STATI[redacted]AL VIEWPOINT IN ENGINEERING EXPERIMENTATION

by

Walter A. Shewhart
Bell Telephone Laboratories, Inc.

Address to be given before Mathematical Statistics Colloquium, Columbia University, New York City, Feb. 15, 1950.

THE FOUR ACTS OF MAN (S 36307)

1. Act of Rational Abstraction.
   - 1.1 Random variable.
   - 1.2 Stochastic process.
   - 1.3 Repetitive operation - sampling.

2. Act of Measurement.

   A fundamental role of science is to predict the future in terms of the past.

   Repetitive measurement under the same essential conditions.

3. Act of Human valuation.

   The basic "meter" is the human being.

4. Act of control. (S 23040)
   - 4.1 Scientific control.
   - 4.2 Statistical control.
   - 4.3 Two kinds of factors in control. (S 30569)
   - 4.4 Spiral nature of the act of control.

     20472

     A. Between outer and inner circle the stochastic nature of the variables is important as a basis for prediction. Inside the inner circle, we may hope to approach randomness as a basis for prediction. In the outer area, the statistician can hope to improve predictions. In the inner circle, he can hope to improve control.

B. Too often the statistician contents himself with a limited view of the practical problem. For example, the design of experiment in agriculture starts with the numbers written on a bag of fertilizer for N, P, k.

FUNDAMENTAL ROLE OF SCIENCE

1. Fundamental distinction between science and other forms of organized knowledge lies in the concern of science with the possibility of accurate and valid prediction. (S 25438)

2. Fundamental distinction between pure science and engineering science lies in two facts:

   2.1 The predictions of the engineer are almost certain to be tested. (S 16061)

   2.2 If the predictions are invalid, the engineer may "lose his shirt".

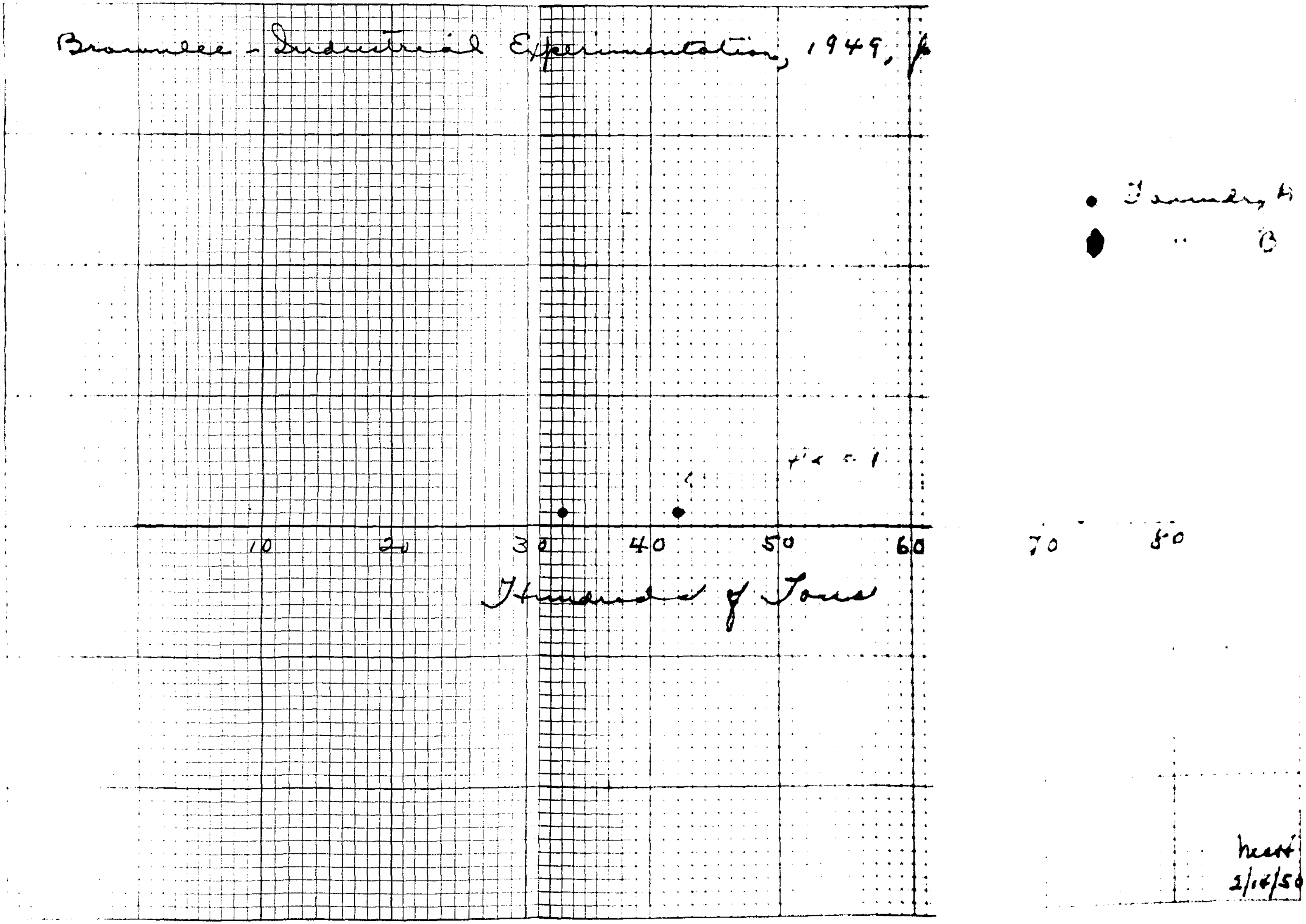
   For example, the engineer's job is to devise operations of using raw materials at his disposal that, if carried out, will give a thing useful in satisfying some human want.

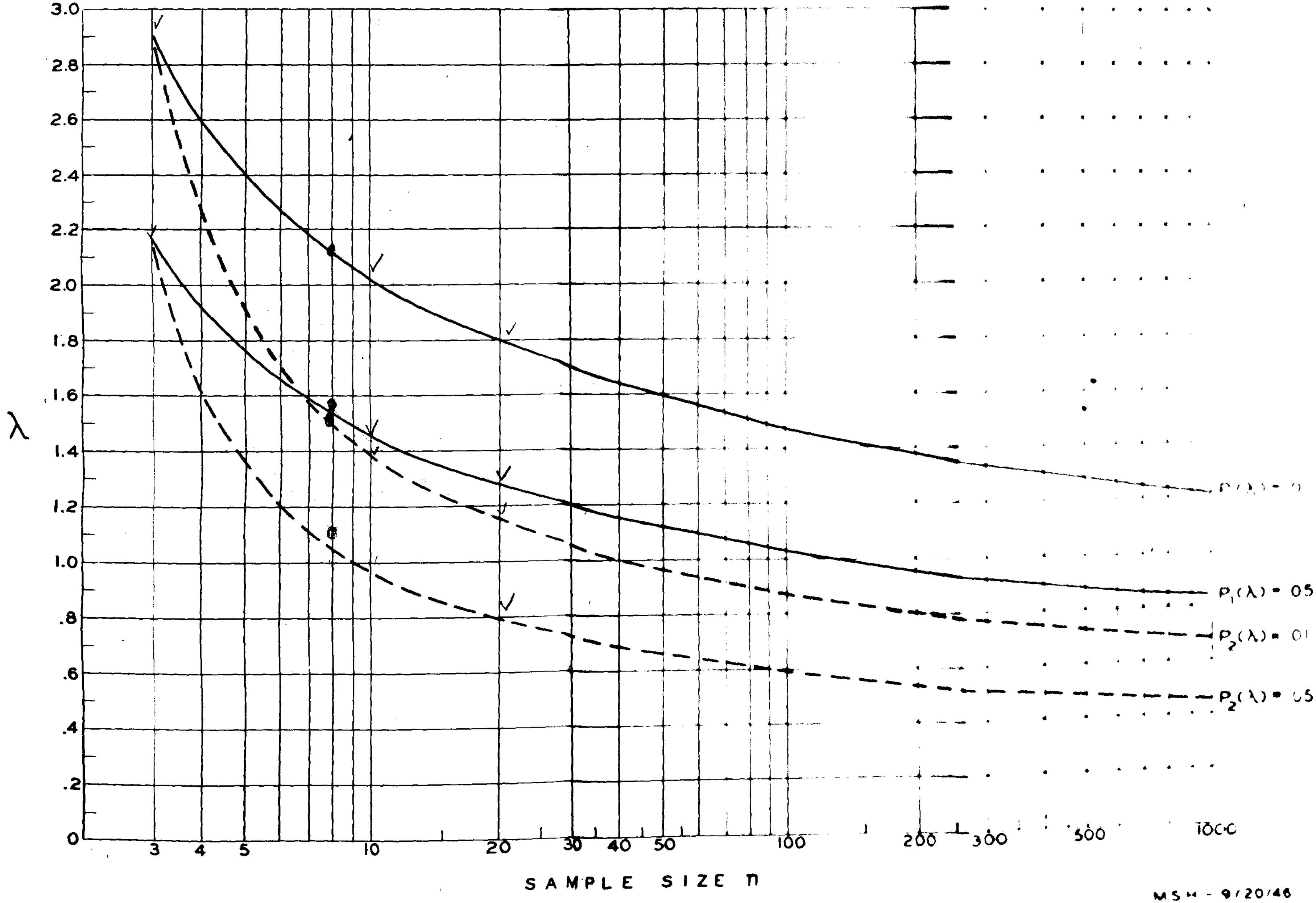
3. Three accepted limitations of applied scientific prediction:

   3.1 It is not possible to know at any given time all of the factors that will prove relevant to human wants at some specified future time.

   3.2 All predictions can never be more than probable.

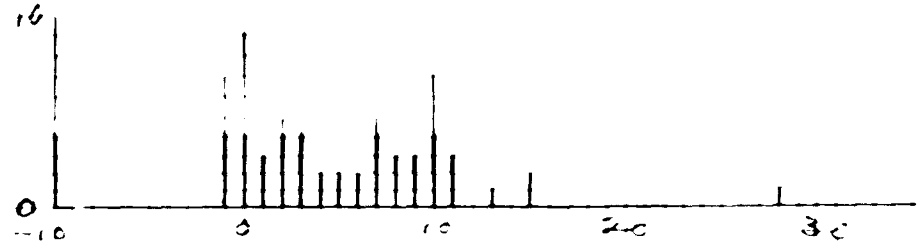
..38 . 10 x 10 o i c
MADE IN U. S. A.

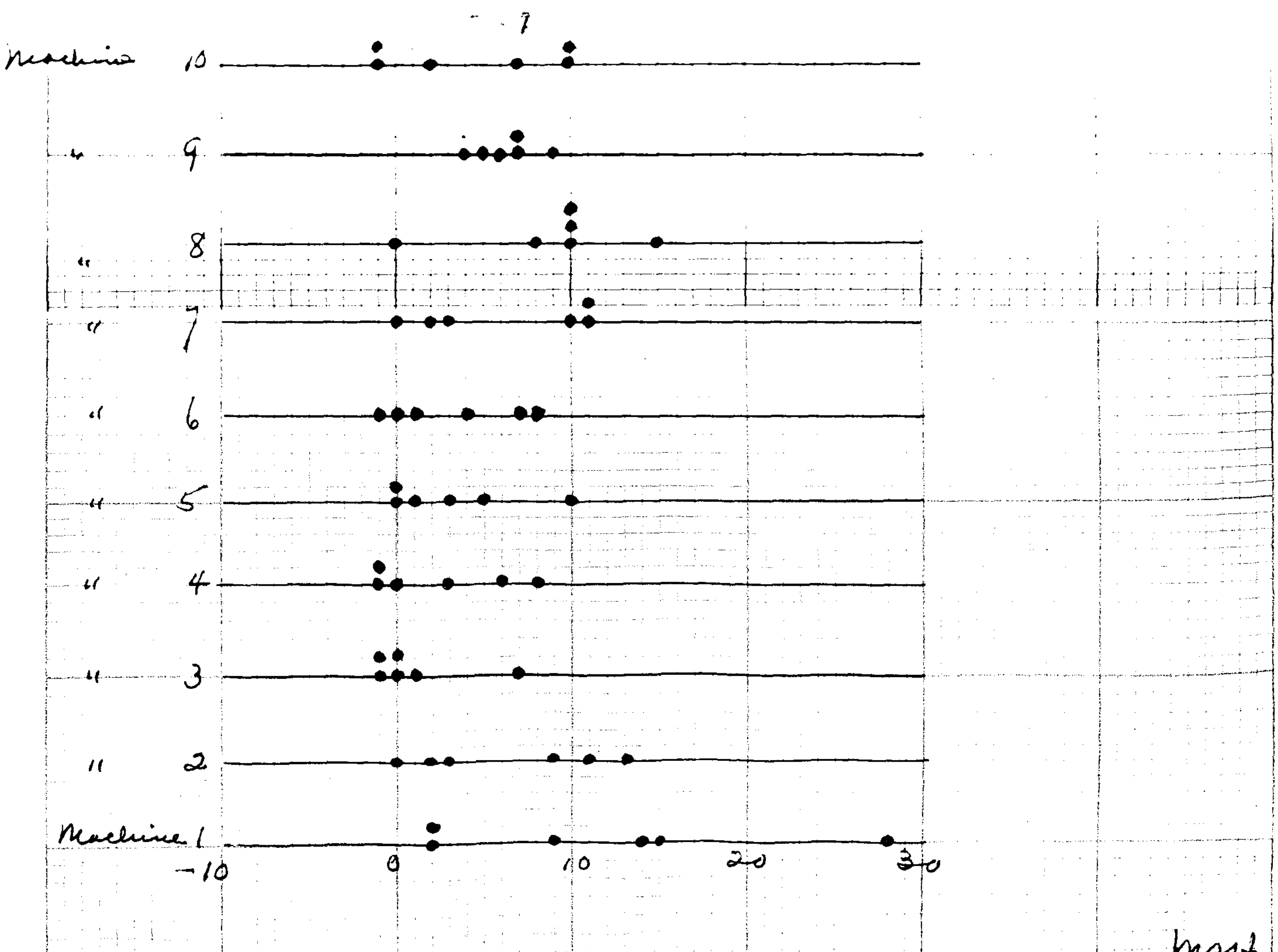




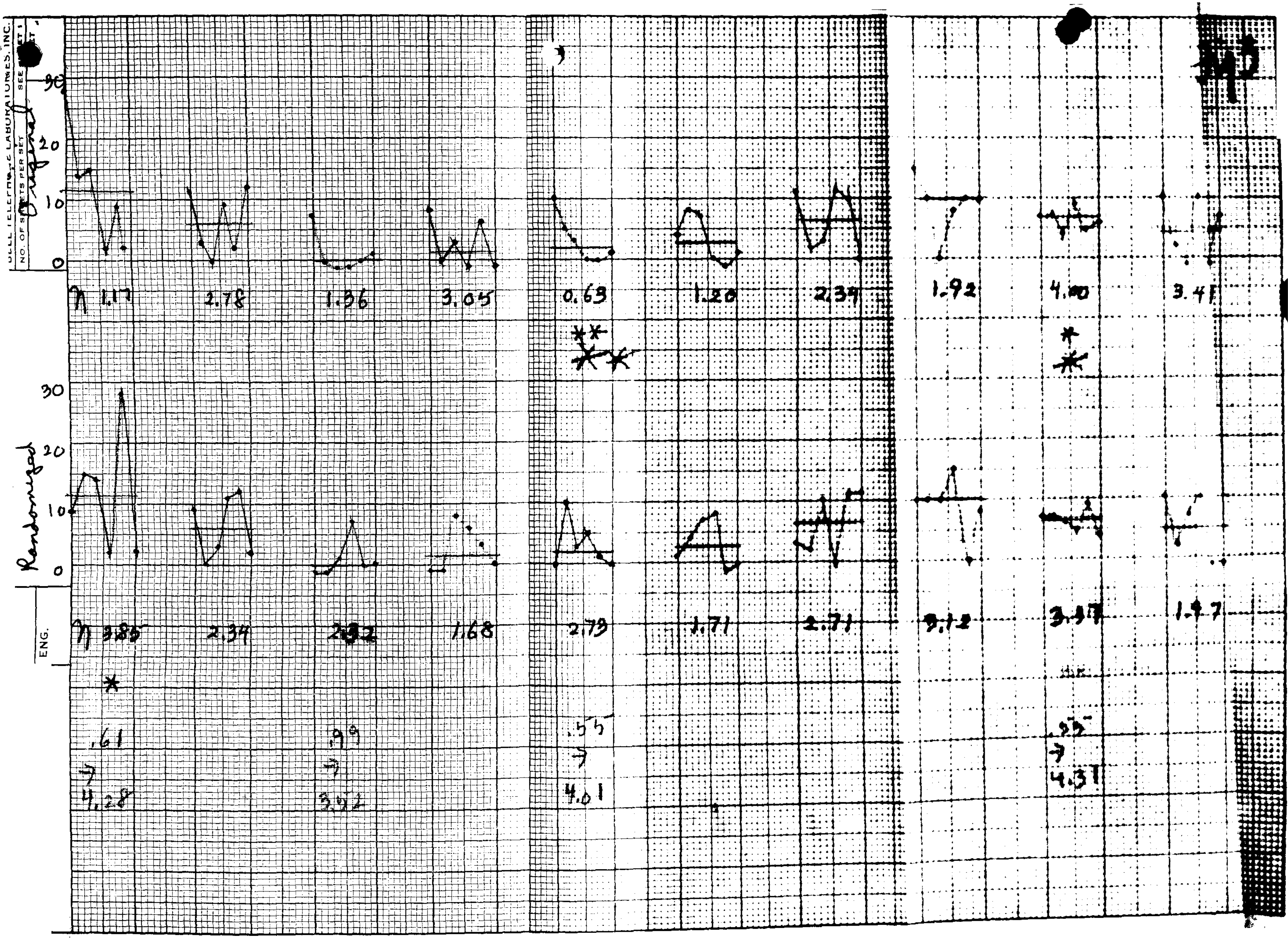
λ
3.0
2.8
2.6
2.4
2.2
2.0
1.8
1.6
1.4
1.2
1.0
.8
.6
.4
.2
0
3
4
5
10
20
30
40
50
100
200
300
500
1000
SAMPLE SIZE n
P1(λ) = 05
P2(λ) = 01
P2(λ) = 05

MSH - 9/20/48

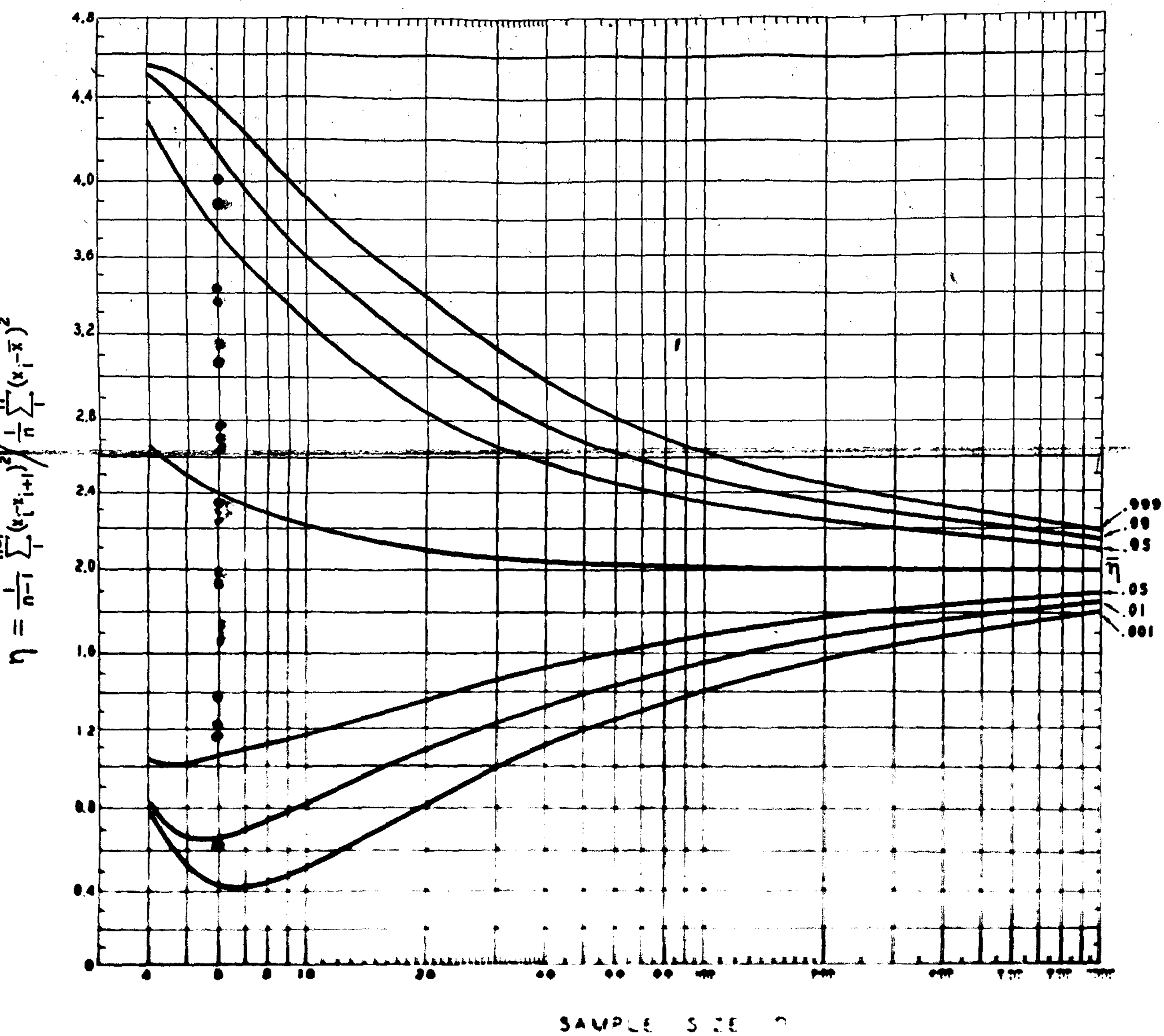




hmf
2/14/50


BELL TELEPHONE LABORATORIES, INC.
Original
30
20
10
0
η 1.17
2.78
1.36
3.05
0.63
1.20
2.39
1.92
3.41
Randomized
30
20
10
0
2.34
1.68
2.79
1.71
1.47
ENG.
.61
→
4.28
.99
→
3.92
.55
→
4.01
→
4.37

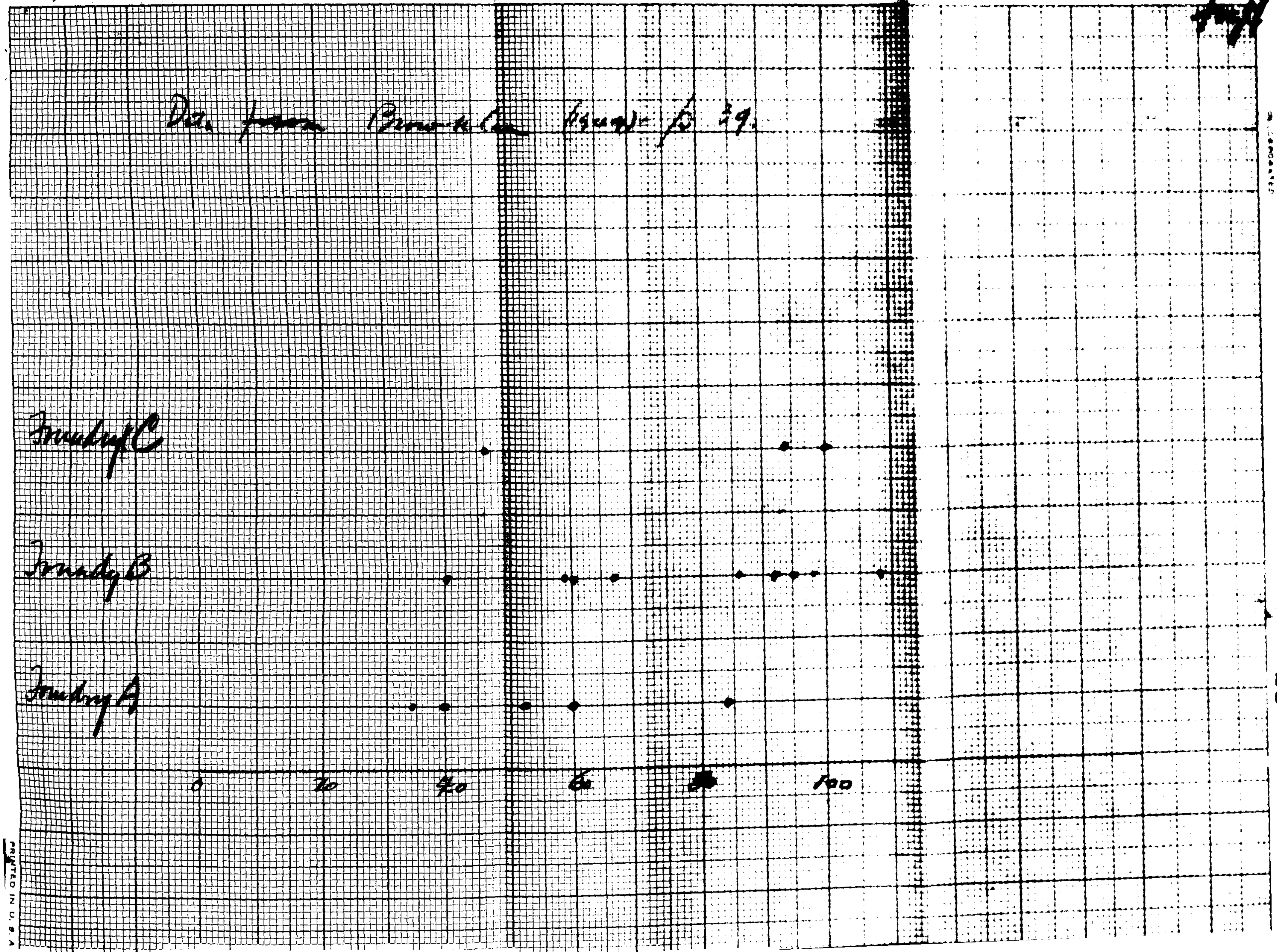
Fig



PROBABILITY LIMITS FOR η

- Original sequence
- Randomized sequences.

7/10/50


Foundry C
Foundry B
Foundry A
0
20
40
60
100

3.3 It is not possible to conduct an experiment in which the operation of measurement does not in some way influence the phenomena to be observed.

This is particularly true in the measurement of preferences. (S 36019)

STATISTICAL EXPERIMENTATION

R. A. Fisher - Statistics for Research Workers
Design of Experiment.

K. A. Brownlee - Industrial Experimentation, 1949.

O. L. Davies - Statistical Methods in Research and Experimentation.

F. Yates - Design and Analysis of Factorial Experiments.

D. Finney - Probit Analysis.

F. Yates - Sampling Methods for Censuses and Surveys.

1. These books usually start with a discussion of the statistical significance of the differences between two observed means or two or more observed variances.

Example 1 - Brownlee, 1949, p. 35.

Foundry A: 71, 67, 33, 79, 42.
Foundry B: 73, 80.

where figures are outputs in hundreds of tons.

Observed t is significant at only 25% level.

Brownlee concludes: "... there is no evidence that here the populations are different". See Fig. 1.

Here there is no consideration of the fact that the application of the test may not be valid, viz.,

1.1 No prior evidence is given to indicate that A and B are random variables and the gap in A would lead me to doubt it.

1.2 Important from control viewpoint to note evidence of lack of homogeneity of A.

Example 2 - Brownlee, 1949, p. 38.

Compares observed variances in output for 10 machines by Bartlett test. Concludes that there is no evidence.

Here again no attention is paid to evidence of lack of homogeneity indicated in Figs. 2 and 3.

Example 3 - Brownlee, 1949, p. 39.

Compares variances of three foundrys and concludes that ~~the~~ evidence of statistical significance.

However, evidence presented in Fig. 4 is of practical significance.

In general, such books consider prediction in terms of confidence limits, assuming randomness,

$$
\begin{array}{cccccc|ccc}
X_1, & X_2, & \ldots & X_i, & \ldots & X_n & X_{n+1}, & X_{n+2}, & \ldots \\
| & | & & | & & | & | & | & \\
C_1 & C_2 & & C_i & & C_n & C_{n+1} & C_{n+2}, &
\end{array}
$$

Past Present Future

but do not examine the past for evidence of lack of randomness.

They assume $C_i \eqcirc C_j$.

ROLE OF STATISTICS

1. Statistical theory provides a rational and valid method of prediction if, and only if, model (or hypothesis) is valid.

   Empirical experience can only reveal when to rely upon predictions. Statistical theory can tell how much time will be wasted looking for assignable causes of a given type when not present.

   Example 1 - Small probability not a criterion for practical significance.

We can make shaded area as small as we please and the occurrence of an observation in this interval is statistically significant but not practically significant.

Example 2

Fig. 6 shows recent work of Bureau of Standards on triplets. Discuss.

Example 3

The occurrence of a sample

$$X_1 < X_2 < X_3 < X_4 < X_5 < \dots X_7$$

is statistically no more significant than any other order, but it is certainly of practical significance.

THE REPETITIVE ACT: CRITERIA FOR PREDICTION ON ASSUMPTION OF RANDOMNESS

1. Engineering Approach

   Look for evidence of:

   1.1 Discontinuous jumps produced by different conditions.

   1.2 Trends or pseudo cycles.

   S 25407

   S 25405

2. The gap test
3. The eta chart S 36394
4. Runs above and below percentile.
5. Serial correlation.
6. Accuracy-Precision chart.

EXAMPLES

1. 18 values of g.

S 28184

S 30527

S 30530

2. 144 observed values of thickness of inlay on relay springs.

Data	S 21612
Distribution	S 30568
Runs	S 30528
Eta	S 28088
Eta (random)	S 28085

3. Birge data

Data	S 31019
Runs	S 30669
Eta	S 38086
Eta (random)	S 30535

4. Chemical measurements

B.L.Clarke	S 30981	30983	30982
Iron alloy	S 28594		
Iron alloy (eta)	S 30584		
	S 28592		

SAMPLING

Nothing is said in Brownlee and elsewhere about sampling alloys and materials for chemical analysis.

No allowance made for lack of accuracy and precision.

S 30666

S 30699

NEED FOR COOPERATION BETWEEN INDUSTRIAL STATISTICIANS AND ACADEMIC STATISTICIANS

1. Concepts coming out of statistical control.

   Errors of first and second kind.
   AOQL
   Control limits.
   Empirical criteria for control.
   - Control chart
   - Eta chart
   - Accuracy-precision chart.
   - Runs

2. Future applications in preference studies, information theory, specification, etc. Needs cooperation of many different groups.

3. Joint committee
   NRC Committee
   ISI committee
   NRC-SSRC committee S 36014

4. Also need seminars in which practical and academic men cooperate.

5. Books in which practical and academic men cooperate.

   NRC on chemical statistics.