REASONS FOR SAMPLING

by

Walter A. Shewhart

Informal talk at Opening Session of Conference on Acceptance Sampling given at Newark College of Engineering on September 27, 1946

INTRODUCTION

I. <u>Every set of observed data is a sample: hence the reasons for sampling are inherently the reasons for taking data*.</u>

<u>Applied Science</u>

Raw materials to satisfaction of human wants. At every step, sampling enters.

Examples of fields of sampling:

Raw Materials	(Coal (Ore of all kinds (Wood (Etc.,
Processes	(Quality of piece parts (Routine analyses (Output of machines (Time studies (Cost studies (Etc.

- - - - - - - - - - - - - - - - - - - -

* Even the result of 100% inspection is a sample in more ways than one as we shall see later.

Psycho-physical including human wants	1 Reaction of human beings to physico-chemical quality S R 2 Determination of sizes for clothes telephone booths telephone handsets etc. 3 Human reaction to drugs, foods, and the like 4 Etc.

Two aspects of these problems:

Sample of those to be satisfied.

Sample of the reactions of a sample of individuals.

Life Tests	1 Piece parts 2 Finished product 3 Etc.

Pure Science

1. Constants of nature are estimates based on sample.

2. So-called laws of nature are derived from samples of observations.

Science (exclusive of mathematics and deductive logic) is based on inference from sample.

Hence the reasons for sampling are in the last analysis the reasons for pure and applied science.

Also note that some of most important inferences to all of us are based upon samples of destructive tests as in

a. Foods you eat
b. drugs you take
c. Failures of material
d. Etc.

What is a Sample?

1.1 Does the set of numbers

1,2,3,4,5

constitute a sample of the infinite population of all positive integers?

1.2 I asked one of my colleagues to give me a set of five numbers all different. This colleague gave me the set

1,2,3,4,5.

Is this set of numbers a sample? If so, what is the population of which the sample is a part?

1.3 Assume that my colleague had given me the set

$-15, \sqrt{-1}, \sqrt{2}, e, o$

Would this set constitute a sample? If so, what is the population?

2.1 A bowl contains 998 chips with numbers written thereon. Half of the chips are green and the other half are white. Otherwise they are "presumably" identical or at least "essentially the same.

Ask a child not over 7 years old, to take five chips from the bowl and to read off the numbers on the chips. Assume that the numbers are

1.1; 1.3; .6; .2; 1.2.

Is this a sample? If so, what is the population?

2.2 Would your answers to 2.1 be different, if instead of a child the one who takes the five chips is a high school student, engineer, physicist, chemist, or statistician?

2.3 Would your answers to 2.1 and 2.2 be different if the instructions were to stir the chips in the bowl thoroughly before drawing the five chips? Would your answers be modified if the person making the drawing was told to make the drawings while blindfolded?

2.4 Draw 5 chips at <u>random</u> from the bowl. (How would you do this?) Is the set of numbers thus obtained a sample? If so, what is the population?

2.5 Draw 5 chips at <u>random</u> from the bowl and with replacement. Is the set of five numbers a sample? If so, what is the population?

2.6 Suppose that I put 1/2 teaspoonful of oil on the chips before the drawings in 2.4 and 2.5 are made. In waht way, if any, would you change your answers.

3.0 In each of the questions 2.1 to 2.6, does the set of numbers

$$x_1, x_2, \ldots, x_n$$

by themselves constitute a sample or must we also know how they were drawn?

Is the set of 5 numbers obtained in 2.6 a sample of the totality of all sets of numbers that might be obtained by repeating again and again the <u>operation</u> of drawing therein defined? <u>In what sense, if any, is it a sample of the 998 numbers in the bowl</u>?

THE OPERATION OF DRAWING A SAMPLE

<u>The operation of drawing a sample from a lot of N elements consists of dividing the N elements into k classes (that may or may not overlap) and the act of selecting one of these k classes</u>

1. Random sampling operation - Any sampling operation for which the probability of selecting any one of the k classes is 1/k.

2. Unrestricted random sampling operation of selecting n items is any operation of selecting one of C_n^N possible sets of n elements in which the probability of drawing any one of C_N^N classes is $1/C_n^N$. Here the C_n^N classes overlap.

3. Restricted sampling operation is one in which the N elements are classified into k classes

   $S_1, S_2, \ldots, S_k$

   where S_i consists of

   $x_i; x_{i+k}, x_{i+2k}, \ldots x_{i+(n-1)k}$

   and selecting S_i at random.

4. Other kinds of sampling operations:

   Cluster sampling
   Stratified sampling
   Representative sampling
   Etc.,

In other words, there are an indefinitely large number of ways of defining an operation of selecting n items from a lot of N. It is conceivable that a given set of n elements might result from any one of these operations.

Having defined a sampling operation that may be repeated again and again at will, the set of all possible groups of n items that may thus be attained if the process is repeated indefinitely is a statistical population.

In the present art of statistical science it is possible to make valid inferences about such a population upon the basis of one or more samples of n. To do so, however, it is necessary that we know the sampling operation as well as the n items in the sample.

RULES OF ACTION

So-called sampling plans - By acceptance sampling plans are usually meant rules of action which specify at the conclusion of the analysis of a sample which of the following three actions should be taken:

1.1 To reject the lot
1.2 To accept the lot
1.3 To take additional observations.

In the general field of science, by a sampling plan we mean a rule of action which specifies at each stage of experiment which of the following three decisions should be made:

1. To accept a specified hypothesis upon the basis of the observed sample.
2. To reject such hypothesis.
3. To take further measurements.

For example, the sequential sampling plan is a rule of action for making one of the following three decisions at any stage of the taking of a sample (at the nth trial for each integral value of n): to accept, to reject, to take further observations.

Such rules of action are, as it were, rules for using samples. It is of the utmost importance, however, to note that the rule of action (as an example, the sequential sampling plan) depends specifically upon the kind of sample. For example, the Wald sequential sampling plan assumes a random sampling operation. In other words, if an experimentalist were to stick into the plan some other kind of a sampling operation, it is most likely that the plan would not work as it is supposed to work, that is to say, the validity of a sampling plan depends not simply upon the items in the sample but upon the fact that the operation of obtaining these items was of a specified kind.

This fact is very likely to be lost sight of by those using a plan.

THE IMPORTANCE OF ORDER IN A SAMPLE

In general, the sampling plans set up certain criteria of action based upon the assumption that the sample used in the plan is of a special type, such as random, for example. They usually neglect, however, to set up criteria on checking the hypothesis that it is random. The most useful criteria for such purposes depend upon the sequential order in a sample of n.

1. Run criteria
2. Eta
3. Serial correlation
4. Etc.

To be most efficient, a plan should therefore be one that takes account not only of certain statistics of the distribution of observed values but also of certain statistics of the ordered sequence.

One of the possible ways of drawing a random sample of n from a finite lot is to draw one of C_n^N such samples as a unit. Anothe is to draw the n numbers of such a sample one at a time. Assume, for example, that the distribution of values obtained in the two cases are identical. The second sample contains more information than the first in that it gives you a sequential order.

Hence a sample derived by an operation consists not only of the distribution of values obtained but also of the order.

We should note therefore that the validity of customary acceptance sampling plan depends upon the assumption that the operation of drawing is random. This, however, depends upon the assumption that there are no assignable causes present in the sense of the theory of control

SOME SUGGESTED QUESTIONS

TO BE ANSWERED IN THIS COURSE

I believe that it is true that most of the sampling plans mentioned in the outline of the course to be given here assume that a random sample is to be obtained in some way. Some of the questions that you may wish to get answered by the staff from time to time are the following:

1. How would you draw a random sample of cable from a large spool containing several tons?

2. How would you get a random sample of permalloy as a basis for chemical analysis?

3. How would you get a sample of lead cable sheath to check for thickness.

4. How would you sample the lengths of pole spans in the United States?

5. How would you sample the people who may be expected to use a handset in order to determine the "best" size?

Can the samples called for in the following be used in each of the sampling plans that you are to consider:

1. Sampling steel, cast iron, open-hearth iron, and wrought iron. ASTM E 59-45.

2. Sampling ferro alloys. ASTM #E32-42.

3. Sampling molybedenum salts and compounds for metallurgical use. ASTM E 66-34 T.

4. Sampling wrought nonferrous metals and alloys for determination of chemical composition. E 55-46T.

In taking a sample for chemical analysis, some of the errors of the second kind to be watched for are:

1. Differences between the composition of fine and coarse particles.

2. The fixation of oxygen, water, or carbon dioxide during grinding, sieving, or drying.

3. Uncertainties as to the water content after drying and weighing the sample.

4. Changes in the composition of the sample during storage.

Those of you interested in the field of foods and drugs, analysis of raw materials, and the like, particularly by chemical means, will do well to consider very carefully the fact that a sampling plan as usually discussed depends upon the assumption that you are dealing with random samples. Not until you have learned therefore how to select a random sample will you be thoroughly justified in following blindly sampling plans.

I am pleased to see that tonight you are going to learn how to draw a sample. It may be helpful to think of some of the questions raised above while you are carrying out the laboratory experiments on drawing such a sample