OBSERVATIONAL SIGNIFICANCE

OF

ACCURACY AND PRECISION

by

W. A. Shewhart

Paper to be presented before the Philosophical Society of Washington on February 26, 1938

1. Introduction.

2. Operational Criterion

3. General Concepts of Accuracy and Precision.

4. Reproducibility of What is Observed.

5. Customary Measures of Accuracy and Precision.

6. Five Concepts to be Given Operational Meaning.
   $-p'$ X' $\bar{X}'$ p'_a (predictability)

7. Valid Predictability

8. Criterion for Valid Predictability

9. Conclusion.

PRELIMINARY COMMENTS

I appreciate the honor of the invitation to talk to you this evening on the subject, "Observational Significance of Accuracy and Precision". That subject, assigned by your Chairman, has in it what the engineer often calls "ten dollar words". The phrase, "observational significance", for example, ought to give any speaker enough rope to hang himself. Dr. Oliver Wendell Holmes, the father of the late Justice, use to say that if he were assigned a subject for a lecture he charged \$300, while if the subject was one of his own choosing, he charged only \$200; but <u>in either case he said exactly the same thing</u>.

If this topic had not been assigned, I might have been tempted to choose the text of the old colored parson: "It ain't dem things you don't know what gets you into trouble, it's dem things you know for sure what ain't so". In either case, I too would say exactly the same thing. In particular, I shall consider some of the things about accuracy, precision, and observability, that many of us may have thought we knew for sure, but what ain't so.

Before tackling our job in formal fashion, let ~~me~~ us picture a few things which ~~I once~~ some of us may have thought ~~I~~ we knew for sure - primarily perhaps because ~~I~~ we had seen them stated in high

Let us consider a problem with which all of us – either as pure or applied scientists must often deal.

[illegible] ~~interested~~ in $x \pm \Delta x$

[illegible] interested in ~~th~~ seeing how far one may go in saying what they mean.

Some[illegible] valid prediction within ranges. $x \pm \Delta x$

Too often accuracy and precision are like Mark Twain weather. We talk a lot about them but don't do anything about them.

places - but which ~~I~~ we have been forced by experience to put in the class "ain't so".

Let us try to interpret the meaning of a physical measurement when tabulated in the form

nity of pure iron is
71 ± 0.002 gm/cc)

$$X \pm \Delta X$$

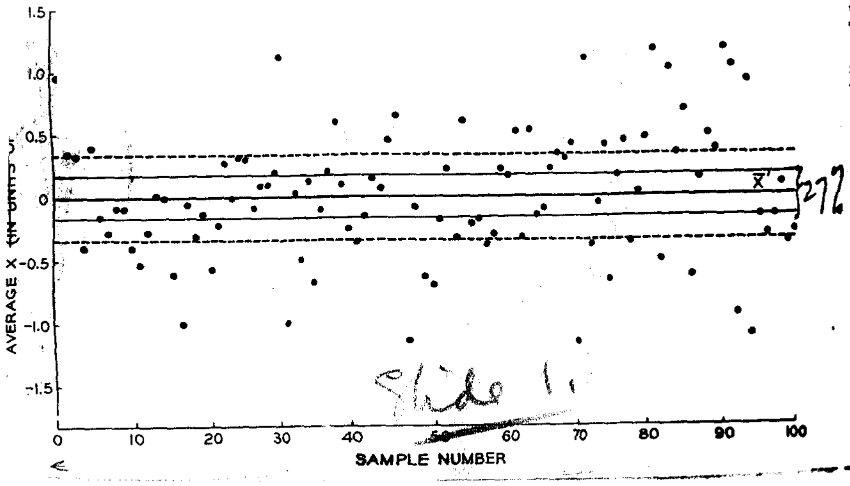
such as found in tables of physical constants. From an engineering viewpoint, one wants to make use of such tabulations; using such information he wants to make valid predictions. To begin, one is confronted with the question:

Is ΔX a measure of precision or of accuracy; or is it a measure of both? If in search for an answer, we turn to the literature on the subject, we find that accuracy and precision are often used as synonyms. But one soon learns from experience that accuracy is not the same as precision.

$100\frac{\Delta X}{X} =$ [illegible]

Obviously we can't ~~the~~ use ~~of~~ a tabulated range $X \pm \Delta X$ intelligently ~~depends upon its interpretation.~~ unless we know what it means. For example, I hold here in my hand a table, "Values of the General Physical Constants", published in 1929 by Dr. Birge. When this table was published ~~At the time I received this table~~, I was inclined to interpret it as giving ranges within which roughly 50% of the true values of the therein tabled constants probably lie. I find that interpretation of mine now in the class of those things "what ain't so".

Of course, in application one is not only interested in tabulating results but also in using them as a basis for guidance in future action. From this angle, it is rather interesting to note that many of the suggested uses in the literature lead to conclusions so much in error as to kill their usefulness ~~invalidate their use.~~ For example, one of the best



ysics published last

ion. This note ex-

error. I hold in

sed it as suggested.

arry out a certain

values will fall

However, on carry-

zed conditions, I

get only 27% within the limits. In commercial mass production, to predict that 50% of a product will be found to lie within a certain range and to find that only 27% lies within that range certainly would get a statistician into hot water.

I have in my hand a paper on error theory by Sir Arthur Eddington. He raises this question: Suppose that I find the following two values of Planck's constant in the literature: Show paper.

$$h \times 10^{27} = 6.551 \pm .013$$

$$h \times 10^{27} = 6.547 \pm .008,$$

which one shall I use? If I understand Eddington's argument correctly, he maintains that "assuming that these are to be taken at their face value", (whatever this reservation may mean), I am justified in choosing the second because the

$\Delta X = .008$ is less than the $\Delta X = .013$. It seems to me that this conclusion belongs in the class "ain't so" unless there is some saving grace implied in the phrase "taken at their face value".

Now with your permission, we will pass to a more formal consideration of some of the principles underlying the observational significance of accuracy and precision. Then in the latter part of the hour I shall show a number of slides to illustrate the principles.

INTRODUCTION

Necessity is the mother of invention. When man became a measuring animal he had to adopt standards of length, mass, and the like. Commerce and industry called for the legalizing of certain standards and the establishment of methods of measuring with requisite accuracy and precision in terms of such standards. Likewise, the introduction of interchangeability about 1787 necessitated accurate measurement and the invention of gauges. The steady increase in the accuracy of interchangeable parts produced under manufacturing conditions has led to the invention of standard length gauges with $\frac{1}{10,000}$ .00001 inch tolerances and pushed the accuracy of test methods out to $\frac{1}{10^6}$ .000001 inch.[1] Both pure and applied science have pushed farther and farther the requirements of accuracy and precision.

- -

[1] Cf. Gauges and Fine Measurements, by F. H. Rolt, The Macmillan Company, London, 1929, Vol. 1, p.10.

Commercial
Two reasons | A Effect Economies.

B Give Quality Assurance

I have said a lot about these elsewhere.

A range $X \pm \Delta X$

$\varphi_1 \leq \varphi \leq \varphi_2$

Element of volume in n Space

Applied science, particularly in the mass production of interchangeable parts, is becoming perhaps even more exacting than pure science in matters of accuracy and precision. It undertakes to make large numbers of things of a given kind with specified degrees of these factors, such as accuracy of 1% or precision of 1%. Failure to meet the requirements may mean rejection and accompanying increase in the cost of production. Such specifications may become the basis of contractual agreement, and any indefiniteness in the meaning of the terms accuracy and precision or in the methods of measuring these may lead to misunderstandings and even legal action. The development of modern methods of mass production through specification is the mother of many changes in our concepts and use of the terms accuracy and precision. ~~One object of the present paper is to set forth some of these changes necessitated by economic production practices.~~

For example, let us consider a specification of the form:

These are of nature of a TARGET

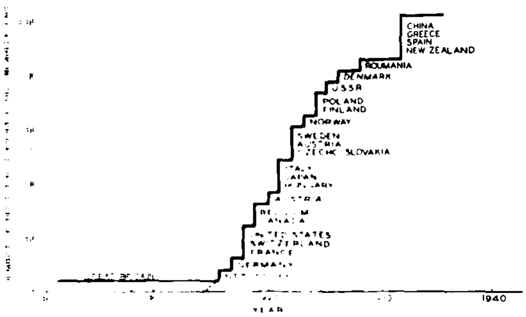
A) The accuracy of the test method shall be ± 1%.
B) The precision of the test method shall be ± 1%.

Under such conditions when is one justified in saying:

a) The accuracy of this test method is ± 1%.
b) The precision of this test method is ± 1%.

These are of the nature of PREDICTION

See Influences
on Operational Theory



#15436

Now suppose that a consumer makes the specification A and B and a producer makes the claim a) and b). How would you as an independent and unbiased observer or scientific judge proceed to verify the producer's statement a) and b)?

Rapidly growing importance of this problem

Slide 2

I suppose a layman has the right to assume that if anyone ever attempts to say just what he means and mean just what he says, that one should be a scientist or engineer when specifying accuracy and precision and when making statements involving these terms. What is to be said therefore, is of interest not only in relation to the special problem of specifying accuracy and precision but also as indicating the limit to which one may hope to go in saying just what he means in a way that is subject to verification - something (also communication and legislation) that is basic to all specification. For example, I think you will agree that the limit to which we can go in specifying the quality of a physical thing is a verifiable manner certainly depends, among other things, upon the limit to which we can go in specifying any simple quality characteristic, such as length, density, or the like, of that thing in terms of quantitative measurements of such quality characteristics.

Call attention to popular interest in this subject now

As a starting point, we must adopt some criterion by which to judge the meaning or significance of the terms accuracy and precision.

Hence the Observational Significance in the title.

(2) OPERATIONAL CRITERION OF MEANING

Since the turn of the century, we have witnessed the rapid development ~~in certain quarters~~, particularly the fields of logic and physics, of what has been termed a new phenomenon, namely, the attempt to apply with rigor the principle that only that which is observable is significant.[1] It is supposed to follow that in order to talk in a meaningful way, one must limit his discourse to concepts and statements that are subject to experimental verification. Let us see what havoc this principle, if accepted without reservation, would wrought with our interpretation of the Probable Values of the Physical Constants (1929) as indicating ranges within which 50% of the true values lie. We certainly cannot observe the true values. Hence, must we discard such an interpretation as meaningless and throw it into the waste basket with other out-moded ideas? [illegible]

We are told that Einstein is responsible for the present virility of the principle that only the observable is significant; at least his discussion of the simultaneity of events occurring at different places started many to think operationally in the field of physics. The outstanding exponent of the theory, in physical science, is Bridgman who, in his book on Logic of Modern Physics, 1928, takes the stand

---

[1]"Science and the Unobservable", by Herbert Dingle, Nature, January 1, 1938, pp. 21-28.

that the concept is synonymous with the corresponding set of operations. Logical positivists at about this time were also saying that the meaning of a statement is its method of verification in terms of some operation that can actually be carried out.

It was soon discovered, however, that the acceptance of such principles as stated would throw out as meaningless many discussions which even adherents of the theory did not wish to throw out. At least as early as 1934, it was pointed out that meaning should include not only practical but also theoretical verifiability.[1]

Under these conditions, we shall take the following criterion of meaning as a background for considering accuracy and precision:

> Every sentence in order to have definite cognitive meaning must be at least theoretically verifiable or confirmable as either true or false upon the basis of evidence theoretically obtainable by carrying out a definite and previously specified operation in the future. The meaning of such a sentence is the method of its verification.

This criterion obviously takes in practical[2] and theoretical verifiability. As stated, this principle allows us to resurrect our interpretation of the Probable Values of the Physical Constants from the waste-basket as being at least theoretically possible provided we can

- -

[1]Lewis, C. I., "Experience and Meaning", Phil. Rev., March 1934, pp. 125-146.
Nagel, Ernest, "Verifiability, Truth, and Verification", Journal of Phil., Vol. 31, March 15, 1934, pp. 141-147.

[2]Sometimes called physical and logical verifiability.

find a satisfactory theoretical operational interpretation of the concept "true value".

(3) General CONCEPTS OF ACCURACY AND PRECISION

These two terms have long been and continue to be used in the discussion of both pure and applied science - they are among the most common terms in scientific literature. But try to find out just what either term means or the difference between them and, as is so often true with terms of common parlance, we find that the meanings are not clear cut. In fact, the two terms are given as synonyms in some of the best dictionaries.

Books on the theory of errors also often use them as synonyms. The author of one of the most widely known books on the precision of measurements, however, bemoans the fact that the two terms are used thus carelessly and indiscriminately.[1] By precision or precision measure of a result he refers to what he terms the best numerical measure of its reliability which can, as he says, be obtained after eliminating or correcting for all known sources of error. By the accuracy of a result he refers to the degree of concordance between it and the true value of the quantity measured. These definitions, however, introduce several new terms and phrases such as "reliability" and "concordance" which are not adequately explained. Hence they don't get us very far.

[1] Goodwin, H. M., Precision of Measurements and Graphical Methods, McGraw-Hill Book Company, New York, 1913, pp. 7-8.

Obviously, if the terms accuracy and precision are synonymous, one cannot readily see any difference between the specification of an accuracy of 1% and a precision of 1%. If, on the other hand, we were to adopt the definitions made by Goodwin, we would first have to define reliability and concordance as well as the numerical measures of these before we could make intelligent specifications. It would appear, therefore, that one of the first things to be done is to decide whether or not the concepts are, as it were, synonymous. I personally feel that the concepts which these two terms try to define are fundamentally different insofar as they have applicability to the affairs of every-day life in engineering and science. however We shall~~, however,~~ see a little later why ~~see that~~ these terms have tended to become confused in the literature ~~of the theory of errors and statistics~~ because of certain assumptions that have been made to simplify the theoretical problem. But Later ~~abundant~~ evidence will be ~~introduced~~ Given to show that in practice there is little justification for believing that the assumptions hold. Hence there is little reason for confusing the two concepts from the viewpoint of practice.

Etymologically the term "accurate" has a Latin origin meaning "to take pains with" and refers to the care bestowed upon a human effort to make it what it ought to be. Likewise, "accuracy" in common dictionary parlance implies freedom from mistakes or exact conformity to truth. "Precise", on the other hand, has its origin in a term meaning "cut off", brief, concise. Likewise, precision is supposed to imply the property of determinate limitations or that of being exactly or sharply defined.

In what follows, therefore, I shall take as a starting point the following distinction:

I. Accuracy in some way or other involves the concept of a difference between what is observed and what is true.

II. Precision involves the concept of reproducibility of what is observed.

4 REPRODUCIBILITY OF WHAT IS OBSERVED

Let us start with the simple concept of length, such as the length of the line A B.

A________________________B

Observable length exists in terms of one or more operations of measurement. We have no way of observing length as such. Given any method of measuring length, the operation of measuring can be repeated again and again thus giving rise to an infinite sequence of numbers:

$$X_1, X_2, \dots X_i, \dots X_n, X_{n+1}, \dots X_{n+i}, \quad \dots. \qquad (1)$$

These differ among themselves. Thus ten measurements of a line with an engineer's scale graduated to 0.01" gave the following results expressed in inches.

4.000	3.994
3.996	3.990
3.996	3.992
3.994	3.994
3.996	3.992

TABLE 1

In what sense is there an observable length of the line A B for a given operation of measurement? In what sense is the observable length verifiable?

Length in this sense is representable by the infinite sequence (1). At the best it can only be theoretically verifiable because it is not possible from a practical viewpoint to observe more than a finite portion of an infinite sequence.

There is another point that we should note. We may divide the infinite sequence (1) into two parts:

Observed		Theoretically Observable
$X_1, X_2, \cdots X_i, \cdots X_n,$		$X_{n+1}, \cdots X_{n+i} \cdots$
Past	Present	Future

We usually think of the length of the line A B as existing as some constant value. However, insofar as observability goes, the length exists only in the sense that: a) it is possible, having taken any finite number n of measurements, to take as many more measurements in the future as we like, and b) knowing the n measurements, it is possible to make a more or less valid prediction in the probability sense as to what one may expect to get in future measurements. Schematically the probability relationship can be represented:

$$X_1, X_2, \cdots X_i, \cdots X_n, \qquad X_{n+1}, X_{n+2} \cdots X_{n+i}, \cdots$$

Probability

Ideal Observable Reproducibility - Random

Let us assume as an ideal case the reproducibility exhibited by a sequence of numbers formed by drawing one at a time with replacement from an experimental bowl. It is assumed that maximum validity is attainable in drawing probable inferences based upon a sample of n observed values drawn under such conditions. An infinite sequence of numbers obtained in this way will be said to be random and to have arisen from a physical state of statistical control. Definition

How Characterize Infinite Observable Sequence

Let us consider the nature of the operationally verifiable predictions that may be made about the unobserved but observable portion of such an infinite sequence. Three choices must be made:

1. Portion or portions of the sequence involved in the prediction.
2. Function or functions of the numbers in the chosen portion of the sequence.
3. Criteria to be satisfied by the numerical values of these functions.

Obviously we cannot say anything definitely verifiable about the as yet unobserved portion of an infinite sequence unless we make the first of these three choices. Schematically we may illustrate one such portion by the terms between the two vertical lines,

$$X_1, X_2, \cdots X_i, \cdots X_n, X_{n+1}, \cdots \;\Big|\; X_{n+i}, X_{n+i}, \cdots X_{n+i+j}, \;\Big|\; X_{n+i+j+1}$$

consisting of the j+1 terms between X_{n+i} and X_{n+i+j} inclusive of the end points. That is to say we must prescribe the manner in which these terms are to be determined.

Under such conditions we see that:

If $(i+j) < \infty$, the chosen portion is practically observable.

If $(i+j) = \infty$, the chosen portion is only theoretically observable.

Customarily $(i+j)$ is chosen equal to infinity corresponding to the concept of true value. In fact it is only in the sense of the whole of the infinite sequence that the property of length <u>exists</u> operationally. Obviously, therefore, operational <u>existence</u> is not verifiable once and for all from the practical viewpoint.

Table 2 shows some of the infinite number of different functions that may be chosen. The average $\bar{X}$, standard deviation σ and precision constant $h = \frac{1}{\sigma\sqrt{2}}$ are three common choices. The values of these for the infinite sequence will be referred to as $\bar{X}'$, σ' and $h' = \frac{1}{\sigma' \sqrt{2\Pi}}$. Of the symmetric moment functions there are an infinite number of choices of which σ and h are only two.

Sample	1.7 1.4	.3 .4	.4 .1	-1.8 -.6	.5 .8	-1.1 .1	-.7 -.4	.3 -.1	-.5 .7	-.1 0	-
	.2 .5	.3 .4	.4 .4	-.9 1.7	-.7 1.0	.7 -.3	1.5 .7	-1.8 1.3	-.8 .3	.2 -1.7	-
Average $\overline{X}$	.950	.350	.325	-.400	.400	-.150	+.275	-.075	-.075	-.400	-
Standard Deviation σ	.618	.050	.130	1.290	.660	.654	.879	1.119	.602	.758	- -
$z = x/\sigma$	1.537	7.000	2.500	-.310	.606	-.229	.313	-.067	-.125	-.528	- -
h											
μ_3	0	0	-.0025	1.596	-.2625	-.0578	.1626	-.6202	-.0151	-.4725	-
μ_4	.179	0	.001	5.839	.405	.335	.849	3.112	.172	.755	-
μ_1											
μ_∞											
aximum	1.700	.400	.400	1.700	1.000	.700	1.500	1.300	.700	.200	
inimum	.200	.300	.100	-1.800	-.700	-1.100	.700	-1.800	-.800	-1.700	
edian	.350	.350	.400	-.750	.650	-.100	.150	.100	-.100	-.050	
Range	1.500	.100	.300	3.500	1.700	1.800	2.200	3.100	1.500	1.900	

3 Having made the first two choices, we must make a choice of the limits within which the values of the chosen functions must lie. For each such function θ, the verifiable prediction must be in the form that the observed value for the chosen portion of the sequence will lie within specified tolerance limits

$$\theta_1 \text{ to } \theta_2.$$

That is to say, if we are to make a statement about the length of a line A B that can be decided definitely

YES or NO,

we must limit our statement to a finite portion of the potentially infinite sequence and must imply specific choices of function and of tolerance limits. In so doing, however, we have let the "existence" of the length of the line slip through our fingers.

<u>Example</u>

Let us consider the statement: "Michelson has during the last few years been making new measurements, and as a result of these announces the velocity of light to be

299,796 kilometers per second,

with a possible error of one kilometer only!"

---

1) Cf Mechanism of Nature by E. N. da C. Andrade, G. Bell and Sons, Lon. 1930 p. 80.

How would you verify this statement? One can conceive of there being a true constant value of the velocity of light. So far as I can see, however, this true value is not observable at least in any practical sense. All that we can do operationally is to take more measurements by Michelson's or some other method. To make an observably definite statement we must state how this quoted value is to be compared with future observations. How many n measurements are to be made? Shall we take the average or some other function of these n measurements as a basis for comparison? These are typical of the questions that must be decided before any statment about the velocity of light becomes operationally definite.

For example, let us consider the statement that future measurements of the velocity of light will not differ from that quoted above by more than one kilometer. Is this statement operationally verifiable in the sense that it can be decided unequivocally Yes or No? If it is to be answered in the affirmative, we must first limit ourselves to a finite future period of time. Then we must specify unequivocally what is to be meant by a measurement. Shall the experimentalist be allowed to group his single observations as he sees fit? To reject so-called discordant observations? To weight his observations as he chooses? To choose at will the number of single observations to be considered as the basis for what is to be taken as a measurement? To illustrate the effect of grouping we may again refer to Michelson's original data. When he grouped them in the following way according to kinds of mirrors used he got the following results in kw. per sec.

Glass 8	299,797
Steel 8	299,795
Glass 12	299,796
Steel 12	299,796
Glass 16	299,796

TABLE 3

However, the range in single observations is 299,718 to 299,859.

Of course, if we could observe the true value C', we could verify once and for all whether or not

$$\left| 299{,}796 - C' \right| \leq 1.$$

Now let us see how the customary measures of accuracy and precision are related to the observability of an operation or measurement.

5 CUSTOMARY MEASURES OF ACCURACY AND PRECISION

Let us consider the concept of accuracy in the case where the true value X' is given theoretically. Let us lay off on a line a number representing this true value Fig. 1. Then let us locate on this line two other points $X = X' - \ell_1$ and $X = X' + \ell_2$.

Fig. 1

Accuracy with respect to the range $X' - \ell_1$ to $X' + \ell_2$ has to do with the clustering of the numbers in a sequence of type (1) within this range.

Let us consider the corresponding concept of precision. Let $\bar{X}'$ be the average of the numbers constituting the sequence. Locate this average and two other points $\bar{X} = X' - \ell_1$ and $X = \bar{X}' + \ell_2$ on a line Fig. 2. Let us assume a case where the theoretical true value X' is not equal to X' is not equal to $\bar{X}'$.

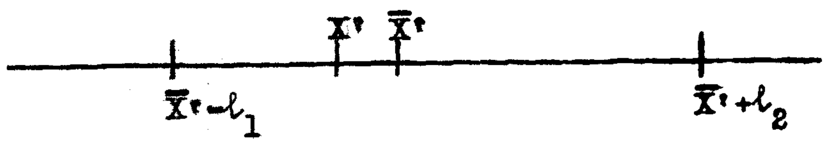


Fig. 2

Precision has to do with the clustering of the numbers in a sequence of measurements within the range $\bar{X}' - \ell_1$ to $\bar{X}' + \ell_2$.

If we view accuracy and precision in this way, it is obvious that for a sequence in which

$$X' = \overline{X}', \qquad (2)$$

or where the theoretical value X' is the same as the average $\overline{X}'$ of the sequence, the concepts of accuracy and precision become the same if the method of measuring the clustering effect is made the same in the two cases. The customary theory of errors measures the clustering in the same way and in effect assumes that equation 26 is satisfied. Hence it is that the two terms are often fused into one and the same meaning.

Let us for the moment take as a measure of clustering the fraction p' of the numbers in a sequence falling within the chosen range. Then we might conceive of comparing two sequences of measurements of the same true value X' in respect to accuracy corresponding to a chosen interval by comparing the corresponding fractions of numbers falling within this range. In the case where $\ell = \ell_1$ is chosen equal to ℓ_2, it is sometimes convenient to speak of the accuracy for a chosen value p' as the ratio

$$\text{percent accuracy} = \frac{\ell}{X'} \times 100. \qquad (3)$$

Such an accuracy expressed as a percentage, however, is obviously dependent upon the value of p' chosen. Presumably we may define precision in much the same way as

$$\text{percent precision} = \frac{\ell}{\overline{X}'} \times 100 \qquad (4)$$

where the precision as thus given corresponds to a chosen value of p'. Equations (27) and (28) constitute the basis for defining accuracy and precision in percentage. It should be noted that if the true value X' is equal to the expected value $\overline{X}'$, then the percentage measures become the same.[1]

6 FIVE CONCEPTS TO BE GIVEN OPERATIONAL MEANING

In the last section, three concepts were introduced corresponding to the symbols

$$p', X' \text{ and } \overline{X}'$$

All of which were defined in terms of an infinite sequence. These call for operational interpretation.

---

1. Of course it is sometimes assumed in the literature of the theory of errors that not only does $X' = \overline{X}'$ but also that the distribution in the sequence is random and follows the normal law of error with standard deviation σ'. Then the probable error which is $.6745\ \sigma'$ is taken as a measure of accuracy and $h' = \dfrac{1}{\sigma'\sqrt{2}}$ is taken as a measure of precision, σ' being the standard deviation of the law of error. Here again perhaps we find another reason why the two concepts are often taken as synonymous in the literature, because they are both expressible in terms of the standard deviation σ'.

OBSERVABILITY OF p' and $\bar{X}'$

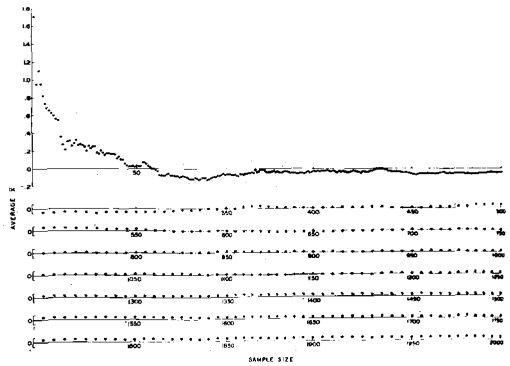
Concept

True probability p' and expected value $\bar{X}'$

Concept of Statistical limit

$$\lim_{n \to \infty} \theta = \theta'$$

Show case for average of sample of size n drawn from a normal bowl universe.



#1545

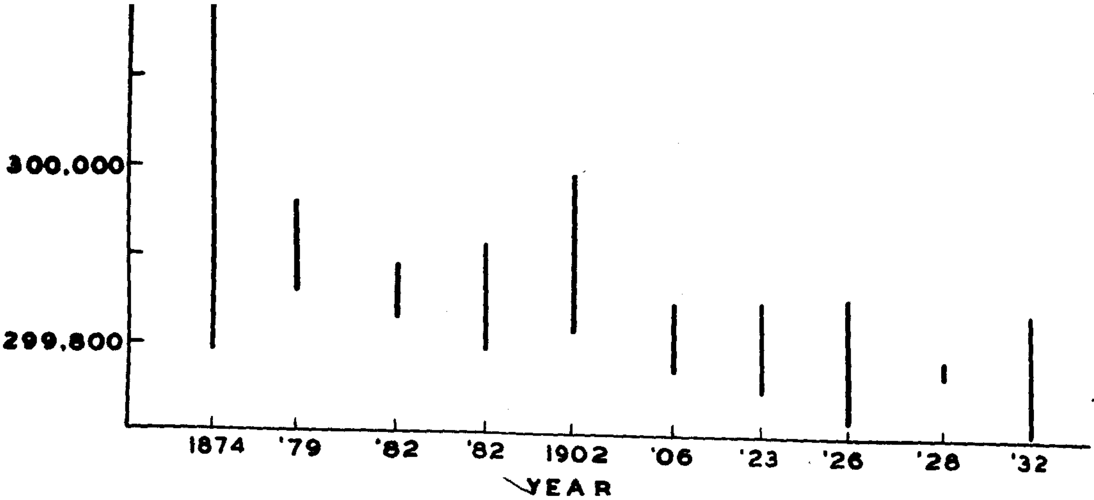
FIG. 8

We can never reach a value of n such that we can say

$$|X' - X| \leq \varepsilon$$

~~Where~~ ε is chosen at pleasure.

Hence p' and $\bar{X}'$ are not observable in practice.



bility

ed

re-

or

in

FURTHER DIFFICULTY - Non predictability

Moreover in discussing reproducibility, we called attention to the fact that all infinite sequences are not the same from the viewpoint of being able to make valid predictions. In the measurement of the length of the line for example, it is tacitly assumed that all the measurements in the sequence are to be carried out under the <u>same essential conditions</u>. Strictly speaking then we must think not only of the sequence itself but also of the conditions surrounding the taking of each measurement. We may represent this situation schematically as follows:

$$
\begin{array}{cccccccccccc}
X_1, & X_2, & \cdots & X_i, & \cdots & X_n, & X_{n+1}, & \cdots & X_{n+i} & \cdots \\
| & | & & | & & | & | & & | & \\
C_1 & C_2 & & C_3 & & C_n & C_{n+1} & & C_{n+i} &
\end{array}
$$

where the C's stand for what we mean by the term condition as used in the phrase, the same essential conditions. What we need is an operationally verifiable criterion that must be met by sequences supposed to arise under the same essential condition. This we shall call the criterion of control.

Finally we need to consider the fact that statements such as

The accuracy of this test method is 1%.

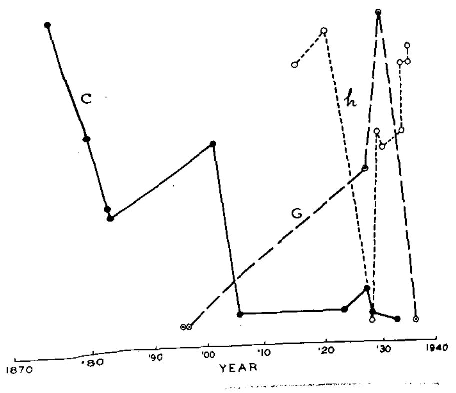
The precision of this test method is 1%.

are probable inferences. In any probable inference P on evidence E there is presumably an objective degree p_b' of belief which one should place in the inference, and this inheres in the relation between E and P. Schematically we have:

$$E \qquad P$$

$$p_b'$$

How shall we give p_b' an operationally verifiable meaning? We shall find this question of particular importance in our discussion of accuracy.



~~Where ε is chosen at pleasure. Hence p' and $\bar{X}$' are unobservable in practical sense.~~

OBSERVABILITY OF X'

So far our discussion of observability has been confined to a consideration of a single type of operation or measurement. In general there are many ways of measuring length, mass, time, c, e, h, G and all other physical properties. What we need to consider is a potentially infinite number of methods of measurement, particularly when we allow for differences between measuring instruments (and observers) of the same kind. The situation is that shown in Fig. 9.

$$
\begin{array}{ccccc|ccc|cc}
\cline{6-8}
x_{11}, & x_{12}, & \cdot\cdot & x_{1i}, & \cdots\ x_{1n}, & x_{1n+1}, & \cdots & x_{1n+i}, & x_{1n+i+1}, & \cdots \\
x_{21}, & x_{22}, & \cdot\cdot & x_{2i}, & \cdots\ x_{2n}, & x_{2n+1}, & \cdots & x_{2n+i}, & x_{2n+i+1}, & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
x_{k1}, & x_{k2}, & \cdot\cdot & x_{ki}, & \cdots\ x_{kn}, & x_{kn+1}, & \cdots & x_{kn+i}, & x_{kn+i+1}, & \cdots \\
\cline{6-8}
\multicolumn{10}{c}{\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots} \\
\multicolumn{10}{c}{\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots}
\end{array}
$$

~~FIG.~~ 9 Slide 6

Any definitely verifiable statement about accuracy must involve three choices:

a) Some finite portion of each sequence from each of a specified number of different operations.

b) Some function Ψ

c) Some tolerance limits Ψ_1 and Ψ_2 such that

$$\Psi_1 \leq \Psi \leq \Psi_2'$$

is a definite yes-no criterion.

Then there is the question of predictability. Already illustrated by measurements of physical constants.

VALID PREDICTABILITY

Concept

a) System of constant chance causes.

b) Same essential conditions.

c) Physical state of statistical control.

d) Random sequence.

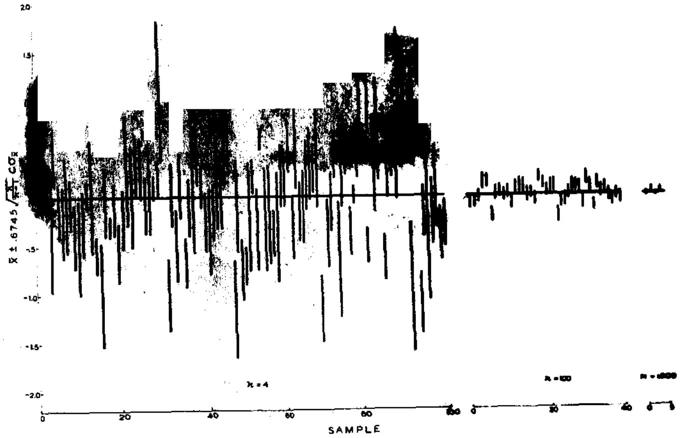
The first three refer to physical conditions or causes. The last refers more to the observable numbers.

Operational Meaning

It is assumed that drawings with replacement and thorough mixing from a bowl satisfy conditions a, b and c. The observed sequence is assumed to satisfy d.

Valid Prediction: Type 1

When distribution in bowl is normal, "Student" theory enables one to make valid prediction as indicated in Fig. 1. This assumes that X' is known. Validity of prediction is the same for all sample sizes.



PREDICTION – TYPE I

FIG. 1

Valid Prediction: Type II

Given	Predict
$X_1, X_2, \ldots\ldots X_n$	$X_{n+1}, X_{n+2}, \ldots X_{n+i} \ldots$

It should be noted that this constitutes prediction in terms of observables. To make it definite we must choose:

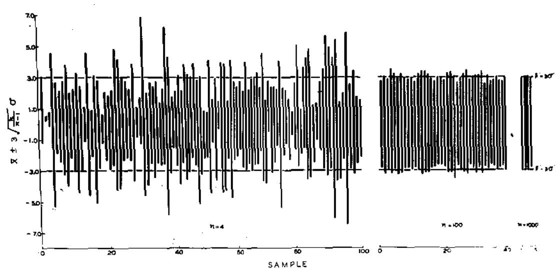
a) Portion of future sequence

b) Function

c) Tolerance range.

All of which have previously been noted. In addition we must choose:

d) The allowable error



PREDICTION – TYPE II

FIG. 2

Note that errors are reduced as n increases.

For example we must often go to n = 1000 or more before errors can be reduced enough to justify setting of tolerances.

Now reconsider Eddington's comment about

$$h \times 10^{27} = 6.551 \pm .013$$

$$h \times 10^{27} = 6.547 \pm .008$$

or in general the results of measurement expressed in the form

$$X \pm \Delta X$$

Even assuming that variable X is in a state of statistical control. Obviously interpretation depends on size of n.

8 CRITERION OF VALID PREDICTABILITY

In general we have infinite sequence which may or may not be like the "random" one drawn from a bowl. Let us represent such a sequence as follows:

$$\begin{array}{cccccc} X_1, & X_2\cdots & X_i\cdots & X_n & X_{n+1}\cdots & X_{n+i}\cdots \\ | & | & | & | & | & | \\ C_1 & C_2 & C_i & C_n & C_{n+1} & C_{n+i} \end{array}$$

Such a sequence should satisfy criterion I when broken into subgroups of 4 (or small sample) by means of the C's.

A. Random Sequence Satisfies Criterion

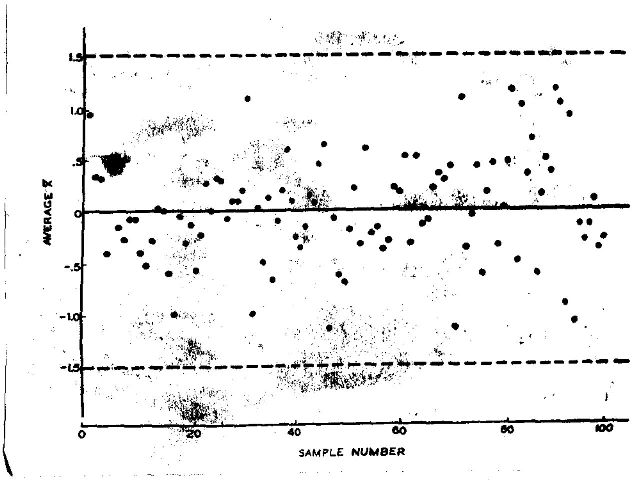


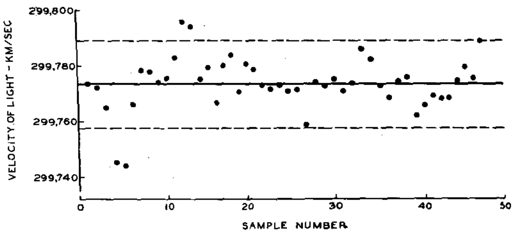
FIG. 3

All points are within.

Empirically found that we must have from 25 to 250 samples of 4 that show control before we can be reasonably sure that we have physical state of statistical control.

B. Most Sequences Do Not Satisfy Criterion

Show application to measurements of the velocity of light.



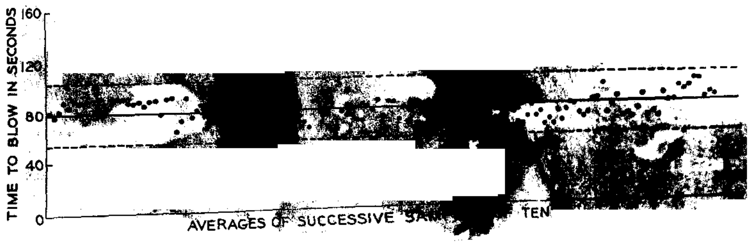
#17718

S 10

FIG. ~~4~~

C. Show That We Can Get Sequences That Satisfy Criterion I

Blowing time of fuses.

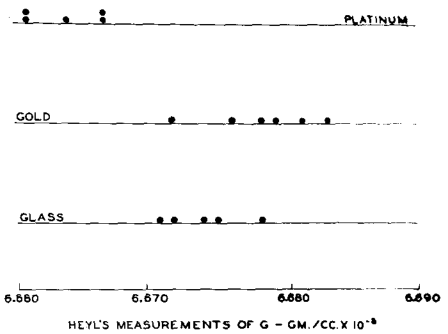


#16047

~~FIG.~~ 5 S 11

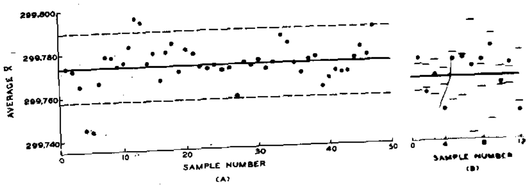
Accuracy.

More difficult in that we must take into account all kinds of operations.



#17161

S 12



#17795

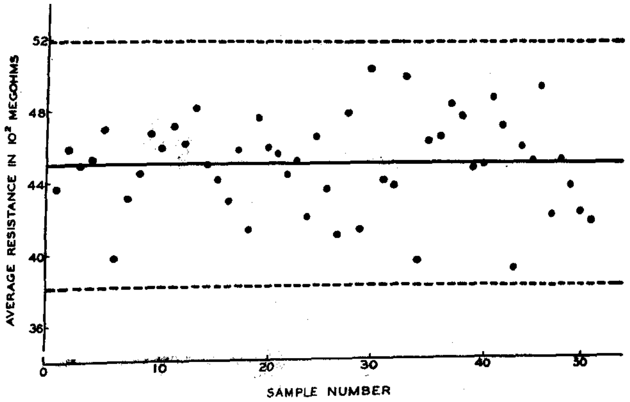
S 13

D. Wipe Out C's and Criterion Fails

5045	4565	4895	4335	4550	4685	4850	4700	4500
4350	4410	4255	5000	4700	4430	4450	4500	4765
4350	4065	4170	4615	4310	4300	3635	4840	4500
3975	4565	3850	4215	4310	4690	3635	5075	4500
4290	5190	4445	4275	5000	4560	3635	5000	4850
4430	4725	4650	4275	4575	3975	3900	4770	4930
4485	4640	4170	5000	4700	2965	4340	4570	4700
4285	4640	4255	4615	4430	4080	4340	4925	4890
3980	4895	4170	4735	4850	4080	3665	4775	4625
3925	4790	4375	4215	4850	4425	3775	5075	4425
3645	4845	4175	4700	4570	4300	5000	4925	4135
3760	4700	4550	4700	4570	4430	4850	5075	4190
3300	4600	4450	4700	4855	4840	4775	4925	4080
3685	4110	2855	4700	4160	4840	4500	5250	3690
3463	4410	2920	4095	4325	4310	4770	4915	5050
5200	4180	4375	4095	4125	4185	4500	5600	4625
5100	4790	4375	3940	4100	4570	4770	5075	5150
4635	4790	4355	3700	4340	4700	5150	4450	5250
5100	4340	4090	3650	4575	4440	4850	4215	5000
5450	4895	5000	4445	3875	4850	4700	4325	5000
4635	5750	4335	4000	4050	4125	5000	4665	
4720	4740	5000	4845	4050	4450	5000	4615	
4810	5000	4640	5000	4685	4450	5000	4615	

Table of Data

Application without reference to C's.



#15438

FIG. 6

Negative result.

Now apply with use of C's.

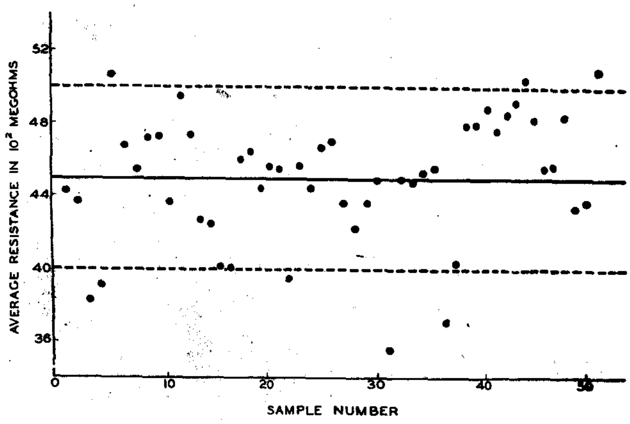


FIG. 7

Positive result.

Hence operational criterion of predictability in terms of observation involves a <u>human element</u>.

<u>Conclusion</u>

There is an operational criterion of valid predictability. Its use involves human ingenuity in use of C's and the empirical fact that a certain quantity of data is necessary.

9) CONCLUSION

A. Meaning of both precision and accuracy verifiable only in a limited sense & involves at least three choices.

B. Meaning of accuracy and precision not operationally definite from practical viewpoint when expressed in terms of expected and true values, as is usually done.

C. Significance of both accuracy and precision is their relation to a valid predictability within a specified range.

D. Valid prediction – in sense of [illegible] – not in general possible until assignable causes have been eliminated.

E. Criterion of Valid Predictability available at least for precision.