Walter A. Shewhart
% Bell Telephone Laboratories
463 West St., N.Y., N.Y.

MATHEMATICAL STATISTICS

IN

MASS PRODUCTION

by

W. A. Shewhart

Paper to be delivered at the Symposium on Applied Mathematics of the American Mathematical Society at Columbia University, February 21, 1941

SOME PRELIMINARY COMMENTS

Some Definitions

It has been said that:

Physicist is one who has a clean mind and works with dirty things,
Chemist is one who has a dirty mind and works with clean things,
Engineer is one who has a dirty mind and works with dirty things.

I might add:

Mathematician is one who has such a clean mind that he must work only with the abstract symbols of clean things.

"Mathematics is the subject in which one never knows what he is talking about nor if what he says is true".

Not so long ago a well-known physicist defined a mathematical physicist as one who among physicists is considered a mathematician and among mathematicians is considered a physicist. In the same way, it might be said about a mathematical statistician in the engineering field that he is one who among engineers, is a mathematician and among mathematicians, is an engineer.

I

INTRODUCTION

Historical

1. England "Student" (W.S.Gosset). Brewing, about 1900. First company report 1904. Slide 1

2. Germany Karl Daeves. Metallurgy. First known publication, 1922. I first learned of Grosszahl-forschung about 1924.

3. United States E.C.Molina. Telephone trunking theory. Malcolm Rorty memo 1903. Molina began internal application 1905; first patent 1906; two important contributions Dec., 1907; publication, 1913.

Contrast

Student 1900	Beer	Error theory and elements of design of experiment	1. Error of the mean. 2. Elements of design of exp.
Daeves 1922?	Steel	Causes of variability in metals.	1. Practical importance of evidence of multimodal freq. curves.
Molina 1905	Telephone switching systems.	a priori design	1. Telephone trunking theory.
Quality control 1924	Manufactured articles	Applications in three fundamental steps: specification [illegible]	1. Sampling plans to meet consumer and producer risks 2. [illegible]

3. Theory for setting tolerance limits.

4. Criteria for studying variation produced by matter in microscopic and even atomic quantities.

5. General theory and technique for control of manufactur[ing] process as an <u>operation</u>.

FUNDAMENTAL CONCEPT

Mass production = repetitive operation

Let X be a quality characteristic of the thing produced. Sequence of repetitions of an operation gives

$$(X_i)_{i=1}^{\infty} = X_1, X_2, \ldots, X_i, \ldots X_n, \ldots \quad (1)$$

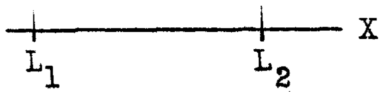
Desire control of causes of variability in X's.

Three Steps in Process of Control	Corresponding three Steps in Scientific Method	Contribution of of Statistics
Specification	Hypothesis	Statistical hypothesis
Production	Experiment	Statistically designed experiment
Inspection	Test of hypothesis	Statistical test of Hypothesis

Bird's-eye View of Operation of Mass Production

I. Specification

1. Tolerance range



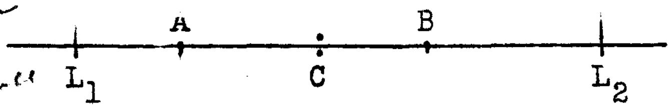
Research Development

2. Operation

Conditions for { Maximum assurance / Minimum tolerance range

II. Production

Since one cannot carry Step 1 to completion before start of production, we must introduce operation of statistical control.



III. Inspection

Sampling plans { 1. Consumer risk / 2. Producer risk

Detect assignable causes of variability and provide evidence for modification of standard.

3 steps like a chain: no stronger than weakest link. Often hundreds and even

FUNDAMENTAL CONTROL HYPOTHESES

Hypothesis I - Some repetitive operations exist in nature that obey mathematical laws of probability.*

Hypothesis II - The maximum attainable degree of validity of prediction** that an operation will give a value X lying within any previously specified tolerance range is that based upon the prior knowledge that the probability of this event is q' or more generally, upon the prior knowledge of the mathematical law of chance underlying the operation.

Hypothesis III - The maximum degree of attainable control*** of the cause system underlying any repetitive operation in the physical world is that wherein the system of causes produces effects in accord with a mathematical law of probability.

Hypothesis IV - Some criterion or criteria may be found and methods developed for their application to the numbers obtained in a sequence of repetitions of any operation such that whenever a failure to meet the criterion or criteria is observed, it is worthwhile to look for and try to remove an assignable cause of variability from the operation. As these causes are removed, a

- -

* For example, drawing from a bowl is such an operation.

** Or, in engineering terms, maximum quality assurance.

*** Hence minimum tolerance limits and most efficient use of materials.

state of statistical control is approached where the results of repetitions of the operation behave in accord with a mathematical law of chance.

It is not the object here to discuss the available evidence supporting these physical hypotheses because that has been done elsewhere, but rather to show the prominent part played by mathematical laws of probability in the fundamental assumptions and to emphasize the point that the testing and use of these hypotheses implies that the engineer must keep his eyes on the physical operation as well as on the mathematics.

CRITERIA OF CONTROL

1. Relative Effects of Causes

Criteria based upon frequency distribution of variable X in terms of elemental effects of system of m elemental causes in a constant system of chance causes.

If one of the m causes produces a very large effect in comparison with that produced by any one of the (m-1) remaining causes, it may be possible to find and remove it and the presence of such a cause will likely be revealed by bimodality of the distribution.

2. Lack of Constancy in Probability

Criteria based upon order of occurrence in the sequence (1) revealing lack of constancy in the cause system, i.e., lack of constancy in the probability $f(x)dx$.

This may result in multimodality that may be detected and will always modify runs in a way that can likely be detected.

TWO KINDS OF ERROR IN TESTING ANY HYPOTHESIS

Contrary {Statistical / Statistical Control

1. (Statistical Hypothesis) Example 1

The significance of the mean of a unique random sample $X_1, X_2, \ldots, X_n$.

Slide 2 - Table from Fisher

Compute $\bar{X} = \frac{1}{n}\Sigma(X) = 1.58$

$$\frac{s^3}{n} = \frac{1}{n(n-1)}\Sigma(X-\bar{X})^2 = .1513$$

$$t = \frac{\bar{X}}{\sqrt{\frac{s^2}{n}}} = 4.06$$

$$P(4.06) < .01$$

Hence reject hypothesis.

Example 2 Significant difference between two means.

$$\pi_1 \qquad \pi_2$$

$$m_1\sigma_1 \qquad m_2\sigma_2$$

H_1: $m_1 = m_2$ and $\sigma_1^2 = \sigma_2^2$ against

$m_1 \neq m_2$ and/or $\sigma_1^2 \neq \sigma_2^2$.

H_2: $m_1 = m_2$ against $m_1 \neq m_2$

H_3: assume $\sigma_1^2 = \sigma_2^2$, test $m_1 = m_2$ against

$m_1 \neq m_2$.

Student test is uniformly most powerful for H_3.

Engineering Comments

a. Differences are all positive and test does not change engineer's action. Better to have example where it does.
b. Engineer seldom if ever has unique sample.
c. Sample is almost never random.
d. Engineer interested in two kinds of error.
   A. Chance of rejecting hypothesis when true.
   B. Chance of accepting hypothesis when false.

2. (Statistical Control) Hypothesis

pp. 39 and 40 attached

Contrast {
1. Formal hypothesis: i.e.
If — then —
2. Empirical hypothesis
Such as that of control.
Empirical Rules of Action.

Two kinds of errors in the operation of control. Since a scientific inference about experience can never be more than probable, it is always subject to two general kinds of errors which we may write as follows:

e_1 Sometimes when a scientific hypothesis H is rejected, the hypothesis H is nevertheless true.

e_2 Sometimes when a scientific hypothesis H is accepted, the hypothesis H is nevertheless false.

Neyman and Pearson have considered specific instances of these two general kinds in testing certain statistical hypotheses.[22] They consider the problem of having been given a sample consisting of the first n terms of an infinite sequence considered without respect to order, to determine whether it came from a universe π (hypothesis A). Representing the set of n values as a point Σ in hyperspace, they say—

> *Setting aside the possibility that the sampling has not been random* or that the population has changed during its course, Σ must either have been drawn randomly from π or from π', where the latter is some other population which may have any one of an infinite variety of forms differing only slightly or very greatly from π. The nature of the problem is such that it is impossible to find criteria which will distinguish exactly between these alternatives, and whatever method we adopt two sources of error must arise:
>
> e_{11} Sometimes when Hypothesis A is rejected, Σ will in fact have been drawn from π.
>
> e_{21} More often, in accepting Hypothesis A, Σ will have been drawn from π'.

These two kinds of errors are called by Neyman and Pearson "errors of the first and second kinds," and are obviously somewhat like two different pairs of errors encountered in using the operation of statistical control.

The first of the two pairs of errors (e_1 and e_2) is encountered in interpreting a criterion of control applied to a given finite sequence of observations, and may be written in the following form—

e_{12} We may reject the hypothesis that there existed, at the time the finite sequence was obtained, one or more assignable causes in the process giving rise to that finite sequence, when this hypothesis is nevertheless true.

e_{22} We may accept the hypothesis that there existed, at the time the finite sequence was obtained, one or more assign-

[22] J. Neyman and E. S. Pearson, "On the use and interpretation of certain test criteria for purposes of statistical inference," *Biometrika*, vol. 28A, pp. 175–240, 1928; and in particular, p. 177. The italicizing in the quotation is mine. I have also introduced the symbols e_{11} and e_{12} instead of their numerals 1 and 2.

able causes in the process giving rise to that finite sequence, when this hypothesis is nevertheless false.

It should be noted that the hypothesis in this instance pertains to the existence of assignable causes during the time the finite sequence was being obtained.

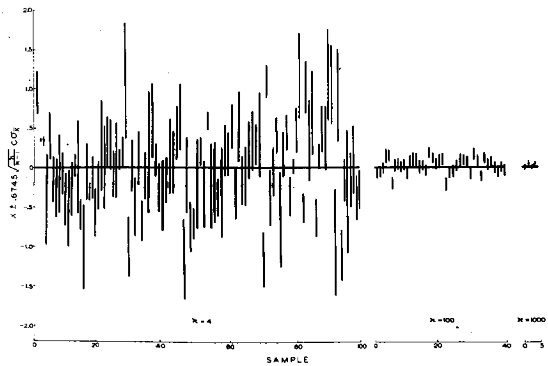
The pair of errors e_1 and e_2, so far as they are encountered in interpreting the operation of control as a whole, may be stated similarly—

e_{13} We may reject the hypothesis that the production process or repetitive operation is in a state of statistical control when this hypothesis is nevertheless true.

e_{23} We may accept the hypothesis that the production process or repetitive operation is in a state of statistical control when this hypothesis is nevertheless false.

In this instance, we should note that the hypothesis pertains to the condiions existing within a repetitive operation *throughout the time required to roduce an infinite sequence.*

These three pairs of errors are alike in general form, but they differ in ıe hypotheses involved. They also differ in that Neyman and Pearson's rrors e_{11} and e_{21} of the first and second kinds are essentially formal, whereas ıe other two pairs are expressed in empirical terms. For example, Neyıan and Pearson can theoretically build an exact mathematical model ıat enables them to compute with any desired degree of exactness the probabilities of occurrence of their two kinds of errors. It will be noted ıat their hypothesis involves the assumption that the observed data onstitute a *random* sample, and we have already considered some of ıe difficulties involved in trying to give this term an empirical and operaionally verifiable meaning. In fact, we may think of the whole operation statistical control as an attempt to give such meaning to the term random. ut just as soon as we pass from the concept of the errors e_{11} and e_{21} of eyman and Pearson, which may be defined in terms of a mathematical ıodel, to errors of the general form e_1 and e_2 expressed in terms of scientific ypotheses about observable phenomena, we can no longer compute with nathematical exactness the probabilities associated with any pair of errors or a given hypothesis. As a background for the development of the peration of statistical control, the formal mathematical theory of testing a atistical hypothesis is of outstanding importance, but it would seem that *must continually keep in mind the fundamental difference between the rmal theory of testing a statistical hypothesis and the empirical testing of ypotheses employed in the operation of statistical control.* In the latter, one



PREDICTION – TYPE I

ESTABLISH TOLERANCE RANGE

Simplest Case

Assume normally distributed quality X with unknown mean $\bar{X}'$ and standard deviation σ'.

Problem - Set up tolerance range $X = L_1$ to $X = L_2$ that will include 100P percent of product where P = .997, let us say.

From a sample $X_1, X_2, \ldots, X_n$, set up a tolerance range

$$\theta_1 \pm t\theta_2 = \bar{X} \pm t\frac{\sqrt{\Sigma v^2}}{\sqrt{n-1}}$$

for normal law.

$$\theta_1 \pm t\theta_2 = 2.02 R$$

for n = 10, rect. universe where R = range.

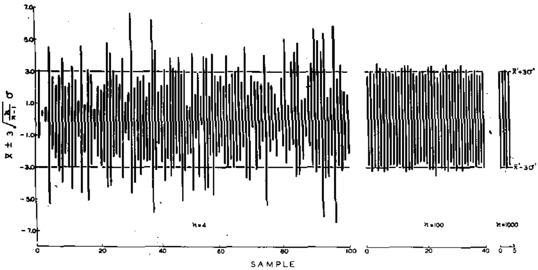
Contrast with Fiducial Range

Given a random sample $X_1, X_2, \ldots, X_n$, under certain conditions it is possible to compute from the sample two limits θ_1 and θ_2 such that the probability P is $1-\alpha$ that a parameter θ lies between two limits, $\theta_1 \leq \theta \leq \theta_2$. For $\bar{X}'$ this is given by Student.

Slide 3 - Fig. 14 in book.

True independent of n.

In case of tolerance range, however, the range is improved as we increase n. Mathematics helps us to choose shortest range.



PREDICTION – TYPE II

Slide 4

What we would like to know is how large a sample n so that

$$P_1(p_1 \le p \le p_2) = 1 - \alpha_1$$

$$P_2(R_1 \le R \le R_2) = 1 - \alpha_2$$

Propagation of Error in Setting Overall Tolerances

Consider pile-up of m = 30 as in relay.

Consider pile-up of m = 100 as in condenser.

Let n = 100. Results shown graphically in

MSH2

Slide 5

	Average cut off by 40 ranges	Minimum percentage	Max. $\frac{R}{\overline{R}'}$
m = 1	.9967	.9923	
m = 30	.9922	.9624	
m = 100	.9732	.8186	

Contrast $m\overline{X} \pm \sqrt{m}ts$ with $m\overline{X}' \pm \sqrt{m}3\sigma'$

Practical Significance

1. Mathematics plays important role but more to be done.
2. Efficient use of material demands mass production.
3. The greater m, the larger n must be as basis for setting tolerance range.

III

OPERATION OF CONTROL

"Experiment without imagination or imagination without recourse to experiment, can accomplish little, but for effective progress, a happy blend of these two powers is necessary."

Lord Rutherford
"The Electrical Structure of Matter",
Science, V.58, 1923.

Operation of mass production not random	1. Hence control limits in production process (Step 2). 2. Hence sampling plans in inspection (Step 3). 3. Hence operation of control in research (Step 1).

Perhaps the greatest potential value of statistics is Guide to Experiment.

Operation of control applied to Step 2 consists of the following five steps:

Slide 6 - Steps in control

Use of operation of control - Step 2	1. Reduction in cost of inspection 2. Reduction in cost of rejections 3. Minimum tolerance-max. efficiency 4. Maximum assurance

As an example - blowing time of fuses

Thickness of Paper in mils

Horizontal - Across sheet

Type of Paper	Method of Measurement	Company 1 Av.	Company 1 Variance Areas	Company 1 Variance Sheets	Company 1 Variance Residual	Company 2 Av.	Company 2 Variance Areas	Company 2 Variance Sheets	Company 2 Variance Residual	Company 3 Av.	Company 3 Variance Areas	Company 3 Variance Sheets	Company 3 Variance Residual
A	Ratchet	1.28	.0016	.0149*	.0025	1.26	.0007	.0212*	.0014	1.32	.0072*	.017[illegible]*	.0017
	Dial	1.29	.0013	.0151*	.0014	1.29	.0009	.0174*	.0015	1.29	.0007	[illegible]	[illegible]
B	Ratchet	1.63	.0064	.1136*	.0050	1.62	.0096	.0987*	.0063	1.64	.0073	[illegible]	[illegible]
	Dial	1.60	.0084	.0956*	.0050	1.59	.0097	.1068*	.0062	1.61	.0084	[illegible]	[illegible]
C	Ratchet	1.81	.0071	.0854*	.0040	1.79	.0093	.0825*	.0057	1.80	.0075	[illegible]	.006[illegible]
	Dial	1.80	.0061	.0993*	.0045	1.81	.0114	.0711*	.0059	1.80	.0058	.0769*	.0063
D	Ratchet	1.92	.0049	.0053	.0025	1.97	.0027	.0136*	.0022	1.99	.0051	.0146*	.002[illegible]
	Dial	1.98	.0044	.0137*	.0018	1.99	.0058	.0138*	.0027	1.98	.0038	.0144*	.002[illegible]
E	Ratchet	1.92	.0020	.0080*	.0016	1.91	.0122*	.0100*	.0025	1.93	.0128*	.0072	.0034
	Dial	1.91	.0092*	.0072*	.0020	1.93	.0069*	.0051*	.0021	1.92	.0081*	.0020	.0022
F	Ratchet	3.66	.0029	.0023	.0030	3.63	.0054	.0065	.0052	3.57	.0220*	.0062	.0064
	Dial	3.63	.0057	.0054	.0033	3.69	.0066	.0085	.0044	3.66	.0048	.0075	.004[illegible]

* Statistically significant

Slide 8

Research Technique

1. Analysis of variance in comparing error of measurement of different machines and laboratories.

Slide 8 - Paper data

2. Need for New Technique of Research

Three difficulties arise when the scale of physical and chemical operations is reduced:

1. New physico-chemical hypotheses
2. New methods of laboratory operations.
3. New techniques for analyzing data and testing hypotheses.

New technique embodies principles, points of view, and objectives that make it differ from classical technique sufficiently to make it a new kind of analysis.

Examples

a) Newtonian vs. quantum mechanics.
b) Quantitative vs. microchemical and micro-gas analysis.
c) Classical statistical criteria ignoring order vs. criteria based on order.

EXAMPLE OF DIFFERENCE BETWEEN TECHNIQUES

[illegible]	[illegible]	[illegible]	29	43	24	45	52	39	62
[illegible]	[illegible]	[illegible]	45	26	30	28	42	38	50
[illegible]	10	[illegible]	41	2	28	38	53	49	70
[illegible]	[illegible]	[illegible]	47	17	71	28	53	38	60
[illegible]	[illegible]	[illegible]	38	28	57	23	42	35	36
[illegible]	[illegible]	[illegible]	56	51	37	18	29	35	65
[illegible]	[illegible]	[illegible]	39	19	43	39	43	22	50
[illegible]	[illegible]	41	[illegible]	42	58	52	17	30	75
[illegible]	[illegible]	[illegible]	[illegible]	35	33	43	39	24	90
[illegible]	[illegible]	[illegible]	[illegible]	3	62	31	48	67	96
[illegible]	[illegible]	[illegible]	[illegible]	27	27	36	58	54	80
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	26	42	59	73	88

[illegible] in Arbitrary Units of Inlay
[illegible] Relay Springs

e sample
m sample
tention paid to order.

old and new drill
- data from paper
ırt

EXAMPLE OF DIFFERENCE BETWEEN TECHNIQUES

Small sample test of significance	1. Unique sample 2. Random sample 3. No attention paid to order.

Example - Comparison of old and new drill on ten plots - data from paper by John Wishart

Example 1 of Use of New Technique - Thickness of Rolled Inlay on Relay Spring

1. Show relay spring and strip from which it was cut.
2. Show table of data on 144 springs.

3. Show plot of these data for

a. Observed order
b. Random order.

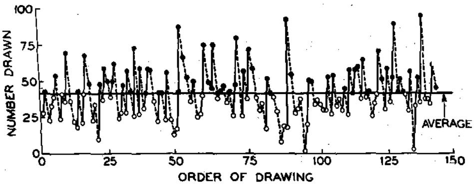


FIG. 3a SEQUENCE OF NUMBERS DRAWN AT RANDOM

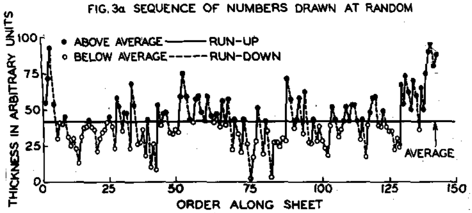


FIG. 3b SEQUENCE OF VALUES OF THICKNESS OF INLAY ON RELAY SPRINGS

4. Interpret figure.

Theory

1. Physico-chemical hypothesis.
   - 1.1 Statistical states
   - 1.2 Transition states
   - 1.3. Transitory states
2. Probability of occurrence of runs in random state.

5 Fields of use (few)

5.1. Corrosion
5.2. metallurgy.
5.3 properties of materials.
5.4 Contact phenomena

Length of Runs	Drawings from Bowl: Runs Above and Below Average, Observed Number	Drawings from Bowl: Runs Above and Below Average, Expected Number	Drawings from Bowl: Runs Up and Down, Observed Frequency	Drawings from Bowl: Runs Up and Down, Expected Frequency	Measurements of Inlay Th: Runs Above and Below Average, Observed Number	Measurements of Inlay Th: Runs Above and Below Average, Expected Number	Measurements of Inlay Th: Runs Observ Frequenc
1	42	37			39	37	
2	18	15	65	61	13	16	90
3	10	9	22	27	3	9	28
4	4	4	10	8	5	4	8
5	4	2	1	2	0	2	2
6	0	1	0	0	4	1	
7	0	0	0	0	4	0	
8	0	0	0	0	1	0	

Example 2 of Use of New Technique - Contact Resistance

1. Show relay.

2. Show 40 observations of contact resistance of special kind.

Slide 12 - 40 obs. of resistance

3. Interpret.

Discussion of technique

1. Property of distribution of runs above and below median independent of numbers.
   Property of randomness.
   Simplicity of test.
2. Need physico-chemical hypothesis of kinds of causes.
   2.1 Slippage at cleavage planes.
   2.2 Kind of breakdown of film may give rise to excess of runs of size m.
   2.3 Transient phenomena - dist.

ABSTRACT.

›plication of statistical methods in mass production makes possible the
›nt use of raw materials and manufacturing processes, effects economies
ıon, and makes possible the highest economic standards of quality for
ctured goods used by all of us. The story of the application, however,
broader interest. The economic control of quality of manufactured
›rhaps the simplest type of scientific control. Recent studies in this
light on such broad questions as: How far can Man go in controlling his
ıvironment? How does this depend upon the human factor of intelli-
ıow upon the element of chance?

Verification

3.1 Changing contact material

3.2 Changing surroundings or conditions of use.

Background of technique is the neglected theory of runs of different kinds in some of which the median appears to play an important role.

CONCLUSION

Jim Pade III

Malvern Ave

Ruxton, Maryland

Univ. of Maryland