# NOTES ON SOME FUNDAMENTAL INSPECTION PROBLEMS

By

W. A. SHEWHART
Bell Telephone Laboratories, Inc.

## INTRODUCTION

In this paper, I attempt to survey the field of inspection, to set down some of the more important problems, to see how far we have gone in solving these problems and to see how far we have to go.

Part I of this paper outlines three fundamental objects of inspection, Part II outlines four steps that must be taken in attaining each one of the three objects and Part III outlines some of the fundamental elements of inspection theory practically sufficient for attaining the first object of inspection and requisite for attaining the other two objects.

I am indebted to several of my colleagues for many helpful criticisms, the consideration of which has indicated the advisability of adding some rather lengthy footnotes to clear up certain points of a technical nature. Furthermore I have used footnotes and appendices to present material which is essential to a clear understanding of some of the theoretical problems involved and to indicate some of the limitations of the theory.

The three objects of inspection are expressed in terms of quality: i.e., detection of non-uniformity of quality, finding causes of non-uniformity of quality and setting economic standards of quality. Naturally it is essential that we have a clear understanding of the meaning of quality.

In general quality has been used in either of two senses: one referring to those characteristics which make a thing what it is, the other referring to the goodness or value of a thing. In either sense, however, quality is fixed when the magnitude of those characteristics required to define the article are fixed.[1]

It is customary practice to specify that units of a given kind of product shall possess certain characteristics whose magnitudes shall lie within fixed limits[2] or tolerances. The statement of the required characteristics

---

1. Thus quality Q is some function $\Psi$ of those characteristics X,Y,Z... required to define a thing.

2. Thus a unit is of standard quality if its characteristics X,Y,Z... fall within intervals $X'\pm\delta X$, $Y'\pm\delta Y$, $Z'\pm\delta Z$,... where $\delta X$, $\delta Y$, and $\delta Z$ are the tolerances. Another meaning of tolerance will be given later.

and their respective tolerances for a unit of any given kind of product defines the quality standard for that unit.

## PART I

## THREE OBJECTS OF INSPECTION

### 1ST OBJECT: DETECTING NON-UNIFORMITY OF QUALITY.

No matter how much care is taken in defining the production procedure, a manufacturer realizes that it is impossible for him to make all units of a given kind of product identical one with the other. This is equivalent to assuming the existence of non-assignable causes of variation in quality of product.[3] Of course random fluctuations in such factors as humidity, temperature, wear and tear of machinery, and the psychological and physiological conditions of those individuals engaged in carrying out the manufacturing procedure may give rise to some of these apparently uncontrollable variations. Knowing these things, the manufacturer effectively contents himself with trying to produce a product which is uniform - one which does not vary with time or rather one which does not vary from one period to another by more than an amount which may be accounted for by the system of non-assignable causes producing variations independent of time.

I want to pause at this point to make clear the significance of the terms "assignable causes" and "non-assignable causes" as they will be used in this paper. Suppose you and I were each given an opportunity of firing 100 rounds at a target. We all know what would probably happen - none of us would hit the bull's-eye every time. Thus charts A and B in Fig. 1 might represent two of our targets. Possibly some of the shots would hit within the first ring, others within the second ring, and so on. Each of us has a more or less

---

3. Let us consider the case where quality is determined for all practical purposes by one characteristic X. The standard quality will then be anything within the range $X' \pm \delta X$. Even when we do our best to control the manufacturing process, units of product will be produced having characteristics $X_1$, $X_2$, ... etc. where, in general, $X_i \neq X_j$ and we attribute the observed differences between units to non-assignable causes.

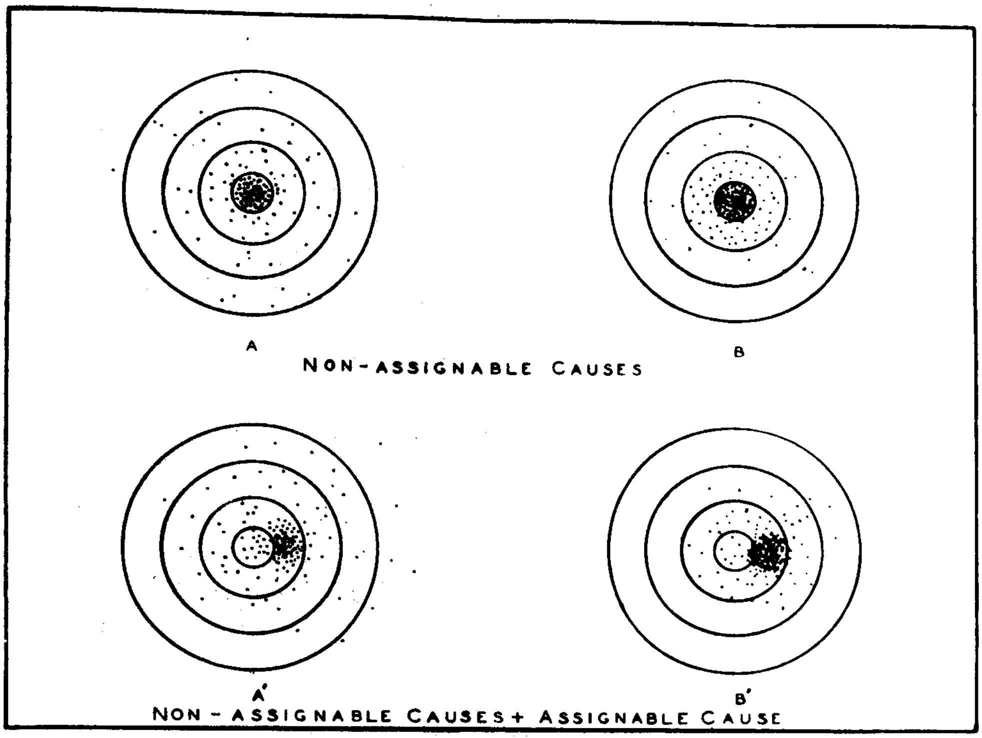



Fig. 1 - SCHEMATIC DIAGRAM SHOWING EFFECTS OF CAUSES.

definite picture of some of the possible reasons why none of us would be successful always in hitting the bull's-eye. We probably cannot assign the reasons or causes for our missing the bull's-eye in any particular instance - the causes of our missing are non-assignable. Suppose, however, that as a result of our experiment in shooting at the target, we found that every one of us tended to shoot to the right of the bull's-eye as is shown in charts A' and B' of Fig. 1. Naturally we would feel that there was some discoverable cause for this general tendency, i.e., we would feel that the observed effect could be assigned to some particular cause.

Of course the object in trying to detect assignable causes is obvious, because it is only through the control of such factors that we are able to improve the product. On the other hand there is no excuse for trying to ferret out or assign some cause for a fluctuation in product which is no greater than that which can be accounted for by the non-assignable causes just as there would be no excuse for trying to find the exact manner in which all of the causes contributed to our missing the bull's-eye in the analogous case of target practice just considered.

Here, then, is the problem. <u>When do the observed differences between the product for one period and the product for another period indicate non-uniformity?</u> When do the differences in quality of manufactured product observed from one period to another indicate fortuitous, chance or random variations produced by non-assignable causes which we cannot reasonably hope to control without radically changing the whole manufacturing process; and when do the observed differences in quality indicate the possible existence of assignable causes which we can reasonably hope to find and control: i.e., when is a product uniform and when is it non-uniform?

An example may serve to make this problem clear. Twelve hundred and fifty instruments were selected each month from a product manufactured in quantities of approximately 2,000,000 per year. The quality, as defined by a characteristic X, was measured on each of the instruments in a month's sample and the results of such measurements are presented in the frequency polygons Fig. 2. Are we to judge from the information given in Fig. 2 that the product was uniform throughout the year?

Obviously no two polygons are the same in respect to average, dispersion and shape, but of course we wouldn't expect them to be the same even though the product were uniform any more than we would expect to find two targets showing the same distribution of shots even though the same individual fired at both targets.[4] How, then, are we to decide whether or not the product has been uniform?

---

4. In other words, non-assignable causes introduce certain differences in the average, dispersion and shape of the observed polygons from one month to another, and we must set up some method of differentiating the effects of assignable from those of the non-assignable causes. Therefore we have the question, do the observed differences shown in Fig. 2 indicate the presence of effects of assignable causes - causes that it should be possible for us to discover - or are the differences attributable to chance, random or non-assignable causes - causes that possibly cannot be discovered easily?

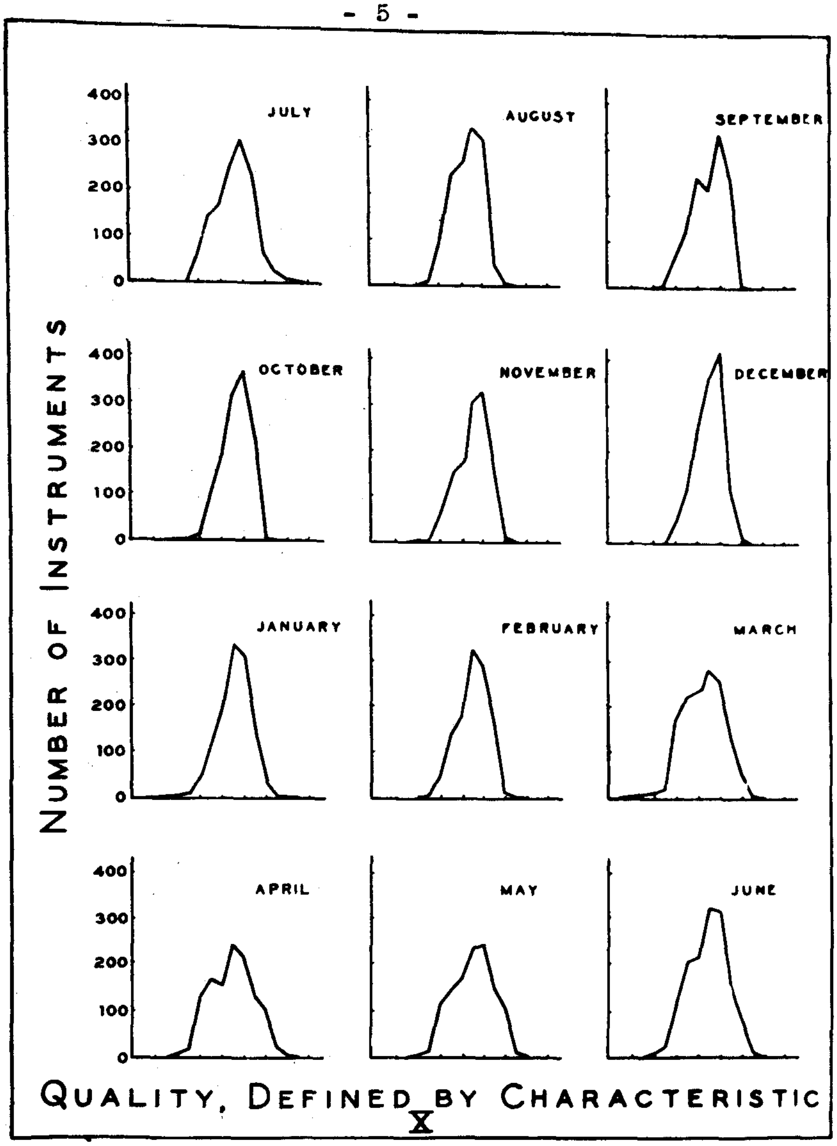



Fig. 2 - MONTHLY POLYGONS SHOWING DISTRIBUTION IN QUALITY FOR SAMPLES OF APPROXIMATELY 1250 UNITS OF PRODUCT. DO THESE DATA PRESENT ANY EVIDENCE OF NON-UNIFORMITY - ANY EVIDENCE OF EFFECTS OF ASSIGNABLE CAUSES?

Let me now illustrate another way in which non-uniformity of product may arise. Fig. 3 shows the frequency polygon for 15,050 instruments in respect to quality defined by a characteristic X. The instruments were selected from the year's product at regular intervals and then grouped together as shown in Fig. 3. Is there any indication in the data given that the product had not been uniform thorughout the 12-month period in which the 15,050 instruments had been selected?[5]

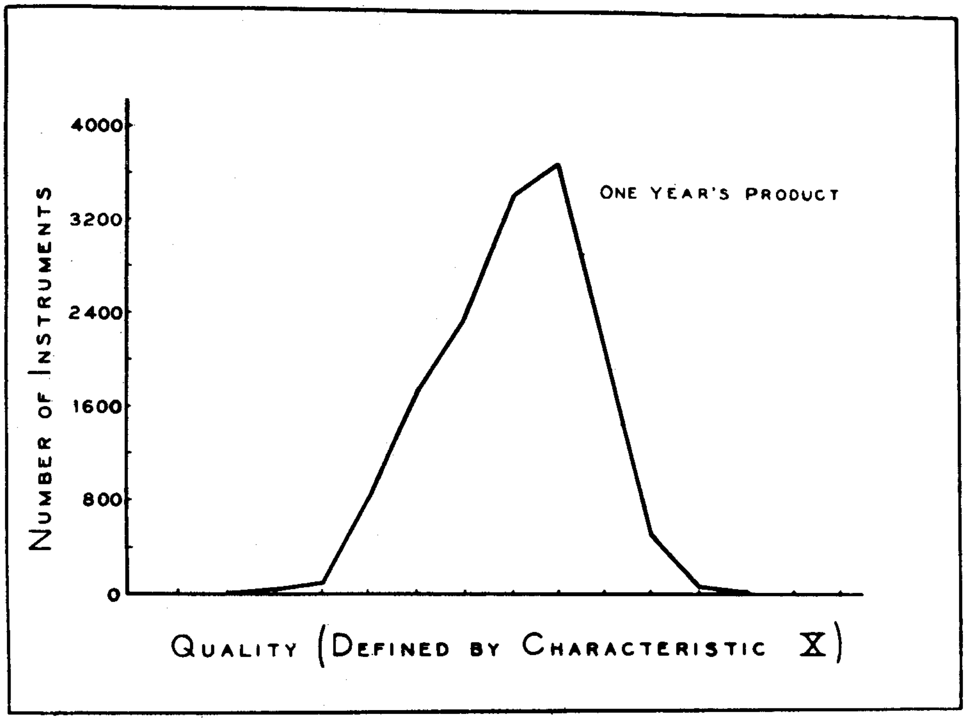



Fig. 3 - POLYGON SHOWING DISTRIBUTION IN QUALITY FOR 15,050 UNITS OF PRODUCT. DO THESE DATA PRESENT ANY EVIDENCE OF NON-UNIFORMITY - ANY EVIDENCE OF EFFECTS OF ASSIGNABLE CAUSES?

2ND OBJECT: FINDING CAUSES OF NON-UNIFORMITY OF QUALITY.

Of course it is desirable to know how to detect non-uniformity of quality, but that information alone won't enable us to produce a product of uniform quality unless we discover the causes of non-uniformity. Now, let us consider three typical illustrative cases where we want to know the causes of variation.

One illustration arises in determining the insulating properties of silks used in covering wires and cables. Here, we wish to find how the insulating property is affected by such factors as the acidity, total water soluble

---

5. Naturally this problem could be broken down into the type previously given, providing information as to the results of monthly inspections were available. Nevertheless, for one reason or another, the inspection engineer is often confronted by the problem stated in the form connected with the data in Fig. 3.

content, non-volatile content and ash content of the material, because we want to find out how accurately these factors must be controlled in the process of producing silk insulation.

A second illustration arises when we try to compare the results of different procedures for creosoting telephone poles. In this instance the depth of penetration of the oil is the quality under consideration and we wish to find how such factors as thickness of sapwood, temperature of oil bath, water content of the poles and so on, affect the quality.

As a third illustration I have in mind the manufacture of a certain article where we know that the quality of the material depends upon numerous factors such as the temperature of roast, the surrounding gas, and the properties of the material such as absorptive capacity, porosity, ash content, specific gravity and so on. I reproduce in Fig. 4 a scatter-diagram showing 90

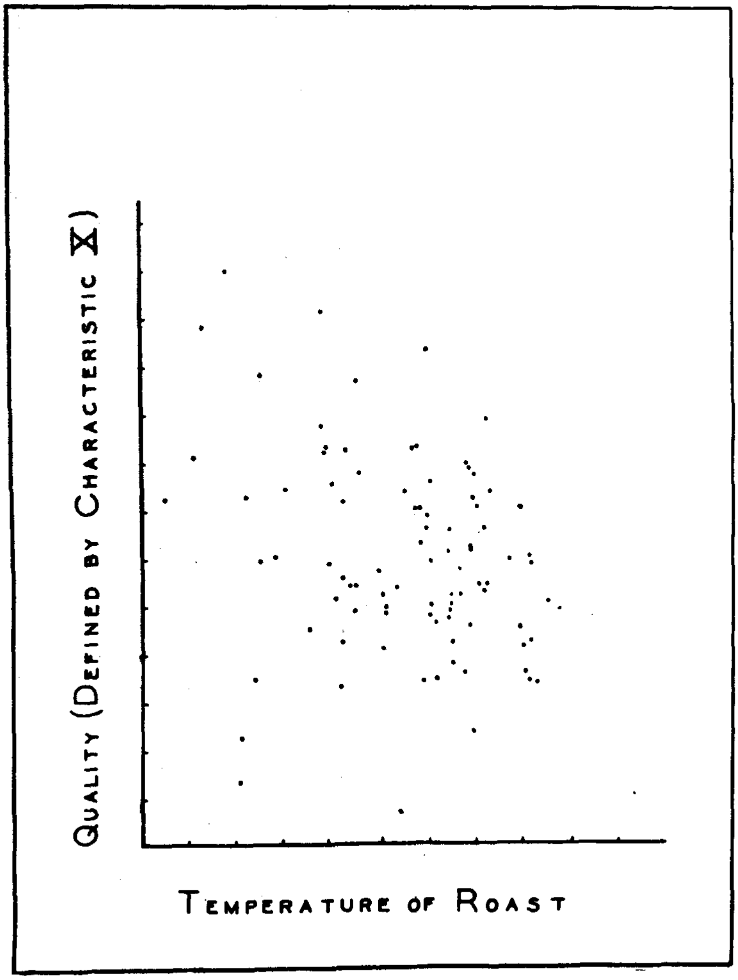



Fig. 4 - SCATTER-DIAGRAM OF QUALITY VS. TEMPERATURE OF ROAST. IS TEMPERATURE VARIATION IN THIS INSTANCE AN ASSIGNABLE CAUSE OF FLUCTUATION IN QUALITY?

pairs of simultaneous observations of temperature of roast and quality of the material as determined by some characteristic X. Are we to conclude from the data presented in this figure that there is any relationship between the quality of the material and the temperature of roast? If there is a relationship, we conclude that the temperature of roast is one of the assignable manufacturing factors which must be controlled more carefully in the future than it has been in the past. If there isn't any relationship, we conclude that no greater effort need be exercised in controlling the temperature of roast in the future than has been exercised in the past.

Such charts are often used, but I fear that many times the results are interpreted incorrectly. For example considering the data given in Fig.4 alone, we might conclude that there is little correlation between the quality and the temperature of roast. This conclusion, however, may not be justified, because it is possible that the correlation, though it exists, is masked by the effects of other factors. To obtain the true picture, therefore, it is necessary to consider the simultaneously observed values of all of those factors which we expect to influence the quality of the manufactured article. In the general case where there are several of these factors, the problem is somewhat complicated.[6]

The importance of the second object of inspection can scarcely be over-emphasized.[7]

<u>3RD OBJECT: SETTING ECONOMIC STANDARDS OF QUALITY.</u>

Naturally we wish to do everything within our power to give quality at a minimum of cost, but to what limit should we go in improving quality if we thereby raise the cost of the product?

---

6. For example let us assume that $X = a + b_1Y_1 + b_2Y_2 + \ldots + b_nY_n$ where the Y's are assignable groups of causes affecting X. It is obvious that, if we study the simultaneous pairs of X and only one of the factors Y, we may observe a correlation between X and Y which is in fact the effect of correlations between X and certain of the other factors. The natural procedure is therefore, to use the theory of partial correlation, which will be considered in a later paper.

7. Not only could this list of illustrations be extended indefinitely because they arise in the production of every manufactured article, but, also, because the problems involved are theoretically the same as arise in the calibration of machines used in testing transmitters and receivers and certain other kinds of apparatus.

To start the discussion let us assume that the appreciation or value of quality is the same for all individuals and that apparatus of only one quality is to be manufactured. Quality (Fig.5) is represented by the horizontal axis, and both the cost and value of quality expressed in dollars are represented by the vertical axis. Of the two curves given in this figure, one represents the value of quality to the individual; i.e., the amount that an individual would be willing to pay for an article of a given quality rather than to do without that article; the other represents the cost of a given quality, including interest, depreciation and insurance charges.

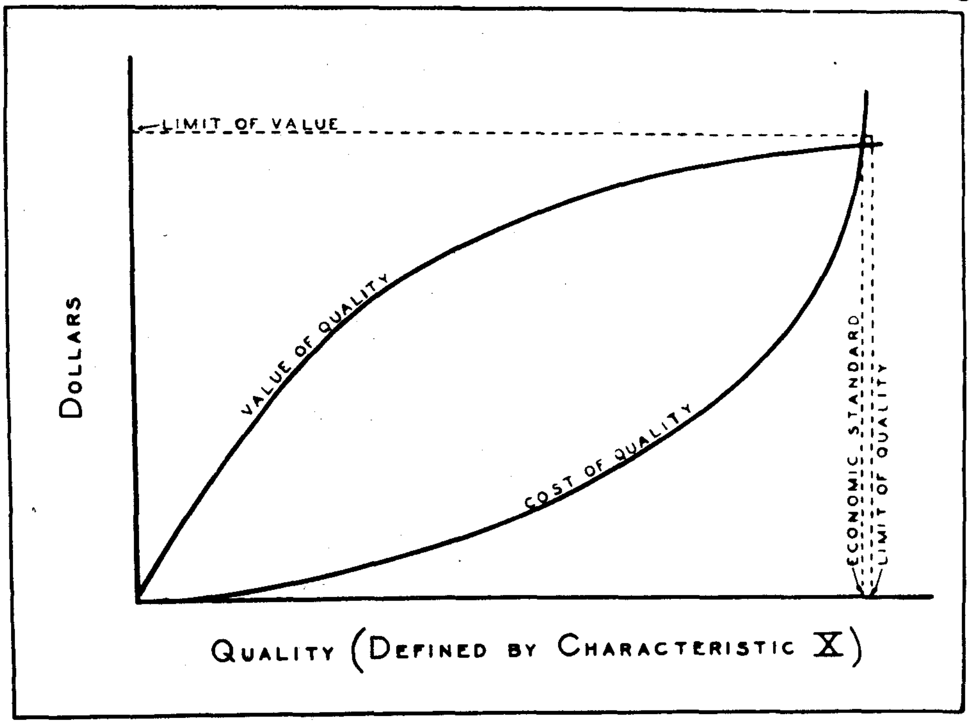



Fig. 5 - SCHEMATIC REPRESENTATION OF THE FACTORS INFLUENCING THE ECONOMIC STANDARD.

We want to supply the public with the maximum of quality which it wants and for which it is willing to pay the actual cost. Obviously this quality corresponds to the point on the quality axis where the cost curve crosses the value curve, because quality better than this would cost more than it would be worth to the consumer whose value curve is such as shown in Fig. 5, and hence the consumer would rather do without than buy such quality. We may think, therefore, of the quality corresponding to the intersection of the cost curve[8] and the value curve as being the economic standard of quality for the conditions assumed.

Let us now try to define the economic standard in a case nearer practice. Neither the value curve for an individual nor the cost curve of the manufacturer remains constant with time. Hence the economic standard as defined above actually fluctuates with time. We must, therefore, determine from a study of the economic conditions of the country what the range of fluctuation in this economic standard may be expected to be and choose some point on the quality axis as an economic standard, which allows for the random fluctuations in the value and cost curves.

Furthermore I assumed in drawing Fig. 5 that all people have the same estimate of the value of quality. Obviously, however, there are practically as many different value curves as there are individuals. We may picture the condition somewhat after the manner indicated in Fig. 6 where each

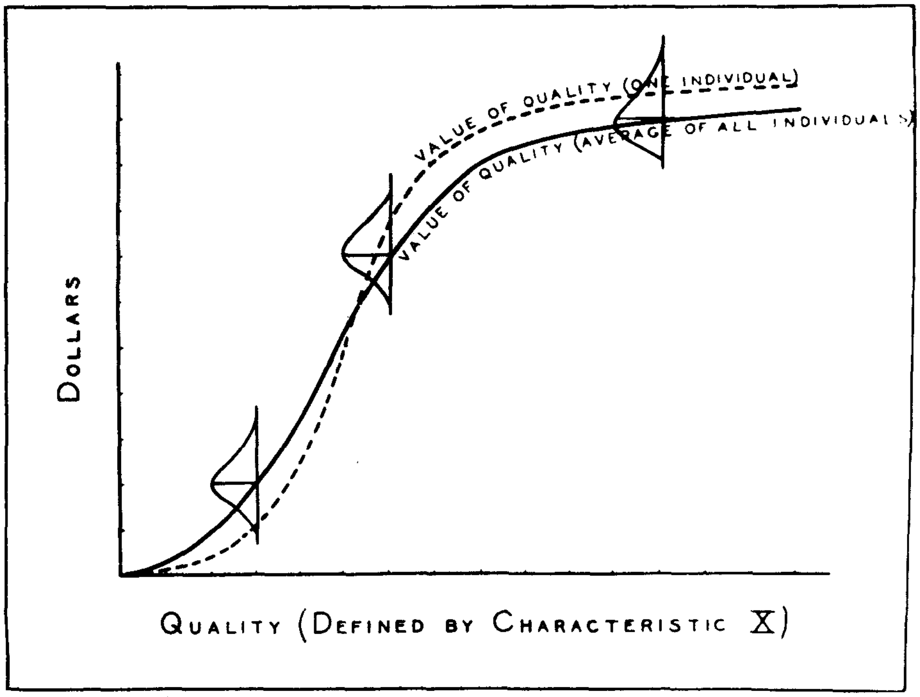



Fig. 6 - SCHEMATIC REPRESENTATION OF VALUE CURVES.

8. Probably the more typical cost curve would be discontinuous at zero quality and would start at some cost appreciably greater than zero for a quality just greater than zero.

point on the solid curve represents the average of the frequency distribution of the value estimates of all individuals, and where an individual's value curve might look like the dotted one shown in the figure. In the actual case the distribution curves would probably extend over greater ranges than shown in the figure. In fact the value curves of some individuals might fall on the X axis. Hence the distribution of values must be taken into account in setting the economic standard (or standards, providing different groups of individuals are to be supplied with the quality which they demand).

Thus the problem of setting economic standards is inherently very important and incidently involves many questions of a statistical nature.[9]

## PART II

## FOUR STEPS IN INSPECTION

Having considered the three objects of inspection, let us next look at four of the fundamental steps which must be taken in attaining each of the three objects, and let us consider a few typical problems encountered in taking each of the four steps.

### FIRST STEP: ESTABLISHMENT OF INSPECTION PROCEDURE

Naturally we must decide first what characteristic (or characteristics) to measure and why we want to measure it (or them) in order to specify the quality. Our decision will be determined by the object in view. Once this decision is reached, we must establish the inspection procedure outlining the method of measurement which will achieve our object at a minimum of cost. Of course this involves the choice of the best measuring apparatus for each particular job and calls for the exercise of engineering judgment based upon known scientific laws and past experience. But this phase of the problem need not be considered here. Certain other types of problems arise, however, which I do wish to discuss in some detail.

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

9. Obviously, in the general case, where the quality Q is some function $\Psi$ of many characteristics X, Y, Z, ..., the problem of setting economic standards can be attacked in a manner similar to the simple case considered.

EXAMPLE 1: HOW LARGE A SAMPLE?

One of the first questions which arise in determining the procedure is how many units of product must we inspect from a given lot so that we may insure ourselves and the customer that the risk of the tolerance fraction[10] being exceeded in the lot is not greater than some fixed value for each kind of apparatus. This question is always present in inspection work and arises many times in connection with the inspection of all kinds of apparatus.

The answer to this question may be based on either à posteriori or à priori probability but is subject to different limitations in the two cases.[11] Obviously the answer must be given in terms of the size N of the lot from which the sample of size n is drawn and in terms of the tolerance $p_t$ set on the apparatus.[10] A little reflection shows the necessity of introducing two other factors in reaching the final answer, because the choice of the best sample size calls for a balance between the cost of inspection and the value of inspection. Hence it is necessary to introduce the conceptions of Producers' and Consumers' risks. Obviously it is necessary to prepare extensive tables and sets of curves which can be used by those engineers engaged in devising methods to cover the practically unlimited number of special cases.

A simple discussion of some phases of this problem is given in Appendix 1 of this paper. The discussion in this appendix is by no means complete, because it is not sufficient merely to say how many should be inspected. Instead we also must specify the procedure to be followed when a

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

10. Sometimes the non-assignable causes produce a unit of uniform product falling outside of one or more of the ranges $X'\pm\delta X$, $Y'\pm\delta Y$, $Z'\pm\delta Z$, ..., and hence such pieces are classed as defective. Sometimes the quality standard of the lot of N pieces is set by saying that not more than $p_tN$ of a standard lot may be defective where $p_t$ is the tolerance fraction. A defective lot is one which contains more than $p_tN$ defective units.

11. à posteriori - influence of causes from effects. à priori - influence of effects from causes. Of course the application of any theory, in a sense, involves à posteriori reasoning unless it can be proved that no other than the assumed set of causes could give the observed data.

sample is found defective: i.e., we must specify whether the inspector is to reject the lot or to extend the inspection. Furthermore the choice of procedure depends upon the acceptance number,[12] upon the psychological effects of additional inspection on the inspectors, and upon the breakage and disproportionate cost accompanying extended inspection.

EXAMPLE 2: WHAT IS THE CHEAPEST WAY TO MEASURE QUALITY OF PRODUCT WHEN THE METHOD OF MEASUREMENT IS SUBJECT TO ERROR?

Suppose we select n units of a given kind of apparatus and measure the quality of each unit by a method subject to error. The true qualities probably are not identical, and the observed qualities show a greater dispersion than the unknown true qualities would, providing the method of measurement were not subject to error. Now, we can increase the precision of our estimate of the quality of product either by making more than one measurement on each unit or by increasing the number of units measured. Customarily the cost of making two or, in general, b measurements on a single unit is less than the cost of making one measurement on each of two or, in general, b units of apparatus. The most economical sample size and number of measurements to be made on each unit in the sample must be sought.

Another form of the same type of problem is: Assume that we are manufacturing loading coil cases, each case containing several units, and that one of the steps in check inspection is the measurement of the quality of each of 10% of the number of units in each of the cases manufactured. Could we obtain, more economically than by the assumed present practice, the same degree of precision of check upon the quality of product if we measure the quality of, say d units in each case, but only inspect a fraction of the total number of cases?

Analytically the last two questions involve the same theory as does the question, how can observed data be corrected for errors of measurement?[13]

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

12. The number of defective pieces which can be found in a sample without indicating an unsatisfactory condition of the lot is termed the acceptance number.

13. Shewhart, W.A., Correction of Data for Errors of Measurement, Bell System Technical Journal, Jan. 1926.

EXAMPLE 3: HOW SHALL WE MEASURE QUALITY - AS A VARIABLE OR AS AN ATTRIBUTE?

Let us make this question clear by assuming that we are inspecting condensers for capacity. We can measure the capacity of each condenser either as so many microfarads or simply as above or below a certain capacity, say $X_0$. The first method treats capacity as a variable, the second treats it as an attribute.

Let us consider only one phase of this question and assume in doing so that the object of inspection is to detect the evidence of assignable causes of variation in product. Thus suppose that we inspect the same number of units of product for the months of March, April and May and find the same fraction of the month's product below $X_0$ in capacity for each month. Upon the basis of this measurement the product is running uniformly. It is perfectly possible, however, that the product for these three months may be distributed in respect to quality (capacity) as shown in Fig. 7 where the areas to the left of $X_0$ are equal. Even to the unskilled eye, differences

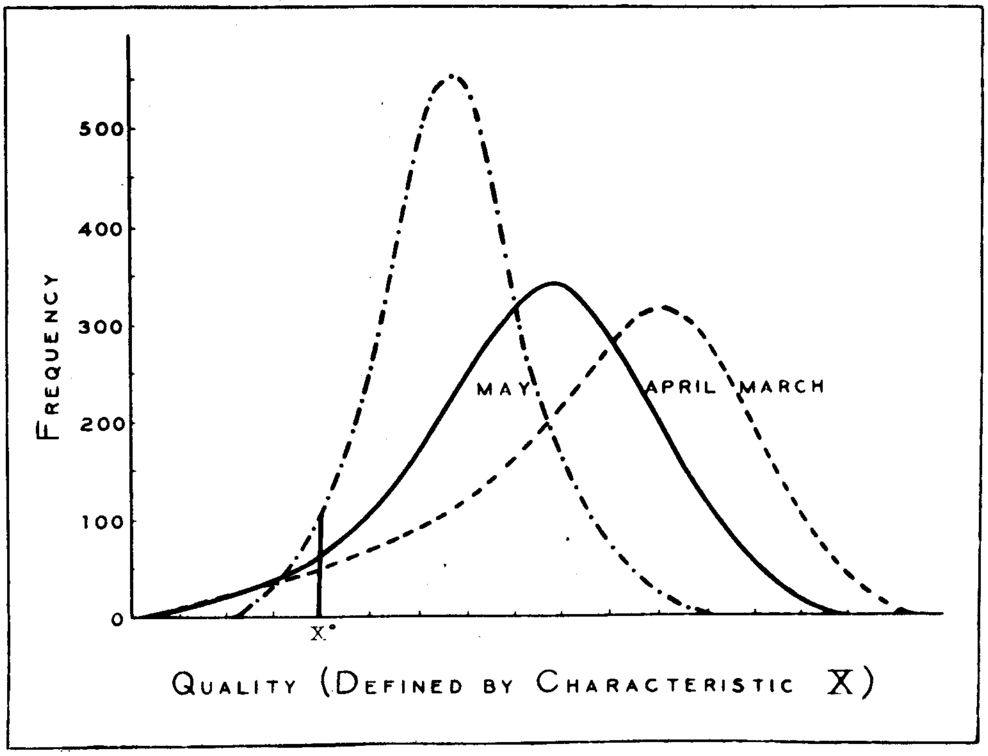



Fig. 7 - SCHEMATIC DIAGRAM INDICATING SOME DISADVANTAGES OF INSPECTING BY METHOD OF ATTRIBUTES.

in frequency distributions as great as those indicated in this figure would month to another. In other words, the method of variables would not whereas the method of attributes would suggest that the product is running uniformly.[14]

Other illustrations could be added to indicate the inefficiency of the method of attributes as compared with the method of variables for measuring the quality of product, particularly when it becomes necessary to correct data for errors of measurement. On the other hand it usually costs less to measure the quality of a unit of apparatus as an attribute than it does to measure it as a variable. Now, in a practical case the value of a measurement as a variable will always be greater than that as an attribute, but the cost of a measurement as an attribute may be less than the corresponding cost of the measurement as a variable. We must strike a balance between these values and costs. To do this often involves the solution of certain statistical problems to be discussed elsewhere.

SECOND STEP: COLLECTION OF DATA

Turning now to the collection of data we find certain principles which must be kept continually in mind. Mention will be made of a few of these.

OBTAIN ALL NECESSARY INFORMATION BUT NO MORE

To prepare a form calling for all necessary information isn't quite so easy to do as it may appear. The preparation of such a form requires at least two things on the part of the engineer in charge: First, a clear understanding of the physical and engineering details of the measurements involved and, second, a clear vision of the steps to be followed in analyzing and interpreting the results of inspection. For example, if some

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

14. It is sometimes argued that we are only interested in the percentage below some value of quality such as $X_o$ in the above example. There are many advantages, however, in maintaining a controlled product - one in which all assignable causes are found and controlled. In the case assumed above there would be real cause for alarm in the indicated downward trend in the average quality. Who knows but what the next month might bring serious trouble?

difficulty in maintaining a uniform product is anticipated, engineering judgment must be called upon to suggest those factors which may possibly be assignable causes for the anticipated non-uniformity, and a knowledge of theory must be called upon to specify details of inspection method such as the number of measurements to be taken on the various factors supposed to be assignable causes. Sometimes the lack of information such as the size of the sample or the size of the lot may make it impossible to determine whether or not observed fluctuations in product can be attributed to assignable causes, and thus render the data almost useless. Hence the preparation of a suitable form for the collection of inspection data requires the co-operation of many groups in a manufacturing organization. One illustration may serve to show some of the questions which must be considered in preparing a report form.

Assume that two machines are used in manufacturing the product and that the distribution of quality given by one machine is that represented by the broken curve on the left of Fig. 8 and the distribution of quality given by the other machine is that represented by the broken curve on the right of Fig. 8.

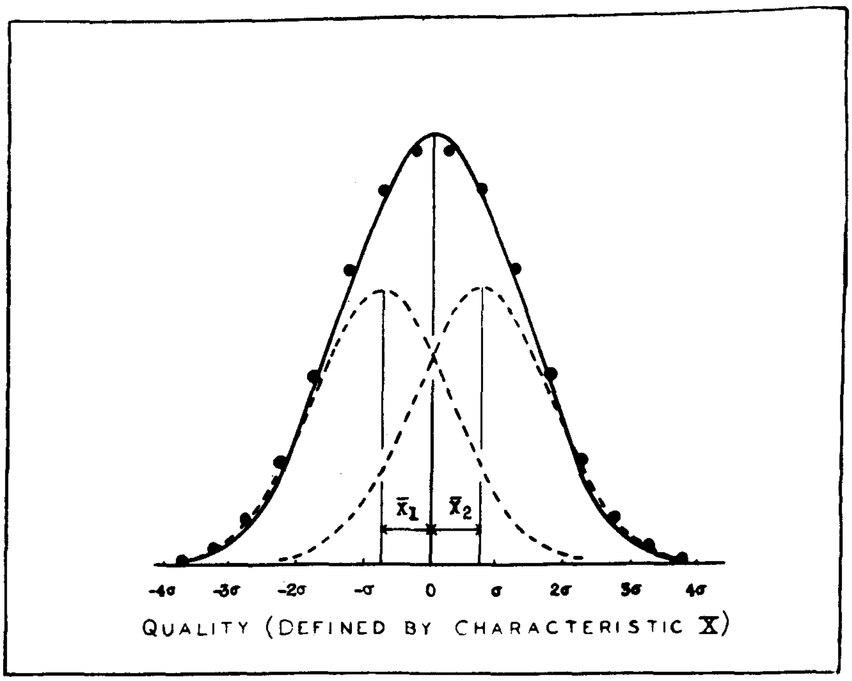



Fig. 8 SCHEMATIC ILLUSTRATION SHOWING HOW NON HOMOGENEOUS PRODUCT [illegible]

Product from these two machines, when combined in equal quantities, would give the distribution of quality outlined by the dots in Fig. 8 and approximated by the solid curve. Analysis of the combined results of inspection of product from the two machines would give no indication of the non-homogeneity of the product which actually exists.[15] If, however, the inspection results for the two machines were kept separate, it would be very easy to detect the non-homogeneity. Therefore, whenever an inspection engineer wishes to detect non-homogeneity of product, he should use a report form classifying the data according to the causes of non-homogeneity. This is only one of many ways in which the problem of detecting the effects of component systems of causes arises in practice but it serves to illustrate the necessity of the use of foresight in collecting the data.

In addition to the principles already noted, care should be taken in recording the data so as to make the cost of analysis a minimum. Also the data should be recorded in sequence wherever possible so that the analysis may reveal the presence of any cyclic or long time trend fluctuations in the quality. In those instances where several characteristics are measured on each instrument, it is often highly desirable to record the data so that all of the characteristics belonging to a given instrument may be kept separate from the others, because a study of the data may, then, reveal a correlation between some of the factors which will make it possible to reduce the total amount of inspection.

THIRD STEP: ANALYSIS OF DATA

How useless is an unclassified set of several thousand (or for that matter a set of only 100) data before they are classified and analyzed! It is seemingly difficult for the mind to grasp the significance of a large number of figures. Therefore, we must try to reduce the observed data to a few figures or statistics which give us the essential information. This

---

15. See Appendix 2 for a more complete discussion of the problem of detection of non-homogeneity, and for an explanation of the notation on Fig. 8.

calls for the development and standardization of methods for analyzing data to determine the essential statistics at a minimum of cost.

FORMS FOR ANALYSIS OF DATA

Customarily the mind seeks measures of the central tendency, dispersion and asymmetry (skewness) of a set of data. Commonly accepted measures of these are the arithmetic mean $\bar{X}$, root mean square deviation $\sigma$, and skewness $k, \sqrt{\beta_1}$. These factors are calculated as shown in the sample analysis sheet[16] presented in Fig. 9. Another statistic $\beta_2$ given in Fig. 9 is often used as a measure of the degree of flatness of the distribution of the observed data. This statistic is technically known as a measure of kurtosis.

WESTERN ELECTRIC CO., INC., INSPECTION ENGINEERING DEPT., NEW YORK.

INSPECTION ENGINEERING ANALYSIS SHEET

OBSERVATIONS

| | |
|---|---|
| CHARACTERISTIC INSPECTED (a) | QUALITY X |
| UNITS (b) | - |
| TO NEAREST | 0.1 |
| SOURCE OF DATA | - |

SUBJECT: 1000 DRAWINGS BY MBC and VLM — 1st Sample

$f(x)$ NORMAL LAW

| CELL | Quality (a) IN X (b) | CELL BOUNDARIES LOWER | UPPER | CELLS FROM ORIGIN AT $\delta$ = -3.0 X | OBS. FREQ. y | yX | $yX^2$ | $yX^3$ | $yX^4$ | THEOR. FREQ. $y_1$ | $y-y_1$ | $(y-y_1)^2$ | $\frac{(y-y_1)^2}{y_1}$ | THEOR. FREQ. $y_2$ | $y-y_2$ | $(y-y_2)^2$ | $\frac{(y-y_2)^2}{y_2}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | -3.0 | -3.25 | -2.76 | 0 | 5 | 0 | 0 | 0 | 0 | 3 | 2 | 4 | 1.333 | | | | |
| 1 | -2.5 | -2.75 | -2.26 | 1 | 9 | 9 | 9 | 9 | 9 | 10 | - 1 | 1 | .100 | | | | |
| 2 | -2.0 | -2.25 | -1.76 | 2 | 36 | 72 | 144 | 288 | 576 | 29 | 7 | 49 | 1.690 | | | | |
| 3 | -1.5 | -1.75 | -1.26 | 3 | 55 | 165 | 495 | 1485 | 4455 | 66 | -11 | 121 | 1.833 | | | | |
| 4 | -1.0 | -1.25 | - .76 | 4 | 123 | 492 | 1968 | 7872 | 31488 | 121 | 2 | 4 | .033 | | | | |
| 5 | - .5 | - .75 | - .26 | 5 | 165 | 825 | 4125 | 20625 | 103125 | 173 | - 8 | 64 | .370 | | | | |
| 6 | 0 | - .25 | .24 | 6 | 203 | 1218 | 7308 | 43848 | 263088 | 195 | 8 | 64 | .328 | | | | |
| 7 | .5 | .25 | .74 | 7 | 172 | 1204 | 8428 | 58996 | 412972 | 173 | - 1 | 1 | .006 | | | | |
| 8 | 1.0 | .75 | 1.24 | 8 | 123 | 984 | 7872 | 62976 | 503808 | 121 | 2 | 4 | .033 | | | | |
| 9 | 1.5 | 1.25 | 1.74 | 9 | 68 | 612 | 5508 | 49572 | 446148 | 67 | 1 | 1 | .015 | | | | |
| 10 | 2.0 | 1.75 | 2.24 | 10 | 31 | 310 | 3100 | 31000 | 310000 | 29 | 2 | 4 | .138 | | | | |
| 11 | 2.5 | 2.25 | 2.74 | 11 | 8 | 88 | 968 | 10648 | 117128 | 10 | - 2 | 4 | .400 | | | | |
| 12 | 3.0 | 2.75 | 3.24 | 12 | 2 | 24 | 288 | 3456 | 41472 | 3 | - 1 | 1 | .333 | | | | |
| 13 | | | | | | | | | | | | | | | | | |
| 14 | | | | | | | | | | | | | | | | | |
| 15 | | | | | | | | | | | | | | | | | |
| | Σ | | | | 1000 | 6003 | 40,213 | 290,775 | 2,234,269 | | | | 6.612 | | | | |
| | | | | | | | | | | NO. OF CELLS = 13 | | P = .760 | | NO. OF CELLS = | | P = | |

MOMENTS ABOUT ORIGIN $\delta$

$'\mu_1 = \frac{\Sigma yX}{\Sigma y} = \frac{6003}{1000} = 6.003$

$'\mu_2 = \frac{\Sigma yX^2}{\Sigma y} = \frac{40213}{1000} = 40.213$

$'\mu_3 = \frac{\Sigma yX^3}{\Sigma y} = \frac{290775}{1000} = 290.775$

$'\mu_4 = \frac{\Sigma yX^4}{\Sigma y} = \frac{2234269}{1000} = 2234.269$

NOTE: THE ORIGIN $\delta$ IS THE MID-VALUE BETWEEN THE BOUNDARIES OF THE CELL CHOSEN.

$m$ = UNITS (b) PER CELL = .5

$\bar{X} = \delta + m\,'\mu_1 = -3.0+3.00150 = .002$

CORRECTED MOMENTS

$\sigma = m\mu_2^{1/2} = .5 \times 2.023279 = 1.012$

$k = \frac{\mu_3}{\mu_2^{3/2}} = \frac{-.770593}{8.282612} = -.093$

$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{48.986692}{16.758036} = 2.92$

$\beta_1 = k^2 = (-.093)^2 = .009$

$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{1.012}{\sqrt{1000}} =$ [illegible]

$\sigma_\sigma = \frac{\sigma}{\sqrt{2n}} = \frac{1.012}{\sqrt{2000}} =$ [illegible]

$\sigma_k =$ [illegible]

$\sigma_{\beta_2} =$ [illegible]

UNCORRECTED MOMENTS ABOUT ARITH. MEAN

$\mu_1 = 0$

$\mu_2 = '\mu_2 - '\mu_1^2 = 40.213 - 36.036009 = 4.176991$

$\mu_3 = '\mu_3 - 3\,'\mu_1\,'\mu_2 + 2\,'\mu_1^3 = 290.777 - 724.195917 + 432.648324 = -.770593$

$\mu_4 = '\mu_4 - 4\,'\mu_1\,'\mu_3 + 6\,'\mu_1^2\,'\mu_2 - 3\,'\mu_1^4 = 2234.269 - 6982.137324 + 8694.096180 - 3895.781835 = 51.046021$

CORRECTED MOMENTS ABOUT ARITH. MEAN (SHEPPARD'S CORRECTIONS)

$\mu_2(\text{cor.}) = \mu_2 - \frac{h^2}{12} = 4.176991 - .083333 = 4.093658$

$\mu_4(\text{cor.}) = \mu_4 - \frac{1}{2}h^2\mu_2 + \frac{7}{240}h^4 = 51.046021 - 2.088496 + .029167 = 48.986692$

$h = 1$

(10-9-)

Fig. 9.

The graphical significance of these four statistics is pictured in the schematic diagrams presented in Fig. 10. Fig. 10-a shows two symmetrical distributions differing in the average $\bar{X}$ but not in the root mean square de-

---

16. This sheet represents one of several forms which have been standardized for the analysis of inspection data. I am indebted to Mr. H.F. Dodge for assistance in the preparation of this form.

The data given on this sheet will be referred to in a later section of this paper.

viation $\sigma$ (k=0 and $\beta_2$ = 3 for both). Fig. 10-b shows symmetrical distributions differing in the root mean square deviation σ but not in the average $\bar{X}$ (k=0 and $\beta_2$=3 for both). Fig. 10-c represents non-symmetrical distributions differing in skewness k but not in either the average $\bar{X}$ or the root mean square deviation $\sigma$ ($\beta_2$=3 for both). Fig. 10-d represents symmetrical

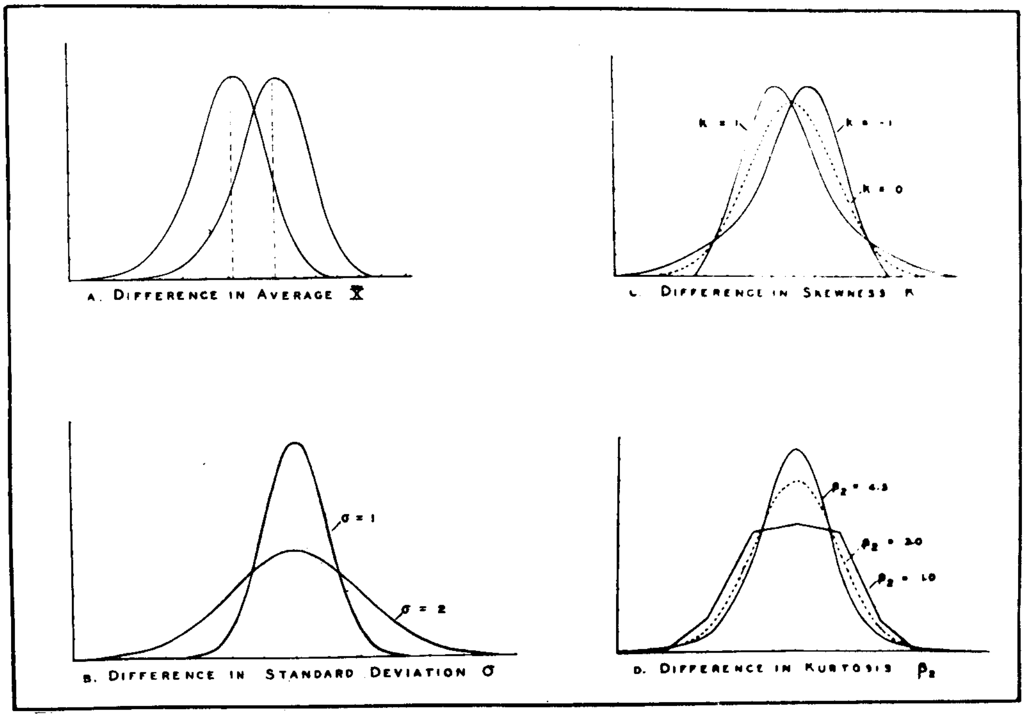



Fig. 10 - SCHEMATIC DIAGRAMS SHOWING GEOMETRICAL SIGNIFICANCE OF COMMONLY USED STATISTICS.

distributions differing in kurtosis $\beta_2$ but not in the average $\bar{X}$, skewness k and standard deviation $\sigma$. These[17] curves may be used for reference in helping an individual form a mental picture of the distribution of a large number of data from a knowledge of the values of these four statistics.[18]

<u>EXAMPLE 1: CORRECTION OF DATA FOR ERROR OF MEASUREMENT.</u>

Suppose the method of measurement used in obtaining original data is subject to error. In this case the observed distribution of quality is

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

17. These curves were obtained by using the first four terms of the Gram-Charlier series. See Eq. 3 in Part III.

18. In some instances a greater number of statistics may be required to give the essential characteristics of the observed data.

not the true one. This follows from the fact that the number of instruments actually having values of quality within any pair of limits will probably be observed as distributed over much wider limits. Hence the four statistics calculated from the observed data do not represent the true distribution of quality unless corrections are made to allow for the error of measurement. The need for such corrections arises in practically every phase of engineering inspection work and the theory has been given in a paper already referred to.[19] Corrections of this nature are particularly necessary in establishing limits, as we shall see in a future section.

EXAMPLE 2: EFFICIENT USE OF DATA

There is always a best way of analyzing data to make the most efficient use of them. This point can be made clearer by illustration.[20] Suppose we make 10 measurements of the impedance of a coil and get the following values.

TABLE I.

| Number of Measurement | Impedance in Ohms |
|---|---|
| 1 | 100.9 |
| 2 | 100.5 |
| 3 | 99.9 |
| 4 | 101.0 |
| 5 | 101.4 |
| 6 | 102.7 |
| 7 | 100.5 |
| 8 | 99.0 |
| 9 | 99.9 |
| 10 | 99.6 |
| Ave. = | 100.54 |

Textbooks on the theory of errors tell us that the root mean square deviation of the average can be estimated in either of two ways (they might

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

19. See footnote 13. It should be pointed out also that data must be corrected for error of averages obtained from small samples as discussed in my paper in the April 1926 issue of the Bell System Technical Journal on Correction of Data for Errors of Averages Obtained from Small Samples.

20. See Appendix 3 for a more complete discussion.

have said truthfully in an infinite number of ways). The two ways are the so-called root mean square error and the mean error methods as illustrated below:

TABLE II

Showing Two Methods of Estimating the Standard Deviation of the Average.

Root Mean Square Error Method

| Observed Values in Ohms | Dev. from Ave. | $(\text{Dev. from Ave.})^2$ |
|---|---|---|
| 100.9 | .36 | .1296 |
| 100.5 | .04 | .0016 |
| 99.9 | .64 | .4096 |
| 101.0 | .46 | .2116 |
| 101.4 | .86 | .7396 |
| 102.7 | 2.16 | 4.6656 |
| 100.5 | .04 | .0016 |
| 99.0 | 1.54 | 2.3716 |
| 99.9 | .64 | .4096 |
| 99.6 | .94 | .8836 |
| Total | | 9.8240 |
| Average | | .9824 |

Estimate of Standard Deviation = $\sqrt{.9824}$ = .9912

Estimate of Standard Deviation of Average = $\dfrac{.9912}{\sqrt{10-1}}$ = .3304

Mean Error Method

| Observed Values in Ohms | Dev. from Ave. |
|---|---|
| 100.9 | .36 |
| 100.5 | .04 |
| 99.9 | .64 |
| 101.0 | .46 |
| 101.4 | .86 |
| 102.7 | 2.16 |
| 100.5 | .04 |
| 99.0 | 1.54 |
| 99.9 | .64 |
| 99.6 | .94 |
| Total 1005.4 | 7.68 |
| Average 100.54 | .768 |

Estimate of Standard Deviation = $.768\sqrt{\dfrac{\pi}{2}}$ = .9625

Estimate of Standard Deviation of Average = $\dfrac{.9625}{\sqrt{10-1}}$ = .3208

Which one then shall we choose? The obvious answer is that one having the smallest error. Now it turns out that the root mean square method used above gives the best estimate under the condition that the law of error is normal.[21] In general it can be shown that this method is 14% more efficient than the mean error method. In other words, to obtain the same degree of precision with the mean error method as with the root mean square error method would require 1.14 times as many observations as would be required by the root mean square method.

Of course, we often use sampling methods to detect non-uniformity of product. Let us assume a case in which we inspect daily 114 instruments

---

21. See Part III for equation of normal law.

of a given kind and analyze the measurements by the mean error method to determine the dispersion. In this instance we could get just as good an estimate of dispersion by taking a sample of only 100 instead of 114 instruments providing the measurements on these were analyzed by the root mean square method. To make 100 observations serve the purpose of 114 is obviously to be desired.[22]

The above case in which we assume that the true distribution from which the samples were drawn was normal and that the sample sizes were large is a very special one. The study of the most efficient methods of using data is one of the very important problems of modern mathematical statistics and is receiving much attention at the present time. Naturally enough, an organization making millions of measurements per year is keenly interested in results of such work, because the application of the theory may be expected to give maximum efficiency in the use of inspection data.

FOURTH STEP: INTERPRETATION OF RESULTS

This fourth step is the crucial one in inspection work. In fact, interpretation of results is the crucial step in all experimental work. Without it the three other steps are useless, for what good does it do us to make inspection programs, collect data and analyze data if we do not follow these steps by the interpretation of the results.

In general we wish to know whether or not the differences between units of product have been modified by assignable causes. To do this, however, we must set up some basis for detecting the existence of variations in

---

22. The economic importance of efficient use of data is again touched upon in Part III, and in Appendix 3. Of course the computations involved in the root mean square method as indicated in Table II are longer than those involved in the mean error method therein indicated. In practice, short cuts are available for both methods and these minimize the apparent advantages of the mean error over the root mean square method in respect to time required in making calculations. In most cases the cost of analysis of the data by either method is but a small fraction of the cost of taking the data. An example is given in Part III where the yearly saving effected by using the root mean square method instead of the mean error method is approximately $2000 per year.

quality of product which cannot be attributed to random or sampling fluctuations produced by non-assignable causes. This carries us into the theory of sampling treated in Part III of this paper.

## PART III

## OUTLINE OF THEORETICAL BASIS FOR MEASURING QUALITY OF PRODUCT AND DETECTING NON-UNIFORMITY THEREIN

### BASIS FOR DETECTING NON-UNIFORMITY IN PRODUCT

As previously noted, it goes without saying that we cannot control the manufacturing processes so as to make all units of product identical. The best that we can do is to control the product so that the differences which occur between the distribution of product for one period and that for another can be attributed to the action of fortuitous, chance or random non-assignable causes which we cannot reasonably hope to control without radically changing the whole manufacturing process. Typical non-assignable causes previously mentioned are such factors as humidity, temperature, wear and tear on machinery and the physiological and psychological conditions of those individuals engaged in carrying out the manufacturing process.

The first object of inspection, as we recall, is to detect in the results of inspection any effects which cannot be attributed to non-assignable causes. But what is the distinguishing feature between the effects of non-assignable and assignable causes? The basis for answering this question is proposed in the following paragraph and may be justified upon the basis of both à priori and à posteriori reasoning as indicated.

On the one hand it may be shown analytically that a multiplicity of causes acting in a fashion similar to the way in which we have reason to believe the non-assignable causes of variation in the manufacturing processes to act, will give rise to a unimodal distribution of product which can be described quite accurately in terms of the well known probability curves to be considered below. Let us, therefore, define a uniform product as one for which the probability of production of a unit with quality X, lying within the interval X to X + dX, is independent of time. We shall call a system of causes producing such a product constant, because so long as a product is produced by such a set of causes the variations in product from one period to another are independent of time and can be explained as sampling fluctuations. This constitutes the à priori method of laying the basis for detection of non-uniformity and goes back to the theory of causation associated with the names of Laplace, Poisson, Gram, Charlier, Thiele, Edgeworth and others.

On the other hand, we may start from the à posteriori point of view. Following in the footsteps of that great English statistician, Pearson, we find a preponderance of evidence to indicate that frequency distributions of measurable quantities, not affected by cyclic, random or long time trends, are characteristically unimodal and of certain standard types. We may assume therefore that such distributions represent the effects of natural groups of causes.

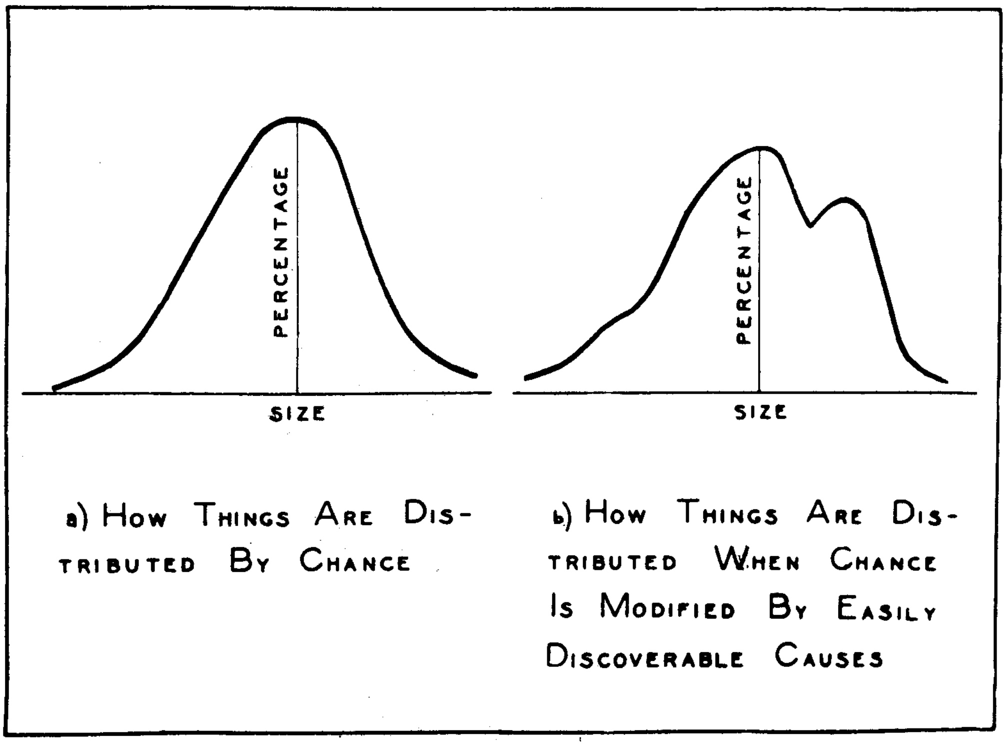



Fig. 11 - SCHEMATIC CONTRAST BETWEEN DISTRIBUTIONS OF EFFECTS OF NON-ASSIGNABLE AND A COMBINATION OF ASSIGNABLE AND NON-ASSIGNABLE CAUSES.

Thus by either the à priori or à posteriori method, we may justify, by taking many long and rather complicated steps, the assumption that the effects of non-assignable causes acting alone will be distributed in a unimodal fashion which can be approximated closely by a smooth curve such as is schematically represented in Fig. 11-a, and that the effects of assignable causes superposed upon the effects of non-assignable causes will be distributed as is schematically represented in Fig. 11-b.[23] We shall return to a quantitative consideration of the differences between the two types of curves shown in Fig. 11, but, before taking up the outline of the general theory for detecting non-uniformity of product according to the basis outlined in this paragraph, let us see how the theory works in a very simple case illustrated with experimental data.

CALCULATION OF SAMPLING FLUCTUATIONS: KNOWN DISTRIBUTION FROM WHICH SAMPLES ARE DRAWN.

In our illustration let us assume that the product is controlled by a constant system of causes for which the probability $dy_{\lambda'}$, of producing a unit with the quality X within the range X to X + dX is given by the bell shaped curve (Fig. 12) which we recognize as the well known normal law of error whose equation is

$$dy_{\lambda'} = \frac{1}{\sigma'\sqrt{2\pi}} e^{-\frac{(X-m)^2}{2\sigma'^2}} dX, \qquad (1)$$

where m is the true average quality and $\sigma'$ is the true standard deviation of quality. Samples of product manufactured under such a constant system of causes will differ in respect to each of the four statistics, average $\bar{X}$, standard deviation $\sigma$, skewness k and kurtosis $\beta_2$(Fig.10).

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

23. This conception of the differences between the effects of non-assignable and assignable causes has been developed more in detail in a recent paper: Shewhart, W.A., "Finding Causes of Quality Variation", Management and Administration, February, 1926.

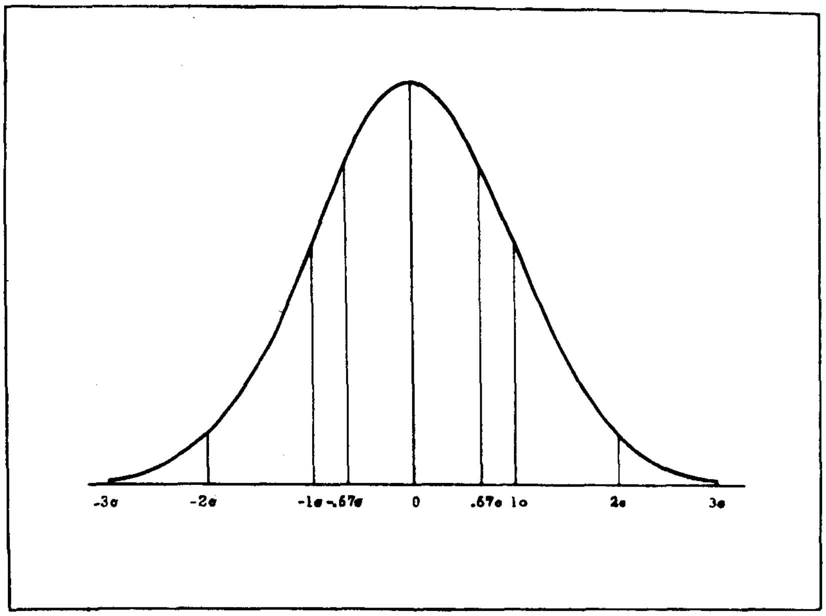



Fig. 12 - CUSTOMARILY ASSUMED LAW OF ERROR CURVE - Normal Law.
50.00000% of area within 0 ± .67449σ
68.26894% " " " 0 ± 1σ.
95.44998% " " " 0 ± 2σ.
99.73002% " " " 0 ± 3σ.

Let us set up an experiment wherein we actually draw samples and compare the observed fluctuations in the four statistics with the ranges of sampling fluctuations which can be easily calculated for these statistics. Suppose we take 998 small circular cardboard chips, half of which are green and half white. Suppose we mark 20 of the white chips with 0, 40 with .1, 39 with .2 and so forth as indicated in Table III. Let us mark the green ones in a similar way except that the numbers are negative. The distribution of numbers on the chips, as indicated in this table, corresponds approximately to the normal curve, Fig. 12.

Suppose we take one out and record the number, put it back and stir the chips thoroughly. Suppose we take out another chip, record its number, put it back in the bowl and mix the chips thoroughly and so continue the process until we have made 1000 observations. Now, the effect of replac-

TABLE III

Distribution[24] from which Samples of 1000 were Drawn with Replacement

| Marking X on Chips | Number of Chips | Marking X on Chips | Number of Chips |
|---|---|---|---|
| .0 | 20 | 1.5 | 13 |
| .1 | 40 | 1.6 | 11 |
| .2 | 39 | 1.7 | 9 |
| .3 | 38 | 1.8 | 8 |
| .4 | 37 | 1.9 | 7 |
| .5 | 35 | 2.0 | 5 |
| .6 | 33 | 2.1 | 4 |
| .7 | 31 | 2.2 | 4 |
| .8 | 29 | 2.3 | 3 |
| .9 | 27 | 2.4 | 2 |
| 1.0 | 24 | 2.5 | 2 |
| 1.1 | 22 | 2.6 | 1 |
| 1.2 | 19 | 2.7 | 1 |
| 1.3 | 17 | 2.8 | 1 |
| 1.4 | 15 | 2.9 | 1 |
| | | 3.0 | 1 |

ing each chip before drawing another obviously keeps the probability of drawing a chip with a given number on it the same for all of the drawings. Hence this method of sampling corresponds to the case where the causes controlling the manufacture of product remain the same from one unit to another in respect to the probability of producing a unit with the characteristic of magnitude within a certain specified interval.[25] The results[26] of drawing four such samples are given in Fig. 13. In general the samples differ

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

24. This distribution gives $\sigma=.9966$, $k^2=\beta_1=0$, and $\beta_2=2.9278$ instead of unity, zero and three which are the respective values of these statistics for the normal law.

25. The sample size 1000 was chosen so that we could use customary error theory and thus avoid complications which arise for smaller samples. (See footnote 19).

26. I am indebted to Miss Victoria L. Mial and Miss Marion B. Cater for obtaining these and other experimental results of a similar nature quoted in this paper. They also constructed the figures and made the calculations.

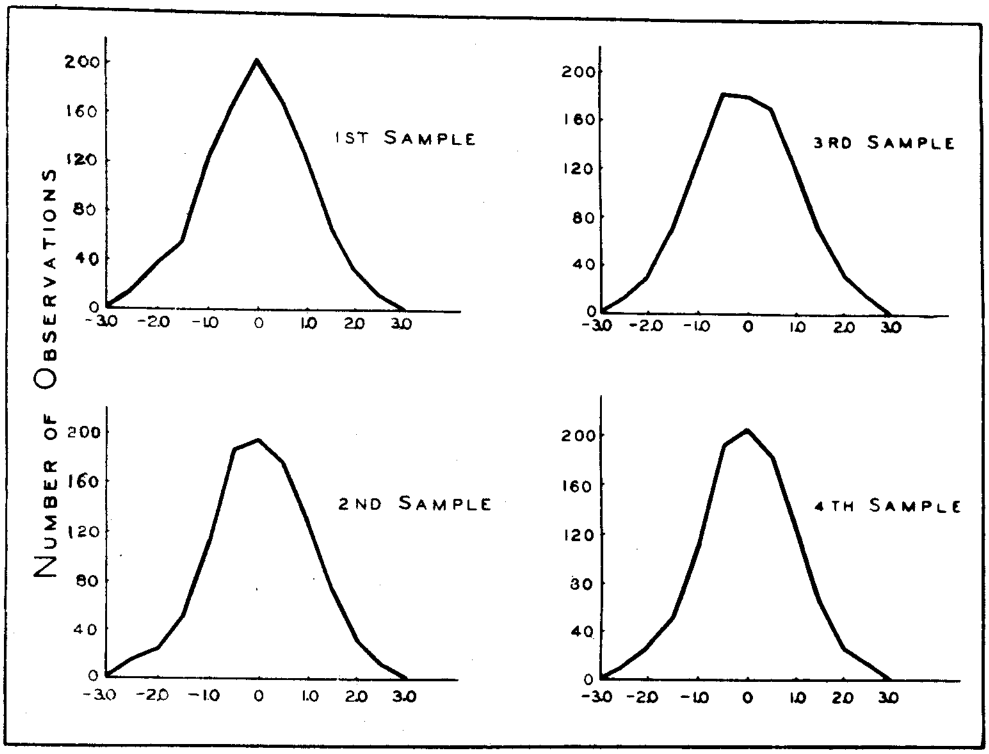



Fig. 13 - DISTRIBUTIONS OF FOUR SAMPLES OF 1000 EACH DRAWN IN SUCH A WAY THAT THE OBSERVED DIFFERENCES IN THE DISTRIBUTIONS WERE PRODUCED BY SAMPLING.

in respect to each of the four statistics $\bar{X}$, $\sigma$, k ($\sqrt{\beta_1}$) and $\beta_2$. Now, statistical theory enables us to calculate the limits of sampling fluctuations of these factors. Thus the theory of sampling tells us, as will be shown below, that the average of samples of 1000 drawn in the way indicated above should not differ from the true average (in this case zero) by more than three times the standard deviation of the average $\frac{\sigma'}{\sqrt{1000}}$ except in about 27 cases out of every 10,000. Limits with similar meaning can be calculated for the fluctuations in each of the other three factors. Fig. 14 shows the observed variations (irregular solid line) in the four statistics to be well within their respective limits (dotted lines).

If we had extended the experiment to include many thousands of drawings of 1000 each, we should probably have found approximately 0.3% of the observed statistics to have fallen outside their respective limits. Carrying this line of reasoning over to the practical case, we should expect occasionally to find a sample of a uniform product possessing one or even more of the four statistics which fall outside the theoretical limits. In other words, the fact that an observed value falls outside these limits does not mean that it is not a sampling fluctuation, but means that it is very probably not a sampling fluctuation. Furthermore, a manufacturing process may actually change, because of some assignable cause, without the effect

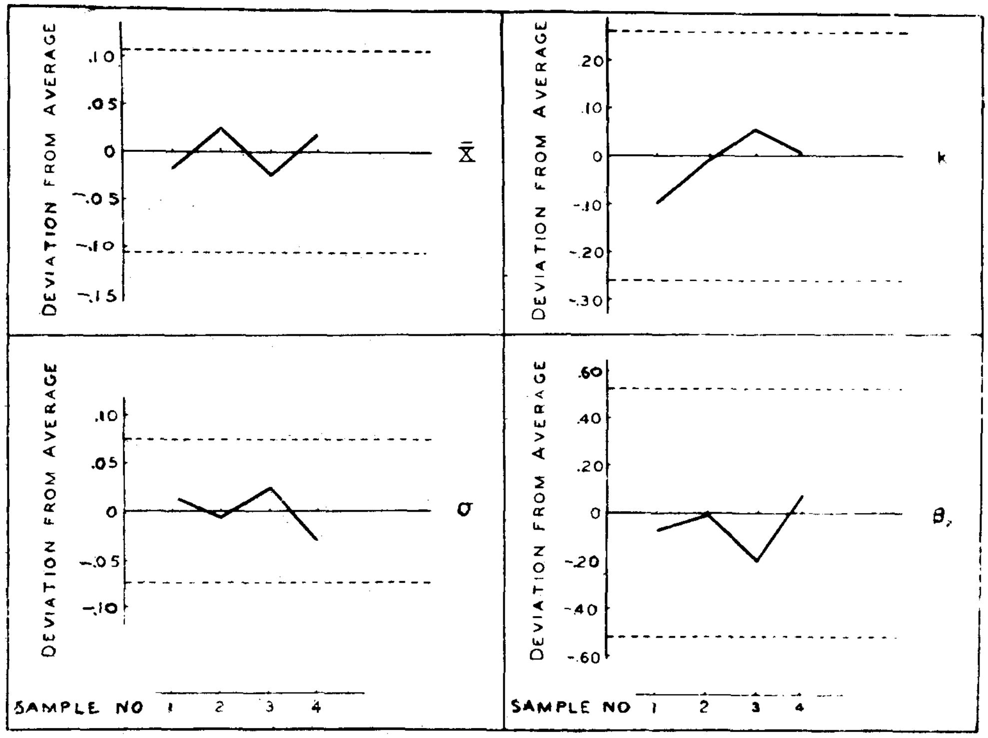



Fig. 14 - OBSERVED FLUCTUATIONS IN THE FOUR STATISTICS [AVERAGE, STANDARD DEVIATION, SKEWNESS AND FLATNESS (KURTOSIS)] FOR FOUR SAMPLES DRAWN FROM THE APPROXIMATELY NORMAL DISTRIBUTION GIVEN IN TABLE III. DOTTED LINES REPRESENT THEORETICAL LIMITS TO SAMPLING FLUCTUATIONS: i.e., LIMITS WITHIN WHICH OBSERVED VALUES SHOULD FALL APPROXIMATELY 99.73% OF THE TIMES THAT THE EXPERIMENT IS TRIED.

of this cause being detected unless the effect is sufficiently large to produce in one or more of the statistics a displacement outside the limits.

CALCULATION OF SAMPLING FLUCTUATIONS: UNKNOWN DISTRIBUTION FROM WHICH SAMPLE IS DRAWN.

The case just considered, however, is slightly different from that which arises in practice, because we have assumed the true frequency distribution, from which the sample is drawn, to be known whereas in practice the true frequency distribution is almost never known. This makes our practical problem a little more complicated, as we can illustrate by means of the above mentioned experiment. To do this, we must start in ignorance of the markings of the chips contained in the bowl except insofar as this information is revealed by the 4000 drawings (with replacement) to which we have referred already. Let us examine the procedure, therefore, to be followed in calculating the sampling fluctuations under these conditions.

Looking at the four polygons in Fig. 13, one would probably assume that the observed distributions might have arisen as samples from a symmetrically distributed universe[27] possibly normal in form. Assuming the universe to be normal for the 1st sample of 1000 data, we find the normal distribution having the same value of average and standard deviation as has the observed distribution. The analysis sheet showing the results was presented in Fig. 9. Column 5 of this sheet gives the observed distribution and column 10 gives the theoretical distribution assuming that the sample came from a normal universe. Column 11 shows the observed differences in cell distribution. How probable is it that such a difference in cell distribution would have been observed if the original distribution had been normal? Pearson's test for goodness of fit gives us the answer, the probability .760. In other words the probability of getting a value of

$\frac{\Sigma(y-y_1)^2}{y_1}$ equal to or greater than 6.612 as a result of random sampling is .760 and our guess that the original distribution was normal in form can be assumed to be reasonably justified.[28] Both the theoretical and observed frequency distributions are presented graphically in Fig. 15 by the smooth curve and the dots respectively. Let us note carefully for the sake of emphasis the meaning of the smooth curve. It represents our best guess as to the distribution of numbers on the chips in the bowl providing we assume

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

27. The distribution from which the sample is drawn is referred to sometimes as the universe or population. In this case the universe is the distribution of numbers on the chips in the bowl.

28. See Appendix 4 for a discussion of the goodness of fit test as it applies to inspection work.

that these are distributed normally and that we do not know the value of the average m and of the standard deviation σ' of the distribution in the bowl

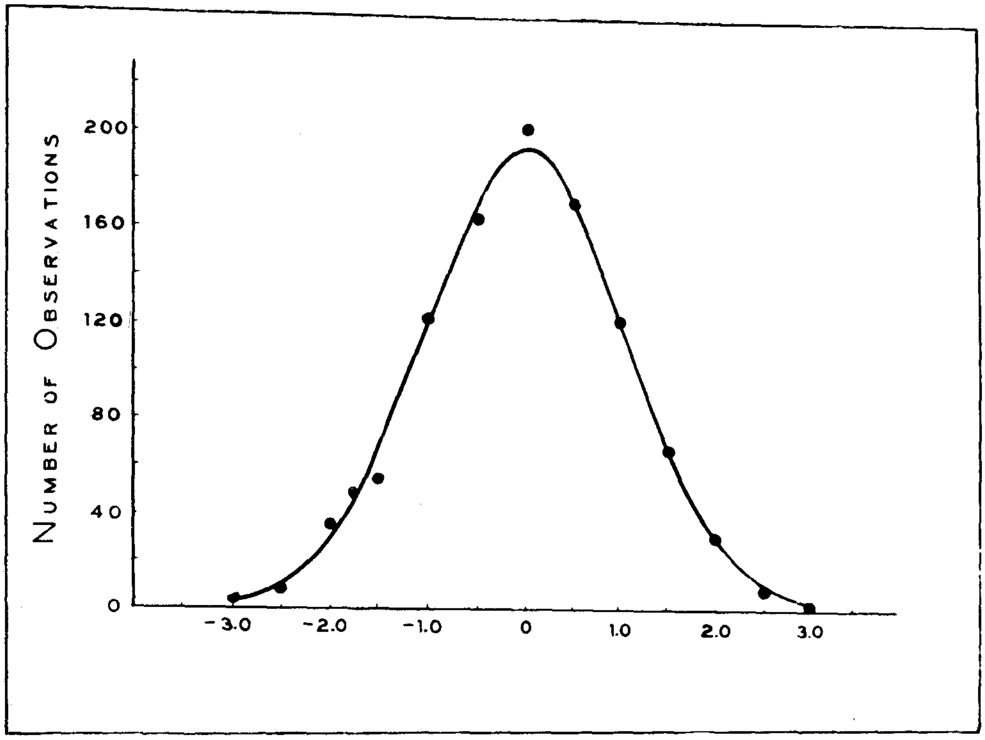



Fig. 15 - COMPARISON OF THE OBSERVED AND THEORETICAL DISTRIBUTION FOR SAMPLE OF 1000 DRAWN FROM A NORMAL UNIVERSE WHOSE EXACT FORM IS UNKNOWN.

except as these are estimated from the sample of 1000. We feel that our guess is justified because the probability of fit is high. To calculate the sampling fluctuations of the statistics $\bar{X}$, $\sigma$, k and $\beta_2$ we proceed as in the previous case except that, since the true values $\bar{X}'$, $\sigma'$, k' and $\beta_2'$ are not known, we may choose to place sampling limits upon the differences to be expected in each of these statistics. This point will be made clearer in footnote 39.

Thus, when the sample is drawn from a population of unknown distribution, we find, for this unknown distribution, an approximate form having a high degree of fit, and then proceed to calculate the sampling fluctuations upon the basis of this assumed form of distribution in the population.

## OUTLINE OF GENERAL THEORY

Let us assume, as before, that uniform product is one such that the probability $dy_{\lambda'}$ of a unit having a quality X within the range X to X + dX is

$$dy_{\lambda'} = f'(X, \lambda_1', \lambda_2', \ldots \lambda_{c_1}')\, dX, \quad \text{---------- } 2$$

where the $\lambda'$ 's represent $c_1$ unknown parameters. Suppose we inspect a sample of n units of apparatus drawn from a uniform product distributed according to Eq. 2. In the general case represented by Eq. 2, there are $c_1$ parameters, all of which are, of course, unknown, whereas in the case treated in the previous section there were only two parameters m and $\sigma'$ (See Eq. 1). In the general case we know only the n observed values of quality and we do not know either the true functional relationship f' or any of the $c_1$ parameters. Our problem is therefore to find the most probable values for both the functional relationship f representing the law of distribution f' and the c parameters of this equation where in general $c_1$ may not be equal to c unless the assumed form f is the true form f'. Theoretically there are three fundamental steps to the solution of this problem.[29] They are:

1. The Problem of Specification: This consists in finding the most probable form f of the distribution of the population.

2. The Problem of Estimation: This consists in finding ways of estimating each of the c parameters from the data given by the sample.

---

29. See, for example, Fisher, R.A., on the Mathematical Foundation of Theoretical Statistics - Phil. Trans. Roy. Soc. of Lon., Series A, Vol. 222, pp. 309-368.
A fourth step should really be included here. This is the application of some test to determine the probability of fit between theory and observation. To outline it in sufficient detail at this point would break the general outline of the theory and hence its discussion is given in Appendix 4. A more specific statement than that given above for the three steps will be given also in that appendix.

3. The Problem of Distribution: This consists in studying the distribution of the estimates of the parameters derived from the sample.[30]

## THE PROBLEM OF SPECIFICATION

Since, in the practical case, we do not know the exact form f' and the values of the $c_1$ parameters of the distribution we must follow the customary scientific procedure of adopting an hypothesis, deducing its consequences and comparing the results with the known facts or experimental data. Thus, when we assumed no à priori knowledge of the distribution of the numbers on the chips in the bowl, we empirically chose the normal law as a trial specification, calculated the theoretical distribution and then compared the theoretical and observed distributions by means of the goodness of fit test. In other words, we chose an assumed form f for the distribution, calculated certain results on the basis of the assumed form and compared the theoretical and observed results in order to see if our assumed form could be justified.

In general the choice of the empirical function to be substituted for f' in Eq. 2 is found to be either of two forms as represented by Eqs. 3 and 4.

$$f(X) = C_0\varphi^0(X) + C_1\varphi^1(X) + C_2\varphi^2(X) + \dots \qquad ----- 3$$

and

$$\frac{1}{f(X)}\frac{d}{dX}f(X) = \frac{X-a}{a_0 + a_1X + a_2X^2 + \dots} \qquad --------- 4$$

where the C's and a's are parameters and $\varphi^i(X)$ represents the ith derivative of the normal error function

$$y = \frac{1}{\sigma'\sqrt{2\pi}} e^{-\frac{(X-m)^2}{2\sigma'^2}} \quad , -------------- 1$$

$\sigma'$ being the root mean square deviation from the average m of the universe from which the sample is drawn.

30. Any estimate of a parameter upon the basis of a finite sample is termed a statistic. In general, primes are used to denote parameters, thus we have $\bar{X}'=m$, $\sigma'$, $k'$ and $\beta_2'$ as parameters and $\bar{X}$, $\sigma$, $k$ and $\beta_2$ as statistics.

The systems of curves given by Eq. 3 and Eq. 4 are well known and need not be considered here.[31] Other forms of specification may be used later, but, for most instances, these have been found quite satisfactory.[32]

THE PROBLEM OF ESTIMATION

How shall we estimate the c parameters? This question has not been studied so extensively as the problem of specification, but one general principle may be laid down: the estimates of the parameters must be symmetric functions of the n observed values of X. We naturally try to use one of the three common types of integral rational symmetric functions. Probably the simplest form of this function, at least from the viewpoint of calculation, is the power-sum. Thus the ith power sum is

$$i = X_1^i + X_2^i + \dots + X_n^i, \quad \text{-------------} \quad 5$$

where $X_1$, $X_2$, ... $X_n$ represent the n observed values of the quantity X and i is a positive integer. This choice of function can be justified, because the other two classes of symmetric functions can be expressed in terms of the power-sums.

Now, in general there are many ways of calculating any parameter of an infinite population. Thus the parameter $\sigma'$ of a normal distribution can be calculated in an infinite number of ways by the equation

---

31. Eq. 3 gives the infinite series which has been studied by Gram, Thiele, Charlier, Edgeworth and others. It is generally known as the Gram-Charlier series.
Eq. 4 is the differential equation of Pearson's generalized system of frequency curves.
One reason why curves belonging to one of these two general types have been used so extensively is that tables which simplify the arithmetical calculations are available through the efforts of numerous investigators of whom Pearson has been the greatest contributor.

32. Thus we have the method of translation and other methods suggested by Edgeworth. See "Untried Methods of Representing Frequency" by F. Y. Edgeworth. Jour. Roy. Stat. Soc., Vol. LXXXVII, Part IV, July 1924.

$$\frac{2}{\sigma' \sqrt{2\pi}} \int_0^{\infty} x^i e^{-\frac{x^2}{2\sigma'^2}} dx = \frac{\sigma'^i 2^{\frac{i}{2}}}{\sqrt{\pi}} \Gamma \frac{(i+1)}{2}, \quad \text{-- 6}$$

where x=X-m and i is any integral number. Why, then, can't we use the same methods to estimate the parameters of the infinite population from the n observed data constituting the sample? For example in the experimental determination of the distribution within the bowl, why can't we estimate the parameters from the sample equally well in any one of the infinite number of ways assuming that there is no difference in the amount of labor involved? The answer is that we can, but all of the estimates do not have the same precision and hence we must use that estimate for which the error of sampling is a minimum, providing this can be found.[33] However, before we can do this, we must solve the problem of distribution.

## PROBLEM OF DISTRIBUTION

In other words, we must find the equation of the distribution of each of the possible estimates for a given parameter and use that estimate with the smallest amount of dispersion. To do this constitutes one of the rather complicated problems of modern statistics, because the distribution of a statistic depends among other things upon the form f' of the distribution of the population and upon the size of the sample.

In practice we customarily use a form of hypothetical frequency distribution which does not involve more than four parameters, $\bar{X}'=m$, $\sigma'$, $k'=\sqrt{\beta_1'}$ and $\beta_2'$. Estimates defined on the analysis sheet Fig. 9 generally are used. The distribution of these four estimates has been shown to be approxi-

---

33. The importance of making efficient use of data is discussed in Appendix 3.

mately normal when the sample size is large[34] and where the distribution from which the samples are drawn is approximately normal. The standard deviations of the four estimates, X, $\sigma$, $k = \sqrt{\beta_1}$ and $\beta_2$ are as follows:

standard deviation of $\bar{X} = \sigma_{\bar{X}} = \frac{\sigma'}{\sqrt{n}}$,

standard deviation of the standard deviation $\sigma = \sigma_\sigma = \frac{\sigma'}{\sqrt{2n}}$,

standard deviation of the skewness $k = \sigma_k = \sqrt{\frac{6}{n}}$,

standard deviation of the flatness (kurtosis) $\beta_2 = \sigma_{\beta_2} = \sqrt{\frac{24}{n}}$.

This gives us the necessary information for setting sampling limits upon each of the four estimates of the parameters, and. although no absolute rules can be formulated to indicate the limitations imposed by the assumptions made in deriving the standard deviations presented in the previous paragraph, the following considerations have proved helpful guides.

a. If the number n of observations is at least 1000, we may assume that all of the statistics are distributed in approximately normal fashion so that the normal law integral table may be used in indicating the significance of deviations measured in terms of $\sigma_k$ and $\sigma_{\beta_2}$.

b. If the number n of observations lies between 100 and 1000 we may assume the distributions of $\bar{X}$ and $\sigma$ to be normal but we may not assume the distributions of the estimates k and $\beta_2$ to be normal.

c. If n is less than 100, none of the four estimates may be assumed to be distributed in normal fashion. Allowance must be made

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

34. Pearson, Karl, and others "On the Probable Errors of Frequency Constants" - Biometrika, Vol. II, 1903, p. 273 - Vol. IX, 1913, p. 1.

Isserlis, L. - "On the Conditions Under which the Probable Errors of Frequency Distributions have a Real Significance" - Proceedings Royal Society, Series A. Vol. XCII, 1915, p.23.

accordingly in the interpretations of deviations measured in terms of the standard deviations of the estimate.[35]

## PRACTICAL APPLICATION OF THE THEORY IN ATTAINING THE FIRST OBJECT OF INSPECTION

Let us return to the questions originally raised in respect to the data presented in Figs. 2 and 3. We shall treat these two cases together, because Fig. 3 represents the distribution of the 12 months' product shown in Fig. 2. We want to use the theory already outlined to show whether or

WESTERN ELECTRIC CO., INC., INSPECTION ENGINEERING DEPT., NEW YORK

**INSPECTION ENGINEERING ANALYSIS SHEET**

OBSERVATIONS

| | |
|---|---|
| CHARACTERISTIC INSPECTED (a) | QUALITY X |
| UNITS (b) | |
| TO NEAREST | |
| SOURCE OF DATA | |

SUBJECT: TYPE A-INSTRUMENT
1. Quality from
2. July 1923 -
3. June 1924 incl.

CALC BY [illegible] CHK BY AO DATE [illegible]-14-31 APPD [illegible] SHEETS SHEET

f(x) Gram-Charlier (3 terms) f(x) Pearson Type IV

| | (a) / (b) | Cell boundaries: Lower | Upper | X (at δ = -5.5) | Obs. freq. y | yX | yX² | yX³ | yX⁴ | Theor. freq. y₁ | y−y₁ | (y−y₁)² | (y−y₁)²/y₁ | Theor. freq. y₂ | y−y₂ | (y−y₂)² | (y−y₂)²/y₂ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | -5.5 | -5.75 | -5.26 | 0 | 13 | 0 | 0 | 0 | 0 | | | | | | | | |
| 1 | -5.0 | -5.25 | -4.76 | 1 | 10 | 10 | 10 | 10 | 10 | | | | | | | | |
| 2 | -4.5 | -4.75 | -4.26 | 2 | 8 74 | 16 | 32 | 64 | 128 | 83 | -9 | 81 | .976 | 101 | -27 | 729 | 7.218 |
| 3 | -4.0 | -4.25 | -3.76 | 3 | 43 | 129 | 387 | 1161 | 3483 | | | | | | | | |
| 4 | -3.5 | -3.75 | -3.26 | 4 | 100 | 400 | 1600 | 6400 | 25600 | 254 | -154 | 23716 | 93.370 | 229 | -129 | 16641 | 72.668 |
| 5 | -3.0 | -3.25 | -2.76 | 5 | 815 | 4075 | 20375 | 101875 | 509375 | 675 | 140 | 19600 | 29.037 | 631 | 184 | 33856 | 53.65[illegible] |
| 6 | -2.5 | -2.75 | -2.26 | 6 | 1761 | 10566 | 63396 | 380376 | 2282256 | 1466 | 295 | 87025 | 59.362 | 1464 | 297 | 88209 | 60.2[illegible] |
| 7 | -2.0 | -2.25 | -1.76 | 7 | 2397 | 16779 | 117453 | 822171 | 5755197 | 2662 | -265 | 70225 | 26.381 | 2708 | -311 | 96721 | 35.717 |
| 8 | -1.5 | -1.75 | -1.26 | 8 | 3431 | 27448 | 219584 | 1756672 | 14053376 | 3679 | -248 | 61504 | 16.718 | 3720 | -289 | 83521 | 22.4[illegible] |
| 9 | -1.0 | -1.25 | - .76 | 9 | 3703 | 33327 | 299943 | 2699487 | 24295383 | 3442 | 261 | 68121 | 19.791 | 3476 | 227 | 51529 | 14.824 |
| 10 | - .5 | - .75 | - .26 | 10 | 2165 | 21650 | 216500 | 2165000 | 21650000 | 2025 | 140 | 19600 | 9.679 | 1997 | 168 | 28224 | 14.133 |
| 11 | 0 | - .25 | .24 | 11 | 510 | 5610 | 61710 | 678810 | 7466910 | 686 | -176 | 30976 | 45.155 | 625 | -115 | 13225 | 21.160 |
| 12 | .5 | .25 | .74 | 12 | 77 | 924 | 11088 | 133056 | 1596672 | | | | | | | | |
| 13 | 1.0 | .75 | 1.24 | 13 | 15 94 | 195 | 2535 | 32955 | 428415 | 78 | 16 | 256 | 3.282 | 102 | - 8 | 64 | .627 |
| 14 | 1.5 | 1.25 | 1.74 | 14 | 2 | 28 | 392 | 5488 | 76832 | | | | | | | | |
| 15 | | | | | | | | | | | | | | | | | |
| | Σ | | | | 15050 | 121157 | 1015005 | 8783525 | 78143637 | 15050 | | | 303.751 | 15053 | | | 30[illegible].706 |

No. of cells = 10, P = .00000; No. of cells = 10, P = [illegible]

MOMENTS ABOUT ORIGIN δ

NOTE: THE ORIGIN δ IS THE MID-VALUE BETWEEN THE BOUNDARIES OF THE CELL CHOSEN

$'\mu_1 = \frac{\Sigma yX}{\Sigma y} = \frac{121157}{15050} = 8.050299$ $\quad m$ = units (b) per cell = .5

$'\mu_2 = \frac{\Sigma yX^2}{\Sigma y} = \frac{1015005}{15050} = 67.442193$ $\quad \bar{X} = \delta + m\,'\mu_1 = -5.500000+4.025150 = -1.474850$

$'\mu_3 = \frac{\Sigma yX^3}{\Sigma y} = \frac{8783525}{15050} = 583.622924$

$'\mu_4 = \frac{\Sigma yX^4}{\Sigma y} = \frac{78143637}{15050} = 5192.268239$

(Corrected moments)

$\sigma = m\mu_2^{1/2} = .5 \times 1.597356 = .798678$

$k = \frac{\mu_3}{\mu_2^{3/2}} = \frac{-1.730023}{4.075727} = -.424470$

$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{22.144232}{6.510387} = 3.401370$

$\beta_1 = k^2 = (-.424470)^2 = .180175$

$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{.798678}{\sqrt{15050}} = .00651224$

$\sigma_\sigma = \frac{\sigma}{\sqrt{2n}} = \frac{.798678}{\sqrt{30100}} = .00460330$

$\sigma_k = \sqrt{\frac{6}{n}} = [illegible] = .0199667$

$\sigma_{\beta_2} = \sqrt{\frac{24}{n}} = [illegible] = .0399334$

UNCORRECTED MOMENTS ABOUT ARITH. MEAN

$\mu_1 = 0$

$\mu_2 = '\mu_2 - '\mu_1^2 = 67.442193-64.807314 = 2.634879$

$\mu_3 = '\mu_3 - 3\,'\mu_1\,'\mu_2 + 2\,'\mu_1^3 = 583.622924-1628.789457+1043.436510 = -1.730023$

$\mu_4 = '\mu_4 - 4\,'\mu_1\,'\mu_3 + 6\,'\mu_1^2\,'\mu_2 - 3\,'\mu_1^4 = 5192.268239-18793.556164+26224.484274-12599.763844 = 23.432505$

$3\sigma_{\bar{X}} = .0195[illegible]$

$3\sigma_\sigma = .0138[illegible]$

$3\sigma_k = .0599001$

$3\sigma_{\beta_2} = .1198[illegible]$

CORRECTED MOMENTS ABOUT ARITH. MEAN (SHEPPARD'S CORRECTIONS)

$\mu_2(\text{cor.}) = \mu_2 - \frac{h^2}{12} = 2.634879-.08333 = 2.551546$

$\mu_4(\text{cor.}) = \mu_4 - \frac{1}{2}h^2\mu_2 + \frac{7}{240}h^4 = 23.432505-1.317440+.020167 = 22.144232$

$h = 1$

(10-24)

Fig. 16

not this product was uniform. We assume that if it had been uniform, it could be represented by some one of the well known types of frequency curves such as given by Eqs. 3 and 4. Our first step in the detection of non-uni-

---

35. Loc. cit. footnote 19.

formity is, therefore, to see if the distribution presented in Fig. 3 can be fitted by such a smooth curve.

The original grouped data are presented on the analysis sheet in Fig. 16. The values of the estimates $\beta_1$ and $\beta_2$ suggest that we try either type IV of the Pearson family of curves, or the first few terms of the Gram-Charlier series. This has been done. Column 10 represents the theoretically derived distribution making use of only three terms of the Gram-Charlier series. Column 14 gives the distribution making use of the type IV Pearson curve. The probabilities of fit in both instances are practically negligible, and hence we have reason to believe that the distribution of product is not uniform.[36] This is assumed to be true even though the theoretical distribution given in column 10 and represented by the solid curve in Fig. 17 may indicate comparatively close agreement.[37]

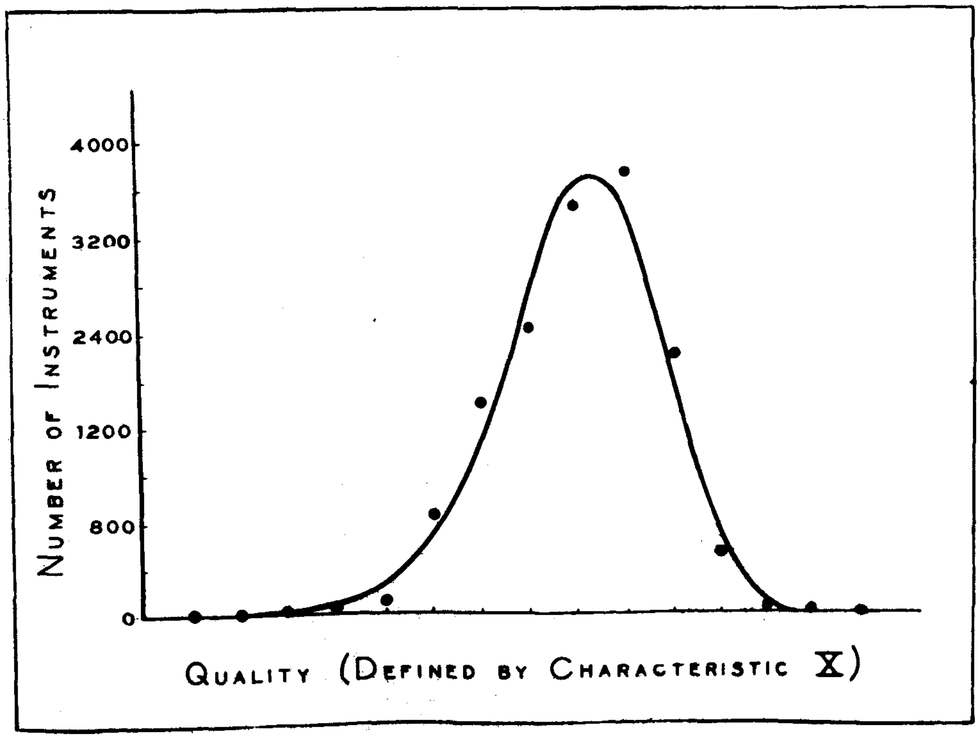



Fig. 17 - DATA OF FIG. 3 FITTED WITH THEORETICAL FREQUENCY CURVE. ALTHOUGH FIT MAY APPEAR TO BE FAIR, THE PROBABILITY OF FIT TEST SHOWS IT TO BE POOR.

---

36. See Appendix 2 for cases where a good fit may be found even though the product is not uniform.

37. The theoretical distribution calculated with the aid of four terms of the Gram Charlier series did not give as high a fit as the one

The attempt to fit each of the monthly distributions also met with failure. Then the statistics X, $\sigma$, k and $\beta_2$ were calculated for each of the 12 monthly samples and the monthly observed values of these four factors were plotted as indicated in Fig. 18. The ordinates are given in terms of observed deviations from the values for these four factors presented on the analysis sheet in Fig. 16. The dotted lines drawn on each of the four charts in Fig. 18 show the limits within which these factors might be expected to

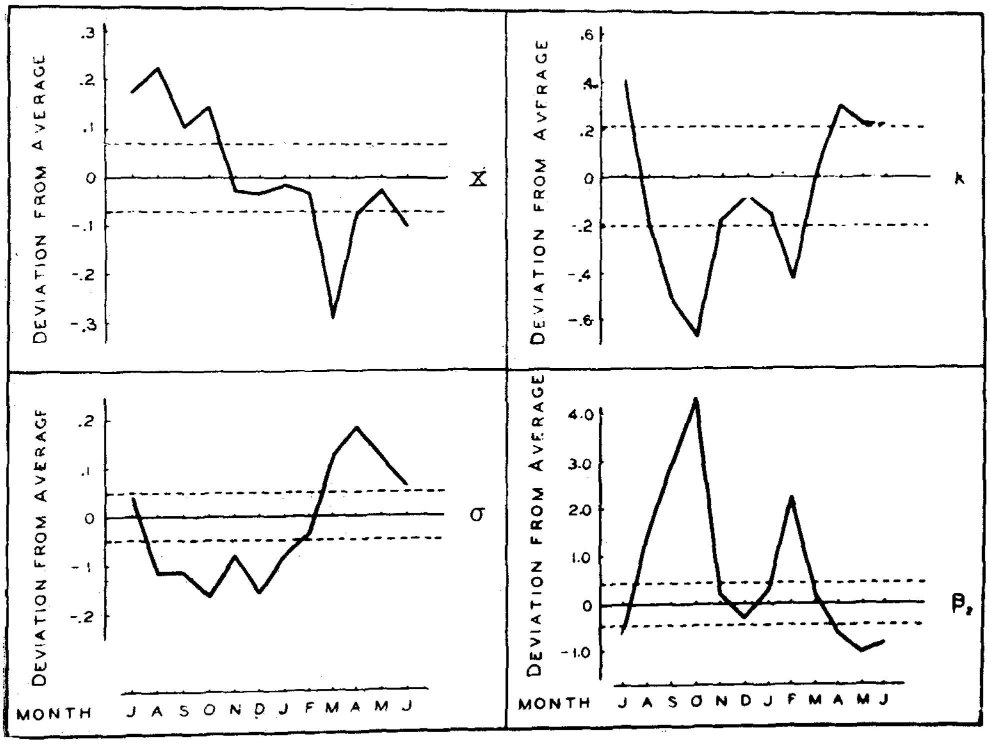



Fig. 18 - VARIATIONS IN THE MONTHLY ESTIMATES OF THE STATISTICS $\bar{X}$, $\sigma$, k and $\beta_2$ FOR THE DATA IN FIG. 2.

fluctuate, because of sampling.[38] But, since the factors fluctuate outside these limits, Fig. 18 presents practically conclusive evidence that

---

38. In other words, 99.73% of the observed values for each factor should lie within the limits for this factor.

the product was not uniform over this period.[39]

CONTROL CHARTS: VARIABLES.

Charts such as that shown in Fig. 18 are thought of as control charts. So long as deviations from month to month fall within the limits shown on this form of chart we may feel quite certain that the variations may have been produced by some complex system of non-assignable causes which it would be very impractical to try to control. If, however, some one or more of the factors shown in the chart fall outside these limits, it is reasonable to believe that there may be some assignable cause for the observed effect. Hence such a chart is helpful in indicating trouble which it should be possible to find and probably control.

CONTROL CHARTS: ATTRIBUTES:

Let us now consider the application of the theory to the establishment of a form of control chart which applies when the method of attributes is used in inspection or, in other words, the chart which applies when a piece of apparatus is classed as being either above or below a certain standard quality. Customarily the quality standard is so chosen that instruments whose quality is less than the standard are termed defective. It usually proves to be economically unsound to refine the production procedure to an extent requisite for insuring that all units of product will conform to the quality standard, and hence even in a uniform product we may expect to find in the long run a certain fraction of the total number of instruments to be defective. This fraction $\overline{p}'$ of the number of units of product must be rejected and later modified or junked. However, the observed number pn found

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

39. Some objection may be raised to the conclusion derived from a consideration of Fig. 18, because it may be argued that the grand average for each of the four factors used as bases for measuring the deviation, would not be the true value even though the product had been uniform. We may, of course, get around this difficulty by finding the standard deviations of monthly differences from the grand average, and using these to set up limits for the differences. This has been done and leads essentially to the same conclusion as to the non-uniformity of product. In the practical case we felt justified in using the chart of Fig. 18, because previous experience based upon many hundred thousands of observations had established averages for each of the four factors which were practically the same as those shown on the chart.

defective in samples of n pieces of apparatus will show variations about the true average $\overline{p}'n$ even though the product is uniform or, in other words, even though the product is being manufactured under a constant system of causes. We must, therefore, set up certain limits within which we may expect observed values of p to fluctuate because of sampling. Essentially this is the same type of problem as the one already considered, and it differs only in certain technical details as to the way of carrying out the calculations. Naturally we do not know the true value $\overline{p}'$, but we do know that the standard deviation of an observed value of this factor is equal to $\sqrt{\frac{\overline{p}'(1-\overline{p}')}{n}}$ where n is the size of the sample which is measured. Thus, in cases where millions of pieces of a given kind of apparatus are inspected each year, the precision with which the process average $\overline{p}'$ can be estimated is very high indeed.[40] Fig. 19 presents a copy of a summary analysis sheet which provides:

a. A summary of past data for as long a period as desired.

b. A monthly record for the current year requiring less than five minutes calculation each month to bring it up to date.

c. A chart similar to the final form of the monthly report.

The middle line in the chart, Fig. 19, is the grand average of the percent found defective between January 1, 1923 and December 31, 1923. The limit lines are so set[41] that, providing the product is running uniform with a process average which is the same as that observed over the period indicated,

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

40. Of course we must substitute the observed value of $\overline{p}'$ in the expression $\sqrt{\frac{\overline{p}'(1-\overline{p}')}{n}}$

41. These limits are generally set with the aid of the law of small numbers.

SUMMARY OF CHECK INSPECTION DATA

APPARATUS TYPE B

| Year 1924 | | MONTHLY REPORT | | | | | Totals from Jan.1,1923 to date | | | | Averages from Jan.1,1923 to date | | | | | Checked 1-23-25 by MPC | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Manu-factured | Inspected | | Defective | | No. | Manufactured | Inspected | Def. | Manu-factured | Inspected | | Defective | | | |
| No. | Month | Quantity | Quant. | % | Quant. | % | No. | Quantity | Quant. | Quant. | Quantity | Quant. | % | Quant. | % | Calc. by | No. |
| 0 | Grand Total from Jan.1,1923 to Dec.1,1923 | | | | | | | 215166 | 1680 | 58 | 17930.5 | 140.00 | | 4.83 | 3.45 | MMD | 0 |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | |
| 1 | JANUARY | 16917 | 182 | 1.08 | 4 | 2.20 | | 232083 | 1862 | 62 | 17852.54 | 143.23 | .80 | 4.77 | 3.33 | | 1 |
| 2 | FEBRUARY | 13260 | 107 | .81 | 11 | 10.28 | | 245343 | 1969 | 73 | 17524.50 | 140.64 | .80 | 5.21 | 3.71 | | 2 |
| 3 | MARCH | 31043 | 185 | .60 | 18 | 9.73 | | 276386 | 2154 | 91 | 18425.73 | 143.60 | .78 | 6.07 | 4.23 | | 3 |
| 4 | APRIL | 21864 | 130 | .60 | 3 | 2.31 | | 298250 | 2284 | 94 | 18640.63 | 142.75 | .77 | 5.88 | 4.12 | | 4 |
| 5 | MAY | 28044 | 140 | .50 | 0 | 0 | | 326294 | 2424 | 94 | 19193.76 | 142.59 | .74 | 5.53 | 3.88 | | 5 |
| 6 | JUNE | 23530 | 135 | .57 | 2 | 1.48 | | 349824 | 2559 | 96 | 19434.67 | 142.17 | .73 | 5.33 | 3.75 | | 6 |
| 7 | JULY | 15349 | 141 | .92 | 1 | .71 | | 365173 | 2700 | 97 | 19219.63 | 142.11 | .74 | 5.11 | 3.60 | | 7 |
| 8 | AUGUST | 15725 | 133 | .85 | 5 | 3.76 | | 380898 | 2833 | 102 | 19044.90 | 141.65 | .74 | 5.10 | 3.60 | | 8 |
| 9 | SEPTEMBER | 15509 | 157 | 1.01 | 2 | 1.27 | | 396407 | 2990 | 104 | 18876.52 | 142.38 | .75 | 4.95 | 3.48 | | 9 |
| 10 | OCTOBER | 10666 | 144 | 1.35 | 0 | 0 | | 407073 | 3134 | 104 | 18503.32 | 142.45 | .77 | 4.73 | 3.32 | | 10 |
| 11 | NOVEMBER | 11256 | 140 | 1.24 | 6 | 4.29 | | 418329 | 3274 | 110 | 18188.22 | 142.35 | .78 | 4.78 | 3.36 | | 11 |
| 12 | DECEMBER | 12100 | 201 | 1.66 | 4 | 1.99 | | 430429 | 3475 | 114 | 17934.54 | 144.79 | .81 | 4.75 | 3.28 | | 12 |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | | |
| 13 | Grand Total from Jan. 1, 1923 to Dec. 31, 1924 | | | | | | | 430429 | 3475 | 114 | 17934.54 | 144.79 | .81 | 4.75 | 3.28 | | 13 |

MONTHLY AVERAGES

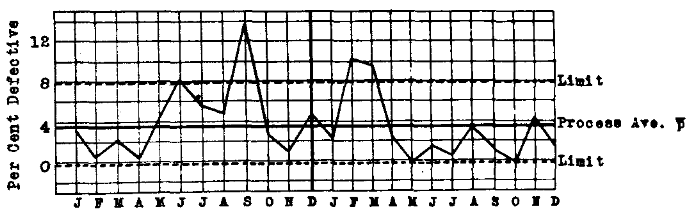



Fig. 19 - ANALYSIS SHEET USED IN PREPARING CONTROL CHART - METHOD OF ATTRIBUTES.

the observed values of $\overline{p}'$ will fall between the limits in approximately 99% of the cases.[42] Of course, if sufficient data are not available to assure a comparatively high degree of precision in the determination of the process average $\overline{p}'$, it is possible, as in the case of variables (footnote 39), to set limits on the differences between the observed value of $\overline{p}'$ for the entire previous period and the monthly observed values of $\overline{p}'$.

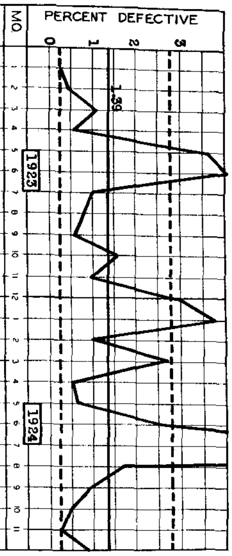
Figure 20 presents six typical examples of control charts prepared from analysis sheets such as shown in Fig. 19. These six examples cover a

---

42. Of course limit lines can be set which would include more or less than 99% of the cases.

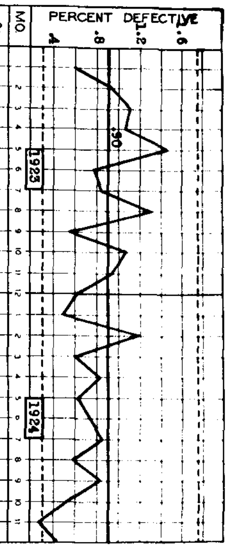
TYPE D

| NO. MFD. | NO. INSP. |
|---|---|
| 4188 | 555 |
| 2681 | 546 |
| 9257 | 689 |
| 10564 | 552 |
| 10780 | 534 |
| 5863 | 540 |
| 6423 | 525 |
| 11996 | 400 |
| 17032 | 744 |
| 21708 | 433 |
| 16094 | 435 |
| 24677 | 385 |
| 30089 | 450 |
| 31686 | 549 |
| 28326 | 755 |
| 23970 | 560 |
| 25567 | 330 |
| 36685 | 500 |
| 30103 | 490 |
| 28688 | 334 |
| 31250 | 609 |
| 24687 | 479 |
| 19474 | 290 |
| 27551 | 487 |



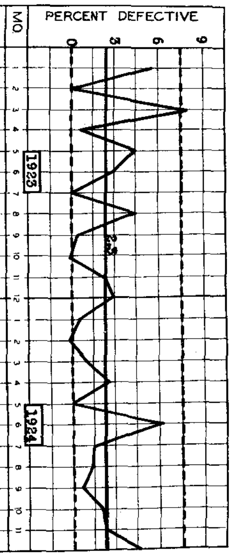

TYPE A - INSTRUMENT

| NO. MFD. | NO. INSP. |
|---|---|
| 86749 | 700 |
| 98692 | 656 |
| 135655 | 1180 |
| 111083 | 757 |
| 122876 | 840 |
| 135959 | 922 |
| 113536 | 958 |
| 113686 | 672 |
| 149879 | 973 |
| 113744 | 628 |
| 109600 | 860 |
| 133777 | 1955 |
| 118135 | 1713 |
| 136701 | 1827 |
| 171054 | 2253 |
| 115023 | 1869 |
| 131671 | 1980 |
| 141153 | 2303 |
| 85049 | 1539 |
| 88169 | 1071 |
| 109702 | 1723 |
| 57737 | 1598 |
| 83066 | 1554 |
| 91113 | 1891 |



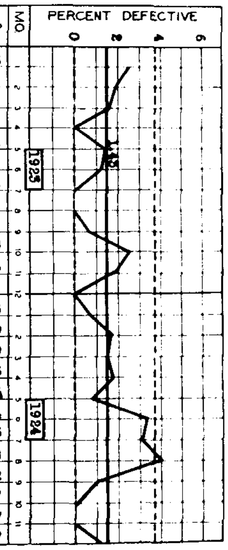

TYPE E

| NO. MFD. | NO. INSP. |
|---|---|
| 4049 | 35 |
| 5850 | 31 |
| 7770 | 20 |
| 2363 | 111 |
| 6628 | 111 |
| 5801 | 140 |
| 2428 | 95 |
| 3324 | 90 |
| 8198 | 122 |
| 4364 | 96 |
| 4434 | 118 |
| 3159 | 104 |
| 3343 | 110 |
| 6769 | 78 |
| 9379 | 89 |
| 7606 | 121 |
| 7032 | 85 |
| 5887 | 80 |
| 3498 | 67 |
| 4584 | 75 |
| 4917 | 89 |
| 4463 | 98 |
| 4797 | 83 |
| 2886 | 62 |



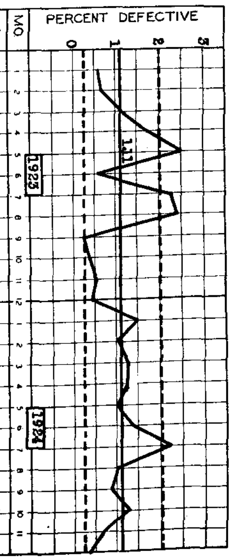

TYPE B

| NO. MFD. | NO. INSP. |
|---|---|
| 11981 | 155 |
| 16613 | 249 |
| 18016 | 244 |
| 13143 | 84 |
| 16584 | 212 |
| 16924 | 169 |
| 13809 | 76 |
| 14512 | 29 |
| 25020 | 137 |
| 22854 | 228 |
| 24051 | 166 |
| 20604 | 187 |
| 20848 | 263 |
| 20870 | 260 |
| 33986 | 269 |
| 21579 | 233 |
| 20921 | 126 |
| 31863 | 184 |
| 21558 | 195 |
| 23947 | 178 |
| 32331 | 192 |
| 17643 | 147 |
| 21312 | 157 |
| 24153 | 157 |



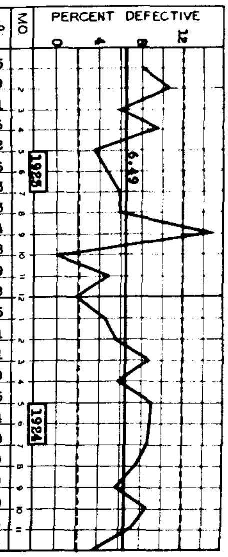

TYPE F

| NO. MFD. | NO. INSP. |
|---|---|
| 73502 | 1274 |
| 47949 | 1152 |
| 84027 | 1883 |
| 75328 | 1442 |
| 47945 | 1157 |
| 74038 | 1246 |
| 73348 | 1465 |
| 72181 | 1413 |
| 105361 | 1825 |
| 106587 | 1257 |
| 99775 | 1516 |
| 90108 | 1044 |
| 89427 | 1173 |
| 96873 | 1153 |
| 148403 | 1683 |
| 114002 | 1275 |
| 112036 | 816 |
| 137098 | 1051 |
| 127309 | 944 |
| 151341 | 899 |
| 13794? | 1102 |
| 88147 | 872 |
| 107193 | 728 |
| 74664 | 1130 |




TYPE C

| NO. MFD. | NO. INSP. |
|---|---|
| 6849 | 195 |
| 8987 | 189 |
| 12047 | 171 |
| 8544 | 186 |
| 7992 | 172 |
| 7108 | 156 |
| 5809 | 105 |
| 8353 | 103 |
| 8163 | 144 |
| 7292 | 118 |
| 4508 | 129 |
| 4493 | 148 |
| 3482 | 135 |
| 5463 | 111 |
| 6828 | 121 |
| 7443 | 138 |
| 7264 | 65 |
| 7771 | 134 |
| 7614 | 130 |
| 8700 | 80 |
| 8561 | 139 |
| 3798 | 90 |
| 3079 | 76 |
| 4578 | 146 |




F... 20 TYPICAL CONTROL CHARTS - METHOD OF ATTRIBUTES.

wide range of conditions. A fluctuation of the observed estimate of $\overline{p}'$ outside the limits indicates the existence of assignable causes of variation in the product whereas fluctuations within the limits can be attributed to sampling.[43]

EFFECT OF ERROR OF OBSERVATION UPON CONTROL CHARTS

Making use of the theory for correction of observed data for errors of measurement,[44] it is possible to determine the effect of errors of mea-

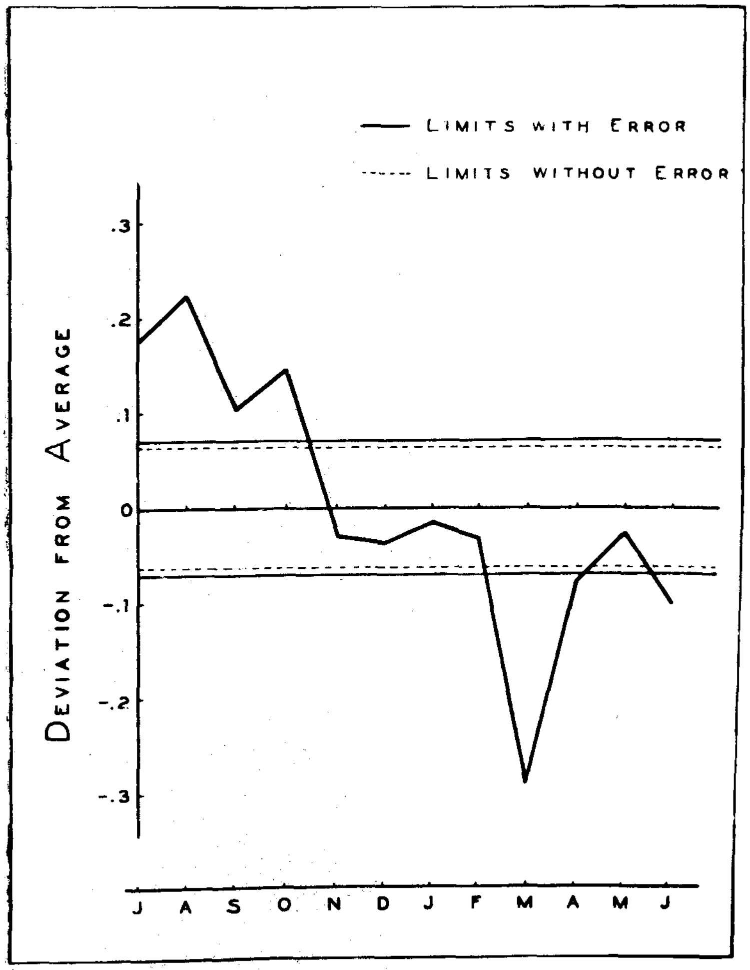



Fig. 21 - CHART SHOWING EFFECT OF ERRORS OF MEASUREMENT UPON SPACING OF LIMIT LINES ON CONTROL CHART.

---

43. It is, of course, necessary in preparation of such control charts to take into account several minor details in the way of calculation but no attempt need be made here to go into these details.

44. Loc. cit. Footnote 13.

surement upon the limits within which product may be expected to vary without indicating the presence of effects of assignable causes. Thus, Fig. 21 shows how errors of measurement modify the spacing between the limit lines for the average shown in Fig. 18. In Fig. 21 the dotted lines show the limits which would actually have existed had it not been that the method of measurement was subject to error. A chart such as that given in Fig. 21 shows the effect of errors of measurement in increasing the range over which product may vary without indicating the existence of assignable causes of variation.

## EFFICIENT USE OF DATA IN ESTABLISHING CONTROL CHARTS

Naturally we are interested in maintaining the best possible control of product at a minimum of cost. Ways of making efficient use of data discussed in some detail in Appendix 3, will now be applied to the problem of establishing control charts. As already pointed out in connection with the study of analysis of data, one or the other of two methods for calculating the estimate of $\sigma'$ of the population is used customarily. Now, in the preparation of the limit lines for the average in Fig. 18 the root mean square method was used. Approximately 1250 units of apparatus had to be inspected each month in order to maintain the limits indicated in this figure. If we had used the mean deviation method we should have had to make on an average (1.14) x (1250) or 1425 measurements each month in order to maintain the same degree of precision, i.e., the same limits.[45] For Type A apparatus, for which the control chart Fig. 18 was made, the cost of making one measurement is approximately one dollar so that the yearly saving accruing from using the root mean square deviation instead of the mean deviation estimate of $\sigma'$ is $2000. Such savings become even more significant when many kinds of apparatus are being inspected.

## POSSIBLE MODIFICATIONS OF THE CONTROL CHARTS

Naturally certain modifications of the control charts suggest themselves. For example it is conceivable that control charts may prove

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

45. Similarly, if we had used other powers in accord with Eq. 5 we should have had to make correspondingly more observations.

useful even when regular or long time trend movements are superposed upon the effects of some constant system of causes. For example we may have a seasonal variation such as represented schematically in Fig. 22. It should be possible to discover such a trend, providing data are available over a sufficiently long time, and to allow for the effects of sampling fluctuations as indicated schematically in this figure. Other instances may arise where even irregular charts must be used. Thus the product may be known to change because of some assignable cause. Such a change would necessitate

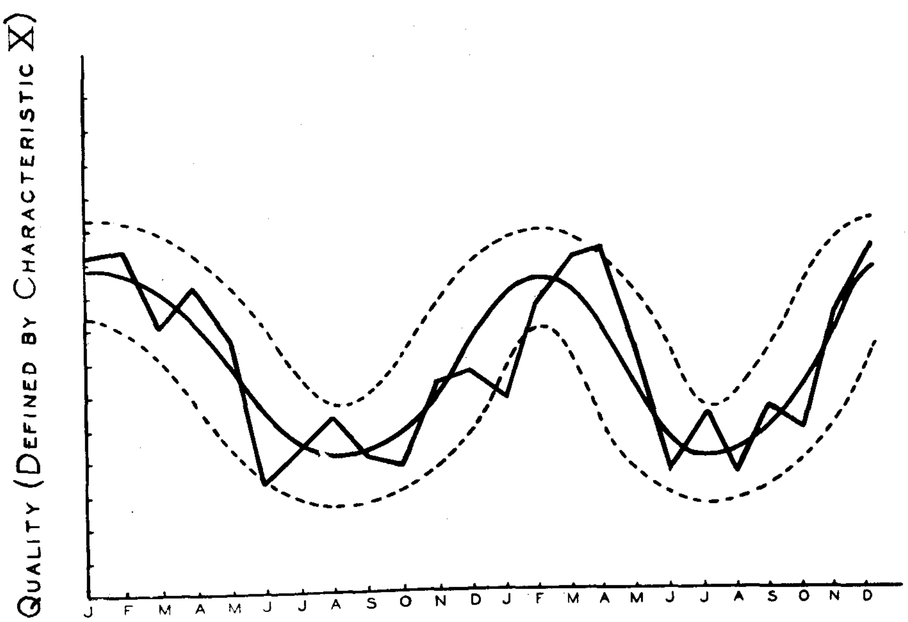



Fig. 22 - SCHEMATIC FORM OF CONTROL CHART TO TAKE SEASONAL OR CYCLIC FLUCTUATIONS.

the establishment of a new set of parameters and this fact would need to be taken into account in the preparation of the chart. As an example let us consider the manufacture of timber products such as poles, cross-arms, and pins. Of course the quality of such products depends upon certain characteristics of the standing timber. But the quality of standing timber may not be uniform even over comparatively small areas. Thus the wood in trees grow-ing on a mountain range may depend upon whether the trees grow on the east or west slope. Product manufactured from trees on one slope may, therefore, give one set of limit lines and product manufactured from trees on the other slope may give another set. Similar conditions may arise in a manufacturing plant where the source of raw material is not regular or where some step in the manufacturing process must of necessity be modified at times throughout the year. All of these elements must be taken into account in the preparation and interpretation of the control charts.

## CONTROL CHARTS BASIS FOR RATING

Not only do control charts indicate the presence of assignable causes of variation in quality of product, but they also form the basis for rating quality.

It may not be out of place to outline the way in which these control charts form the basis for rating. In the general case quality is determined by those characteristics which make a thing what it is and, customarily, more than one characteristic is considered necessary to determine the quality of a unit of apparatus. Thus if there be e different characteristics we may say that quality Q is a function of these e characteristics, as for example, $Q = \Psi(X,Y,Z,\ldots)$ where X, Y, Z, .... represent the e characteristics. Even for a controlled product, X, Y, Z, ... will vary within certain limits, because of the effects of non-assignable causes. Hence the quality Q will be subject to sampling fluctuations, the determination of which depend upon whether or not X, Y, Z, ... are independent.

Now, the rate is in turn a function of the quality Q, and hence the effects of non-assignable causes upon the quality Q are mirrored in the rate. Therefore we must use the information revealed in the control charts as a basis for establishing a control chart for the rate: i.e., limits within which the rate may be expected to fluctuate, because of the effects of non-assignable causes.

## PART IV - SUMMARY

The viewpoint taken in respect to inspection is considerably different from that which looks upon inspection as the routine process of weeding out defective pieces of apparatus.[46] The viewpoint taken is that inspection is the scientific process of collecting and interpretating data in terms of the controlling causes constituting the manufacturing process. A theory of inspection must necessarily guide each of the four steps in attaining any one of the three objects of inspection.

Inspection is essentially different from the so-called exact sciences insofar as the data cannot be completely explained in terms of assignable causes or known natural laws. Instead the interpretation of inspection data involves the application of laws of chance. Essentially an inspector believes that every observed effect or phenomenon is a necessary consequence of a previous state of things even though he is not able to trace this connection in terms of assignable causes. Therefore, to make the best use of inspection data, it is necessary for the inspection engineer to be acquainted not only with the known natural laws but also with the theory of statistics.[47] The essential characteristic of the theoretical basis proposed for attaining the first object of inspection consists in the definition of a uniform product as one whose variations are controlled by non-assignable causes. It is assumed that the probability $dy_{\lambda'}$, of producing a unit having a quality X within the range X to X + dX will be given by

$$dy_{\lambda'} = f'(X, \lambda'_1, \lambda'_2, \ldots \lambda'_{c_1})\, dX, \quad \text{- - - - - - - - - -} \quad 2$$

so long as the product is uniform. To detect the effects of non-uniformity we must show that the observed distribution of product is not consistent

---

46. Not only should inspection weed out defective pieces, but it should also weed out the causes of defective pieces.

47. This idea is amplified in the paper referred to in footnote 23.

with such an assumed law.

It is hoped that the outline of the theory given in this paper will be helpful in making possible the application of this method to inspection problems, particularly through the use of control charts.

Much remains to be done by way of amplification of the statistical methods referred to in the discussion presented in this paper. In certain instances the mathematical theory developed for analyzing and interpreting data in fields of biology, psychology, meteorology, physics, etc., may be carried over to the field of engineering inspection. In other instances engineering inspection offers problems which apparently have not been considered elsewhere.

We have considered the three objects of inspection:

1. The detection of the existence of assignable causes of variation in product.
2. The discovery of the assignable causes.
3. The establishment of economic standards.

We have considered four steps which must be taken in attaining each of the three steps:

a. Establishment of the inspection procedure.
b. Collection of data.
c. Analysis of data.
d. Interpretation of data.

Inspection theory makes it possible to take each step at a minimum of cost and hence to attain the objects of inspection at a minimum of cost. It shows how large a sample to choose, i. e., it shows when the sample is just large enough to give the desired precision but not larger than is necessary to give this precision; it shows how many measurements to make on each unit; it shows whether or not any saving can be effected by treating the measurements as variables instead of attributes. It shows how to keep from taking unnecessary data, to record data so as to save time in analysis, to make possible the separation of the effects of assignable from those of non-assignable causes and to detect trends and cyclic causes of variation; how to analyze data in order to make the most efficient use of them, thereby reducing to a minimum the number of measurements required; how to correct data for errors of measurement; how to make allowance for the size of the sample and how to rate the quality of product.

In other words Inspection theory makes possible a comparison of the efficiencies of different inspection methods and a calculation of the risks and advantages of sampling inspection. The theory helps to detect the effects of assignable causes of variation in product and to find those causes at a minimum of cost.

## APPENDIX 1[48]

## SIZE OF SAMPLE

To reduce the cost of inspection of manufactured product, careful consideration of the factors involved indicates that partial or sampling inspection frequently may be employed. From each group or lot of N items, a sample of n items may be selected and examined. If the sample satisfactorily meets the tests made upon it, the entire lot may be approved. If the sample does not meet the tests, then either the lot may be rejected or additional inspection be made.

When any scheme for sampling inspection is to be established, the first question which arises is: How much inspection should be made; should 50% of the product be inspected or only 10% say? This is often a difficult question to answer. Even when one's judgment is backed by a considerable amount of experience, it is not easy to state decisively that a certain sample size is economically the best. The reason for this, of course, is that the most economical sample size is not the one that results in the cheapest inspection but the one which provides the best balance between cost of inspection and the benefit derived from inspection.

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

48. I am indebted to Mr. A. O. Beckman for the preparation of this Appendix

The benefit which results from inspection lies in the information which is obtained. When each item in the lot of product is inspected, the benefit is the greatest, for the information is complete. But, at the same time, the cost is a maximum. With sampling inspection the cost is reduced, but the information obtained is incomplete. Thus, if we inspect a sample of n units of a product taken from a lot containing N units, we know only the status of the n units and can merely guess at the nature of the remaining N-n units which are not inspected. Obviously the guess is not a blind one, for usually the uninspected portion of the lot is similar to the portion which is inspected, but we can never be certain of this fact without complete inspection. The sample may, and usually will, because of errors of sampling, have a composition somewhat different from that of the lot, and we may easily make a serious mistake in accepting a lot merely because the sample is acceptable. There is this chance that the sample may prove satisfactory although the lot from which the sample is drawn contains actually an excessive number of defective units and is therefore not acceptable. This chance always exists in sampling inspection, and detracts from the value of this kind of inspection.

Thus we see that the saving effected by sampling inspection is not a pure profit, for sampling inspection always involves a potential loss in the risk that the sample will represent a bad lot as being good.

## RISK IN SAMPLING INSPECTION

What is the risk encountered in sampling inspection, and how is it to be evaluated?

As an example let us consider the following situation: A consumer is about to purchase a lot of N units of a certain product. He has stipulated that a certain requirement must be met. Any unit failing to meet this requirement is defective. Now, a certain number of defective units in the lot will be tolerated, but, if this number is exceeded, the lot will be rejected. Let us call this number the tolerance number $p_tN$, where $p_t$ is the tolerance fraction for defective units. Obviously, if the product is completely inspected, the consumer assumes no risk, for he knows exactly the number of defective units in the lot. If he examines only a sample of n units, however, he knows only the number of defective units in the sample, and he must guess at the number in the remainder of the lot. The uncertainty attendant to this guess is his risk.

Let us define the risk in a more precise fashion. Suppose the consumer selects a sample of n units under the condition that, if c or less defective units are found in the sample, the lot is to be accepted, but, if more than c defective units are found, the lot is to be rejected. This is, of course, the condition usually occurring in practice. The consumer's risk, then, is the chance or probability that c or less defective units will be found in a sample of n units selected from a lot of N units containing pN defective units, where p is greater than $p_t$; i.e., the chance that the sample will indicate a bad lot to be good. The calculation of the risk so defined involves à posteriori probability theory. We can, however, find the maximum value of this risk by simple à priori probability theory as follows: Assume that we know the value of p. In this case the probabilities that 0, 1, 2, ... c ... n defective units will be found in the sample are given by the successive terms of the series

$$1 = \frac{1}{C_n^N}\left[C_n^{qN} + C_{n-1}^{qN}\,C_1^{pN} + \dots + C_{n-c}^{qN}\,C_c^{pN} + \dots + C_n^{pN}\right] \quad --- \quad 7$$

where $C_j^i$ is the number of combinations of i things taken j at a time. The sum of the first c+1 terms of this series is greater when $p=p_t$ than the similar sum for any values of p greater than $p_t$. Hence we shall use this sum as the maximum value of the consumer's risk throughout the present discussion.

## FINDING THE CORRECT SAMPLE SIZE

In many cases N is fixed by tne size of the lot which the consumer is to buy, $p_t$ is set by the standard of quality he desires to maintain and the risk P is taken as the maximum risk which he may safely assume. The problem is to find the size of the sample to be inspected.

This, of course, is influenced also by the choice of c, the maximum number of defective units to be allowed in a satisfactory sample. If two defective units are to be allowed, then the sample will be larger than the sample for which only one defective unit is allowed. We shall not enter here into a discussion of the factors underlying the best choice of c except to point out that for the very simple case where the lot is accepted or rejected on the basis of one sample, c should be the minimum, namely 0, for the most economical inspection. If, as in many cases, additional inspection is made when the sample indicates that the lot is unsatisfactory, the most economical inspection often demands a value of c larger than 0. This subject will be reserved for future discussion. For the present let us assume the value of c to be zero. This means, of course, that the presence of one or more defective units in the sample stamps the lot as unacceptable.

The problem of finding the proper sample size is easily solved by the use of Eq. 7. We need only the first term, since we are concerned only with the probability of finding 0 defective units in the sample. We have then the relation

$$P_o = \frac{C_n^{qN}}{C_n^N} = \frac{\underline{|qN} \quad \underline{|N-n}}{\underline{|qN-n} \quad \underline{|N}}, \text{ - - - - - - - - - - - - - } \quad 8$$

where $\underline{|qN} = (qN)(qN-1)(qN-2) \ldots 3.2.1$, etc. By solving for n, the sample size is determined.

A convenient method is to establish curves for the values of N, $p_t$, and P which occur in practice. For example, a risk of P=.10 is frequently assumed. If we plot curves for this value of risk where c=0, we obtain the chart shown in Fig. 23.

As an illustration of how the curves can be used, let us assume a lot of N=1000 units and a tolerance fraction of $p_t$=.02. How many units

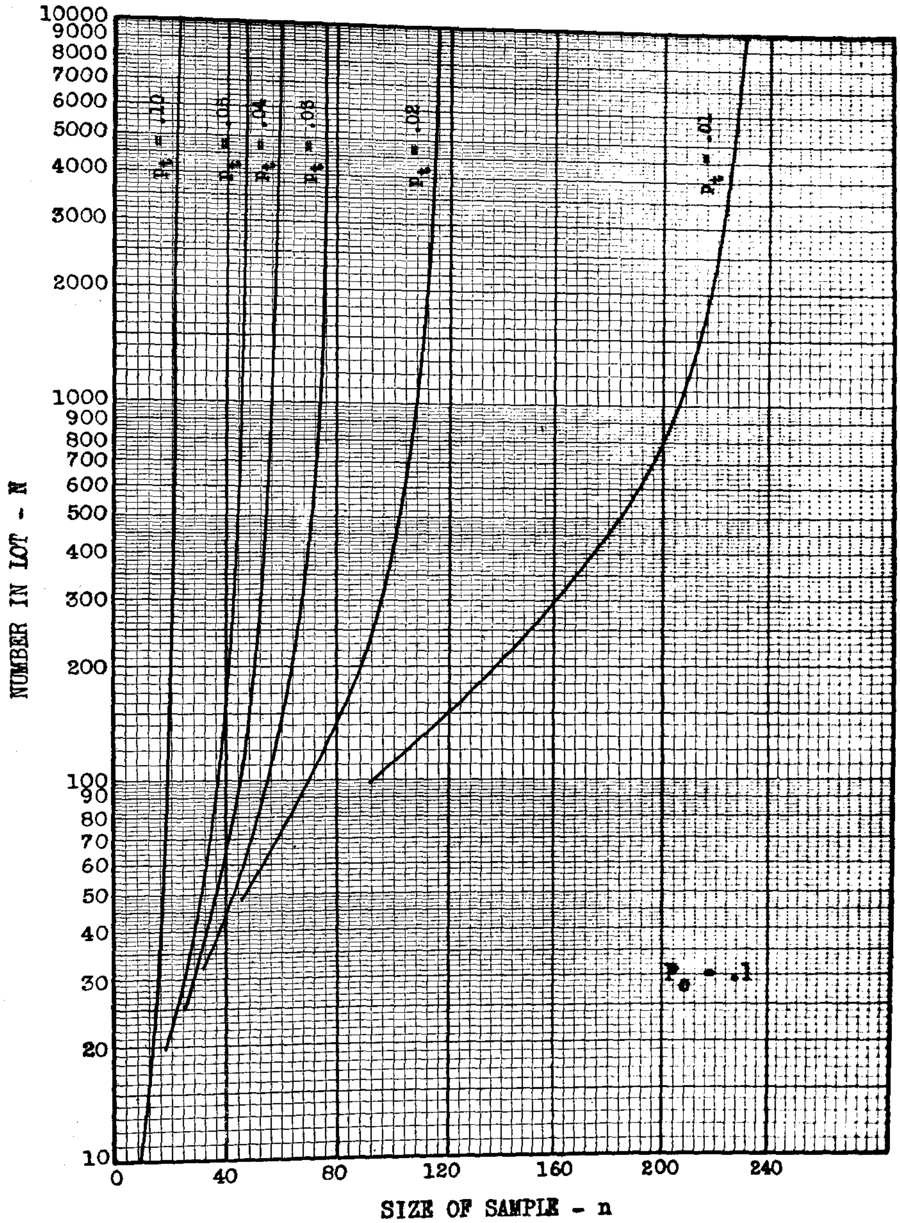



Fig. 23 - CURVES SHOWING CORRECT SAMPLE SIZES.

N = Number of units in the lot.

n = Number of units in the sample.

$p_t$ = Tolerance fraction for defective units.

$P_o$ = Maximum risk of failing to find defective units in sample when the lot contains $p_tN$ or more defective units.

c = 0

shall be inspected so that the maximum risk of accepting a lot which contains a fraction of defective units greater than the tolerance is $P_c$=.10? The chart gives the answer immediately, n=107. That is, 107 units are to be selected from the lot of 1000 units. If no defective units are found in the sample, the lot is to be accepted. If one or more defective units are found, the lot is to be rejected.

## SAVING AND RISK WHEN REJECTED LOTS ARE COMPLETELY INSPECTED

Let us consider a slight extension of the preceding simple sampling inspection scheme. Let us assume that the consumer is buying many lots of 1000 units each and is selecting a sample of 107 units from each lot. In addition, however, to avoid rejecting good lots which the samples indicate are bad, the rejected lots are completely inspected; i.e., the remaining 893 units in each lot are examined. What will be the saving and the risk? Fig. 24 is a schematic representation of the saving in the cost of inspection and the consumer's risk (both in percentage) when the fraction of defective units in the lots ranges from zero to the tolerance, $p_t$, and beyond. That is, if we assume that the manufacturer produces uniform lots in which

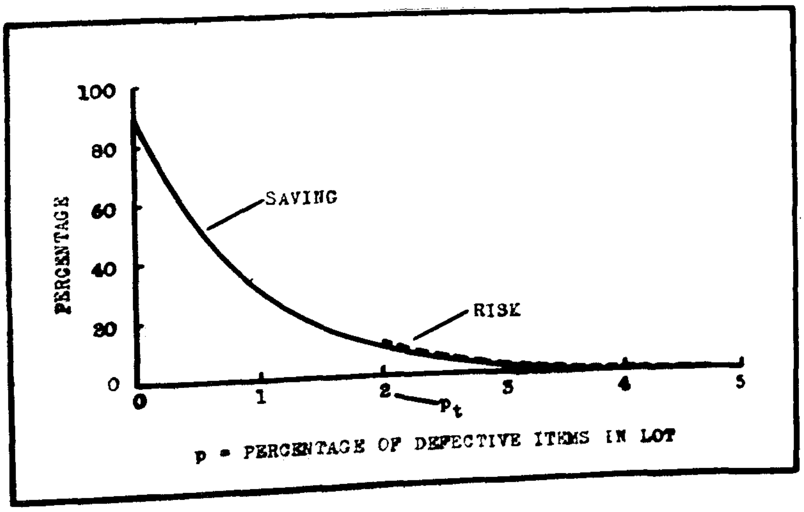



Fig. 24 - CURVES SHOWING HOW THE CONSUMER'S RISK AND THE SAVING IN THE COST OF INSPECTION DEPEND UPON p, THE FRACTION OF DEFECTIVE UNITS IN LOTS, WHEN REJECTED LOTS ARE COMPLETELY INSPECTED. N=1000, n=107, $p_t$=.02, and $P_c$=.10.

the actual fraction of defective units is p, then, as shown in the figure:

a. If there are no defective units (p=0), the total amount of inspection is only $\frac{107}{1000}$ of that involved in complete inspection, or a saving of 89.3%. The consumer's risk is obviously zero, for all the lots are good and there is no possibility of getting a bad lot.

b. If the fraction of defective units exactly equals the tolerance ($p=p_t=.02$), then 10% of the lots will be accepted on the basis of the sample and 90% must be completely inspected. The saving is therefore only 8.93%. The risk is again zero, for the tolerance has not been exceeded. When the tolerance is just exceeded, however, the risk as we have defined it becomes the maximum or 10%.

c. As the fraction of defective units increases beyond the tolerance fraction, both the saving and the risk become less. In the limit, when all of the units are defective (p=1.00) all of the lots will be rejected and must be completely inspected. Hence the saving and risk are both zero. Actually, for all practical purposes, this condition is reached at p=.05.

Incidentally, the "Saving" curve in Fig. 24 has another point of interest. In case the manufacturer has to pay the cost of inspection, this curve shows the incentive for the manufacturer to produce uniformly good product. In the long run, $\overline{p}$ is the average percentage of defective units in the product. If this is kept well below the tolerance percentage, the manufacturer saves considerably on the cost of inspection, but, if on the average, the product contains nearly the tolerance percentage of defective units, the saving is slight.

PRODUCER'S RISK

In this article we have considered the consumer's risk in sampling inspection; i.e., the risk that bad lots will be accepted as good. Obviously there is an analogous risk or chance that good lots will be rejected as bad. This is a risk of the manufacturer or producer. It cannot be entirely neglected by the consumer, however, for if a large amount of good product is rejected, the cost of the accepted product must necessarily be increased. To avoid this, elaborate inspection schemes are often provided for reinspecting rejected product. In such cases the relations existing between the various factors are not so simple as in the case considered here, but the application of the principles of probability usually enables one to choose the best scheme and to determine the best conditions for its successful operation.

## SUMMARY

The amount of inspection which should be made in a given case depends primarily upon the risks to be assumed, the standard of quality to be maintained and the amount of product under investigation. When these factors are fixed, it is possible to compute the proper sample size.

## APPENDIX 2

## DETECTION OF NON-HOMOGENEITY OF PRODUCT.

Sometimes the conclusion is drawn that product is homogeneous if it is distributed in accord with one of the well-known types of probability curves in Eqs. 3 and 4. In general this is justified, but important instances may arise when this conclusion is not justified as we shall now see.

Given a set of $n$ observed values of some quantity $X$ we proceed, as indicated in Parts II and III, to calculate certain statistics. Knowing the values of $\beta_1$ and $\beta_2$, we use standard methods for choosing the type of probability curve with which to fit the data. As an example, for a normal population $\beta_1'$ and $\beta_2'$ are 0 and 3 respectively. Hence, if a pair of statistics $\beta_1$ and $\beta_2$, are found to be approximately 0 and 3, we usually try to fit the observed distribution with a normal curve. We then apply the criterion for measuring the goodness of fit to give us some idea whether or not our assumption of a normal universe is justified. If the fit is good, we usually assume homogeneity.

Now let us suppose that a product is produced by two machines, that the product from one machine is distributed normally about an average quality $\bar{X}_1$ with standard deviation $\sigma_1$ and that the product from the other machine is distributed in the same manner about another average $\bar{X}_2$.

If $a\bar{X}_1 = \bar{X}_2$, $\Sigma y_1 = aN_1 = a\Sigma y_2 = aN_2$ and $\mu_{1i} = \mu_{2i}$;

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{\left[\dfrac{a(\mu_{31}+3\bar{X}_1\mu_{21}+\bar{X}_1^3)}{1+a} + \dfrac{(\mu_{32}+3\bar{X}_2\mu_{22}+\bar{X}_2^3)}{1+a}\right]^2}{\left[\dfrac{a(\mu_{21}+\bar{X}_1^2)}{1+a} + \dfrac{\mu_{22}+\bar{X}_2^2}{1+a}\right]^3}, \text{ - - - - - - } 9$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{\dfrac{a(\mu_{41}+4\bar{X}_1\mu_{31}-6\bar{X}_1^2\mu_{21}+\bar{X}_1^4)}{1+a} + \dfrac{(\mu_{42}+4\bar{X}_2\mu_{32}-6\bar{X}_2^2\mu_{22}+\bar{X}_2^4)}{1+a}}{\left[\dfrac{a(\mu_{21}+\bar{X}_1^2)}{1+a} + \dfrac{\mu_{22}+\bar{X}_2^2}{1+a}\right]^2}, \text{ --- } 10$$

where all symbols refer to the population.

Fig. 8 shows two component distributions (equal in area or $N_1 = N_2$) with averages separated by $1\frac{1}{2}\,\sigma$. Using Eqs. 9 and 10 we can find $\beta_1$ and $\beta_2$ for the compound distribution. These calculated values are approximately those of the normal curve. Hence fitting a normal curve to the compound distribution (dots) we get the smooth curve shown in Fig. 8.

Now Fig. 25 shows two component distributions with averages separated by $2\frac{1}{2}\,\sigma$ and with areas in the ratio 10 to 1. The solid curve is the Gram-Charlier curve having the same values of standard deviation, $\beta_1$ and $\beta_2$ as the compound distribution.

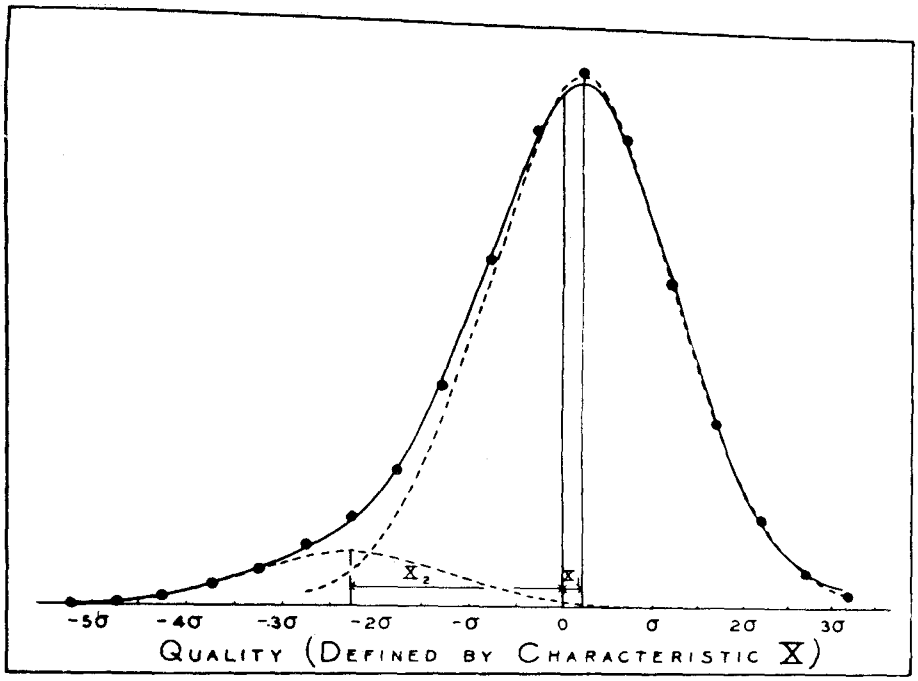



Fig. 25 - SCHEMATIC DIAGRAM SHOWING HOW NON-HOMOGENEITY MAY ESCAPE DETECTION. COMPONENT DISTRIBUTIONS OF UNEQUAL SIZE.

Obviously the fit in the two cases is very good indeed, and it is doubtful if we should detect non-homogeneity by analyzing the data representing the compound distribution. On the other hand it would be a very simple matter to detect non-homogeneity if the data for one of the components were kept separate from those of the other.

Of course the practical case of detecting non-homogeneity is much more complicated than the one considered. The illustrations should indicate, however, the necessity of carefully planning the inspection procedure to detect non-homogeneity of the character noted.

## APPENDIX 3

## DISTRIBUTION OF STATISTICS

Given a set of $n$ observed values of quality $X_1, X_2, \ldots X_n$, we try to find the equation of the curve

$$dy_{\lambda'} = f'(X, \lambda_1', \lambda_2', \ldots \lambda_{c_1}')\, dX, \quad \text{- - - - - - - - 2}$$

where $dy_{\lambda'}$ represents the probability of producing a unit with quality X within the range X to X + dX and the $\lambda'$'s represent the parameters. As indicated in the paper, our first step consists in finding the distribution f and the second step consists in estimating the c parameters from the n observed results.

Now, in general, there are many ways of estimating each of the parameters, and, naturally, we want to know which method to use. Two criteria must be used in reaching a decision as to the method of estimate to use in a particular case. These are:

1. The cost of the analysis involved in making the estimate.
2. The precision of the estimate.

It is an easy matter to compare two or more methods of estimating a parameter in respect to cost, but it is not such an easy matter to compare them in respect to precision.

To determine the precision of an estimate, we must find the distribution of that estimate or statistic. In general, the distribution of a statistic depends upon two factors, the distribution f' of the population, and the size n of the sample used in calculating the statistic. Let us illustrate the effect of the size of the sample upon the distribution.[49]

Let us assume that the population is normal, Eq. 1, and hence that it involves only two parameters m and $\sigma'$. We shall consider only two of the

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

49. I hope to present later a general discussion of this subject. The present note is intended only to call attention to the problem in its relation to inspection. Of course the experimental results are interesting in themselves to the extent indicated in the text.

many possible estimates for each of these parameters. They are the median and arithmetic mean for m and $\sigma_1 = \sqrt{\frac{\pi}{2}} \frac{\Sigma|X-\bar{X}|}{n}$ and $\sigma_2 = \sqrt{\frac{\Sigma(X-\bar{X})^2}{n}}$ for σ', where the summation Σ extends over all of the n values in the sample and $\bar{X}$ is the average of these n values.

When n is large, the four distributions of these statistics are all approximately normal with standard deviations $\frac{1.2533\sigma'}{\sqrt{n}}$, $\frac{\sigma'}{\sqrt{n}}$, $\frac{\sigma' \sqrt{\pi-2}}{\sqrt{2n}}$ and $\frac{\sigma'}{\sqrt{2n}}$ respectively.[50]

Hence the arithmetic mean $\bar{X}$ of a large sample is the more efficient estimate of the mean m of the population, because the same precision can be obtained from the mean of 100 observed values of X as can be obtained from the median of 157 observed values of X. The median of large samples is therefore only $\frac{1.00}{1.57}$ = 63.7% efficient.

Obviously, however, the efficiency of the median is 100% when n=2. What, then, is the efficiency of the median for small samples, but larger than 2. No general answer to this question is apparently available so the 4000 drawings given in Fig. 13 have been used to determine the efficiency for n=4. This was done in the following way.

The 4000 drawings were available in the sequence in which they were taken. These observations were divided into 1000 groups of four each, taking the first 4 observations as the first sample, the second 4 observations as the second sample and so on. The arithmetic means and medians were calculated for each sample. From normal law theory we know that the distribution of the means should be given by the equation

$$dy = \frac{2}{\sigma' \sqrt{2\pi}} e^{\frac{-4(X-m)}{2\sigma'^2}} dX = \frac{2}{\sqrt{2\pi}} e^{-2(X-0)^2} dX. \quad - - 11$$

---

50. Whittaker and Robinson, "The Calculus of Observations", Chapter VIII.

A comparison of the theoretical and observed distribution of the means gives a probability of fit of .92 (Fig. 26-a). No theoretical distribution is known for the medians, but the observed distribution (Fig. 26-b) gives a probability of fit of .60 with the normal law

$$dy_1 = \frac{1}{.55\sqrt{2\pi}} e^{-\frac{(x_1-0)^2}{2(.55)^2}} dx_1, \text{ - - - - - - - - - - } 12\text{-a}$$

where $x_1$ is the median value and .55 is the observed standard deviation of the medians.

Now the probable error of the observed standard deviation .55 of the medians is .012, and hence the empirically determined efficiency of the median is $(\frac{50}{55})^2 \times 100 = 83\%$ with the 3 doubtful. Furthermore, even for samples of size 4 or greater, the arithmetic mean is more efficient than the median, providing, of course, the population is normal.[51]

Let us next consider the distributions of $\sigma_1$ and $\sigma_2$ for small sized samples. Since for large samples, the standard deviations of the estimates $\sigma_1$ and $\sigma_2$ are $\frac{\sqrt{\pi-2}\,\sigma'}{\sqrt{2n}}$ and $\frac{\sigma'}{\sqrt{2n}}$ respectively, the efficiency of $\sigma_2$ as compared with $\sigma_1$ is $(\sqrt{\pi-2})^2 \times 100 = 114\%$. Hence the advantage of $\sigma_2$ over $\sigma_1$ in the way of efficiency, usually far outweighs the disadvantage of $\sigma_2$ compared with $\sigma_1$ accruing from the slightly more involved computation.

In practical inspection work, however, we often have small samples and hence we want to know the efficiency of $\sigma_2$ under these conditions.

The theoretical distribution of $\sigma_2$ was first given by "Student"[52] and later checked by Pearson[53] and R. A. Fisher.[54] For our case this theo-

---

51. Of course, the median is more efficient than the arithmetic mean if the population is distributed according to the law

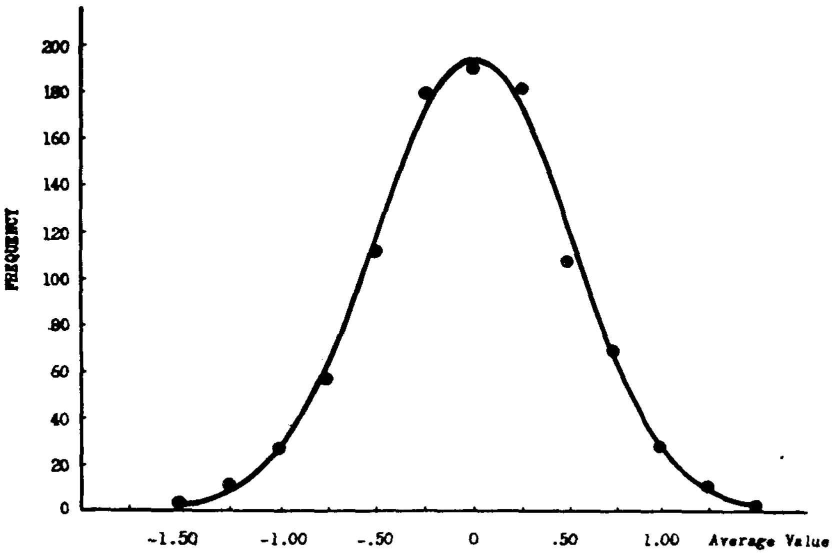
$$dy = ce^{-h|x|} dx$$

where c and h are constants.

52. "Student", "Probable Error of the Mean", Biometrika, Vol. 6, pp. 1-25, 1908.

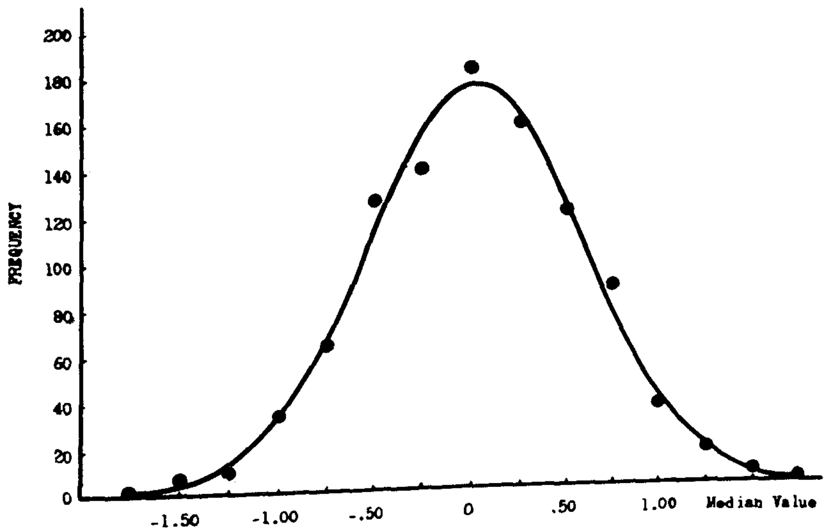
53. Pearson, K.P., "On the Distribution of Standard Deviations of Small Samples", Biometrika, Vol. 10, pp. 522-529, 1915.

54. Fisher, R.A., "The Frequency Distribution of the Values of the Correlation Coefficient in Samples from an Indefinitely Large Population", [illegible] Vol. 10, p. 507, 1915.




FREQUENCY DISTRIBUTION OF OBSERVED VALUES FOR AVERAGES OF 1000 SAMPLES OF SIZE 4

● Observed Frequencies
—— Theoretical Distribution




FREQUENCY DISTRIBUTION OF OBSERVED VALUES FOR MEDIANS OF 1000 SAMPLES OF SIZE 4

● Observed Frequencies
—— Theoretical Distribution

Fig. 26

retical distribution becomes

$$dy_2 = 6383.064\sigma_2^2 e^{-2\sigma_2^2} d\sigma_2 . \text{ - - - - - - - - - 12-b}$$

By empirical methods I find the distribution of $\sigma_1$ to be a Pearson Type IV or

$$dy_3 = 203.44\left(1 + \frac{\sigma_1}{3.4766}\right)^{2.4385} \left(1 - \frac{\sigma_1}{10.776}\right)^{11.767} d\sigma_1 . \text{ - - - -12-c}$$

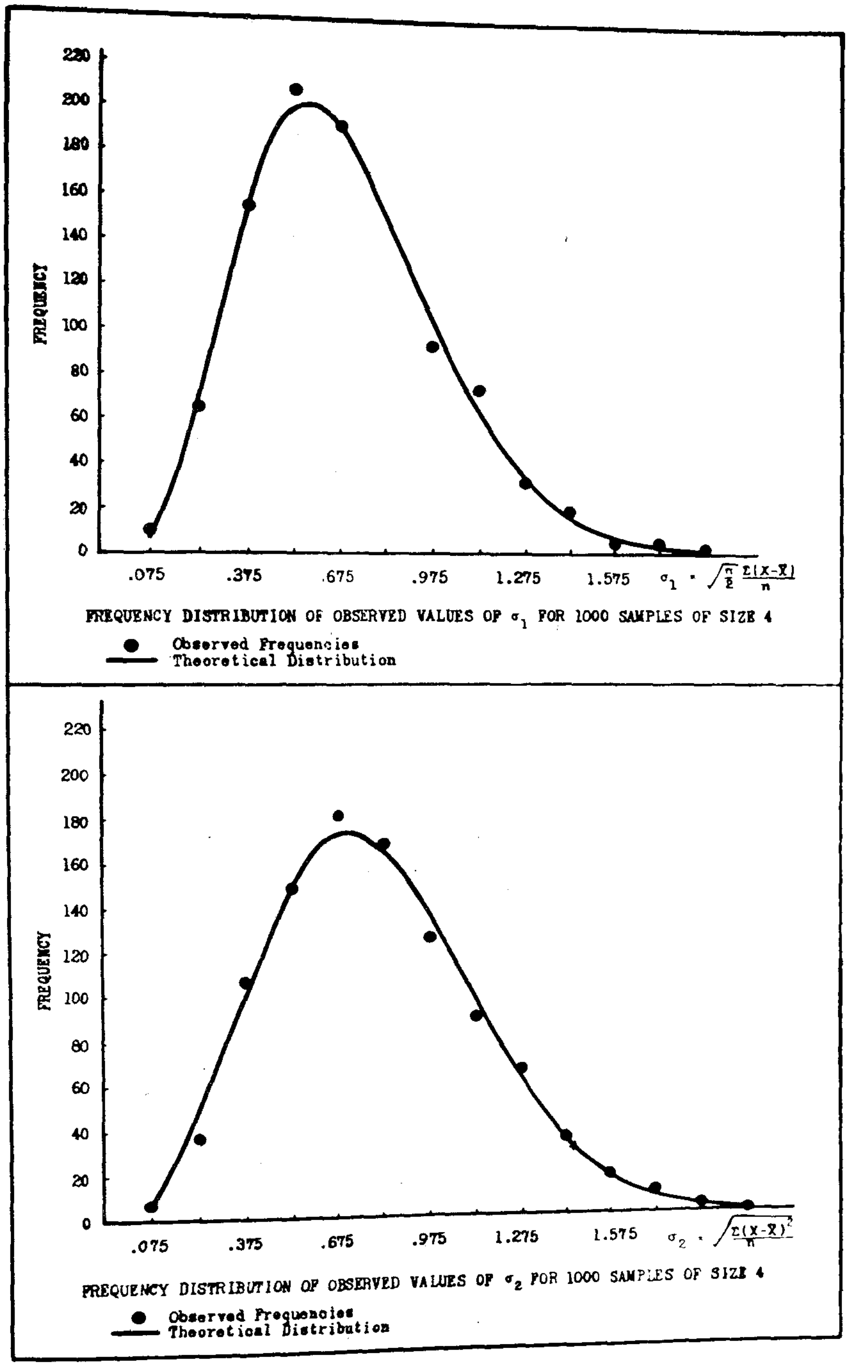
The results are shown graphically in Fig. 27. The probabilities of fit in the two cases are .84 for $\sigma_2$ and .69 for $\sigma_1$. The observed standard deviations of $\sigma_1$ and $\sigma_2$ were .375 and .337 with probable errors of .0084 and .0075 respectively. Hence the observed efficiency of $\sigma_2$ as compared with $\sigma_1$ is 100 $\left(\frac{.375}{.337}\right)^2$ = 124%, and, even for samples of size n = 4, the $\sigma_2$ method is more efficient that the $\sigma_1$ method.

This, however, is not the only point of interest as we shall now see. Suppose we wished to establish a control chart for the two estimates $\bar{X}$ and $\sigma_2$ of m and σ' respectively, for the above case where n = 4. As we have already seen (Part III) the bases of the limits on these two charts are $\frac{\sigma'}{\sqrt{n}}$ and $\frac{\sigma'}{\sqrt{2n}}$ respectively. Would we be justified in substituting for σ' the average of the 1000 observed values of $\sigma_2$? The answer is no, because, as we see from Fig. 27 and from the equation of the distribution of $\sigma_2$, the most probable observed value of $\sigma_2$ is less than σ'. In fact

$$\sigma' = \frac{\Gamma\left(\frac{n-3}{2} + 1\right)}{\Gamma\left(\frac{n-2}{2} + 1\right)} \sqrt{\frac{n}{2}}\ \bar{\sigma}_2$$

where $\bar{\sigma}_2$ is the mean of the 1000 observed values of $\sigma_2$. This is a comparatively large correction for small values of n(see footnote 19).

These experimental results are useful in showing the nature of the problem involved, and in showing how to prepare the best form of control chart.




FREQUENCY DISTRIBUTION OF OBSERVED VALUES OF $\sigma_1$ FOR 1000 SAMPLES OF SIZE 4

FREQUENCY DISTRIBUTION OF OBSERVED VALUES OF $\sigma_2$ FOR 1000 SAMPLES OF SIZE 4

Fig. 27

## APPENDIX 4

## APPLICATION OF GOODNESS OF FIT TEST TO INSPECTION PROBLEMS

### How Need for the Test Arises

Suppose we measure the quality X of each unit in a sample of n pieces of apparatus and then ask ourselves the question: Is the product uniform, and, if so, what is the probability $dy_{\lambda'}$ that a unit of product will be produced with a quality within the range X to X + dX where as before we write

$$dy_{\lambda'} = f'(X, \lambda_1', \lambda_2', \ldots \lambda_{c_1}')\, dX, \quad \text{-------- } 2$$

f' and $\lambda_i'$ being unknown? For example this question arose in our study of the measurements presented in Fig. 3. The data given in this figure represent the distribution of quality observed in a sample of 15,050 instruments of a given type. In answering the question, the first step was to assume a trial form f for the function f' of Eq. 2. Let us write this assumed true probability $dy_\lambda$ that a unit of product will be produced within the interval X to X + dX in the symbolic form

$$dy_\lambda = f(X, \lambda_1, \lambda_2, \ldots \lambda_c)\, dX. \quad \text{-------- } 13$$

The next step was to obtain the best estimate or statistic for each of the c parameters in Eq. 13. This gave us a theoretical curve

$$dy_\theta = f(X, \theta_1, \theta_2, \ldots \theta_c)\, dX \quad \text{-------- } 14$$

where $dy_\theta$ was our best theoretical estimate of the probability that a unit of product will be produced within the interval X to X + dX based upon the assumption that f is the true form f' of the probability distribution for the uniform product ($\theta_1$ being the best estimate or statistic for $\lambda_1$).[55]

---

55. Obviously the number c of parameters in Eq. 13 need not be the same as the number $c_1$ in Eq. 2, and, even if $c=c_1$, $\lambda_i'$ need not be the same as $\lambda_i$ unless f is the same as f'.

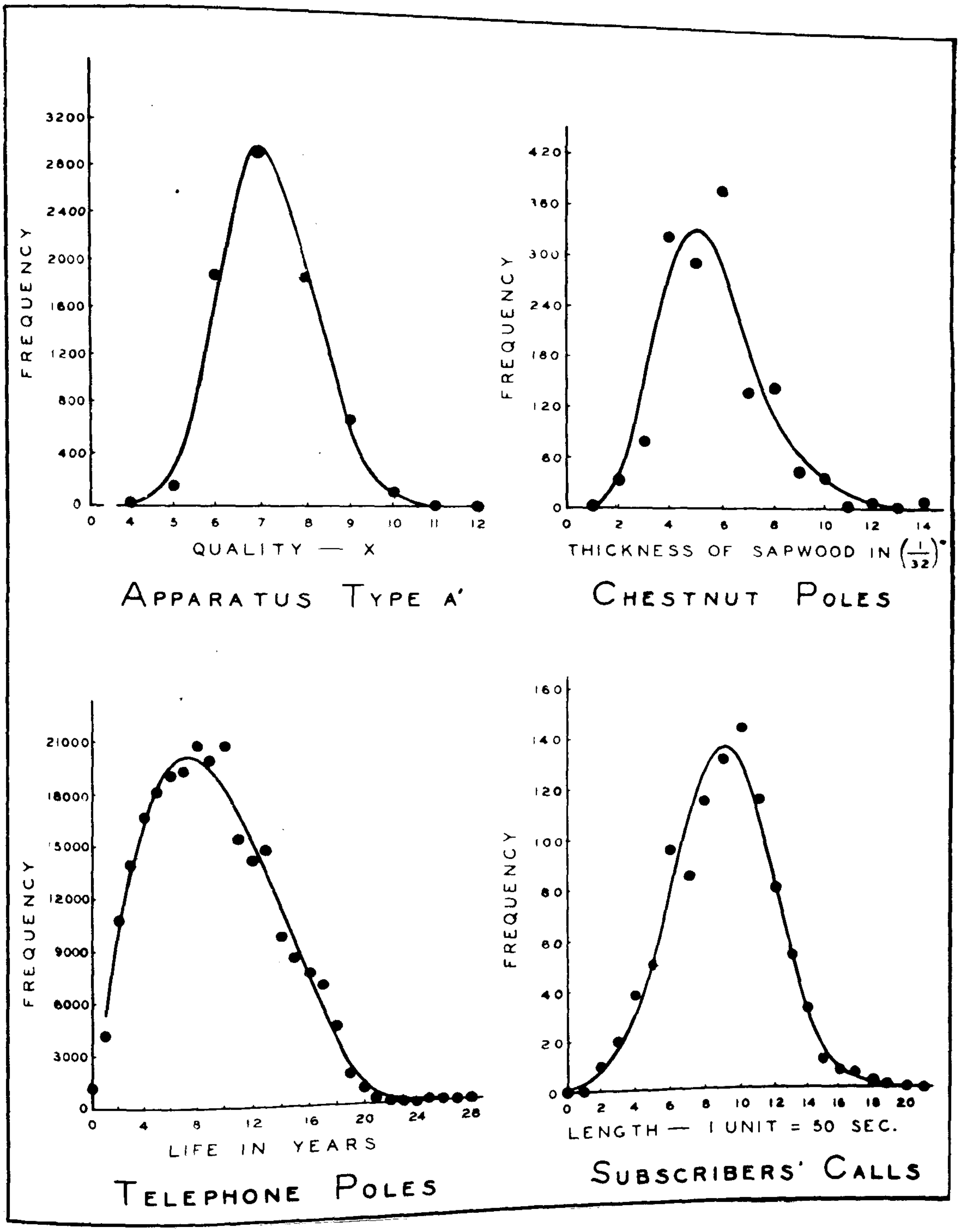



Fig. 28 - IS THE FIT BETWEEN THE SMOOTH THEORETICAL CURVES AND THE OBSERVED POINTS GOOD: DOES IT INDICATE UNIFORMITY IN EACH CASE?

Fig. 28 presents four other typical observed distributions (dots) chosen from the field of telephone practice. In this case the assumed forms f for the distributions of quality of type A' apparatus, thickness of sapwood in chestnut poles, life of telephone poles[56] and length of subscriber calls[57] are three terms of the Gram Charlier series (Eq. 3) for the first two and Pearson Types I and IV (Eq. 4) respectively for the last two. The smooth theoretical curves in Fig. 28 are of the type shown in Eq. 14 where estimates of the assumed true parameters are substituted in the forms of f.

Is the fit between the observed and theoretical distributions good? Does a good fit necessarily mean a uniform product? These are the questions to be settled. Let us consider the steps necessary to answer these questions assuming the product to be uniform from which a sample of n units is drawn.

Let us assume the following distributions into m+1 cell frequencies:

$$ny_o, \quad ny_1, \quad ny_2, \quad \ldots ny_m \quad \text{---- observed}$$

$$ny_{o\lambda'}, \quad ny_{1\lambda'}, \quad ny_{2\lambda'}, \quad \ldots ny_{m\lambda'} \quad \text{---- true (Eq. 2),}$$

$$ny_{o\lambda}, \quad ny_{1\lambda}, \quad ny_{2\lambda}, \quad \ldots ny_{m\lambda} \quad \text{---- assumed true (Eq.13),}$$

$$ny_{o\theta}, \quad ny_{1\theta}, \quad ny_{2\theta}, \quad \ldots ny_{m\theta} \quad \text{---- theoretical (Eq.14).}$$

The first of the above questions calls for some method of comparing the observed and theoretical cell frequencies. Largely thru the labors of Pearson[58] such a method has been developed. Applied with care this test for goodness

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

56. Original data given in the article, Replacement Insurance, Administration, July 1921, p.55.

57. Original data given in A. T. & T. Company's Statistical Bulletin, Introduction to Frequency Curves and Averages, Table c, p.2.

58. Pearson, Karl - On the Criterion that a Given System of Deviations from the Probable in the Case of a Correlated System of Variations is such that it can be Reasonably Supposed to have arisen from random sampling - Phil. Mag. S. 5, Vol. 1, 1900, p.157-175.

Pearson, Karl. - On a Brief Proof of the Fundamental Formula for Testing the Goodness of Fit of Frequency Distributions and on the Probable Error of "P" - Phil. Mag. and Journal of Science. 1916, p.369.

of fit is a very powerful tool for the inspection engineer as will be shown below, but too much emphasis cannot be laid upon the care which must be taken in applying this test. A few paragraphs will now be devoted to a discussion which should prove of value as a popular supplement to the original highly technical memoirs cited. Such a digression is necessary to give an understanding of the applications to follow.

## Description of the Test.

Pearson's first contribution[58] was a method of comparing the true distribution with the observed distribution when the former is assumed to be known à priorily. Let us form the differences, $ny_{i\lambda}, - ny_i = x_i$. Then, obviously,

$$x_0 + x_1 + x_2 + \dots + x_m = 0, \qquad \text{--------- 15}$$

so that there are m degrees of freedom for the set of correlated differences between the observed and true cell frequencies. If no probability of the type $y_{i\lambda}$, is small, we may assume that the deviations in $x_i$ will be normally distributed. Hence the probability surface for the simple case where there are three cells is that shown in Fig. 29, where $z \, dx_1 \, dx_2$ is the probability of occurrence of a value of $x_1$ within the range $x_1$ to $x_1 + dx_1$ simultaneously with the occurrence of a value of $x_2$ within the range $x_2$ to $x_2 + dx_2$. The equation of this surface is of the form

$$z = z_0 \, e^{-\frac{1}{2}\chi^2} \qquad \text{---------- 16}$$

where $\chi^2$ is a function of the values of x For, $\chi^2$ = constant, z is constant, say $z_1$, and the plane, $z = z_1$, Fig. 29 cuts the probability surface in an ellipse. For values of $x_1$ and $x_2$ lying within the projection of this ellipse upon the base of the solid, the value of z will be greater than $z_1$. Similarly for values of $x_1$ and $x_2$ lying outside this same ellipse, the

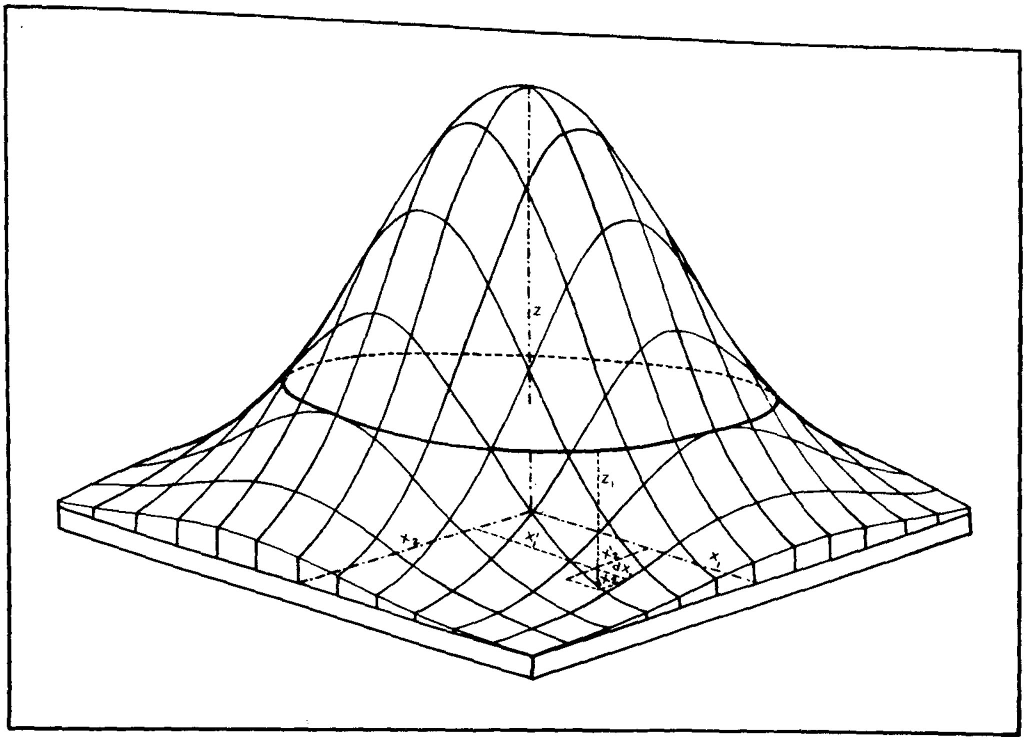


Fig. 29 - PROBABILITY SURFACE FOR INTERPRETING GOODNESS OF FIT FOR THE SIMPLEST CASE.

values of z will be smaller than $z_1$. Suppose, now, that $z_1$ corresponds to an observed set of values $x_1'$ and $x_2'$. The plane $z = z_1$ cuts off, as it were, the cap of the surface. The volume under this cap and extending to the base $z = 0$ represents the probability of getting a set of values of $x_1$ and $x_2$ more probable than the observed set $x_1'$ and $x_2'$ whereas the remainder of the volume under the surface represents the probability of getting sets of values less probable than the observed set $x_1'$ and $x_2'$. This second probability is taken by Pearson[59] as a measure of the Goodness of Fit.

A simple illustration will serve to make clear the method of applying the test. It should be noted that the test calls for an à priorily

---

59. The case where there are m+1 cells is treated in a similar manner except that the geometrical representation calls for m dimensional space. Tables of values of probability of fit P covering a wide range of values of $\chi^2$ and values of m+1 from 4 to 30 are available in Pearson's Tables for Statisticians and Biometricians.

known law of distribution, hence we may apply it to the results of the 4000 drawings from the known law of distribution given in Table III of this paper. Table IV below presents the results. The third column gives the values of $\chi^2$ and P for a comparison of the observed frequencies in the 13 cells with the corresponding known cell frequencies. The observed values

TABLE IV

GOODNESS OF FIT: SAMPLES DRAWN FROM DISTRIBUTION GIVEN IN TABLE III

| | | 13 Cells à priori Theory 12 Degrees of Freedom | 13 Cells à posteriori Theory Assuming 12 Degrees of Freedom | 13 Cells à posteriori Theory Assuming 10 Degrees of Freedom |
|---|---|---|---|---|
| 1st Sample | $\chi^2$ | 9.983 | 6.612 | 6.612 |
| | P | .617 | .880 | .760 |
| 2nd Sample | $\chi^2$ | 11.579 | 8.694 | 8.694 |
| | P | .481 | .728 | .562 |
| 3rd Sample | $\chi^2$ | 4.184 | 3.360 | 3.360 |
| | P | .979 | .991 | .969 |
| 4th Sample | $\chi^2$ | 8.776 | 6.146 | 6.146 |
| | P | .721 | .908 | .802 |
| Four Samples Combined | $\chi^2$ | 9.940 | 6.737 | 6.737 |
| | P | .622 | .873 | .749 |
| Average | $\chi^2$ | 8.892 | 6.310 | 6.310 |
| | P | .684 | .876 | .768 |

of $\chi^2$ lie within the interval 4.184 to 11.579 and the corresponding values of P lie within the interval .979 to .481, the average values of $\chi^2$ and P being 8.892 and .684. Now, what values of $\chi^2$ and P should we expect to find from the theory? Are the observed values of these two factors consistent with those which we should expect to find?

To answer these questions we must consider the probability distribution of $\chi^2$ for the given number of cell frequencies m+1, in this case 13.

Common sense tells us that, if we were to take a very large number of samples, we should expect to find for some sample or samples every value of $\chi^2$ between 0 and $\infty$, that we should not find either 0 or $\infty$ very often and that we should find some value of $\chi^2$ more often than any other value. Theory checks common sense. The distribution of $\chi^2$ is skew, the degree of skewness decreasing as the number of cells m+1 increases. Hence the most probable value of $\chi^2$ is not equal to the average value of $\chi^2$ as determined from the distribution of $\chi^2$. In general the most probable value is less than this average value, but as m+1 increases the distribution curve for $\chi^2$ approaches normality so that the most probable value of $\chi^2$ approaches the average value and incidently approaches that value of $\chi^2$ for which P = .5. Hence for large values of m+1 the value of P to be expected most often is .5, whereas for small values of m+1, values of P to be expected most often are greater than .5. The distribution curve computed for the case, m+1 = 30, is shown in Fig. 30 by the solid line. The average, standard deviation and skewness of

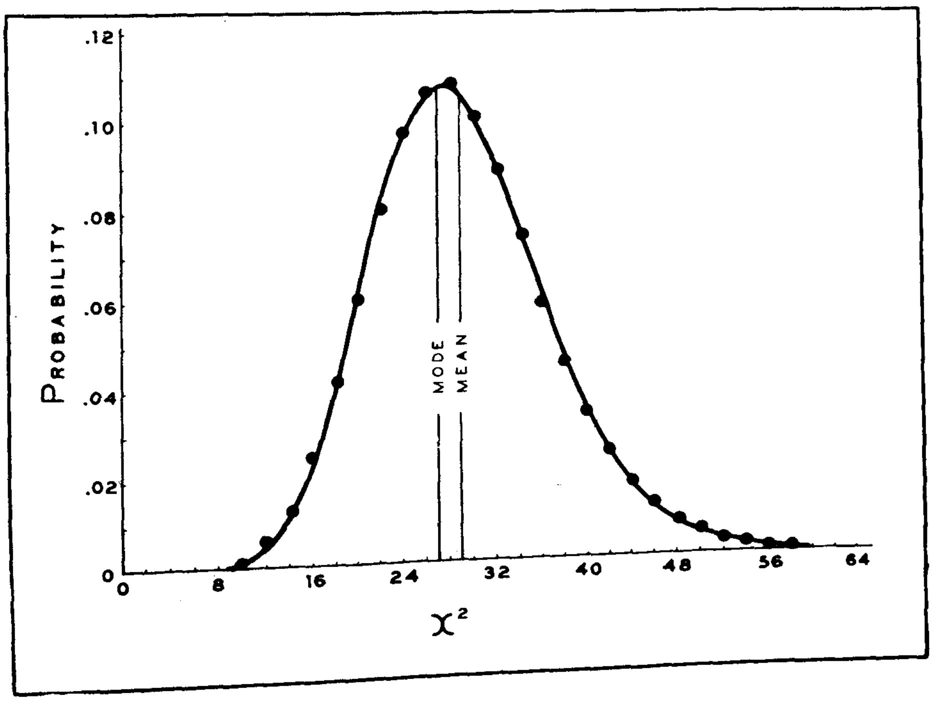



Fig. 30

_____ Distribution of $\chi^2$ for, m+1 = 30.

.....Gram-Charlier Series of three Terms Fitted to the Distribution.

this distribution of $\chi^2$ are respectively 28.98, 7.57 and .48. When these values are substituted in the first three terms of the Gram-Charlier series we get the dots shown in Fig. 30. We see that the most probable value of $\chi^2$ is displaced approximately 1.83 units to the left of the average of the distribution indicating that even for, $m+1 = 30$, the most probable value of P is slightly greater than .5. We see, however, that either very large values or very small values of P are very improbable. Hence we see that the observed values of $\chi^2$ and P shown in column 3 of Table IV are consistent with theory.

Let us now assume that we do not know the true distribution with which to compare the five observed distributions of Table IV. Let us instead find the theoretical distributions assuming the distribution, from which the samples were drawn, to be normal. We then get an equation of the type given by Eq. 14. Column 4 of Table IV gives the observed values of $\chi^2$ upon this basis. Values of P given in this column were obtained by entering the goodness of fit tables with a value of $m+1 = 13$ as in the previous case. We note, however, that the observed values of P are on an average much larger than those given in column 3. This is accounted for by the fact that the theoretical equations each contained two statistics calculated from the data, thereby decreasing the number of degrees of freedom of the set of values of x by two.[60] Hence we should enter the tables for a value of, $m+1 = 11$, and doing this we get the set of values of P shown in column 5. These values more nearly check the set of values given in column 3. This illustrates a very important principle[61] in the application of the $\chi^2$ test which has only recently been established; i.e. we should always enter the goodness of fit tables with a value equal to the number of cell frequencies (m+1) minus the number of statistics used in calculating the theoretical

---

60. Fisher, R.A., On the Interpretation of $\chi^2$ from Contingency Tables and the Calculation of P, Proc. of Roy. Stat. Soc., Vol. LXXXV, p.93, 1922.

61. This is particularly true in engineering work for we seldom, if ever, know à priorily the distribution from which the sample is drawn.

distribution.

Let us compare the physical interpretation of P calculated from differences $(x_i = y_{i\lambda'} - y_i)$ between true and observed cell frequencies with the interpretation of P calculated from differences $(x_i = y_{i\theta} - y_i)$ between theoretical and observed cell frequencies. In the first case P is the probability of occurrence of as likely or less likely complex system of the values of cell deviations x being produced by random sampling from the true distribution (Eq. 2). In the second case P is the probability of occurrence of as likely or less likely complex system of the values of cell deviations x being produced by random sampling from the assumed true distribution (Eq.13), for which our best guess is an equation of the type of Eq. 14.

Before passing to the applications[62] we must consider one other extension of the use of the $\chi^2$ test. Assume that we draw two samples from supposedly the same universe. We can use the $\chi^2$ test to determine the probability that the two samples came from the same universe[63] or we can measure the probability that two samples came from a uniform product. We need not go into details of the calculation of the value of $\chi^2$. Obviously, however, we may calculate the limits within which the observed values of $\chi^2$ may be expected to fluctuate because of sampling variations. If we wish to compare several samples, one way is to add the first two and compare this resultant set of frequencies with those given by the third sample and so on.

---

62. Note should be made of certain other facts which must be taken into account in the application of the test. For example the following reasons may be given for an abnormally large value of $\chi^2$.

    1. The assumed true distribution Eq. 13 is not the true distribution.
    2. The method of estimate of the parameters is inconsistent; i.e., the method of estimate should be such that if applied to an infinite sample a statistic $\theta_i$ should become identical with the corresponding parameter $\lambda_i$.
    3. The method of estimating the parameters is inefficient.

63. Pearson, K. P., On the Probability that Two Independent Distributions of Frequency are Really Samples from the Same Population, Biometrika, Vol. VIII, 1911.

    Rhodes, E. C., Are Two Samples Drawn from the Same Population, Biometrika, Vol. XVI, pp.239-248, 1924.

Such a test has been applied to the four samples of 100 drawings already referred to several times. The results are plotted in Fig. 31 on the part

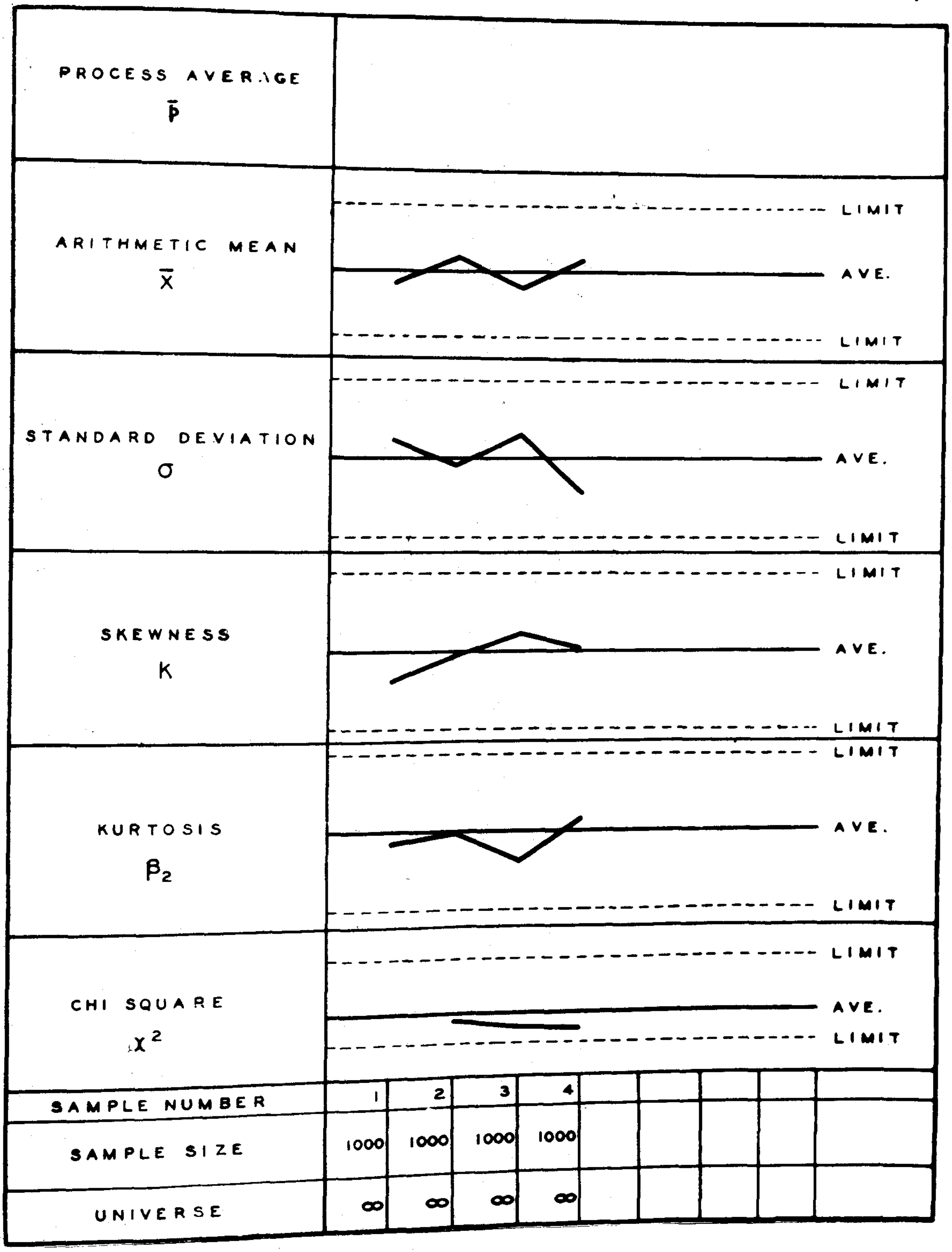



Fig. 31 - CONTROL CHART APPLIED TO CASE WHERE IT WAS KNOWN A PRIORILY THAT THE OBSERVED FLUCTUATIONS SHOULD FALL WITHIN THE SAMPLING LIMITS.

of the control chart marked chi square. The values of $\chi^2$ lie well within the limits (dotted lines) of sampling indicating that the samples probably (or at least possibly) came from the same universe as, of course, they did. The observed fluctuations of the four statistics average $\bar{X}$, standard deviation $\sigma$, skewness k and kurtosis $\beta_2$ given in Fig. 14 are shown also in Fig. 31 to make it complete. This form of chart is called a <u>control chart</u>, because when applied to samples of product it shows us (under limitations set forth in Part III) whether or not the product is controlled. Since the fluctuations in the statistics and in $\chi^2$ resulted from random sampling they should fall within their respective limits, and they do.

## PRACTICAL APPLICATIONS

Applying the $\chi^2$ test to each of the four distributions shown in Fig. 28, we found negligible fit in each case although there appeared to be a close fit between some of the theoretical curves and the corresponding observed distributions (dots). In fact the fit appeared so good in the case of the type A' apparatus that the validity of the $\chi^2$ was questioned by the engineer applying the test. It so happened, however, that this distribution could be broken up into its monthly components. When this was done and the control chart such as that shown in Fig. 31 was constructed, the evidence was very conclusive that the product had not been uniform and hence the $\chi^2$ test gave very valuable results which might have been overlooked otherwise.

Returning, now, to the application of the $\chi^2$ test to the distribution of product shown in Fig. 17, we have already noted the lack of fit as shown on the analysis sheet Fig. 16. Let us go one step further and compare one month's product with the resultant product of all months preceding it as suggested above. Doing this we get the part of the control chart in Fig. 32 marked chi square. In every instance the observed value of $\chi^2$ is outside its sampling limits thus indicating non-uniformity of product. The other forms of charts considered in Part III are reproduced

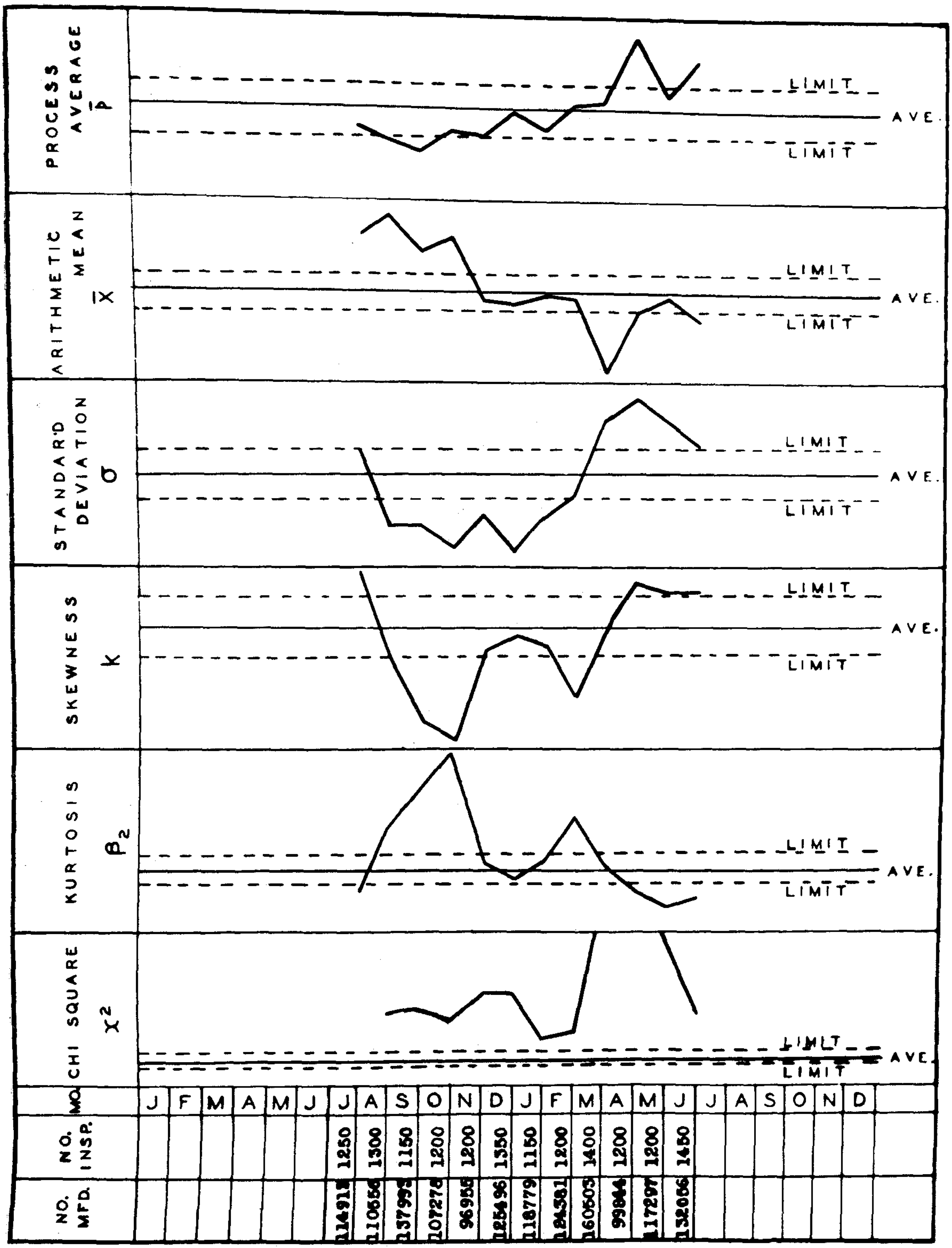



Fig. 32 - CONTROL CHART SHOWING DIFFERENT WAYS OF DETECTING NON-UNIFORMITY OF PRODUCT.

in this figure for reference and comparison with the chi square chart. Obviously there can be little doubt, in the light of the data given in the control chart, that the product from which the 15,050 instruments were drawn was not uniform. Hence there can be little doubt that the observed poor fit shown on the analysis sheet Fig. 16 had a real significance. This illustrates the value of the $\chi^2$ test in inspection engineering. Furthermore, where a poor fit is found, it is usually not difficult to find the assignable causes of the variations.

So far we have considered practical cases only where the fit was bad. Fig. 33 has been included, therefore, to give a typical example of a

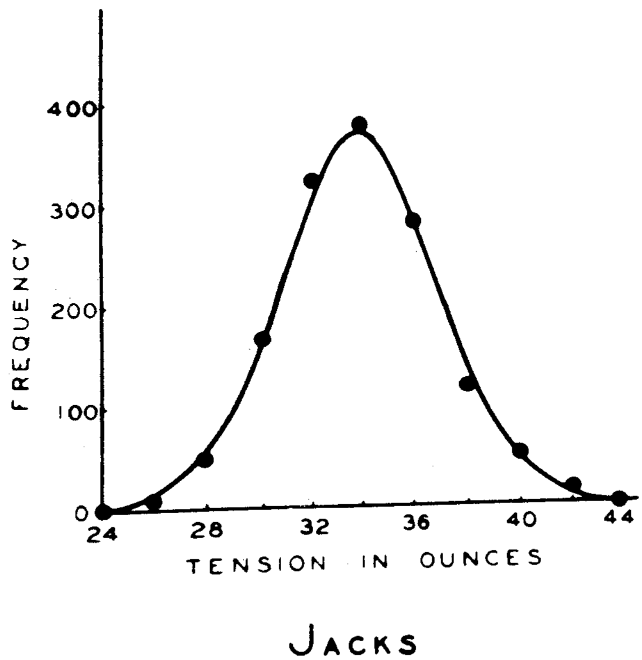



Fig. 33 - EXAMPLE OF UNIFORM PRODUCT AS INDICATED BY THE CHI SQUARE TEST.

distribution of a uniform product (Jacks) for which the probability of fit is good (.495).