numbers may be found described in the directions accompanying those tables.[1] The

[1] See the aforementioned tables by Fisher and Yates, or *Tracts for Computers, No. 15, Random Sampling Numbers*, by L. H. C. Tippett.

methods here described, however, will be adequate for most situations met in educational research.

It may be worth while to point out here the possible defects in a certain method of sampling (of pupils) that has been frequently employed in educational research — the practice of making the selection schematically from a list in which the pupils' names have been arranged in alphabetical order. For example, we might take every 6th name in an alphabetical list of 95 names to secure a sample of 15, or if we wished to split the group into two equal groups, we might do so by selecting alternate names from the list. A selection of this kind should be free from bias (unless the selection is made from only part of the list), so far as measures of central tendency are concerned, but it may nevertheless not be the equivalent of a random selection. This is because of the possibility that alphabetized lists may be "stratified," since pupils with the same last name (or names beginning with the same letter) may be related, or of the same nationality, or otherwise more nearly alike than pupils selected at random. In general, particularly since it involves so little trouble, there is no excuse for failing to use the "random numbers" type of selection in situations of the type described.