FOREWORD

This is the first of a series of Bulletins issued primarily for the use of members of the Inspection Engineering Department.

Inspection Engineering is concerned primarily with the quality of product. Naturally this involves the measurement of quality and the correction of such measurements for items such as errors of observation, methods of measurement, and size of sample, so that we may attain as nearly as possible a true picture of quality. This involves the establishment of methods of measuring the quality of a number of pieces of the same kind in terms of one or more quality characteristics in such a way as to make possible the comparison of the quality of one lot of equipment with that of another. But the function of inspection engineering does not end here. We must be able not only to compare qualities of product from period to period but we must also assist in the establishment of economic standards of quality. That is to say, the information obtained in the inspection of apparatus and equipment should indicate whether or not it is reasonable to expect that the quality of product can be modified economically. This brings us to the subject of our present bulletin, "When Must a Thing Be Left to Chance?".

In the first place we cannot interpret the significance of variations in quality unless we know whether or not the observed variations indicate the presence of causes of variation which should not be left to chance. Furthermore, as we shall show in a later bulletin, one of the first steps in setting up economic standards of quality is to determine whether the observed differences in a series of observed values of quality must be left to chance. If in such a case we find that they need not be left to chance, it is reasonable to expect that we may approach the economic standard of quality without modifying the whole manufacturing process. But if, on the other hand, we find no evidence of causes of variation which should not be left to chance, then the modification of the distribution of product involves an entirely different kind of change in the manufacturing process. Hence, the contents of this bulletin are basic to a large part of the work of inspection engineering.

The four criteria for determining when a thing should be left to chance, as discussed herein, were developed primarily in connection with the study of the quality of finished product, but naturally the results have a far wider field of application than that in which they were originally applied. Quality of finished product depends upon every step of the fabrication process and at almost every stage in production some form of inspection is conducted to determine the quality of the product up to that point. Hence it is but reasonable to expect that the methods which have been found valuable in the inspection of the quality of final product should apply equally well at every stage where quality is measured.

Moreover, the inspection engineer is not the only individual interested in determining whether or not the observed variations in a given set of data are attributable to chance. The very spirit of research and development is to detect and explain all variations which need not be left to chance. Hence the contents of this bulletin have a very broad field of application in the work of the Laboratories. Typical problems involving the use of this theory in research, development and even in commercial work have already come to our attention. For this reason some of the illustrations in this bulletin have been chosen from fields other than inspection of finished product in the hope that by so doing the information contained herein may be made of more value to other departments.

It is the object of this bulletin to make the theory available in such form that it can be applied by those not interested in the details of the theory itself. Hence actual problems are solved and data sheets given wherever necessary to illustrate the details of the calculation. Each of these bulletins is to be complete in itself and will include a brief outline of the theory underlying the methods discussed.

TABLE OF CONTENTS

WHEN MUST A THING BE LEFT TO CHANCE?

Part 1

Introduction

Informal Statement of the Problem

The ultimate aim of research is to reduce everything to known laws, thereby doing away with chance. In the laboratory, where all but one variable can often be quite thoroughly controlled, research progress has been very encouraging. Even here, however, the most refined measurements, such, for example, as the determination of the charge on an electron, indicate the existence of a group of uncontrolled or chance causes. In other words, chance we have with us always. But when do the results of a series of observations indicate that further attempts to reduce the effects of the unknown causes will prove successful?

The proposed answer to this question is: A thing must be left to chance only when controlled by what will be defined as a Constant System of Chance Causes among which no single cause or distinguishable group of causes appears to have a predominating effect.

What we need, then, in any investigation where it is of importance to know whether or not a thing must be left to chance, are criteria for detecting lack of constancy of the cause system and the effect of a predominating cause or group of causes. Monday 24.07.06

Significance of the Problem

A few illustrations will reveal the wide importance of the question considered in this paper. A business executive looks at the sales record of his company and wishes to know if it is possible to increase the sales appreciably by finding and controlling some single factor; a purchasing engineer observes differences in the quality of material supplied from two sources and wishes to know whether or not these should be left to chance; a manufacturer observes wide variability in the quality of his product and wonders if he could materially improve this condition by modifying some one factor; in just this way every industrial and scientific worker often sees differences between observed results which make him wonder if he has overlooked some important factor or cause of variation.

Object

The purpose of this paper is to present in as simple a form as possible the method of applying certain criteria for determining when a thing must be left to chance.

Part I deals primarily with a very brief formal statement of the problem and the definitions of certain fundamental concepts. No time is spent in elucidating the definitions at this point because it is felt that this is sufficiently done later in the discussion of numerous actual problems.

Parts II and III present in outline form the details of the methods of applying the criteria without requiring a special knowledge of statistical theory.

Part IV indicates how these results may be applied in various fields of investigation.

Part V gives a critical discussion of some of the theoretical points of major interest in the paper.

Formal Statement of the Problem

1. Given a set of n observed values

$$X_1, X_2, \ldots X_i, \ldots X_n,$$

of some quantity or thing taken under supposedly the same essential conditions, where, in general, not all of these observed values are the same, to determine whether or not the observed differences must be left to chance.

The n different values of X might be, for example, n observed values of the charge on an electron, the years of life of n telephone poles, measurements of tensile strength on n supposedly identical specimens of a given material , or the percentages of stale bread returned by n different bakeries, or in general, any set of n data where, as in the above cases, it is assumed that they have been produced under the same essential conditions.

2. Given m sets of values such as

$$\begin{array}{l} X_{11}, X_{12}, \ldots X_{1i}, \ldots X_{1n_1} \\ X_{21}, X_{22}, \ldots X_{2i}, \ldots X_{2n_2} \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ X_{m1}, X_{m2}, \ldots X_{mi}, \ldots X_{mn_m} \end{array}$$

representing observed magnitudes of some quantity or thing to determine whether or not the differences between the m sets of values must be left to chance.

For example, the n_1 different values of X might represent the observed values of the charge on an electron taken in one laboratory and the n_2 values of X, the observed values of the charge on an electron as determined in another laboratory and so on; the first set of values might be the years of life of a group of n_1 telephone poles of a given kind in one line and the other (m-1) sets might give similar information about the same kind of telephone poles in (m-1) other lines. That is to say, the m different sets of values are differentiated in some way, although it remains to be determined whether or not the differences between sets are such as to indicate that they were produced by causes which need not be left to chance.

We shall now introduce definitions of "Constant System of Chance Causes" and "Assignable Causes" and then restate our problem in terms of these.

A Constant System of Chance Causes

If a variable X is not continuous but is controlled by a constant system of chance causes, finite in number, we assume that it can take on only a finite number M of different values, say

$$X_1', X_2', \dots X_i', \dots X_M'$$

with the probabilities of occurrence

$$p_1', p_2', \dots p_i', \dots p_M'.$$

Such a variable is said to be controlled by a constant system of chance causes, because the probability p_i' of obtaining a particular value X_i' is the same every time a value of X is observed.

If, however, the variable X is continuous, so that in any small interval it takes on all possible values, then in contradistinction to the case of the discrete variable, the probability dy' of obtaining a value of X between X and X + dX is constant every time a value of X is observed. This probability is defined symbolically by

$$dy' = f(X_1, \lambda_1', \lambda_2', \dots \lambda_{m'}')dX,$$

where f is the functional form defining the distribution of effects of the cause system involving m' parameters.

Any set of n observed values of X, therefore, may be taken as a sample of the possible effects of such a system of causes.

<u>Assignable Cause</u>

1. If the system of causes is not constant, it is said to contain an assignable cause of variation.

2. If a constant system of causes contains one predominating cause or group of causes, we term this one cause or group of causes an assignable cause of variation.

These will be referred to as assignable causes of types I and II, respectively

Very briefly, then our object may be restated as follows:

To present four criteria for detecting the presence of assignable causes of variation.

Part II

Detection of Lack of Constancy of Cause System

Introductory Note

The two criteria to be discussed in Part II can be applied only when the original set of N observed data can rationally be divided into m different groups.

Typical Problem

Let us take a problem from the field of industrial research. In the production of a certain kind of equipment, considerable cost was involved in securing the necessary electrical insulation by means of materials previously used for that purpose. This condition led to a research program to secure a cheaper substitute material. After a long series of preliminary experiments, a tentative substitute was chosen and an extensive series of tests for insulation resistance were made, care being taken to eliminate known causes of variation. One such group of 204 observations taken on the proposed substitute material under supposedly controlled conditions is given in Table 1. Reading from top to bottom in a column beginning at the left row and continuing throughout the table gives the order in which the observations were made.

5045	4565	4895	4335	4560	4485	4080	4700	4800
4350	4410	4255	5000	4700	4430	4450	4600	4765
4350	4065	4170	4615	4310	4300	3690	4840	4800
3975	4565	3860	4215	4310	4690	3635	5075	4800
4290	5190	4445	4275	5000	4560	3635	5000	4880
4430	4725	4650	4275	4575	3075	3900	4770	4580
4485	4640	4170	5000	4700	2965	4340	4570	4700
4285	4640	4255	4615	4450	4080	4340	4985	4890
3980	4895	4170	4735	4860	4080	3665	4775	4425
3925	4790	4575	4215	4650	4425	3775	5075	4455
3645	4845	4175	4700	4570	4500	5000	4925	4135
3760	4700	4550	4700	4570	4430	4850	5075	4100
3300	4600	4450	4700	4685	4840	4775	4925	[illegible]
3685	4110	2855	4700	4160	4840	4500	[illegible]	[illegible]
3463	4410	2920	4095	4325	4310	4770	4915	[illegible]
5200	4180	4375	4095	4125	4185	4500	[illegible]	[illegible]
5100	4790	4375	3940	4100	4570	4770	5075	[illegible]
4635	4790	4355	3700	4340	4700	5150	4450	[illegible]
5100	4340	4090	3650	4575	4440	4850	4415	5000
5450	4895	5000	4445	3875	4850	4700	4565	5000
4635	5750	4335	4000	4050	4125	5000	4405	
4720	4740	5000	4845	4050	4450	5000	4615	
4810	5000	4640	5000	4685	4450	5000	4615	

Table 1 - Electrical Resistance of Insulation in Megohms.
Must Such Variations be Left to Chance?

No a priori reason could be assigned why the measurements forming one portion of this series should be different from those forming any other portion. In other words, there was no rational basis for dividing the total set of data up into groups of a given number of observations except that it was reasonable to believe that the system of causes might have changed from day to day as a result of changes in atmospheric conditions, changes in observers, changes of material on which measurements were made or in other similar ways. In general, if such changes take place, we may more readily detect their effect if we divide the total number N of observations into comparatively small sub-sets. If there is no reason for choosing a particular sub-set or sample under such conditions, the size of sample is taken as four.

Are we justified in leaving to chance the observed differences between the observations in Table 1?

<u>Criterion I</u>

All necessary terms such as average, standard deviation and others are defined in the data sheet of Fig. 1. The method of carrying out the requisite calculations is given and illustrated in Fig. 1, using the data in Table 1.[1] The following steps must be taken:

a. Calculate the arithmetic mean $\bar{X}$, the standard deviation σ, the variance σ^2 and the flatness or kurtosis β_2 of the N observed values.

b. Calculate the expected values $\bar{X}$, $\bar{\sigma}$, $\overline{\sigma^2}$, respectively, of the average, standard deviation and variance in samples of size n drawn from a universe characterized by the four functions given in <u>a</u>.

	Symbols and Method of Calculation	Expected Values in a Sample of Size n	Standard Deviations in a Sample of Size n
Total number of observations	$N = 204$		
Average or arithmetic mean	$\bar{X} = \frac{\sum_{i=1}^{N} X_i}{N}$ $= 4498.18$	$\bar{X} = \bar{X}$	$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$ $= 232.61$
Variance	$\sigma^2 = \frac{\sum_{i=1}^{N} (X_i - \bar{X})^2}{N}$ $= \frac{\sum_{i=1}^{N} X_i^2}{N} - \bar{X}^2$ $= 216416.82$	$\overline{\sigma^2} = \frac{n-1}{n}\sigma^2$ $= 162314.12$	$\sigma_{\sigma^2} = \frac{\sigma^2}{n}\sqrt{\frac{n-1}{n}[(n-1)\beta_2 - n + 3]}$ $= 164631.38$
Standard deviation	$\sigma = \sqrt{\sigma^2}$ $= 465.21$	$\bar{\sigma} = c_1\sigma$ (where c_1 is given by Fig. 2-a) $= .80\sigma = 372.17$	$\sigma_\sigma = c_2\frac{\sigma}{\sqrt{2n}}$ (where c_2 is given by Fig. 2-b) $= 156.65$
Flatness or kurtosis	$\beta_2 = \frac{\sum_{i=1}^{N} (X_i - \bar{X})^4 / N}{\sigma^4}$ $= 4.448359$		

FIG. 1 - DATA SHEET FOR CRITERION I FILLED OUT FOR DATA OF TABLE 1

c. Calculate the standard deviations σ, $\sigma_{\bar{X}}$, σ_σ and σ_{σ^2} respectively of a single observation, average, standard deviation and variance in samples of

---

1. I am indebted to Miss Marion B. Cater and Miss Miriam S. Harold for carrying out the necessary calculations, drawing the figures and assisting with the proof of this paper.

size n.

d. Calculate the average, standard deviation and variance of each of the m samples of size n.

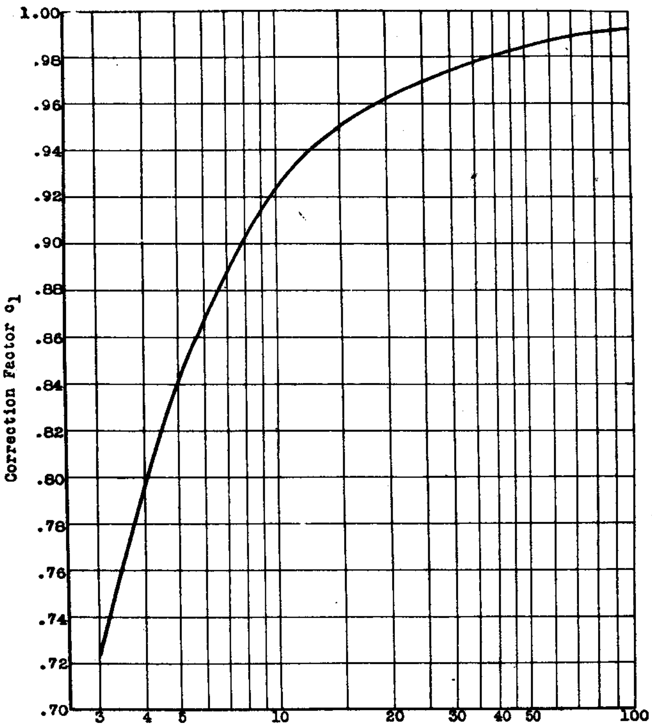


FIG. 2-a - CORRECTION FACTOR c_1 FOR OBTAINING EXPECTED STANDARD DEVIATION $\overline{\sigma}$ OF SAMPLE OF SIZE n

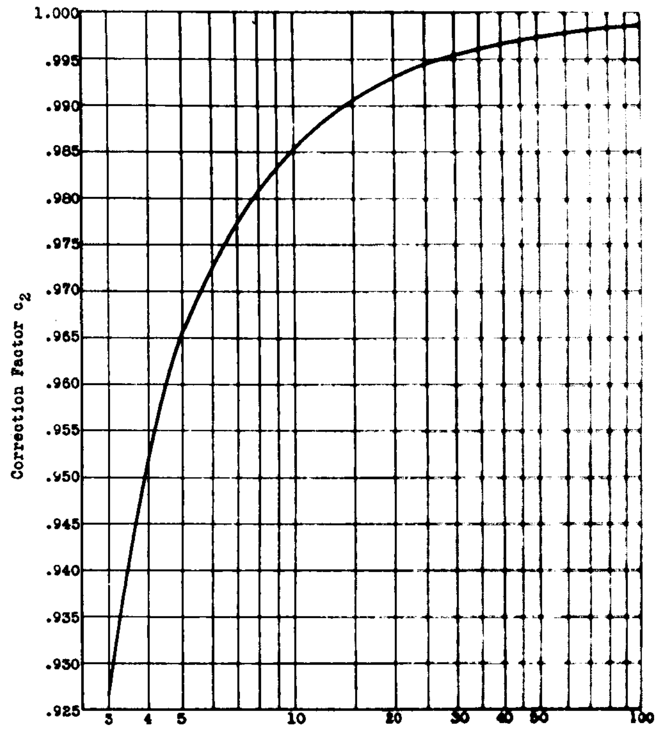


FIG. 2-b - CORRECTION FACTOR c_2 FOR OBTAINING STANDARD DEVIATION σ_σ OF SAMPLE OF SIZE n

If the observed values of a single observation, the average, standard deviation and variance in samples of size n fall outside the respective ranges $\overline{X} \pm 3\sigma$, $\overline{X} \pm 3\sigma_{\overline{X}}$, $\overline{\sigma} \pm 3\sigma_\sigma$ and $\overline{\sigma^2} \pm 3\sigma_{\sigma^2}$, it is taken as a positive indication of lack of constancy of the cause system, or, in other words, of the existence of an assignable cause of Type 1.

We may present the results of the application of this criterion in graphical form by means of the control charts shown in Fig. 3. Four of the 204 observed points and two of the 51 averages of four fall outside of their respective control limits. In other words, we get a positive test indicating the presence of assignable causes of variation for this set of data.

It will be noted, however, that the standard deviations and variances fall within their respective limits.

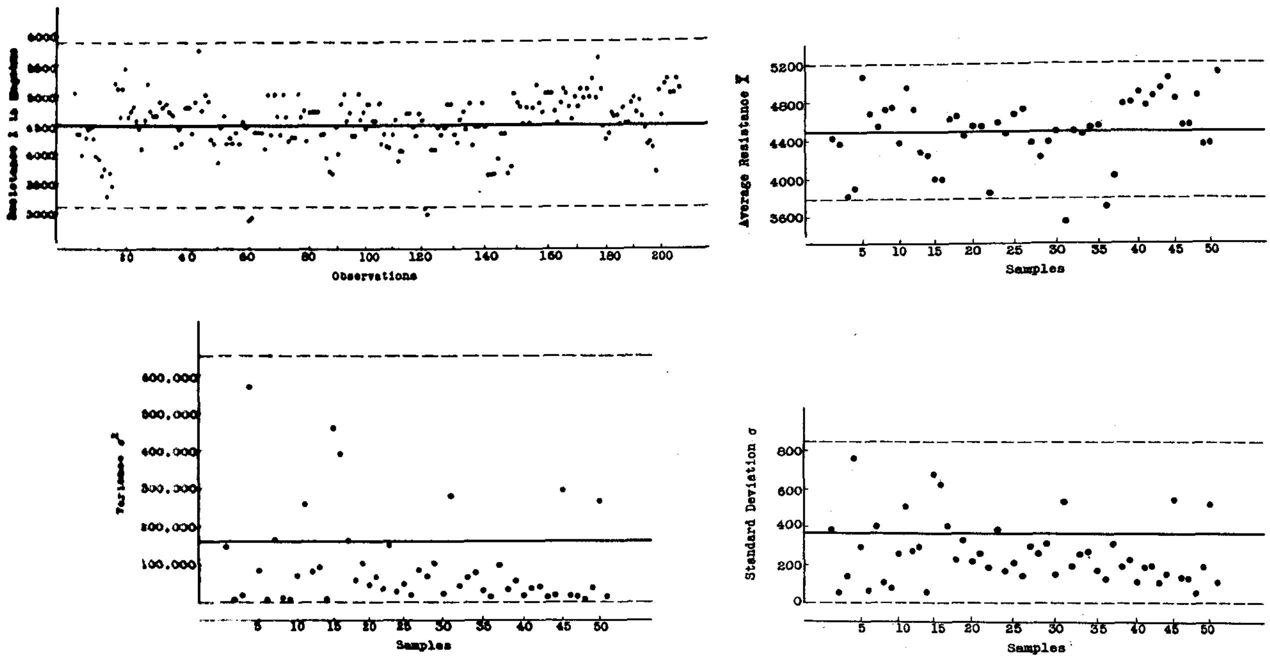


FIG. 3 - HOW CRITERION I INDICATES A VARIATION IN THE CAUSE SYSTEM WHICH SHOULD NOT BE LEFT TO CHANCE

Criterion II

Carry out the computations called for in the general data sheet of Fig. 4.

If the ratio $\frac{|d|}{\sigma_d}$ is greater than three, the test gives a positive indication of lack of constancy of cause system.

The method of applying this criterion to the data of Table 1 is indicated in Fig. 4. We find that the ratio $\frac{|d|}{\sigma_d}$ in this particular case is 12.72 and hence we get a positive indication of lack of constancy in the cause system underlying the data of Table 1 or, in other words, we again get a positive test for the existence of an assignable cause of Type 1.

Summary Statement

Two criteria have been defined. The requisite calculations are indicated in the data sheets of Figs. 1 and 4. These criteria have been applied to the data of Table 1 and both gave positive indications.

Calculation of $\frac{|d|}{\sigma_d}$

Number of observations $N = 204$
Size of sub-group $n = 4$
Number of sub-groups $m = 51$

Sample Number	Average $\bar{X}$ of Sample	$\bar{X}^2$	Variance σ^2 of Sample
1	4430.0000	19,624,900.0000	149,512.5000
2	4372.5000	19,118,756.2500	7,606.2500
3	3827.5000	14,649,756.2500	17,656.2500
- - -	- - -	- - -	- - -
51	5100.0000	26,010,000.0000	11,250.0000
Σ	229,407.0000	1,038,119,072.0700	4,832,876.1050
Av.	4498.1765	20,355,275.9229	94,762.2766

$$\sigma_{\bar{X}}^2 = \frac{\sum_{i=1}^{m} \bar{X}_i^2}{m} - \bar{\bar{X}}^2 = 20{,}355{,}275.9229 - (4498.1765)^2$$

$$\sigma_{\bar{X}}^2 = \underline{121{,}684.3586}$$

$$\overline{\sigma^2} = \frac{\sum_{i=1}^{m} \sigma_i^2}{m} = \underline{94{,}762.2766}$$

$$d = \frac{n}{n-1}\overline{\sigma^2} - \frac{m}{m-1} n\sigma_{\bar{X}}^2 = \underline{-370{,}122.4810}$$

$$\sigma_d = \left[\sqrt{\frac{2\ (mn-1)}{m(m-1)(n-1)}}\ \left(\frac{n}{n-1}\overline{\sigma^2}\right)\right] = \underline{29{,}107.6083}$$

$$\frac{|d|}{\sigma_d} = \frac{370{,}122.4810}{29{,}107.6083} = \underline{12.7157}$$

FIG. 4 - DATA SHEET FOR CRITERION II FILLED OUT FOR DATA OF TABLE 1

PART III

Detection of Predominating Cause or Group of Causes

Introductory Note

The two criteria to be discussed in Part III are of particular value when it is not possible to sub-divide the total set of N ($N \geq 500$) observations rationally into m different groups. It is obvious that the two criteria previously described cannot be used in such cases. We shall see later, however, particularly in Part V, that the two criteria to be described are less powerful than the two already given. Furthermore, it will be found that, in general, they cannot be applied when we have less than a certain number of observations. These facts emphasize the important point that, in the taking and recording of data, we should always give all available information that will assist in making it possible to divide the total group of observations rationally into sub-groups.

Typical Problem

A certain kind of instrument was manufactured at several different shops. A number of these from each shop were sent to a central testing laboratory and in this way a total of N = 7686 instruments were gathered together, no care being taken to keep those coming from a given shop separate from the others. The efficiency of each instrument was then measured and recorded, after which all of the data were grouped into nine cells, giving the frequency distribution reproduced in Column 4 of the data sheet in Fig. 5. Starting with this set of data just as they are given, how can we analyze them to determine whether or not the variations between instruments are such that they must be left to chance?

Criterion III

The requisite calculations are all indicated on the data sheet of Fig. 5. The steps are as follows:

a. Calculate the average $\bar{X}$, the standard deviation σ and the skewness k from the N observations and use these in the expression

$$\int_{X_1}^{X_2} \frac{1}{\sigma\sqrt{\pi}} \left[1 - \frac{k}{2}\left(\frac{(X-\bar{X})}{\sigma} - \frac{1}{3}\frac{(X-\bar{X})^3}{\sigma^3}\right)\right] e^{-\frac{(X-\bar{X})^2}{2\sigma^2}} \, dX \qquad (1)$$

to calculate[1] the theoretical frequencies within the c cell intervals into which the original data have been grouped.

b. Calculate the function

$$\chi^2 = \sum_{i=1}^{c} \frac{(y_i - y_i')^2}{y_i'} \qquad (2)$$

c. Determine, as indicated on the data sheet, the probability of obtaining a value of χ^2 as large or larger than that observed.

<u>If the probability P is less than .001 this is taken as a positive indication either that the cause system has not been constant or that there is a predominating cause or group of causes.</u>

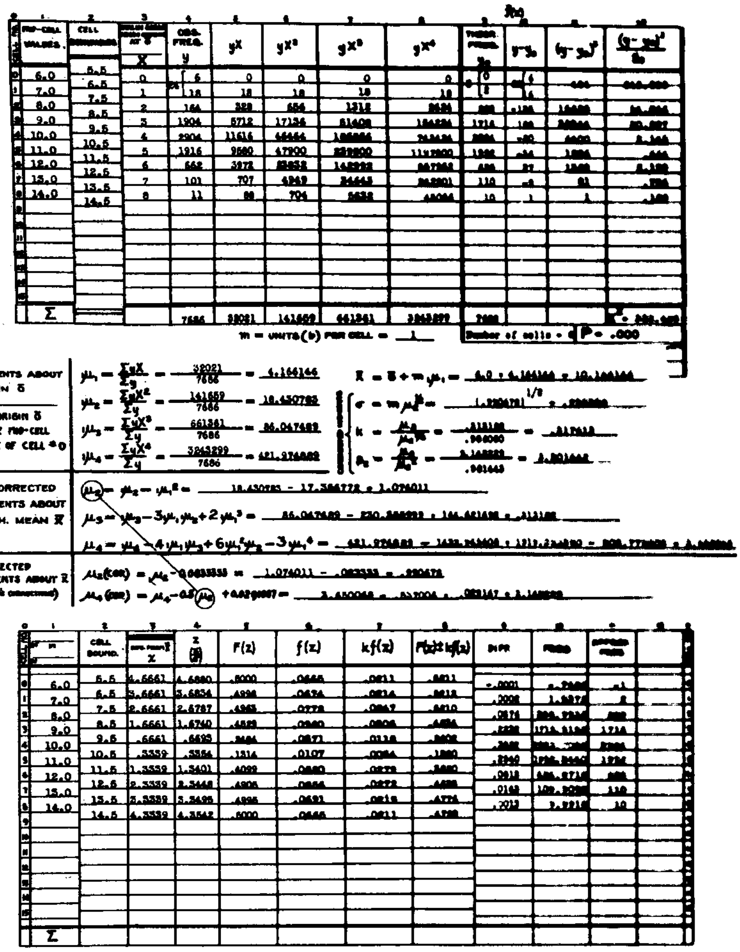


FIG. 5 - DATA SHEET FOR CRITERION III

The observed distribution given in Column 4 is represented graphically by the black dots in Fig. 6. The theoretical frequencies are represented by the smooth curve. There appears to the eye to be a very close fit between the theoretical curve and observed frequency. However, Criterion III detects what the eye does not see. Since the probability of getting a value of χ^2 as large or larger than that actually observed in this problem is exceedingly small, it was concluded that the differences between the instruments were such that they need <u>not</u> be left to chance. Wednesday 02.08.06

---

1. By means of tables for F(z) and f(z) referred to in data sheet and given in Bowley's "Elements of Statistics".

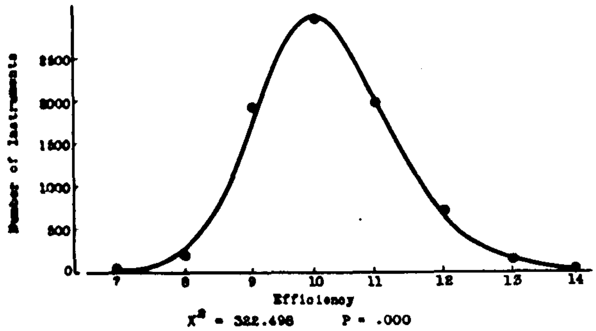


FIG. 6 - HOW CRITERION III INDICATES THE PRESENCE OF CAUSES WHICH SHOULD NOT BE LEFT TO CHANCE

Although the observations as originally presented were all mixed together without reference to whether they came from one shop or another, it later became possible to sub-divide these measurements according to the particular shop. It was then found that there was a definite evidence of a significant difference existing between the sub-groups of instruments as is indicated by the control chart for averages shown in Fig. 7. Later the assignable causes of the differences were discovered. Wednesday 23.05.06

Criterion IV

This test applies when two characteristics, say X and Y, have been measured on each of N (N greater than or equal to 25) pieces and it is known that any existing correlation between the two characteristics is attributable to one known group of causes. To determine if the known group of causes is an assignable group in respect to the variations in either X or Y, calculate the correlation coefficient

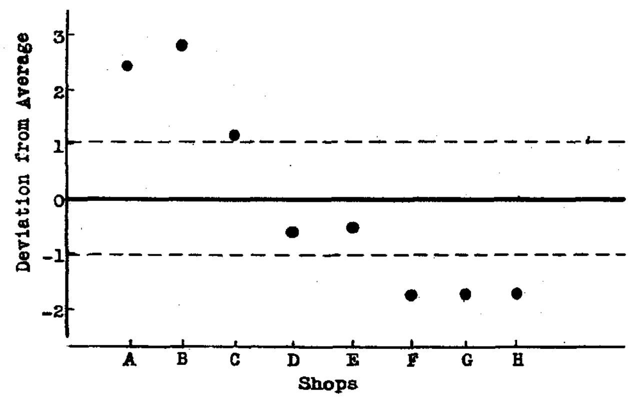


FIG. 7 - HOW THE CONTROL CHART HELPED INDICATE EVIDENCE OF TROUBLE

$$r = \frac{\dfrac{\sum_{i=1}^{N} X_i Y_i}{N} - \bar{X}\,\bar{Y}}{\sigma_X \sigma_Y} \qquad (3)$$

where $\bar{X}$ and $\bar{Y}$, σ_X and σ_Y are the arithmetic means and standard deviations respectively of the observed N data.

If r is greater than or equal to .5, it is taken as a positive indication that the possible common group of causes of variation in either X or Y is an assignable group.

Practical Application of Criterion IV

In a particular kind of apparatus made in large quantities and used extensively in the telephone plant, it is very desirable to secure a hardness of material in each instrument which differs from that in every other instrument by as small an amount as possible. In other words, it is desirable to leave no fluctuation in hardness to chance that can reasonably be controlled.

Since there were less than 500 observations we were not justified in applying Criterion III, although the irregularities in the frequency distribution indicated to the trained eye the possibility of the lack of constancy of the cause system. In this instance, however, it was possible to obtain more definite information because hardness measurements had been taken on two parts of each piece of apparatus. The N = 59 pairs of values of hardness on Parts 1 and 2 are given in Table 2. Since the material going into one part of a piece of apparatus came from different sources than that going into the other part and since, in general, there was nothing in common about the treatment of the material going into the two parts prior to the heat-treatment given these parts after they had been welded together, it was reasonable to believe that the only source of common cause for variation in hardness entered through the heat-treatment.

Sample Number	Hardness Part 1	Hardness Part 2	Sample Number	Hardness Part 1	Hardness Part 2
1	50.9	44.3	31	47.6	36.9
2	44.8	25.7	32	45.0	37.6
3	51.6	39.5	33	46.6	32.4
4	43.8	19.3	34	48.0	39.8
5	49.0	43.2	35	44.5	35.3
6	45.4	26.9	36	48.2	38.3
7	44.9	34.5	37	46.0	38.1
8	49.0	37.4	38	48.9	35.0
9	53.4	38.1	39	46.3	34.9
10	48.5	33.0	40	46.1	32.9
11	46.0	32.6	41	47.9	36.7
12	49.0	35.4	42	46.8	36.3
13	43.4	36.2	43	47.9	35.5
14	44.4	32.5	44	45.8	35.1
15	46.6	31.5	45	49.1	33.2
16	50.4	38.1	46	50.0	36.1
17	45.9	35.2	47	47.3	35.9
18	47.3	33.4	48	46.9	35.2
19	46.6	30.7	49	49.1	38.1
20	47.3	36.8	50	48.2	35.9
21	48.7	36.8	51	44.9	33.8
22	44.9	36.7	52	49.0	37.6
23	46.8	37.1	53	44.7	35.6
24	49.6	37.8	54	51.7	36.2
25	51.4	33.5	55	45.2	34.4
26	45.8	37.5	56	44.8	27.5
27	48.5	38.3	57	42.4	31.1
28	46.2	30.7	58	49.5	36.8
29	49.5	33.9	59	50.1	34.6
30	50.9	39.6			

Table 2 - Hardness of Each of Two Distinct Parts on Each of 59 Pieces of a Given Kind of Apparatus - Should Such Variations in Hardness be Left to Chance?

According to Criterion IV, the existence of a value of correlation coefficient equal to at least .5 would be taken as a positive test for the existence of an assignable group of causes of variation in the hardness of both parts of each piece of apparatus arising from the process of heat-treating. As can be readily verified from information given in Table 2, the correlation coefficient as derived from Equation 3 is .513. Hence Criterion IV gives positive evidence that the final heat-treatment constituted a predominating or assignable group of causes. Thursday 24.08.06

Part IV

General Fields of Application

Three Problems of Research and Development

We have already considered in Part II a typical problem of this nature which arises in the development of new materials as carried on in any industrial laboratory. For example, we may wish to obtain substitute materials or alloys to replace precious metals in contacts or we may wish to compare the deterioration of different materials under slightly different conditions such, for example, as the pitting of conduits or the years of life of telephone poles or comparative yields of different plants and so on. In fact, in securing any series of experimental observations, we realize full well that assignable causes of variation may enter.

The first problem which we shall consider is one typical of those proposed by two committees of the American Society for Testing Materials. It is introduced here primarily to illustrate the method which has been followed in preliminary analyses in the study of the physical properties of certain alloys.

Problem 1 - Each of five producers, C, D, G, W, and S furnish from five to seven specimens of twelve different special alloy die castings to each of eight different testing laboratories who in turn are to make tests for tensile strength, hardness, elongation and other physical and chemical characteristics. One of the objects of this very comprehensive series of tests to be extended over a period of several years is to provide data for setting standards upon the characteristics measured on each of the different alloy die castings.

Naturally we shall limit our consideration to only one of the many phases of this problem but the method itself applies in all of the different phases. The particular problem to be considered is that of examining the data for tensile strength obtained by one laboratory on the same number N pieces of material from each of the m=5 different producers to determine if the differences between the material furnished by one supplier and that of the others are such that they must be attributed to chance.

Since we have the total of N data rationally sub-divided into m sets of n each, we may use either Criterion I or II. The first of these, however, gives us not only an indication of the presence of assignable causes but also tells us what sets of data differ from the grand average by an assignable amount. Hence it is obviously desirable to use Criterion I. Before doing this, however, we shall indicate a way of making it somewhat more sensitive for the detection of assignable causes of Type I where each group of n observations comes from what may be a constant system of causes different, in general, from that from which each succeeding sample comes. In other words, this modified criterion is more sensitive when the essential conditions under which one group of n observations is taken differ from those under which succeeding groups of n observations are taken.

In the present instance one group of n measurements of tensile strength were made on one alloy obtained from one supplier. Another set of n measurements were made on the same alloy from another supplier and so on for each alloy. It is reasonable to believe that the essential conditionsof manufacture of the alloy may have varied from supplier to supplier and hence we are justified in using the Modified Criterion I, now to be described.

Modified Criterion I - Instead of calculating the standard deviation σ and the expected standard deviation $\overline{\sigma}$ as shown in the data sheet of Fig. 1, calculate the standard deviation of each of the m groups of n observations and find the average $\overline{\sigma}$ of the m values of the standard deviations thus obtained. To calculate σ used in the data sheet of Fig. 1, multiply the arithmetic means of the m values of standard deviation by $\frac{1}{c_1}$, where c_1 is the correction factor for a sample of size n as read from the curve of Fig. 2-a. The remainder of the calculations are carried out in the same way as for Criterion I except that we do not use the variance.

The results of such an analysis in the problem under consideration give for averages and standard deviations the control charts of Fig. 8 which show at a glance that there appear to be differences between the alloys supplied by certain producers which should not be left to chance.

<u>Problem 2</u>

In the development of methods of preservation of timber it becomes of interest to know the distribution of thickness of sapwood to be expected in poles of a given kind. It is also of interest to know whether or not there are assignable causes of variation in sapwood thickness traceable to such cause groups

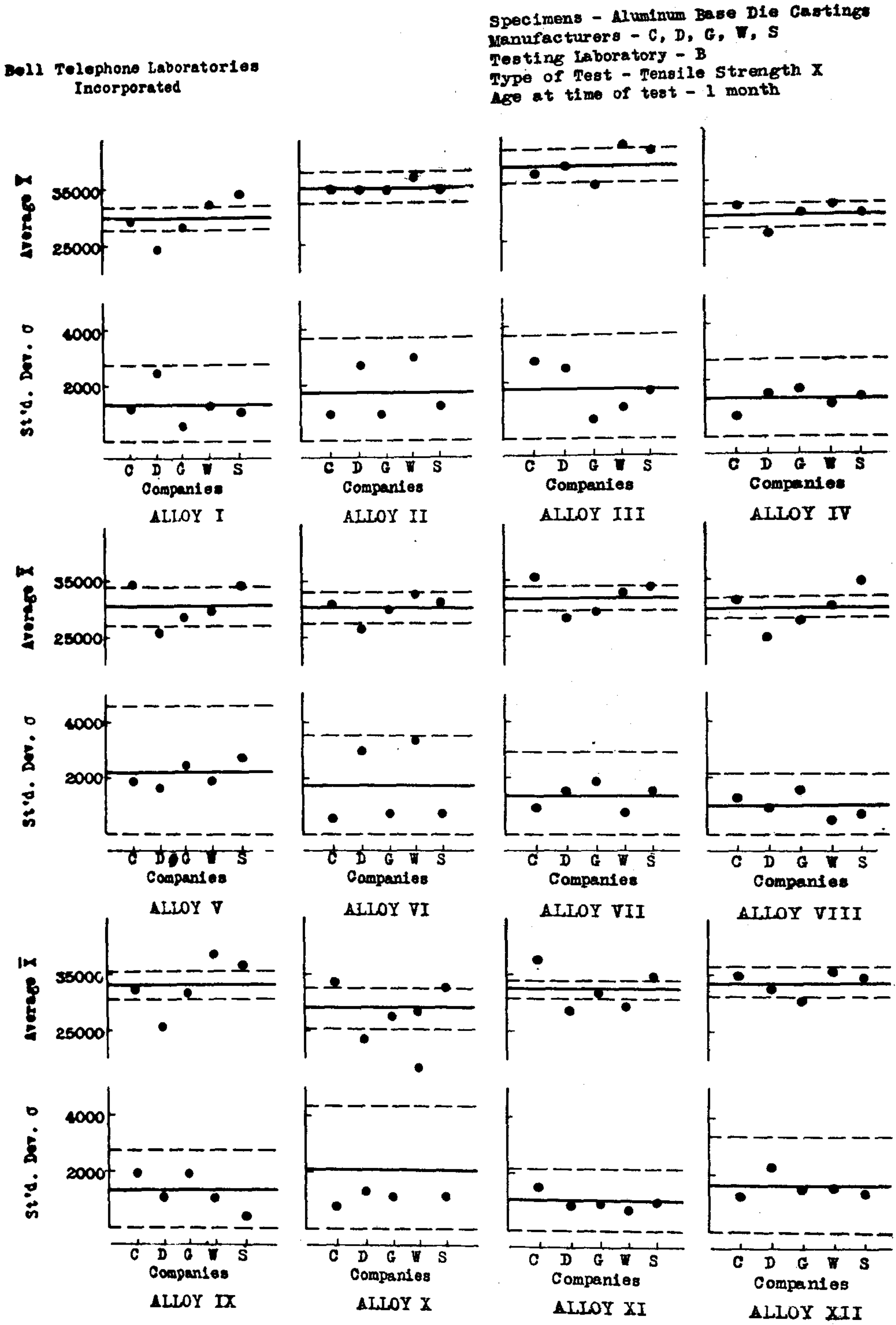


FIG. 8

as soil and climatic conditions.

At an early stage of the study of this problem a set of 1528 measurements of thickness of sapwood on as many chestnut poles became available. At the time these observations came to us, no additional information was available which would make it possible to rationally sub-divide the group of data. Hence it was necessary to make use of Criterion III for the purpose of detecting the evidence of assignable causes of variation. Monday 07.06.66

The smooth distribution obtained in the application of Criterion III is shown by the solid line in Fig. 9, whereas the observed distribution is represented by the black dots. It was found that the probability of obtaining a value of χ^2 as large or larger than that observed was much less than .001, thus giving a positive test for the presence of assignable causes.

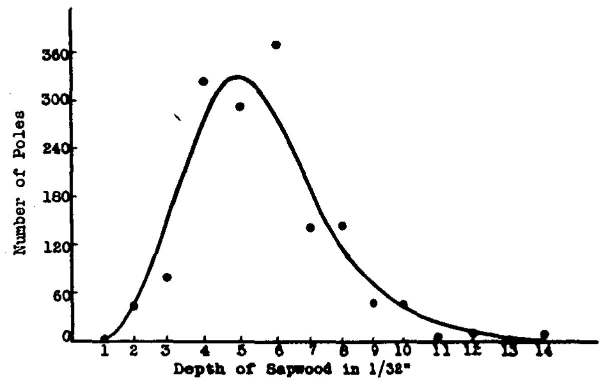


FIG. 9 - CRITERION III INDICATED PRESENCE OF ASSIGNABLE CAUSES AND THREE WERE FOUND

Upon the basis of these results, further investigation was carried on by making additional measurements and by going back to the original records which made it possible to rationally sub-divide the total group of data so as to check certain hypotheses. By these methods, three assignable causes were found as follows:

1. The men who made the measurements favored even numbers.

2. The thickness of sapwood was determined from borings. A considerable period of time elapsed from the date when the measurements were started to that when they were completed. During this time, there was definite shrinking of the borings, tending to decrease the thickness of sapwood.

3. The chestnut poles used in this experiment were taken from two sides of a mountain slope and when the measurements were sub-divided upon the basis of the locality from which they came, differences in the average depth of sapwood of the two groups were easily shown to be significant.

Problem 3 - A very important problem in civil engineering is that of constructing reservoirs or dams adequate to hold back flood waters.[1] In such instances we have a series of observations representing the run-off of a given area for a given interval of time, such as a day or a week, over a considerable period and from this series of say N observations we are to determine the maximum flood condition which must be provided for in the construction of the dam or reservoir.

- -

1. Transactions of the American Society of Civil Engineers, Paper #1622.

Stated in more general terms we are given a set of N measurements of some quantity controlled by an unknown system of causes. That is to say, the differences between the observations themselves are attributable to this unknown system of causes. Upon the basis of this set of N observations, we are called upon to predict the probability that a future observation will fall within a given range. As already noted, this problem is foremost in the setting of engineering standards or measurements of physical quantities in general.

It is not our intention here to discuss the details of the method of solution of this general problem or of the specific one except to point out that the value of our prediction of the future fundamentally rests upon our success in detecting the assignable causes of variation. For example, if we knew the system of causes underlying the variations in the N values of the measured quantity, we could use modern statistical theory to estimate the probability of future observations falling within the given range, although this particular problem is not discussed in this paper. If, on the other hand, the system of causes is not constant and the way in which it varies is not known, we cannot use modern sampling theory to estimate the probability that a future observation will fall within a given range. Hence, it becomes obvious that the first step in the solution of such a general problem as that stated above, is to make use of criteria for detecting lack of constancy of the cause system or, in other words, the presence of assignable causes of Type 1.

From what has already been said earlier in the paper it is evident that an attempt should be made to rationally sub-divide the data so that Criterion I or II can be applied. It is believed that when such steps are taken to detect the existence of assignable causes of Type 1 and when account has been taken of those that have been found, much better estimates, in particular, of the probability of a flood reaching a certain height, may be obtained than can be made otherwise.

Three Problems in Production

Problem 1 - The producer of manufactured goods wants to eliminate, insofar as possible, assignable causes of variation or at least to know what these causes are. Under modern conditions of quantity production, this means that the producer needs some easily adaptable method for the detection of the existence of assignable causes of variation through the application of criteria to observed measurements of quality of product.

This problem has been treated at some length in two reprints from these laboratories.[1] We have, in these two papers, a thorough explanation of the way in which Criterion I has been applied successfully in assisting in the control of manufactured telephone equipment. We shall not consider this problem further at this point but shall mention another practical problem in which the data are given in such a way that they cannot be rationally subdivided to make the application of Criterion I feasible.

Problem 2 - In the initial stages of the production of a certain kind of equipment in which the resistance was one of the important quality characteristics, the observed frequency distribution of results was that shown in Fig. 10-a, that is to say, the number of instruments having a given resistance were as indicated by black dots. The problem was to detect any evidence of the presence of assignable causes of variation in the method of production. If such causes existed, it was essential that they be found and controlled.

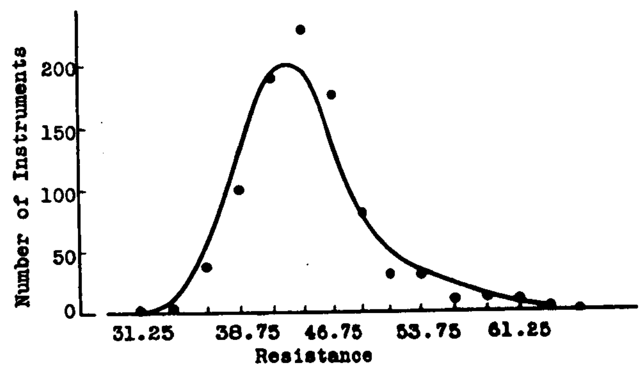


FIG. 10-a - POSITIVE TEST GIVEN BY CRITERION III INDICATED TROUBLE

This situation called for the application of Criterion III with the result that the probability found was less than .001, thus giving a positive indication of the existence of assignable causes of variation. An investigation was then started and soon indicated that one portion of the data had come from a different set of causes than another. The small group of instruments, supposed to have been affected by this change in the constant system of causes which took place during the latter part of the production of

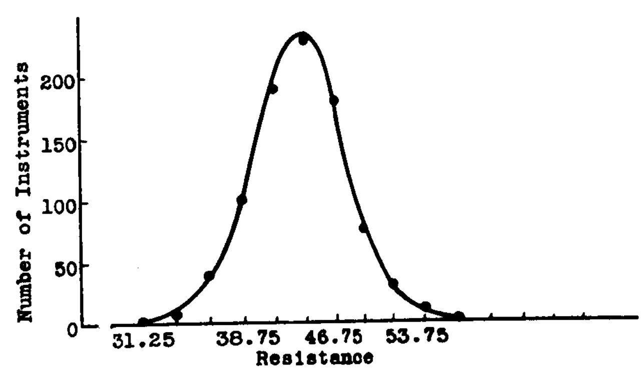


FIG. 10-b - TROUBLE REMOVED, CRITERION III GIVES NEGATIVE TEST INDICATING NO TROUBLE

---

1. Bell Telephone Laboratories Reprints B-223 and B-277.

the total group of instruments, was eliminated from the original group leaving us with the observed frequency distribution indicated by the black dots in Fig. 10-b. The application of Criterion III to this modified distribution gave a probability greater than .001 or, in other words, a negative test for the existence of assignable causes. Production was then continued upon the assumption that, having removed the one assignable cause or condition, the variations which existed between the instruments in respect to resistance were such that they should be left to chance.

Problem 3 - Not only is a producer interested in detecting the presence of assignable causes of variation so that he may, therefore, eliminate them or take them into account in the production of his material, but he is also interested in making reports on the quality of the product manufactured. For example, in the measurement of any quality characteristic X, the method of measurement may introduce a comparatively large error which in the sense of the present paper is an assignable group of causes. In any report on the quality of material, it is desirable that the original quality data be corrected for assignable errors of observation. This particular point has been treated in a separate paper, available in the reprint series of these laboratories.[1]

Before the man in charge of production passes the information along for the consideration of others interested primarily in getting a correct general picture of the run of quality, he usually desires to have some means of indicating the variations in quality which are sufficiently large to indicate that they should not have been left to chance. In other words, he is interested in presenting the quality data, corrected for errors of measurement, in such a way that the existence of any assignable cause of variation in the quality of the product would be apparent. He then adds information indicating the nature of the assignable causes wherever the cause has been found.

In other words, we have occasion to again make use of Criterion III on the original data corrected for errors of measurement, for the purpose, in this case, of giving a picture of the quality of product.

Two Problems of Business Research

We shall consider two typical problems, one in purchasing and one in

- -

1. Bell Telephone Laboratories Reprint B-186.

sales distribution.

Problem 1 - A purchasing engineer is interested:

a. In determining whether or not the product is controlled,

b. In detecting assignable differences between the products of different suppliers, and

c. In discovering the assignable causes of differences.

Material of a given kind used extensively within the Bell System was obtained from six different suppliers. For our purposes we may call the particular quality of this equipment X, and in this case it happened to be a measure of the effect of treatment on the material furnished to the different suppliers. In other words, a certain kind of material was distributed among six different treating plants and after treatment the product of each plant was measured for the particular quality X given it by the treatment of this plant.

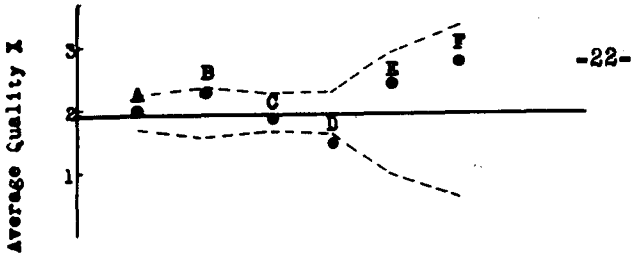
Fig. 11-a illustrates the use of Criterion I to indicate that at least one plant was assignably different from the others.[1]

Of course up to this time we had no indication as to whether or not there were assignable causes of differences only between the plants or whether or not the results given by a single plant indicated the presence of assignable causes. If the latter was the case, it might explain the observed differences between the plants. The application of Criterion III, Fig. 11-b, to the entire group of data from one of the plants gave a probability less than .001 and thus a positive test for the existence of assignable causes of variation within a single plant was obtained.

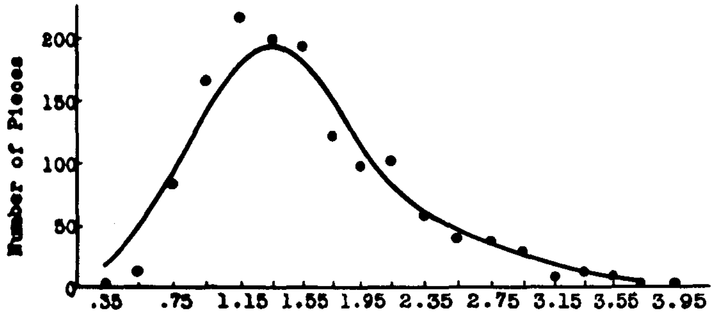
It occurred to the men who were in charge of the investigation to look for the assignable causes in the quality Y of the material furnished to the treating plants. A study of the correlation between the quality X after treatment and the quality Y of the material given the treatment, as illustrated in Fig. 11-c, proved to be the right type and the right magnitude to indicate that the quality Y was probably the source of the assignable difference between plants, and, therefore, possibly that the differences in the plants, in respect to the inherent property of treating the material, were no greater than

- -

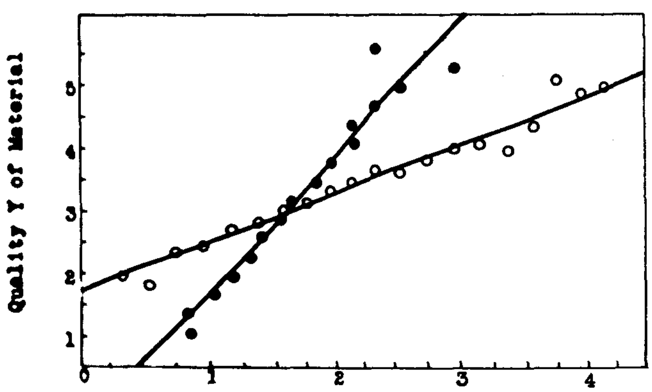
1. The variation in the width of limit lines arises from the fact that the sample sizes were different in the six different cases.



11-a - Criterion I Indicates Plants to be Assignably Different



11-b - Criterion III Indicates Presence of Assignable Causes in a Single Plant



11-c - Correlation Reveals What Proved to be an Assignable Cause

FIG. 11 - A PURCHASING ENGINEER'S PROBLEM

should be left to chance. Additional information and further study revealed this lead to be correct.

Problem 2 - The work of the Food Research Institute of Stanford University[1] shows that the loss from stale bread constitutes an important item of cost for a great number of wholesale bakeries and some retail bakeries as well. They estimate that this factor costs the people of the United States millions of dollars per year. Every sales manager of a baking corporation is interested, therefore, in detecting and finding assignable causes of variation in the returns of stale bread so that he may reduce to a minimum the loss arising in this way. In other words, he wishes to make use of his data to determine the variations which must not be left to chance.

Some time ago it became possible to secure the weekly record of return of stale bread for ten different bakeries operating in a certain metropolitan district. These observed results are shown by the heavy irregular lines in Fig. 12. An application of Criterion II showed that the differences between the bakeries were too large to be left to chance. Furthermore, the application of Criterion I to each of the bakeries taken as a unit gave positive indication of the existence of assignable causes within each plant. For example, in the control charts shown in Fig. 12, it is obvious that many of the observed percentage returns fall outside of the dotted limit lines, thus indicating the existence of assignable causes.

- -

1. Davis, J. S. and Eldred, E. W., Stale Bread as a Product of the Baking Industry.

It follows from theroretical considerations that the average percentage return should, in general, be least for that bakery freest from assignable causes. It is, therefore, interesting to note that for the bakery having the lowest percentage return, 1.99%, the number of points outside of the control limits is, in general, less than the corresponding number for the other plants.

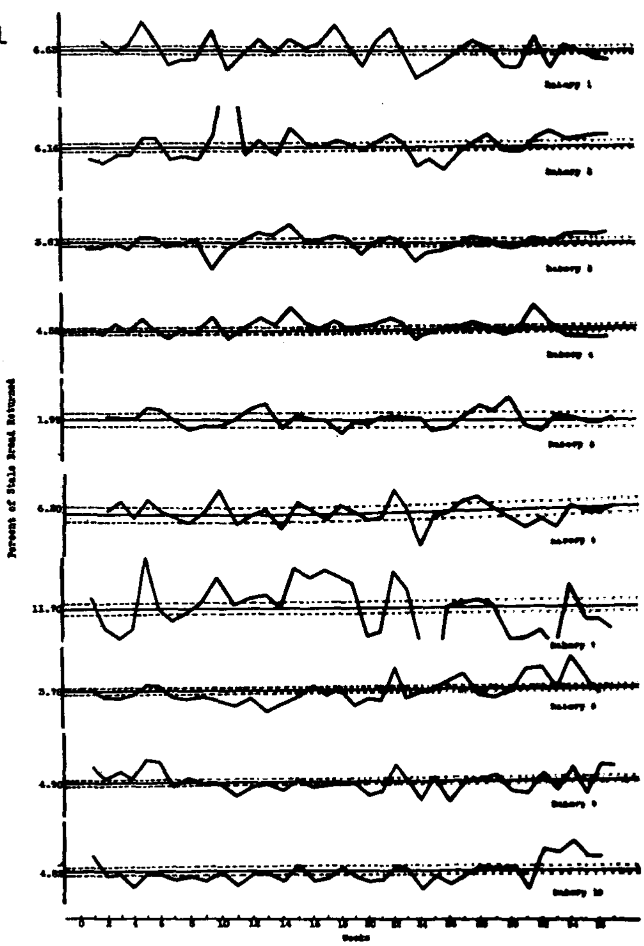


FIG. 12 - HOW CRITERION I SHOWS THAT THE DIFFERENCES IN THE OPERATION OF THE TEN PLANTS SHOULD NOT BE LEFT TO CHANCE

The daily use of this simple form of chart would reveal to the manager in charge evidence of assignable causes which probably could be found and possibly eliminated, thus, in general, making possible a reduction in the average percentage return of stale bread, thereby effecting an appreciable economy. The use of the chart of Fig. 12 as a report on the quality of product is also obvious, for it gives definite indication of the existence of assignable causes of variation which should not, in general, be left to chance. Monday 07.08.04

PART V

Critical Discussion of Results

Thus far we have considered the method of applying certain criteria to determine whether or not a thing should be left to chance and have illustrated the methods of analyzing the data so that even a novice in the field can apply the criteria. No attempt has been made to give the details of the theory underlying these methods because to do so is beyond the scope of this paper. On the other hand, those familiar with statistical theory as well as those who apply the criteria given above may be interested in an outline of some of the more recent important developments in the theory of sampling which justify the use of the criteria as outlined in this paper.

The reader who follows through the discussion given below will learn to his satisfaction that, whereas on the one hand criteria outlined in this paper may fail to detect the presence of assignable causes and thus indicate that a thing should be left to chance where it might not be wise to do so, on the other hand, the likelihood of the criteria indicating that a thing should not be left to chance even though it should be so left is very small. In fact, during several years of commercial experience, no instance has come to our attention where a test has indicated the existence of assignable causes where further investigation did not reveal the justification of the assumption that assignable causes existed.

We shall first consider the four criteria one by one, outlining the theory sufficiently to justify the use of the criteria subject to indicated limitations. We shall then consider certain generalized problems in the theory of sampling to indicate that other methods of setting up criteria are available, although these methods are not so efficient, in general, and are for the most part considerably more involved than the very simple tests discussed above.

Fundamental Basis for All Criteria

If a variable X is continuous within limits L_1 to L_2 and can take on a value within the interval X to X + dX with a probability p' given by

$$p' = f'(X, \lambda_1', \lambda_2', \ldots \lambda_{m'}') \, dX,$$

where
$$\int_{L_1}^{L_2} f'(X)\,dX = 1,$$

we say that $f'(X, \lambda_1', \lambda_2', \ldots \lambda_m')\,dX$ is the probability distribution of X. Sometimes it is referred to as a continuous chance variable. In a similar way, if a variable X may take on only a finite number N different values, $X_1', X_2', \ldots X_i', \ldots X_N'$, with corresponding probabilities $p_1', p_2', \ldots p_i', \ldots p_N'$, such that

$$\sum_{i=1}^{N} p_i' = 1,$$

we have defined the discontinuous probability distribution of X or, as it is sometimes called, the discontinuous chance variable. It is also customary to speak of a chance variable as a universe and it is in this way that it is pictured in Fig. 13.

In the development of the methods considered in this paper, we have defined a constant cause system to be one such that it gives rise to a chance variable or, in other words, such that it gives one or the other of the typical universes shown in Fig. 13.

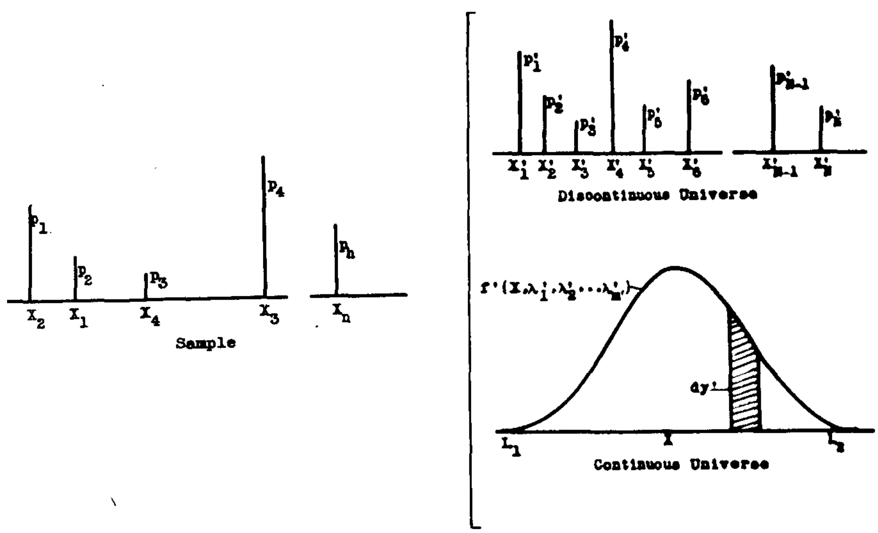


FIG. 13 - A FUNDAMENTAL CONCEPTION IN SCIENTIFIC WORK

Now, if we make n observations of X as given by such a cause system, we get an observed probability distribution such as that indicated to the left of Fig. 13, where, of course, p_i represents the fraction of the total number n observations having the value X_i, if the variable is discontinuous, or p_i

represents the fraction falling within the corresponding differential range about X_i if the variable is continuous. In our future considerations we shall deal only with discontinuous universes or chance variables because, even in practice, it would not be possible to observe a continuous distribution. Beyond this point, we shall not go into a consideration of the theory of causation but instead shall leave this subject for discussion elsewhere. Thursday 10.08.06

We are now in a position to invoke modern sampling theory to tell us what we may expect to get in the way of certain functions of samples of size n provided they are drawn from a known universe. In the first three criteria discussed above, we have assumed that the observed probability distribution for the entire group of data constituted a universe and have calculated the probability of getting certain functions of samples of size n within certain ranges upon the basis of the assumption that they camefrom the universe characterized by the average, standard deviation, skewness and kurtosis of the total group of observed data.

Now, if we have a sample of n observations, we can characterize this sample by means of the above mentioned four functions calculated for the data of the sample. In addition to these, we might use any other function of the observed data, such as the mean or the maximum plus the minimum divided by two or any moment of the distribution and so on. Since we assume that the chance variable is discontinuous and varies only within finite limits at least for all the functions which we have occasion to use, there is only a finite number of possible values of any function obtainable from all of the possible samples of size n that may be drawn from a given universe of effects of a given cause system.

Before going further, let us consider just a very simple illustration. Suppose that the variable X may take on only the values, 1, 2, 3, 4, with equal probabilities of $\frac{1}{4}$ as shown in Fig. 14a and suppose that we take all possible sets of four observations that the cause system giving rise to this chance variable can produce. For example, we might get the four values of X: 1, 1, 1, 1; or 1, 2, 3, 4; or 1, 4, 1, 1 and so on. In fact, just a little computation shows that there are 256 such possible samples of four. Just a little arithmetic shows that the possible distribution of the averages of these samples of four is given by Fig. 14b. The distribution of medians is given by Fig. 14c and so on for the other distributions therein indicated.

For the reader who is not familiar with modern sampling theory it would be intensely illuminating to carry through the simple calculations required to obtain the information given in Fig. 14. If, however, this same reader attempts to carry through this operation for a sample of size n=100, he will find that the arithmetic labor involved is almost prohibitive. It is just this difficulty, of course, that modern sampling theory gets around. In other words, the fundamental concept underlying the theory of sampling is, as illustrated by this example, extremely simple. Hence, in what follows, the reader should find no difficulty in following the logic.

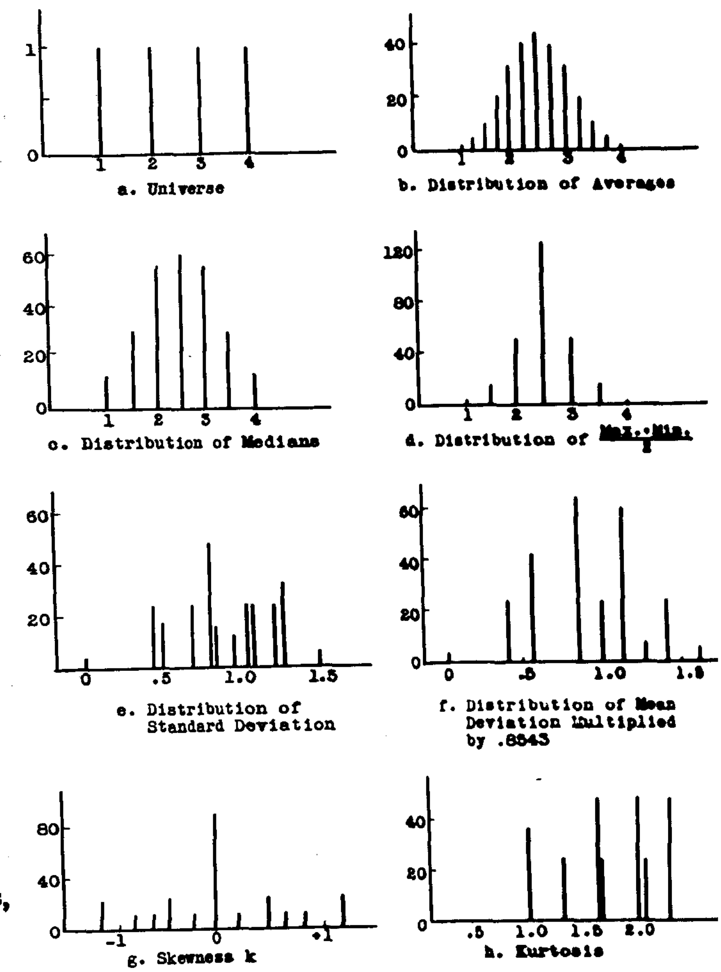


FIG. 14 - TYPICAL OF WHAT THE ANALYST CAN TELL IN ABSOLUTE TRUTH

More specifically, given any universe of effects of a constant system of chance causes, there will be, in general, K different possible samples of size n which we may observe. It is theoretically possible to set down the n values of X associated with each of these K possible samples but, as we have already pointed out, it would take too long to do this by simple methods such as were used in obtaining Fig. 14. Let us assume that θ is any function or statistic of the n data constituting a sample. In general, there will be K values of θ corresponding to the K different samples of size n. As an example, the function θ in the case of Fig. 14b is the average of the sample and the 256 averages are distributed as indicated in Fig. 14b. In a similar way for Fig. 14e, θ is the standard deviation for a sample of size four and the distribution of possible standard deviations is that given by Fig. 14e.

For our purpose, the important thing that sampling theory does is to show us how to calculate the expected value $\bar{\theta}$ of a given parameter of a sample of size n and certain functions of the distribution of this parameter. The most important characteristic of the distribution of a statistic θ so far as our work is concerned is that of the standard deviation σ_θ of the statistic in samples of size n.

Knowing the expected value, $\bar{\theta}$, of a statistic θ and the standard deviation σ_θ of the same statistic in samples of size n, the beautiful Tchebycheff theorem, derivable by the simplest algebra, shows that the probability P of observing a value of θ within the range of $\bar{\theta} \pm t\sigma_\theta$ is always equal to or greater than $1 - \frac{1}{t^2}$. That is to say

$$P \geq 1 - \frac{1}{t^2}.$$

Now, if it so happens that the distribution of θ is approximately normal so that the skewness and kurtosis of this distribution are approximately equal to zero and three respectively, then we can say that the probability that θ will lie within the range of $\bar{\theta} \pm t\sigma_\theta$ is given by the normal law integral.

In the criteria developed above we have considered the range corresponding to t=3. Now we see that, for this particular value of t, the probability associated with the range is always greater than or equal to .8889 and for a statistic distributed normally it is .9973.

Let us now see how the four special criteria are derived.

Criterion I

Obviously we take the observed average, standard deviation and kurtosis of the total of N observations as characterizing the average $\bar{X}'$, standard deviation σ' and kurtosis β_2' of a universe. Taking these values to characterize the universe, sampling theory shows that the expected values $\bar{X}$ and $\bar{\sigma}$ of averages and standard deviations of samples of size n drawn from this universe are those shown in the data sheet of Fig. 1, irrespective of the values $\bar{X}'$, σ' and β_2'. Similarly the standard deviations of these functions are those given on this same data sheet. We cannot, however, calculate the expected standard deviation of the standard deviation in samples of size n except for the case when the assumed universe is normal.

It follows that we may use Criterion I on a single observation X, average $\bar{X}$ and variance σ^2 in samples of size n, irrespective of the universe, and that we may use this criterion also on the standard deviation, if the universe is sufficiently near normal, as it is in most cases, to justify our use of the expected value of the standard deviation and the standard deviation of this standard deviation in samples of size n as known for the case of a normal universe.

The probability associated with the range $\bar{\theta} \pm t\sigma_\theta$ for any one of these statistics of a sample of size n may be estimated by means of Tchebycheff's theorem. Of course, if the distribution in the postulated universe is practically normal, the probability associated with a single observation will be approximately equal to that given by the normal law.

For the case where θ is the average of a sample of size n we have a very interesting situation since, irrespective of the distribution in the postulated universe, the distribution of averages in samples of size n will be approximately normal. For example, if $\beta_{1\bar{X}}$ and $\beta_{2\bar{X}}$ represent the square of the skewness and the kurtosis for the distribution of averages of samples of size n, drawn from a universe characterized by β_1' and β_2', the following relationships will hold:[1]

$$\beta_{1\bar{X}} = \frac{\beta_1'}{n}$$

$$\beta_{2\bar{X}} = \frac{\beta_2' - 3}{n} + 3.$$

Even for n = 4, the values of skewness and kurtosis, as given by the above equations, for averages of samples of size n approach zero and three quite closely for all the assumed universes with which we have had occasion to deal. This fact is very important because it shows that the probability associated with a

1. Church, A.E.R., Means and Squared Standard Deviations of Small Samples, Biometrika, Vol. XVIII, pp. 321-394, 1926. See particularly Page 344.

given value of t is in the case of the average more nearly independent of the characteristics of the universe than it is for any other statistic. For example, it is reasonable to assume that the probability associated with the limits corresponding to t = 3 for the case of averages is approximately .99 unless the universe from which the sample is drawn has very large values of skewness and kurtosis.

The correction factors as given in Fig. 2 for the determination of the expected standard deviation and the standard deviation of the standard deviation in samples of size n drawn from a normal universe are based upon the work of Pearson.[1]

Criterion II

Underlying Criterion II there is a very simple philosophy. We start out with the assumption that on the average a certain function d, as indicated on the data sheet of Fig. 4, is zero if the observed data have been obtained from a constant system of causes, whereas the expected value of this function is less than zero if not all of the m samples come from the same constant system of causes and more than zero if all of the observations in a sample of size n do not come from the same cause system but each of the samples of size n are selected in the same way from the different cause systems. For the purpose of our present discussion, we may say that on the average d is less than zero if the cause system is changing from sample to sample and greater than zero if the cause system changes from observation to observation within a sample but all samples of size n are chosen from the same set of cause systems.

It follows naturally that even though m samples of size n are drawn from a constant cause system, the observed value d may not be zero. In fact, d itself is a chance variable, the distribution of which is not known. However, an approximation to the standard deviation σ_d of d based upon the assumption that X itself is distributed normally can be obtained by making application of some recent sampling theory. It is this value of standard deviation of d that

---

1. Pearson, Karl, On the Distribution of the Standard Deviation of Small Samples, Biometrika, Vol. 10, Part 4, 1915, pp.522-529.

has been used in applying the criterion.[1]

Since, in general, X is not distributed normally, the standard deviation obtained by the method given in the test is obviously in error. Furthermore, since we do not know the skewness and kurtosis of the distribution of d, we must fall back on the use of Tchebycheff's criterion to give us the lower bound to the probability associated with the range corresponding to $\frac{|d|}{\sigma_d} = 3$.

In the face of these limitations, however, the criterion is one of the most powerful in the stock in trade of the analyst.

Criterion III

Of course, it will be recognized that the basis of this criterion is the assumption that a multiplicity of causes, each of which produces practically the same effect as any other upon the variation of a quantity X, will give an approximately normal distribution of X if the number of causes is large enough and all the causes act independently. We must hasten to set down, however, two very definite limitations to the use of this criterion.

1. It can easily be shown that we may have a distribution composed of a set of observations arising from more than one constant system of causes and yet obtain a negative test by means of Criterion III. In other words, it is admittedly easily possible for the method to fail to detect causes which should not be left to chance although such failure is very unlikely as becomes apparent when one considers the engineering conditions which would have to be fulfilled in such a case.

In our own work we have found a negative test only in four instances. In every one of these cases, since the test was not absolute and since the problems were of such great commercial importance a search was made for assignable causes of variation regardless of the fact that the test was negative. In not a single one of these four instances, however, did we find assignable causes of variation.

2. It is also possible to set up a theoretical condition under which we would get a positive test for the presence of causes which should not be left

- -

1. The formula for d may be obtained easily from a slight extension of the work given by Coolidge on Pages 66-70 of his book on "Probability". In obtaining this extension and the standard deviation σ_d, I acknowledge the helpful cooperation of Mr. F. W. Winters.

to chance even though all of the data come from a system of causes which should be left to chance Our own experience has shown, however, that in the numerous cases where we have found a positive test, we have always found causes of variation which should not be left to chance.

The limitations of this test lead us to emphasize once more the point which has already been made viz., that date should be taken in such a way that either one or both of Criteria I and II can be applied, thus practically obviating the necessity of relying upon Criterion III.

Criterion IV

Little more need be said about the use of Criterion IV. It is of interest, however, to know that, if the causes of variation in each of two quantities X and Y produce equal effects and act independently one of another, the correlation coefficient r is a measure of the commonness of causation. In other words, if m causes are common, where there are m+n causes of variation in one of the two quantities and m+s causes of variation in the other quantity, then

$$r = \frac{m}{\sqrt{m+n}\,\sqrt{m+s}}$$

As has already been pointed out, Criterion IV may be used even though all others have failed to reveal the presence of assignable causes of variation in either X or Y. Under these conditions, the correlation coefficient r would have the meaning given by the above equation. It is for this reason that in those cases where we appeal to Criterion IV we are often justified in giving this physical interpretation to r. Of course, it is not possible to give such a simple physical explanation of the correlation coefficient r in the majority of problems where it is applied outside of the field now under consideration.[1]

Assurance Given by the Criteria

It will have been observed that, in the application of Criteria I and II, we said that a thing must be left to chance if the probability that it could have arisen as a chance fluctuation under the assumed sampling conditions was less than a certain amount. In the same way we have made the empirical choice of probability .001 in the use of Criterion III. As a case in point, we say

1. This point and all others dealing with the interpretation of phenomena in terms of systems of causes will be the subject of a paper appearing elsewhere.

that a thing should not be left to chance if $\frac{|d|}{\sigma_d}$ is equal to or greater than 3. Naturally the choice of three is empirical in just the same way that there is no absolute criterion for determining what probability is high enough to make it necessary for us to take action. For example, in the game of chance it is up to the player to determine how much risk he is willing to take. Evidently, if we set limits in such a way that the probability of looking for trouble even though it cannot be found is high, we may in this way waste a lot of time and money. On the other hand, if we set the limits in such a way that we get a negative test, we may readily overlook sources of trouble which should be discovered. Our own experience indicates that in the majority of problems the empirical choice of probability employed by the criteria is satisfactory.

Of course, in certain problems of research one may be interested in reducing the limits of fluctuation of a statistic so that there will be a fifty-fifty chance of finding an indication of trouble even though it does not exist. In certain instances we may be willing to reduce the limits still further.

Let us consider an illustrative case. The operation data for a certain gas plant for one month expressed in terms of arbitrary thermal units per cubic foot of gas produced from oil by cracking are given in Fig. 15, in the order in which they occurred. Ideal operation calls for as high and as nearly constant a value as can economically be obtained.

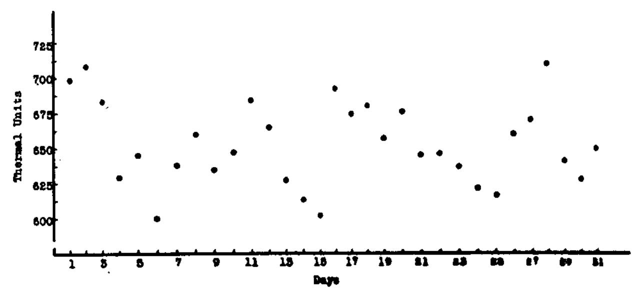


FIG. 15 - SHOULD SUCH FLUCTUATIONS BE LEFT TO CHANCE?

The following question was raised by the Director of Research of the large organization interested in these results:- "If I understand the methods of statistics correctly, it should be possible to determine from these data whether a large or a small number of causes are effective in producing the observed fluctuations and hence whether or not it should be reasonable to expect that a marked improvement in the product could be effected by controlling one or more of the causes of variation. Am I right in this interpretation of the possibilities of statistical methods?"

This problem is quite similar to the first one discussed in Part II, except that we are dealing with fewer observations. Grouping the observations into seven samples of four each, leaving one sample of size three, and making application of Criterion II, we find that

$$\frac{|d|}{\sigma_d} = 1.94.$$

That is to say, we fail to get a positive test for the presence of causes which should not be left to chance in the sense discussed in Part II. However, had the cause system remained constant the probability of getting a value of this ratio as large or larger than that observed is only about .052. Hence, it appears to be very unlikely that the cause system was constant and since it is so important to eliminate all causes which should not be left to chance in this particular problem it was decided to look further for trouble. This action was also qualitatively supported by the fact that the observed distribution of the thirty-one values was such that it might readily have come from a rectangular universe which, if true, would have necessitated the presence of causes which should not have been left to chance.

This problem is only typical of many which arise in engineering research and development. Thus in the early stages of production it often becomes necessary to reduce the limits so that causes which should not be left to chance will be more readily discovered. Then later, after the process of production has become well established, it will likely be found that the criteria as suggested will prove satisfactory.

General Considerations

Obviously other criteria might be established. We may illustrate this point by considering Criterion I. What this criterion really depends upon is the estimate of the probability of getting a value of some function such as average, standard deviation, etc., of the data in a sample of size n lying outside a certain range, provided this sample came from a constant system of causes characterized by $\bar{X}'$, σ', k' and β_2' of the total group of observed data. Of course, it is very unlikely that the values of average $\bar{X}'$, standard deviation σ', skewness k' and kurtosis β_2' observed for the total group of N observations will be identically equal to these same functions of the possible universe of effects of the cause system assuming the cause system actually to be constant.

We might, therefore, attempt to set up criteria which would detect when a sample cannot reasonably have come from the assumed constant but unknown system of causes. Here you see we get into the trouble of trying to correct the observed values of average, standard deviation, skewness and kurtosis so that they will give us the true values corresponding to the unknown constant system of causes. This is a problem fraught with many difficulties and so far as we know the hope of rigorous solution is very faint indeed. Even though we corrected the factors by any chosen method, all that we could rigorously say would be that the conclusions would be true if our method of correction were the right one.

All of this means, simply, that any criterion merely states the following proposition: If so and so is true, then something follows. This is exactly the same situation that we have in all engineering and scientific work. For example, after a man has designed one of the modern suspension bridges, all that he is justified in saying is that the bridge will support the load that he says it will support if material behaves as the assumed laws of physics and engineering say it will behave. Monday 21.08.06