O T T A W A   February 1st, 1943.

REPORT

of the

ORE DRESSING AND METALLURGICAL LABORATORIES.

Investigation No. 1341.

Interpreting Observations.

(Copy No. 12.)

TABLE OF CONTENTS

Abstract.

In the inspection of war materials great numbers of inspection tests are performed. Much useful information can be obtained by analysing large numbers of tests. This phase of industrial research has enabled inspection costs and production costs to be lowered and quality improvements to be made.

Unfortunately, many inspectors and manufacturers still do not know that statistical methods are available for the study of test results. The purpose of this article is to draw attention to the value of statistical methods.

Many contributions to a more efficient war effort can be made by applying the knowledge obtained from rational interpretation of large numbers of test results.

The methods described have been used successfully for many years. Examples of common problems in interpretation of industrial data are given.

A bibliography is attached.

BUREAU OF MINES
DIVISION OF METALLIC MINERALS

ORE DRESSING AND
METALLURGICAL LABORATORIES

CANADA
DEPARTMENT
OF
MINES AND RESOURCES
MINES AND GEOLOGY BRANCH

OTTAWA February 1st, 1943.

REPORT

of the

ORE DRESSING AND METALLURGICAL LABORATORIES.

Investigation No. 1341.

Interpreting Observations.

Introduction.

"No empirical knowledge is ever certain. From the cradle to the grave one must of necessity act on knowledge which is probable only."

– Lt.-Col. L. E. Simon, in "Engineer's Manual of Statistical Methods."

The production and inspection of the materials of war involves thousands of observations. Logical action is generally based on the interpretation of many observations. The success of such action depends upon the accuracy of the observations and the soundness of the interpretation placed upon them.

There are countless instances where the observer "sees but perceives not"; that is, the significance of a group of observations is lost to him, due either to faulty interpretation or to no interpretation at all being placed on the results. This condition is being partly corrected in the United Kingdom and the United States[*] by the promotion of certain principles by Ordnance officials. The term, "Quality Control", has been adopted to describe a system of studying and controlling industrial products by methods of handling observations.

In these pages an attempt is made to outline in a non-technical manner some of the ideas used in the handling of observations. Of course the authenticity of the observations determines the worth of any decisions based on them. No matter how brilliant is the theory built on unreliable sampling it is usually of little practical value.

In some fields it is almost impossible to obtain observations free from bias. Let us suppose that observers of speed in miles per hour were available for judgment: (a) a bystander, (b) a motorcycle policeman, and (c) a motorist accused of speeding. The bystander might be unbiased but probably would be very inaccurate. The motorist is usually biased on the low side. The policeman is the only qualified observer and, if overzealous, he might be biased on the high side. Conclusions drawn from large numbers of biased observations naturally would be of little value.

The collection and study of a great number of observations will often bring out invaluable information

---

* QUALITY CONTROL OF MUNITIONS - G. D. Edwards, War Department, Washington, D.C. "Army Ordnance," 1942.

that one person could not discover for himself in a whole lifetime. Early navigators of the globe could never know what winds to expect. By travelling one route through all seasons of the year, the captain would eventually get to know the prevailing wind for each season for that route only. It was an invalided British Navy man who, for a hobby, requested that vessels sailing to all corners of the globe report wind and weather conditions. From over six hundred log books he compiled a map of the trade winds on all the oceans. Thenceforward a captain could sail a strange course with some certainty of the winds which would be encountered.

The same principle of using large numbers of observations is as applicable to industrial conditions today as it was to wind conditions two centuries ago. Those with vision find an orderly pattern of relationship where others see a confusing welter of a thousand separate facts. As early as 1924, K. H. Daeves (*) stated: "Statistical research is a logical method for the control of operations for the research engineer, the plant superintendent, and the production executive."

---

* "The Utilization of Statistics," in TESTING, March 1924.

100 Per Cent Inspection Without Interpretation May Not Be Satisfactory.

On destructive tests for projectiles, armour, etc., from the test an estimate of the untested material must be made. If this interpretation is to be done scientifically statistical method should be used. It is frequently stated that where 100 per cent inspection is used no statistical method is needed, since 100 per cent assurance is obtained that no defectives occur. 100 per cent GO-NOGO type inspection does not predict the onset of defective material, as does the Quality Control system. The opinion of the Ordnance Division of the U.S. Army[@] on this subject is as follows:

> "But even where the necessary inspections are not destructive, 'inspection fatigue' steps in to prevent one hundred per cent inspections from providing one hundred per cent insurance of conformance to specification requirements. If you have before you a hand truck containing 15,000 cartridges, and you are given the job of inspecting and gauging them visually one hundred per cent, they probably will all look alike to you after you have examined about 9,000 of them, and you won't know whether the discoloration which evidences necessary shoulder anneal, for example, is there on the 9,001st cartridge or not. This is no insult to your intelligence; it is just a plain illustration of experience.
>
> "So 100 or 200 or even 500 per cent manual inspections are not the answer where large quantities of material are involved, even if the resulting production delays could be tolerated. Mechanical gauging and photoelectric-cell gauging are being used in the inspection of ordnance material wherever possible to circumvent inspection fatigue, but even the best of these substitutes have their own margins of error. In other words, it must be recognized that the element of risk just can't be eliminated from quality considerations in mass production, and the real problem is how to reduce the chances which must be taken to a minimum without unduly impeding output. Quality control techniques are built around limiting such risks to a predetermined degree, and they are thus admirably adapted to the problem in hand."

(Continued on next page)

@ "Quality Control of Munitions" - (G. D. Edwards).

(100 Per Cent Inspection Without Interpretation May Not Be Satisfactory, cont'd) -

On February 12th, 1943, inspection of materials was discussed at the annual meeting of the Engineering Institute of Canada, in Toronto. Mr. T. W. Vroom, chief inspector for the Northern Electric Company of Canada, stated that after many investigations, extending over at least five years, they had determined that on 100 per cent inspection inspectors would pick out about 85 per cent of the defective material. This referred to experienced inspectors.

The experience of the Northern Electric Company is borne out by statements made by the inspectors in the Westinghouse Manufacturing Company and in the Picatinny Arsenal, U.S.A.

There is a great amount of evidence, therefore, to show that 100 per cent inspection does not give 100 per cent assurance that defective work is all detected. Statistical analysis of a number of tests can probably predict the occurrence of defective work more accurately than can be done by ordinary "100 per cent" inspection.

---

1. The Significance of Two Observations Differing in Magnitude.

Let us suppose that an ordnance inspector is examining test results which represent two lots of castings. One test bar is recorded at 110,000 p.s.i. yield strength and the other at 126,000 p.s.i. yield strength. What does this mean? Should the manufacturer be asked to take corrective action? Should the work be rejected? The following is a demonstration of how such an occurrence should be interpreted.

Without a background of experience, interpretation of observations is impossible. By experience we mean a collection of facts arranged in orderly manner so that some pattern of behaviour is evident. The facts may be retained mentally, or they may be recorded.

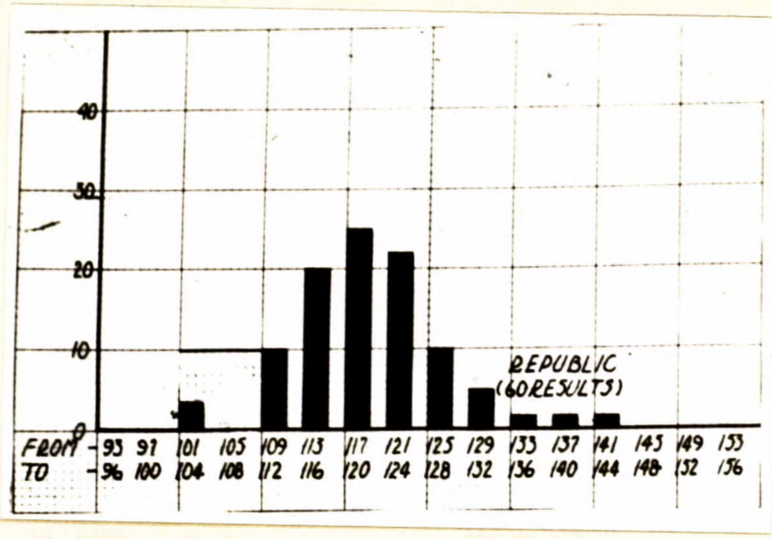
The inspector should first acquaint himself with the normal behaviour of the observation for the source being studied. The frequency distribution chart is a convenient graphic method of showing how observations have occurred. Examining Figure 1, we see that yield values in the past have occurred around a central value of 117,000 to 120,000 p.s.i. and over a range of 100,000 to 144,000 p.s.i. If the process remains unchanged, then what has happened before will be expected to happen again, that is - (from Figure 4)-

Approximately 67 per cent of all results will fall within 113,000 to 124,000 p.s.i.

Approximately 87 per cent of all results will fall within 109,000 to 128,000 p.s.i.

(Continued on next page)

Figure 1.



PAST EXPERIENCE WITH YIELD STRENGTH OBSERVATIONS
SHOWN GRAPHICALLY BY A FREQUENCY DISTRIBUTION
(Unit - 1,000 p.s.i.)

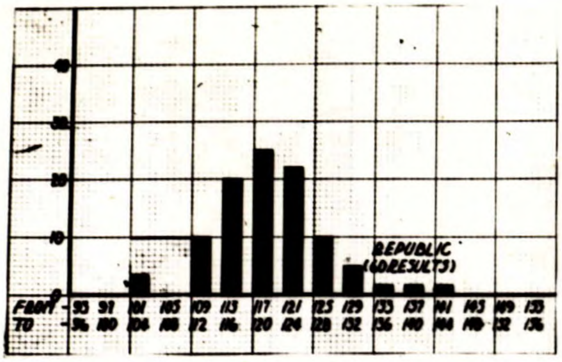
It only common sense, therefore, to say that proof of departure from normal operation requires that an observation well outside the above limits be encountered. The frequency distribution (Figure 1) provides a background of experience to which the new observation can be be compared.

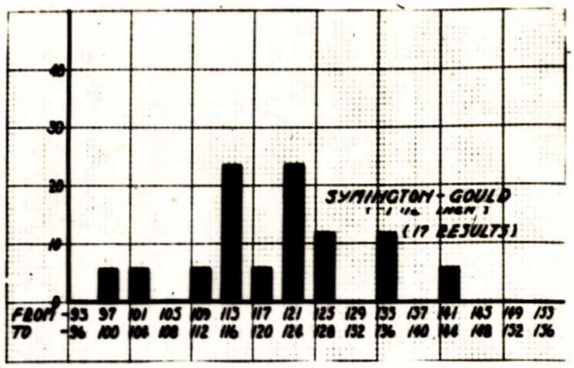
Judgment of an observation must be either that there is no indication of a change in the process or that the process has changed. If the process has not changed, there is a question of whether the process itself is acceptable. This can best be answered by comparing frequency distributions of the observation taken in different industrial establishments known to be producing a satisfactory product.

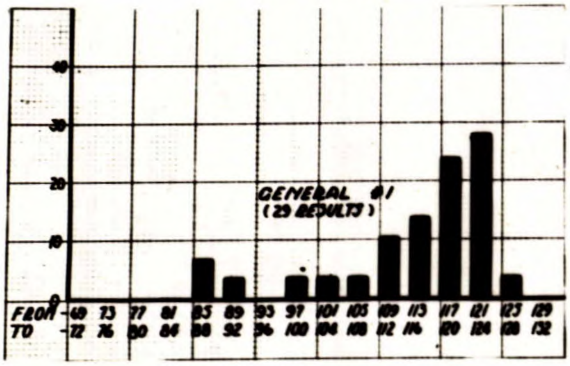
A set of distributions for Yield Point is shown on the following pages. Note that each source has a characteristically different distribution of yield strength observations. Some results were spread out because experimental results were included with production results. These charts show enough background in yield tests to rationally interpret the significznce of a difference of 16,000 p.s.i. between two specimens.

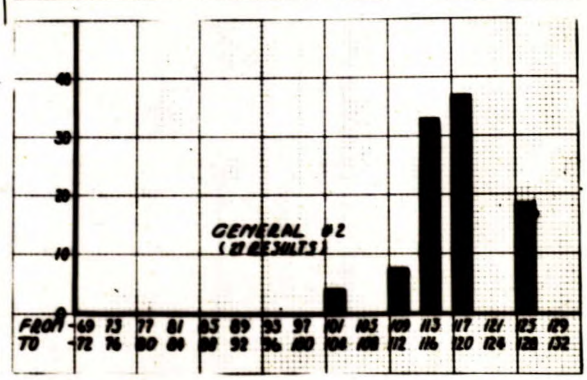
(Pages 6 and 7 are photostats)

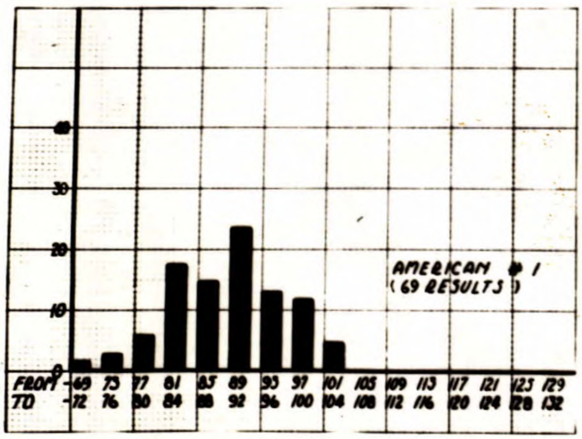
YIELD CHARACTERISTICS

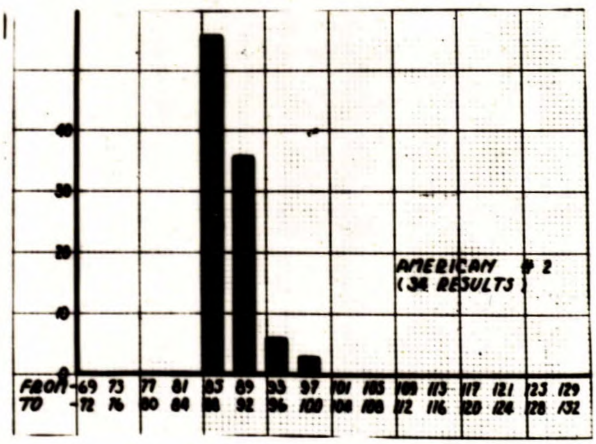


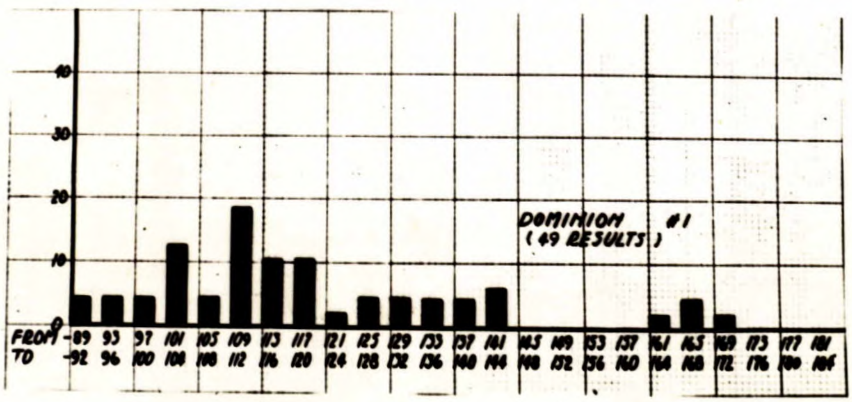




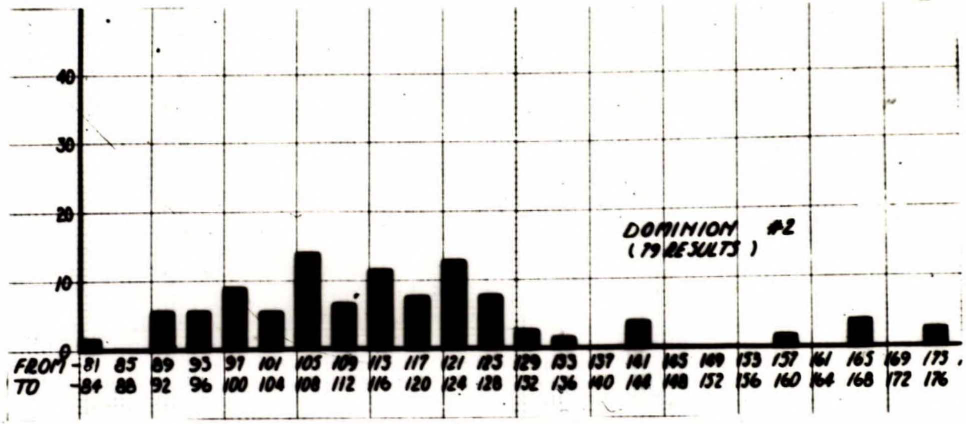


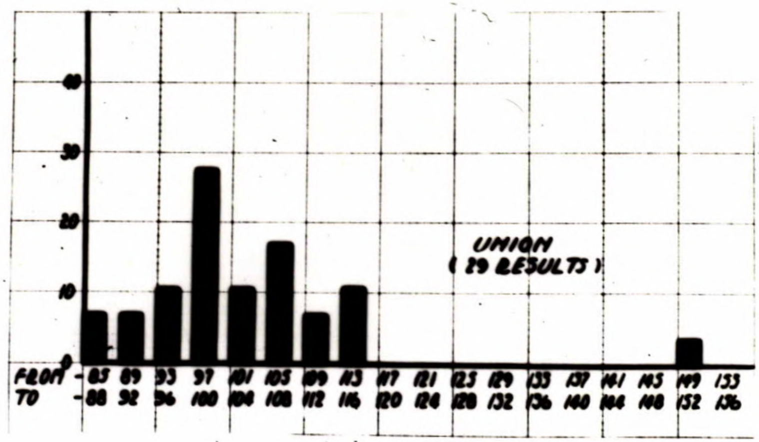


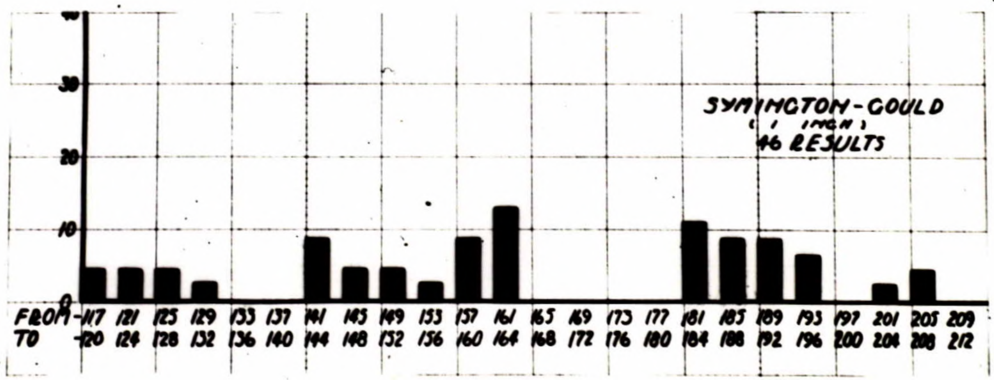




YIELD CONTINUED







2. Conclusions Drawn From A Sample.

A sample is a small part or quantity of anything intended to be used as evidence of the quality of the whole. A common mistake, made by many inspectors, is the assumption that the material is exactly like the sample. This fallacy has been exposed very definitely by L. E. Simon.[*]

Common sense would indicate that the larger the sample taken the more sure is the estimate of quality. Simon proves that if a lot of material were 10 per cent defective and samples of ten were taken at random, then

35 per cent of the samples would contain no defectives,
39 per cent of the samples would contain one defective, and
26 per cent of the samples would contain more than one defective.

Obviously, the material cannot then be exactly like the sample in 60 per cent of the cases.

Q. How, then, can a sample be interpreted if there is no certainty that the sample is like the material?

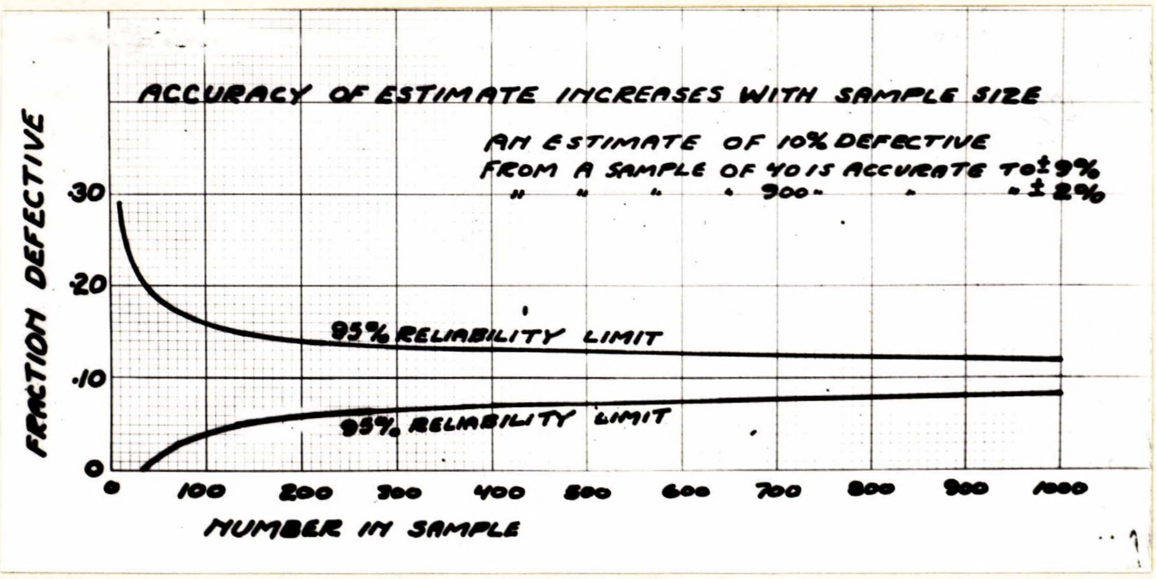
A. From the value obtained from a sample, the range within which the true value lies may be estimated. A value obtained from a sample is merely an estimate, the accuracy of which depends upon the number in the sample.

Page 9 shows how accuracy of a sample found to be 10 per cent defective increases with the size of the sample.

(Page 9 is a large chart)
(Text continued on page 10)

(*) ENGINEER'S MANUAL OF STATISTICAL METHODS, by Lt.-Col. L. E. Simon, Ordnance Department U.S. Army, Assistant Director, The Ballistic Research Laboratory. Pub. by John Wiley & Sons, New York (1941).

Figure 4.


ACCURACY OF ESTIMATE INCREASES WITH SAMPLE SIZE
AN ESTIMATE OF 10% DEFECTIVE
FROM A SAMPLE OF 40 IS ACCURATE TO ± 9%
" " " " 900 " " " ± 2%
FRACTION DEFECTIVE
·30
·20
·10
0
95% RELIABILITY LIMIT
95% RELIABILITY LIMIT
0
100
200
300
400
500
600
700
800
900
1000
NUMBER IN SAMPLE

The accuracy of averages, percentages and deviations is also subject to the same effect of number in sample. Any report of properties of a lot of material should be qualified by stating the accuracy of the estimate.

On page 9 it can be seen how many observations must be made to obtain an accuracy of ±2 per cent. Statistical organizations questioning individuals must obtain about one thousand observations in order to state the attitude of the whole population within 2 per cent (assuming perfect sampling method). The following example is offered in order to illustrate the calculations involved in assessing a sample:

Example: An estimate of izod impact strength of a steel from Company A is required. Ten values for successive heats were reported as follows:

Izod Impact Strength, in Ft. Lb.
(Number of observations = 10)

69	
61	
63	
73	
54	Average is 61.3
54	
56	
58	
74	
51	
613	

The average value in the sample is 61.3 foot pounds. Values recorded in the sample range from a minimum of 61 foot pounds to a maximum of 74 foot pounds. What is required is an idea of where the average izod for the process lies and the limits within which individual values are expected to fall. This requires that the standard deviation of the results be calculated.

(Continued on next page)

Standard Deviation (Long Method*) -

The term "sigma" and the symbol σ are often used as abbreviations for standard deviation.

$$\sigma = \sqrt{\frac{\Sigma(\bar{X} - X)^2}{N}}$$

Observation X	Deviation $\bar{X} - X$	Deviations[2] $(\bar{X} - X)^2$
69	+ 7.7	59.29
61	- .3	.09
63	+ 1.7	2.89
73	11.7	136.89
54	- 7.3	53.29
54	- 7.3	53.29
56	- 5.3	28.09
58	- 3.3	10.89
74	12.7	161.29
51	-10.3	106.09
613	0	612.10

X = An izod observation.

$\bar{X}$ = Average izod = 61.3.

N = Number of observations = 10.

Σ = The sum of all.

$$\Sigma(\bar{X} - X)^2 = 612.10$$

$$\Sigma\frac{(\bar{X} - X)^2}{N} = 61.21$$

$$\sigma = \sqrt{\frac{\Sigma(\bar{X} - X)^2}{N}} = \sqrt{61.21}$$

$$\sigma = 7.8 \text{ (approx.)}$$

The average of the sample is 61.3 foot pounds and the standard deviation is 7.8 foot pounds.

Reliability of Average -

The sample serves as an estimate of the true average of the process. The reliability of the sample average depends upon the standard deviation of the sample and the number of observations in the sample. It is calculated as follows:

$$\left.\begin{array}{r}\text{Standard}\\\text{error of}\\\text{average}\end{array}\right\} = \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{N}} = \frac{7.8}{\sqrt{10}} = 2.5 \text{ foot pounds.}$$

The standard error of the average, therefore, is 2.5 foot pounds.

(Continued on next page)

* For a shorter method, see Page 18.

The reliability of the average, 61.3 foot pounds, may now be stated as follows:

*Odds are 68 out of 100 that true average lies between 61.3 ± 2.5
Odds are 95 out of 100 that true average lies between 61.3 ± 2 times 2.5.
Odds are 99.7 out of 100 that true average lies between 61.3 ± 3 times 2.5

The average of the process, therefore, lies somewhere between 53 and 69 foot pounds.

From the information given in the sample, individual results may be expected to fall within average ± 3 sigma.

61.3 ± 3 times 7.8,

38 to 85 foot pounds.

The accuracy of the standard deviation is expressed in the following:

Standard error of sigma = sigma divided by the square root of twice the number in the sample. In this case this equals

$$\frac{7.8}{\sqrt{20}} = 1.74 \text{ foot pounds.}$$

The standard deviations of samples of ten should therefore fall within:

*7.8 ± 1.74 68 per cent of the time,
7.8 ± 2 times 1.74 95 per cent of the time, and
7.8 ± 3 times 1.74 97.8 per cent of the time.

Thus, from a sample of ten observations, general conclusions as to the nature of the behaviour of observations can be made. The accuracy of any statement is qualified by probabilities that will occur within a definite range. As the number of observations increases the accuracy of any estimate

---

* See A.S.T.M. Manual on Presentation of Data, p. 23.

becomes greater.

The only calculation required for this type of work is that of standard deviation. For precise work the standard deviation of a sample should be corrected for sample size. This has been omitted here in order to prevent complicating calculations. For simple methods of handling data it is not necessary to make this correction. Students of statistical methods will follow this question in standard texts.

---

3. The Significance of Difference between Samples.

Often it is necessary to compare two samples to determine whether they are from the same lot or from different lots, or an inspector wishes to determine by sample whether or not the process has changed. Many mistakes are made in comparing samples. Unless there is a sufficient background of experience, or statistical methods are used, faulty conclusions may be drawn. The significance of the difference between samples depends upon the reliability of the values determined from the samples.

For a full discussion of this problem the reader should refer to "Applied General Statistics" (by Croxton and Cowden). The following example is given here to show a practical problem.

Example:

Let us assume that a comparison of two types of tank track pins is made. 168 pins of type "A" and a like number of type "B" are placed in the tracks of a tank so that they will be subject to the same conditions. After a standard proving ground test has been carried out, it is found that five "A" and ten "B" pins have broken.

If the significance of the above difference was left to unaided human judgment, there would be a variety of opinions. Some would say that there was no difference; others would say that "A" pins were definitely superior. Errors in this type of judgment occur so frequently that it is considered worthwhile to explain the method of handling this problem.

Given Data -

"A" PINS	"B" PINS
5 failed.	10 failed.
168 tested,	168 tested.
2.975 per cent defective.	5.95 per cent defective.
.02975 fraction defective.	.0595 fraction defective.

Reliability of Given Data -

Standard error of fraction defective = σ_p
Fraction defective = p
Fraction effective = q
Number tested = N

$$\sigma_p = \sqrt{\frac{pq}{N}}$$

For "A" pins, this becomes $\sigma_p = \sqrt{\frac{.0298 \times .9702}{168}}$

$$\sigma_p = .013.$$

Now the reliability of the results of the test on "A" pins can be stated as follows:

There are odds of 95 out of 100[*] that the true fraction of "A" pins which are defective lies between

$$p \pm 2\sigma_p$$

$$.0298 \pm .026$$

$$.0036 \text{ and } .0456$$

Expressed in terms of pins failing out of 168, the reliability of the test is such that from 0 to 8 failures may

[*] From the normal curve of distribution.

normally be expected.

Analyzing the data on the "B" pins, the reliability of this test is such that from 4 to 16 failures may normally be expected.

It is obvious that no estimate can be made without qualifying the probable accuracy of the estimate. Now, if two estimates are so close together that their accuracy limits overlap, the significance of the difference may be small. Without statistical method the observer may draw erroneous conclusions. The method generally used for determining significance rationally follows:

$$\sigma_{P_A - P_B} = \sqrt{\sigma_{P_A}^2 + \sigma_{P_B}^2} = \sqrt{.013^2 + .0183^2}$$

$\sigma_{P_A - P_B}$ = Standard error of difference between fractions.

σ_{P_A} = Standard error of fraction of "A" = .013.

σ_{P_B} = Standard error of fraction of "B" = .0183.

P_A = Fraction defective of "A" = .0298.

P_B = Fraction defective of "B" = .0595.

t is the symbol used to represent significance.

$$t = \frac{P_A - P_B}{\sqrt{\sigma_{P_A}^2 + \sigma_{P_B}^2}} = \frac{.02975 - .0595}{\sqrt{.013^2 + .0183^2}} = -1.32$$

PRACTICALLY SPEAKING, IF t IS LESS THAN 2.0 THE DIFFERENCE IN FRACTIONS OBSERVED IS NOT SIGNIFICANT. "A" PINS HAVE NOT PROVEN DEFINITELY SUPERIOR TO "B" PINS.

If t is less than 2.0 the difference is considered to be of little significance. t refers to distances on the curve of normal error, measured in terms of standard deviation. Thus: "From Standard Mathematical Tables,"* fraction of

---

* HANDBOOK OF CHEMISTRY & PHYSICS, Chemical Rubber Pub. Co.

total area under the curve from t = 0 to t = 1.32 is .4066

Significance then equals 2 x .4066 = .8132

or

approx. .80

This means that the odds are

8 out of 10 that "A" pins are the same as "B" pins.
2 " " 10 " "A" pins are different than "B" pins.

There is only a slight chance therefore that "A" pins are different than "B" pins.

When is action needed?

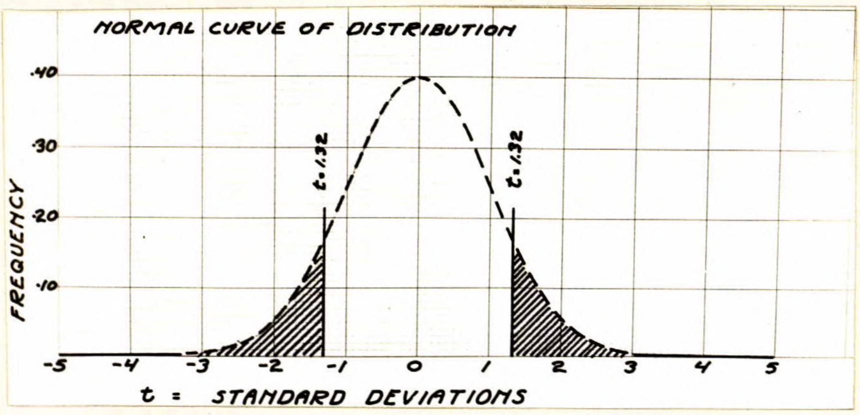
Mathematical methods will determine odds that a venture will be successful. These odds are based on test data which give only a small part of the general overall conditions. Therefore, judgment of intangible conditions supplements the mathematical odds.

In the previous example, the odds are only 1 out of 5 that the "A" pins are superior to "B" pins. Let us suppose that there is reason to believe that the two tests differed and that "A" pins received a more severe test. Then, of course, one would be justified in disregarding the odds calculated from a number of failures.

Thus, in evaluating conditions, statistical methods serve to analyse the quantitative data. The intangible, or qualitative, data on the subject under consideration may outweigh the quantitative date. For example, a heat of steel may have an analysis which has proven satisfactory for a certain grade of steel castings. From the chemical analysis can we predict that castings will be satisfactory? Suppose the foundryman notices that the steel is poured into a wet mold, then he can predict from this practical information that the casting will be defective.

THERE IS NO SUBSTITUTE FOR EXPERIENCE OF PAST CONDITIONS AND KNOWLEDGE OF CURRENT CONDITIONS.

Figure 5.



NORMAL CURVE SHOWING POSITION OF t = 1.32.

Figure 5 shows the normal curve of error of differences. If "A" and "B" pins are actually the same then differences greater than t = 1.32 will normally occur with the frequency shown in the shaded areas in Figure 5. The area of the shaded areas is .1868 of the total area under the curve (from any mathematical handbook of tables). This means that 19 per cent of the time differences greater than the observed would occur due to chance.

Differences are usually considered to be definitely significant when the possibility that they are due to chance is reduced to odds of 5 out of 100 or less. Any risk value can be chosen, however.

The ideas involved when interpreting single observations and samples of observations have been incorporated into the quality control chart method, which relieves the engineer of a great amount of calculation. The section following shows an example of the quality control chart method interpreting the flow of production test results.

---

4. Significance of Variation in Observations During Production.

Inspectors of ordnance are greatly concerned with the fluctuation of observations of ballistic limit, the velocity of projectiles, detonation time of fuzes, dimensions of metal parts, strength of metal, etc., etc. The quality control chart method is being used advantageously to study variations in many kinds of observations.

Standard Deviation, or Sigma -

A brief description of sigma, or standard deviation, is needed in order to explain the quality control system. On Page 11 the standard deviation of 10 izod observations was derived. A simpler method is used on the following group of elongation observations.

(Continued on next page)

STANDARD DEVIATION (SHORT METHOD*)

	X		X^2
	20.5		420.25
	20.5		420.25
	22.0		484
	20.0		400
	23.0		529
	20.5		420.25
	21.0		441
	20.0		400
	20.0		400
	20.0		400
AV'GE	20.75	AV'GE	431.475

N = Number of elongation values.
X = Individual values for elongation.
Σ = The sum of
σ = Standard deviation.

$$\sigma = \sqrt{\frac{\Sigma X^2}{N} - \left(\frac{\Sigma X}{N}\right)^2} = .9$$

$\bar{X}$ = Average elongation.

When the average and standard deviation of a group of test results are known, the following is true if the process is under control:

68 per cent of the results will be within Average ± 1 sigma,
95 per cent of the results will be within Average ± 2 sigma,
99.7 per cent of the results will be within Average ± 3 sigma.

If the process is not under control, the following will hold (Tchebycheff's Theorem):

More than 75 per cent of the results will be within Average ± 2 sigma,
More than 89 per cent of the results will be within Average ± 3 sigma,
More than 94 per cent of the results will be within Average ± 4 sigma.

The accuracy of the above statements, of course, is dependent upon the number of observations used to calculate sigma. The reliability of the standard deviation is determined as follows:

The standard error of sigma is equal to sigma divided by** the root of twice the number of observations.

In this case, then, the standard error is .9 divided by the root of 20, or .2. The reliability of the standard

** This ignores the correction for sigma of population.
* For longer method showing how deviations are obtained, see Page 11.

deviation, .9, is expressed as follows:

Based on the evidence supplied by the sample,

the odds are 68 out of 100 that the true sigma lies between .9 ± .2,
the odds are 95 out of 100 that the true sigma lies between .9 ± 2 times .2,
the odds are 99.7 out of 100 that the true sigma lies between .9 ± 3 times .2.

The effect of a larger sample upon the reliability of a sample can readily be seen.

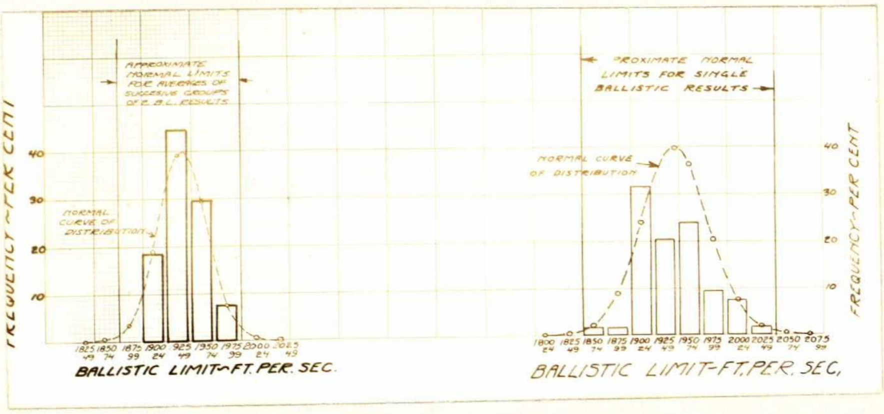
GROUPING.

The use of sigma infers that the distribution of observations is symmetrical about the average, that is, the normal curve of error prevails. However, this is not always the case. Hence, often the 3-sigma range based on individual observations may be in error.

In order to avoid this type of error, the group system has been developed. The chart on Page 20 shows a frequency distribution for ballistic limits of armour plate. Single values have been classified into intervals 1850-1874, 1875-1899, etc., and the per cent falling into each class has been plotted. The observed frequency distribution does not follow the normal curve of error. In assuming the normal curve, therefore, we are taking too much for granted. Note that the normal curve calculated from the data of these results indicates a higher percentage of low results than is actually found.

Now, if each successive pair of results is averaged and these averages plotted in a similar way to the first frequency distribution, it will be found that this new distribution follows the normal curve quite closely. It has been proven both theoretically and practically that by grouping and averaging successive results the normal curve is approached. The number in the group determines how closely the averages will approach the normal curve, that is, the larger the group the closer to normal the distribution of the averages becomes. However, in industrial conditions we cannot wait to obtain a large group. We wish to obtain results at frequent intervals. Therefore, the practice of using groups of 2 to 10 in size has been widely adopted. The choice of a group of 2 in this case is merely to facilitate the interpretation. The technique of calculating normal control limits is followed step by step in the succeeding pages.

Figure 6.



LEFT - AVERAGES OF GROUPS OF 2.

RIGHT - SINGLE VALUES.

Averages of groups of samples tend to a normal distribution even when a distribution of single values is non-symmetrical.

The larger the group, the closer to normal will be the distribution curve of the averages.

QUALITY CONTROL CHART CALCULATION.

P. No.	B.L.	Av'ge	Range	Group No.
25	1928			
26	1917	1922	11	12
28	1942			
29	1962	1952	20	13
30	1959			
32	1900	1929	59	14
33	1925			
34	2016	1970	91	15
35	1927			
36	2003	1965	76	16
37	1914			
39	1948	1931	34	17
40	1955			
41	1940	1947	15	18
42	1940			
43	1901	1945	89	19
44	1910			
45	1981	1945	71	20
46	1955			
47	1940	1942	15	21
48	1861			
49	2025	1943	164	22
50	1899			
51	1985	1942	86	23
52	1965			
53	1944	1954	21	24
54	1900			
55	1947	1923	47	25
56	1910			
57	1952	1931	42	26
58	1958			
59	1906	1932	52	27
60	1905			
61	1903	1904	2	28
62	1914			
63	1986	1950	72	29
64	1955			
65	2008	1981	53	30
66	1961			
67	1949	1955	12	31
68	1905			
69	1955	1930	50	32
70	2002			
71	1983	1992	19	33
72	1949			
74	1903	1926	46	34
75	1912			
76	1907	1909	5	35
77	1955			
78	1951	1953	4	36
79	1903			
80	1902	1902	1	37
81	1952			
82	1954	1953	2	38

(Continued on next page)

(Quality Control Chart Calculation, cont'd) -

Page 21 gives observations arranged in the order of occurrence and placed in groups of two. For the quality control chart method the average and the three sigma limits are calculated. We are indebted to the A.S.T.M. Manual on the presentation of data for factors which enable these limits to be easily calculated using only the

$\bar{\bar{X}}$ = Average of averages.
N = No. of group.
$\bar{R}$ = Average range.
A_2 = Factor for control limits for averages.
D_3 and D_4 = Factors for control limits for range.

The practical man will find that the mechanics of the method can be used without requiring a knowledge of the underlying theory. The student and research worker will want to study the origin of these factors.

$$\bar{\bar{X}} = \frac{\text{SUM OF AVERAGES}}{\text{NO. OF GROUPS}} = \text{GRAND AVERAGE}$$

$$= \frac{1128}{27} = 1942$$

$$\bar{R} = \frac{\text{SUM OF RANGES}}{\text{NO. OF GROUPS}} = \text{AVERAGE RANGE}$$

$$= \frac{1159}{27} = 43$$

CONTROL LIMITS FOR AVERAGE

$= \bar{\bar{X}} \pm A_2 \bar{R}$

$= 1942 \pm 1.88 \times 43$

$= 1942 \pm 81$

CONTROL LIMITS FOR RANGE

$D_3\bar{R}$ and $D_4\bar{R}$

$0 \times 43 - 3.268 \times 43$

$0 - 140$

FACTORS FOR CONTROL LIMITS

THE MANUAL ON PRESENTATION OF DATA GIVES FACTORS FOR DIFFERENT SIZE GROUPS AS FOLLOWS:

NUMBER IN GROUP	A_2	D_3	D_4
2	1.880	0	3.268
3	1.023	0	2.574
4	0.729	0	2.282
5	0.577	0	2.114
6	0.483	0	2.004
7	0.419	0.076	1.924

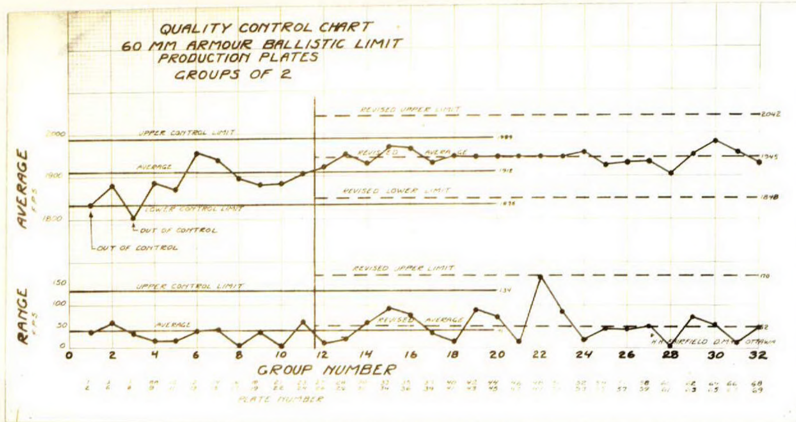
(Quality Control Chart Calculation, cont'd) -

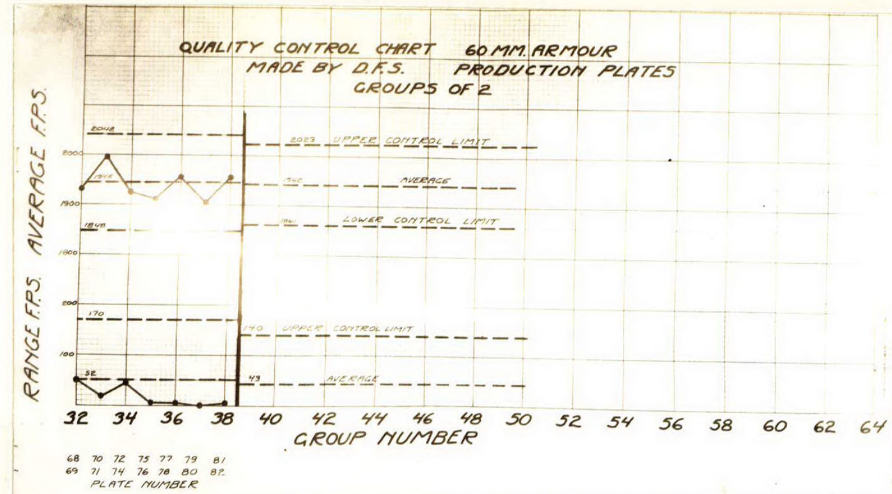
Note on the charts which have been prepared that the first control limits run from 0 to 24 production plate number. The use of such a small amount of data is not generally recommended since the error in calculating three sigma limits is considerable. However, in this case it serves to give a vague outline of the characteristics of ballistic limits. Starting at Group No. 10 a continual upward trend* was evident, hence revision was considered necessary at Group No. 12. The revised limits were based on No. 12 to No. 20 exclusive. The data recently received completed up to Group No. 38. At this point new limits were calculated including data from Groups Nos. 12 to 38. This last calculated limit is more accurate than the previous ones since it is calculated from a greater number of observations. We may state, with a fair degree of certainty, that as long as the process remains under control, averages of successive groups of two observations will fall within 1861 to 2023 f.p.s. and the difference between two successive observations will not exceed 140 f.p.s.

It is very encouraging to note that the lower control limit has increased from 1835 to 1861 and the control limit for range has been narrowed down. These are indications of an improvement in quality and a well controlled process.

* Trends toward change may be detected while the results are still within the control limits. The seven point rule is the safest from the mathematical viewpoint. However, the engineer constantly watching over the process may detect a trend towards change before seven points have been recorded. The seven point rule may be briefly stated as follows: - If seven successive points fall on the same side of the average or form a continuous upward or downward path, then it may be fairly certain that a change is taking place in the process. The users of this type of control chart have found by practical experience that in nearly all cases where the control limits are exceeded this extreme variation is due to a definite assignable cause. Therefore, each time control limits are exceeded an immediate investigation should be made. Variations within the control limits are characteristic of the process. In most cases investigation into the cause of difference between any two such observations will be impractical. The cause of slight variations is found more accurately by the correlation or large number method which is described in Section 2.

Figure 7.





QUALITY CONTROL CHART
BALLISTIC LIMIT OF 60 mm. ARMOUR PLATE.
(Manufactured by D.F.S.).

PRODUCTION RECORD IN QUALITY CONTROL CHART FORM.

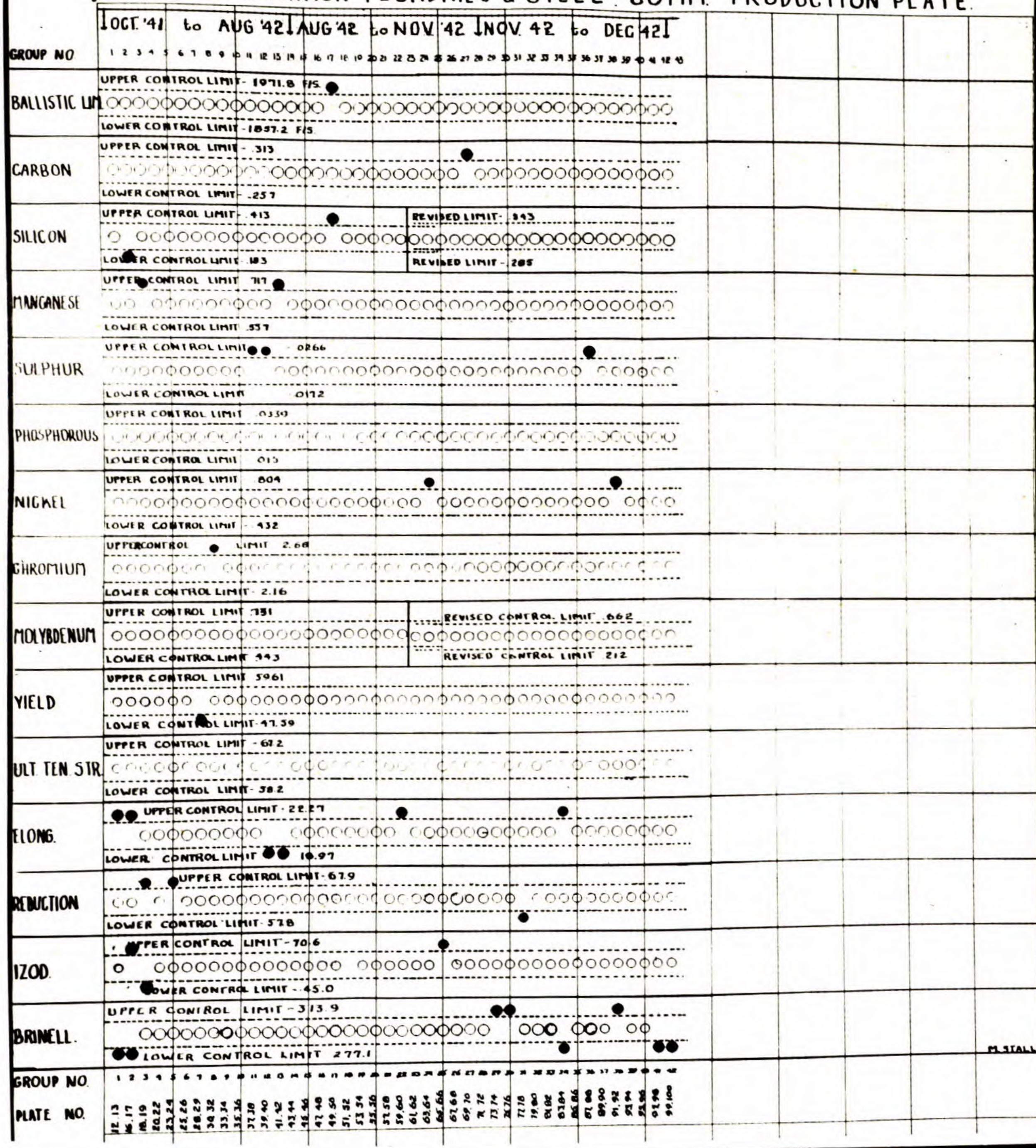
If a large number of control charts are to be kept, an abbreviated form can be adopted. On the following page (Figure 8), fifteen control charts are plotted. The white dot between the lines indicates "in control". The black dot either above or below the lines indicates "out of control".

This type of production quality record focuses attention on the trouble spots in the process. It also saves management and inspection a great deal of time in studying test records.

Figure 8 appears

on next page.

ABRIDGED QUALITY CONTROL CHART
FOR
BALLISTICS CHEMICALS PHYSICALS
OF
DOMINION FOUNDRIES & STEEL 60MM. PRODUCTION PLATE.



Research Work on Industrial Processes -

In order to find the pattern of cause and effect relationships existing in an industrial process, some knowledge of correlation technique is needed.

---

5. Correlation Between Two Types of Observations.

The idea of predicting the occurrence of an event from observations of phenomena in nature has engaged man's attention from earliest times. The phases of the moon, the positions of stars, flights of birds, and countless other omens were assumed to be definitely correlated with certain types of events. Palmistry and phrenology assume correlations between physical measurements and personal characteristics. The persistence of such theories with no foundation of factual evidence into the present day indicates how incapable are many individuals of rational judgment of observations.

Instances of this irrational type of interpretation are frequently encountered even in industries equipped with every known device for making accurate observations but with no system of handling observations for analysis. The following examples deal in a simple way with the problem of finding

(Figures 9, 10 and 11 are shown on Pages 28 to 30. Text continues on Page 31.)

Figure 9.

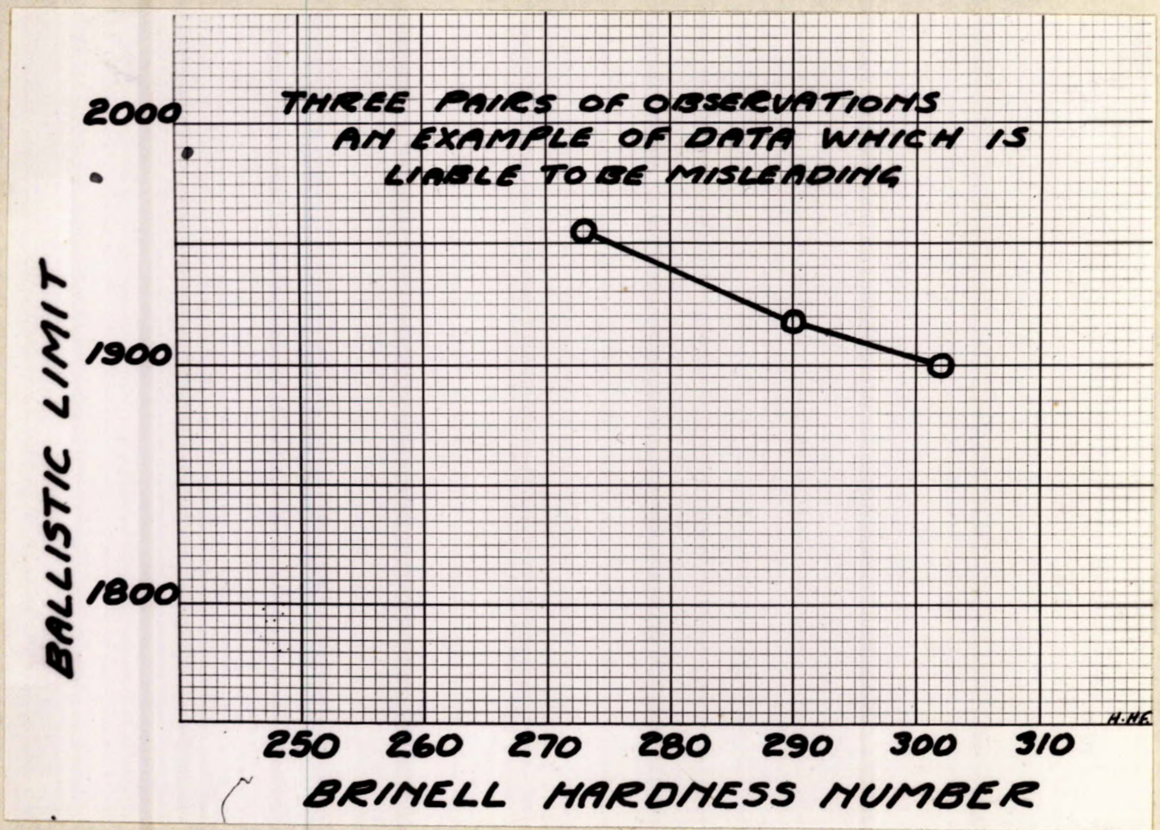

THREE PAIRS OF OBSERVATIONS
AN EXAMPLE OF DATA WHICH IS
LIABLE TO BE MISLEADING
2000
1900
1800
BALLISTIC LIMIT
250
260
270
280
290
300
310
BRINELL HARDNESS NUMBER
H.HE.

Figure 10.

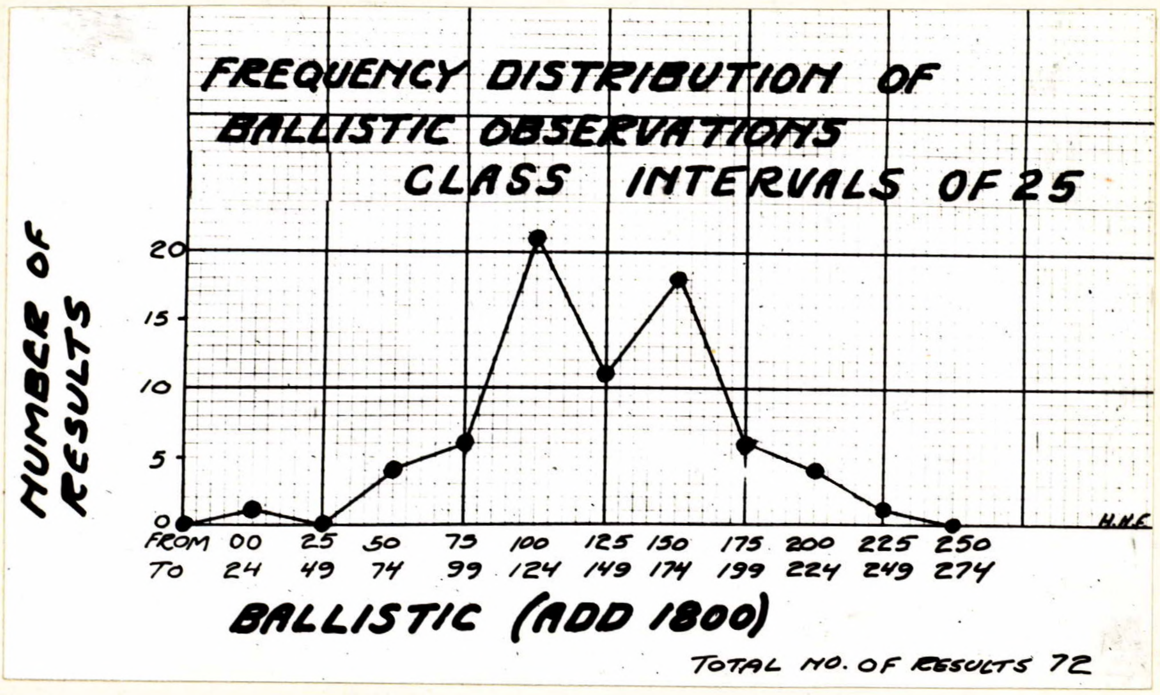
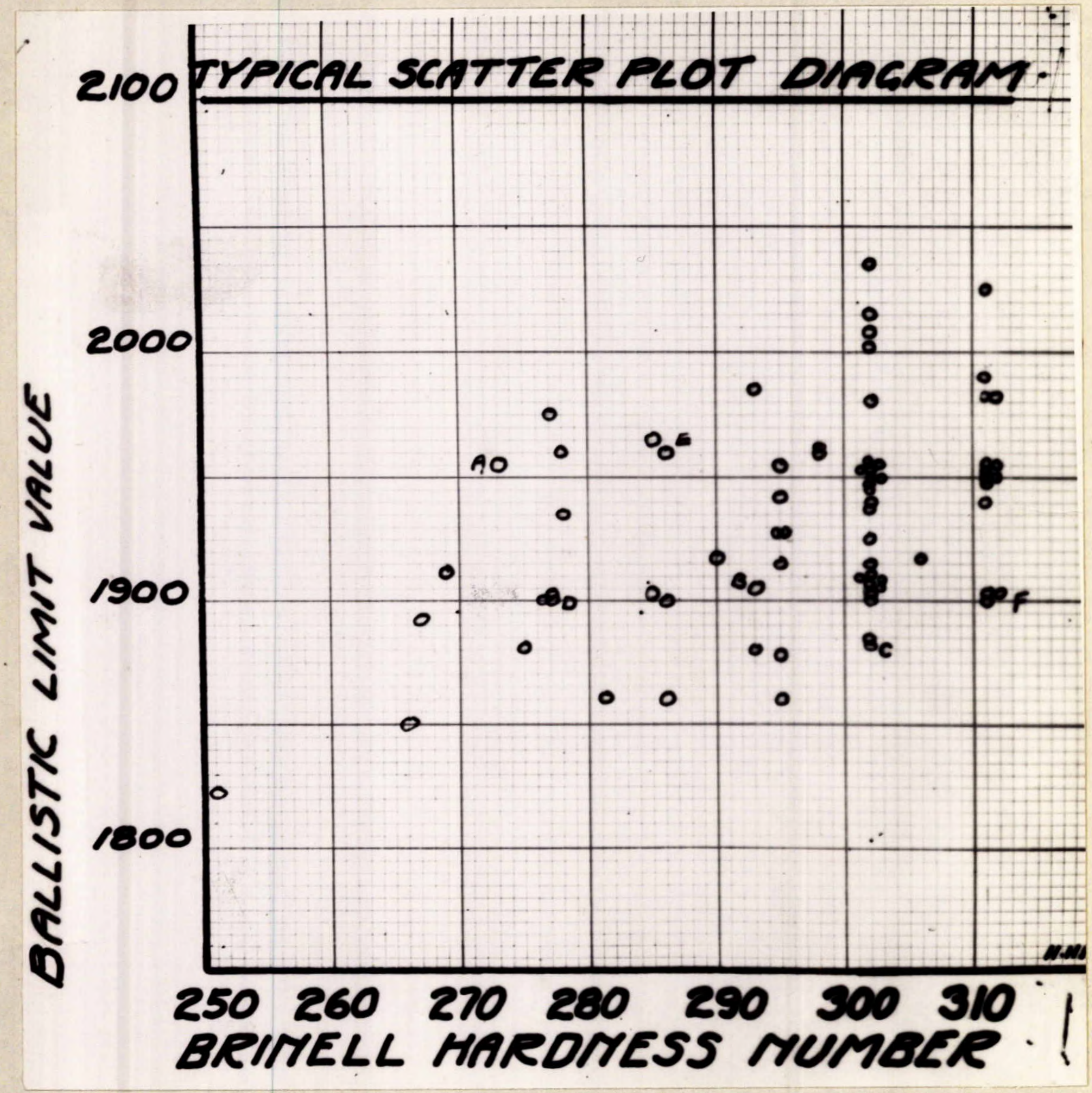

FREQUENCY DISTRIBUTION OF
BALLISTIC OBSERVATIONS
CLASS INTERVALS OF 25
NUMBER OF RESULTS
20
15
10
5
0
FROM 00 25 50 75 100 125 150 175 200 225 250
TO 24 49 74 99 124 149 174 199 224 249 274
BALLISTIC (ADD 1800)
TOTAL NO. OF RESULTS 72
H.H.F.

Figure 11.


TYPICAL SCATTER PLOT DIAGRAM
2100
2000
1900
1800
BALLISTIC LIMIT VALUE
250 260 270 280 290 300 310
BRINELL HARDNESS NUMBER

(Correlation Between Two Types of Observations, cont'd) -

the relationship between two types of observations on a product.

Figure 9 shows three sets of observations of ballistic limit and Brinell hardness. Is a sample of three sets sufficient evidence on which to base the statement inferred in this chart?

Before attempting to judge data similar to Figure 9, a background of experience should be obtained. The normal fluctuation of observations should be known.

Figure 11 shows 72 pairs of observations. This shows that for any given hardness, ballistic limit results vary over a considerable range. The points selected in Figure 9 do not represent the true relationship.

The larger the number of data the more accurately will the relationship be portrayed.

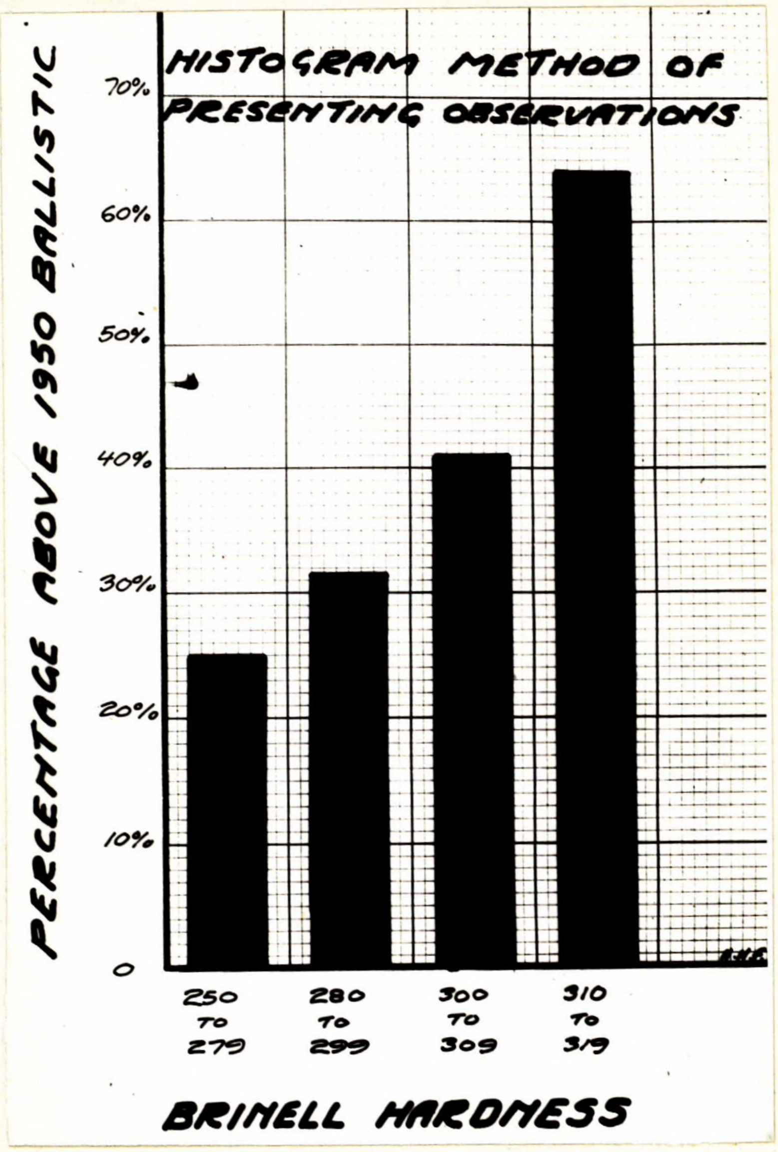
Methods of Determining Relationships between Two Types of Observations.

In Figure 11 a scatter plot diagram for two types of observations has been made. Scanning these data it is apparent that some relationship exists between the two types of observations. However, to attempt to draw a line through these dots would be only a hazardous conjecture. A simple system of analysis may be done as follows: divide the Brinell hardness numbers into groups; select a ballistic limit value near the average; determine the percentage of points above the average ballistic in each group; and plot as in Figure 12.

(Figure 12 is)
(shown on)
(Page 32.)

(Text continues on Page 33)

Figure 12.


HISTOGRAM METHOD OF PRESENTING OBSERVATIONS
PERCENTAGE ABOVE 1950 BALLISTIC
70%
60%
50%
40%
30%
20%
10%
0
250 TO 279
280 TO 299
300 TO 309
310 TO 319
BRINELL HARDNESS

(Correlation Between Two Types of Observations, cont'd) -

Figure 12 is called a histogram. Before accepting this as definite information the reliability of these results should be calculated. This is done as follows:[*]

Since the standard error of percentage is given by the expression,

$$p = \sqrt{\frac{P-(1.00-P)}{N}}$$

(p = Standard deviation of
(the average,
(P = Percentage,
(= 1.00 - P,
(N = No. of results,

it can be readily seen that the smaller the number in a group the less accurate is the value obtained.

Example:

In the Brinell range 250-279, 3 out of a total of 12 observations are above 1950 ballistic. That is, 25 per cent are above 1950.

Q. - How reliable is this?

A. - $\sigma p = \sqrt{\frac{P\ (1.00-P)}{N}} = \sqrt{\frac{.25 \times .75}{12}} = \sqrt{.01565} = 0.1251.$

Since the standard error of this average is 0.125, then the reliability of the percentage 25 per cent may be expressed as follows:

The odds are 68 out of 100 that the true value lies between .25 ± .125;

The odds are 95 out of 100 that the true value lies between .25 ± 2 x .125;

and so on.

- Figure 13 shows the same data as Figure 12, with reliability limits indicated. The odds are about seven out of ten that the experience if repeated would give values

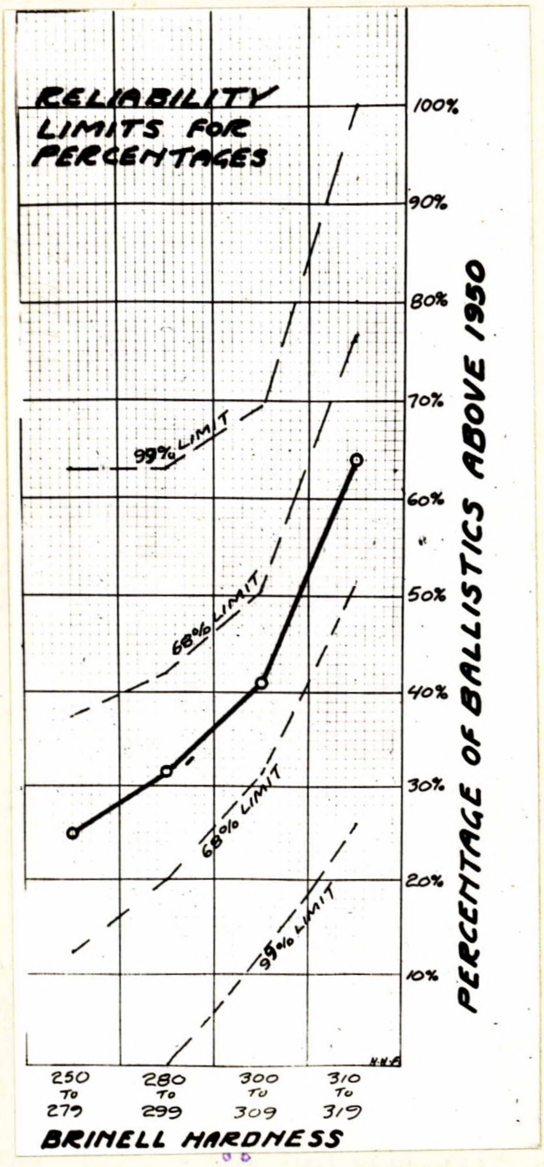
(Figure 13 is)
(shown on)
(Page 34.)

(Text continues on Page 35)

---

[*] See Applied General Statistics, by Croxton and Cowden.

Figure 13.


RELIABILITY
LIMITS FOR
PERCENTAGES
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
99% LIMIT
68% LIMIT
68% LIMIT
99% LIMIT
PERCENTAGE OF BALLISTICS ABOVE 1950
250 TO 279
280 TO 299
300 TO 309
310 TO 319
BRINELL HARDNESS

(Correlation Between Two Types of Observations, cont'd) -

which would fall within the 68 per cent limits.

As the number of observations included increases the reliability limits become narrower and narrower. This is the great advantage of large numbers of data. Figure 14 shows the average ballistic value for each group of hardness observations. The standard error of an average is equal to the standard deviation of the observations divided by the root of the number of observations.[⊕]

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}}$$

(x = Standard error of average,
(σ = Standard deviation of the observations,
(N = No. of results.

Example:

The average ballistic limit of material between 250-279 Brinell hardness is 1906. Standard deviation is 43.3.

Q. - How reliable is this average?

A. - $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}}$

$x = \frac{43.3}{\sqrt{12}} = 12.5$

(σ = Standard deviation of all observations. This is actually unknown; the standard deviation of the sample serves as a rough approximation.
(x = Standard error of the average.
(N = No. of results.

The method of using averages appears to be more accurate than the percentage method. Here again larger numbers of observations would give narrower reliability limit. In Figure 15 hardness of the samples above 1940 ballistic has been plotted as in frequency distribution. The hardness of samples below 1940 has also been plotted. Note that if ballistics above 1940 are desirable, then it would appear that Brinell hardness from 300 to 319 is more desirable than Brinell hardness from 250 to 299.

(Figure 14 and 15)
(are shown on)
(Pages 36 and 37)

(Text continues
on Page 38)

[⊕] Ignoring certain corrections for sigma of population.

Figure 14.

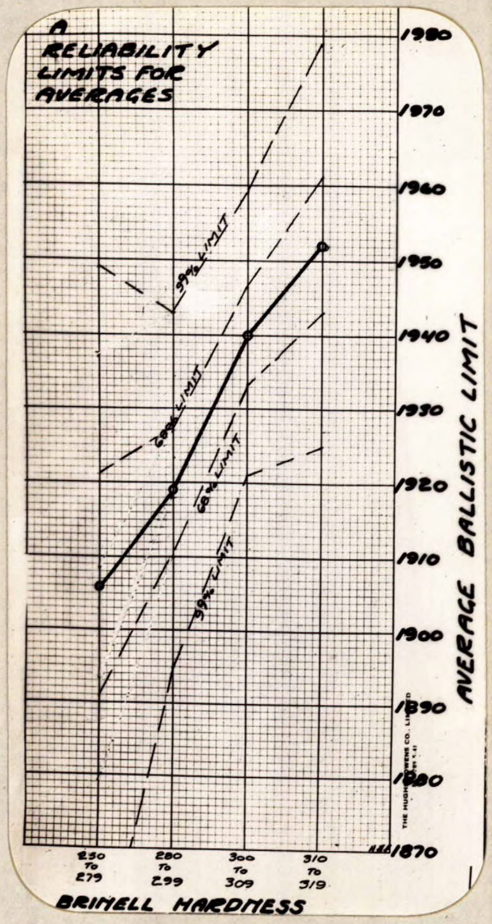
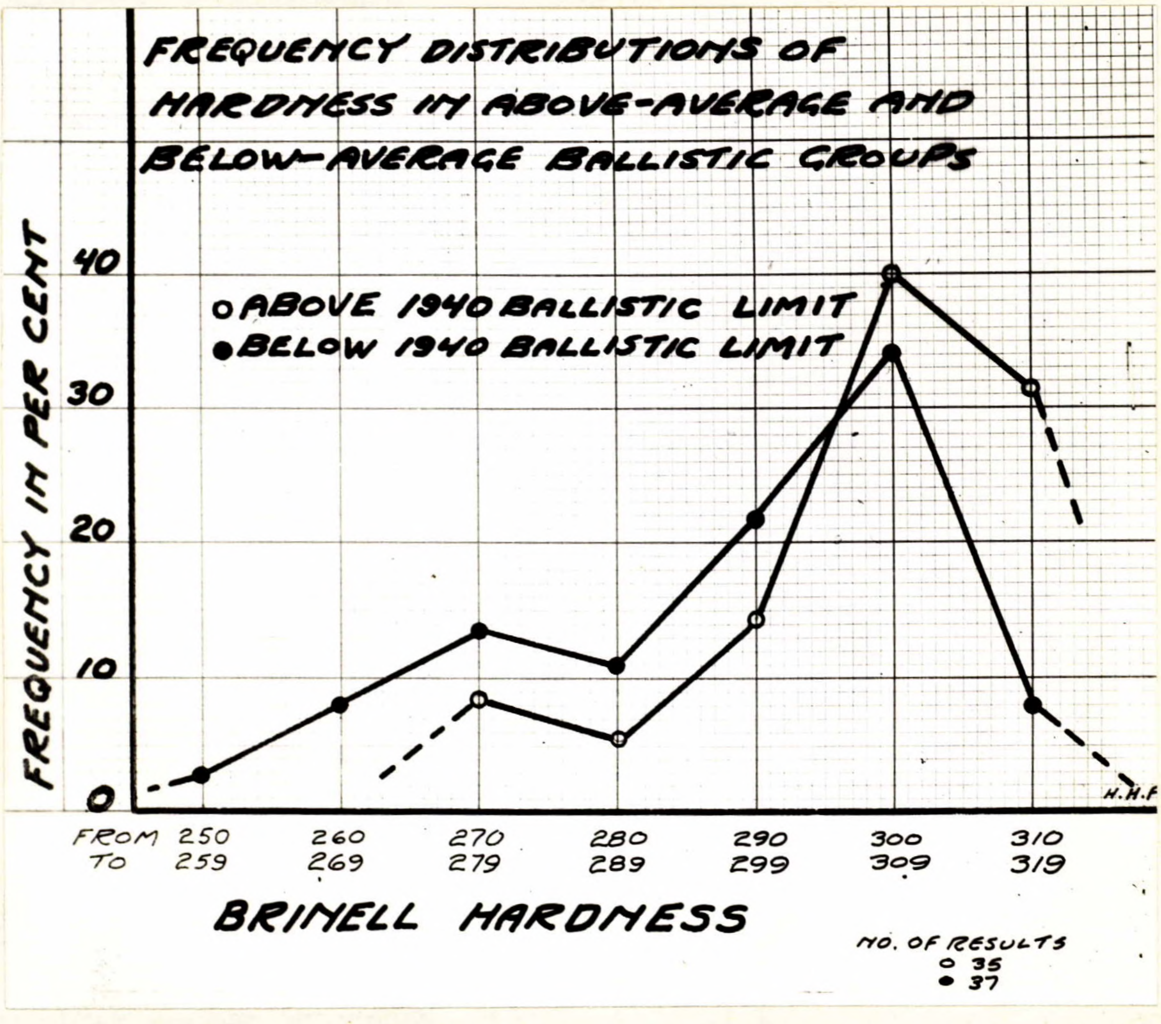

A RELIABILITY LIMITS FOR AVERAGES
99% LIMIT
68% LIMIT
68% LIMIT
99% LIMIT
1980
1970
1960
1950
1940
1930
1920
1910
1900
1890
1880
1870
AVERAGE BALLISTIC LIMIT
250 TO 279
280 TO 299
300 TO 309
310 TO 319
BRINELL HARDNESS

Figure 15.


FREQUENCY DISTRIBUTIONS OF
HARDNESS IN ABOVE-AVERAGE AND
BELOW-AVERAGE BALLISTIC GROUPS
FREQUENCY IN PER CENT
40
30
20
10
0
○ ABOVE 1940 BALLISTIC LIMIT
● BELOW 1940 BALLISTIC LIMIT
FROM 250 260 270 280 290 300 310
TO 259 269 279 289 299 309 319
BRINELL HARDNESS
NO. OF RESULTS
○ 35
● 37
H.H.F

(Correlation Between Two Types of Observations, cont'd) -

There are other methods of determining correlation which involve a considerable amount of calculation. They may be obtained from standard statistical tests.

---

6. Rational Judgment of Statistical Data.

The practical man familiar with a process has a great advantage over the most precise theorist. He may have a wide background of experience in the light of which he can interpret the importance of a set of observations. The statistician may be in error through biased observations, poor sampling, and also the fact that factors of major significance were not considered.

In judging correlation between two types of observations one of the following general interpretations may be made:

1. A cause-and-effect relationship may exist.

   Usually, the cause-and-effect relationship should not be inferred unless there is sound engineering evidence to support this theory.

2. The apparent relationship may be due to a third and unknown variable which controls both of the observed variables. Quenching speed controls both tensile and hardness properties of steel.

3. There may be other correlations of much greater significance and therefore observed correlation is of only secondary importance.

4. The relationship observed may be only a transient one, that is, existing for a short period of time. As lots of raw material vary, the relationship between properties may vary. Properties of malleable iron vary with different lots of pig iron.

(Continued on next page)

(Rational Judgment of Statistical Data, cont'd) -

5. Two values may have no connection with each other and the relationships observed may be due to chance.

6. The relationship is not necessarily a general one. It may hold only for the source of the data.

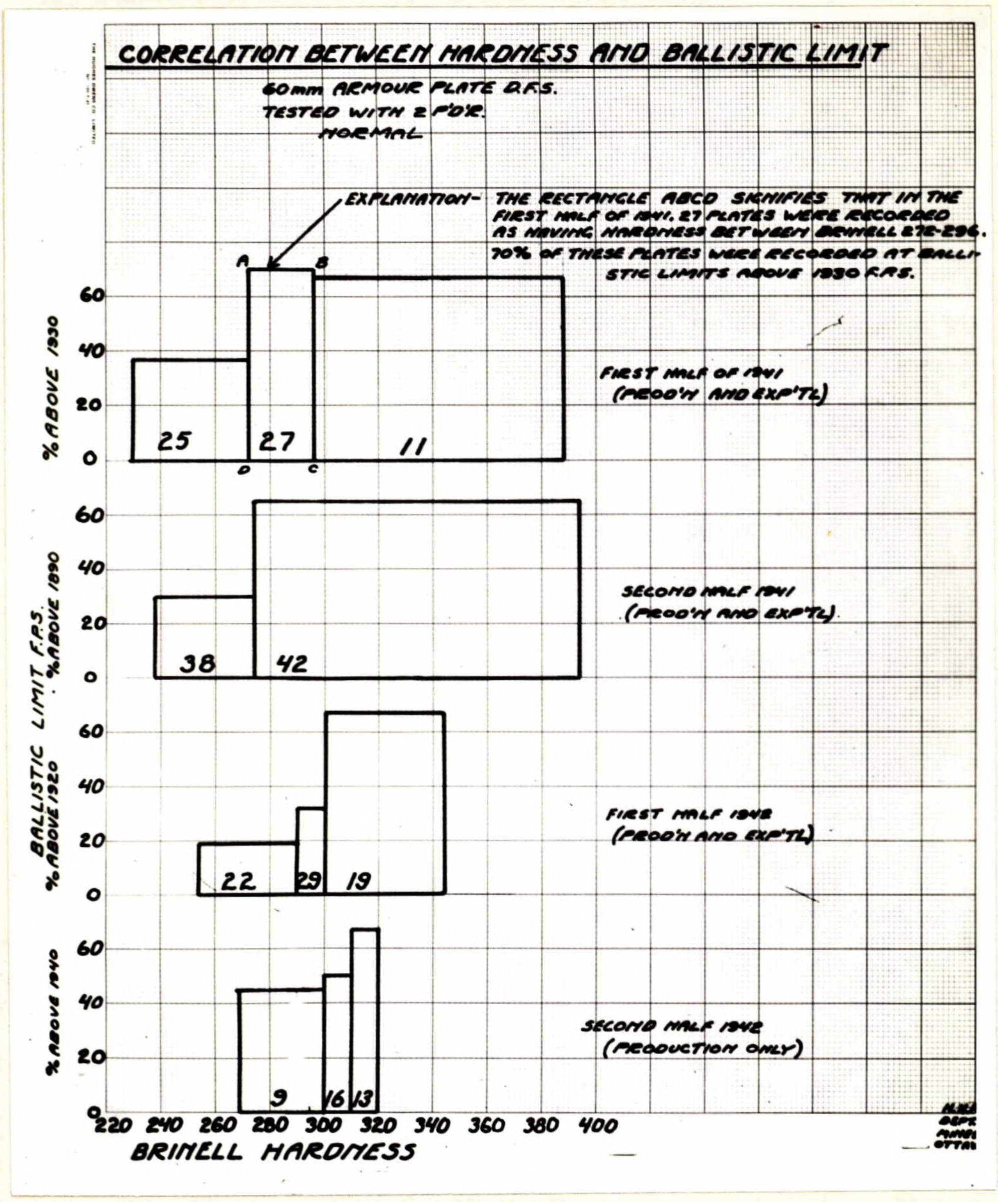
It is obvious that interpretation can best be made by engineers <u>thoroughly familiar with the process and with the methods and with the properties of the material.</u>

A correlation between Brinell hardness and tensile strength is normally expected and considered to be a true cause-and-effect relationship. A correlation between silicon and tensile strength would generally be considered by the metallurgist to be either accidental or transient. The statistician unfamiliar with the process may frequently select observations for correlation which are of little significance when compared to other major controlling variables in the process. However, it is often of interest to study apparently unrelated observations, for important discoveries have been made along this line of investigation.

The best proof of reliability is the fact that the same relationship occurs during several successive intervals of time. Figure 16 illustrates that the relationship between Brinell hardness and ballistic limit has remained the same over the four successive six-months periods. From this we are able to state that the relationship is of a permanent nature. We still do not have sufficient proof to state whether it is a cause-and-effect relationship or whether a third and unknown factor controls both Brinell and ballistic observations.

(Figure 16 is shown on Page 40. Text continues on Page 41.)

Figure 16.


CORRELATION BETWEEN HARDNESS AND BALLISTIC LIMIT
60mm ARMOUR PLATE Q.F.S.
TESTED WITH 2 PDR.
NORMAL
EXPLANATION- THE RECTANGLE ABCD SIGNIFIES THAT IN THE FIRST HALF OF 1941, 27 PLATES WERE RECORDED AS HAVING HARDNESS BETWEEN BRINELL 272-296.
70% OF THESE PLATES WERE RECORDED AT BALLISTIC LIMITS ABOVE 1930 F.P.S.
A
B
C
D
% ABOVE 1930
60
40
20
0
25
27
11
FIRST HALF OF 1941
(PROD'N AND EXP'TL)
BALLISTIC LIMIT F.P.S.
% ABOVE 1890
38
42
SECOND HALF 1941
(PROD'N AND EXP'TL)
% ABOVE 1920
22
29
19
FIRST HALF 1942
(PROD'N AND EXP'TL)
% ABOVE 1940
9
16
13
SECOND HALF 1942
(PRODUCTION ONLY)
220 240 260 280 300 320 340 360 380 400
BRINELL HARDNESS

CONCLUSIONS.

A question frequently asked is HOW CAN WAR MATERIAL BE IMPROVED?

This article has shown one way in which industrial products can be improved. The steps are:

- MAKE TESTS AND OBSERVATIONS DURING MANUFACTURE.
- RECORD PERFORMANCE OF THE MATERIAL.
- STUDY THE OBSERVATIONS AND THEIR FLUCTUATION.
- FIND CORRELATION BETWEEN TYPES OF OBSERVATIONS.
- APPLY INFORMATION SO GAINED.

Unaided human judgment is frequently biased or in error. In handling large numbers of observations, some use should be made of the science of statistics to aid in judging the relationships between test data and variation of observations.

This article serves merely to introduce the subject. Those who intend to utilise statistical methods will refer to standard texts.

As manpower and materials become scarcer, it is of greater importance that industrial processes and inspection of materials become more efficient. When observations are interpreted rationally and statistical methods are used, inspection becomes an engineering science.

A great many of the larger U.S. and U.K. manufacturers are using scientific inspection methods. Reports from users of scientific inspection state that rejects are decreased and at the same time man-hours of inspection are reduced by from 25 to 50 per cent of pre-scientific inspection period. These savings can be a valuable contribution to the war effort.

HHF:PES:GHB.

BIBLIOGRAPHY:

A.S.T.M. MANUAL ON PRESENTATION OF DATA - American Society for Testing Materials.

QUALITY CONTROL CHART METHOD - American Standards Association.

APPLIED GENERAL STATISTICS, by F. E. Croxton and D. J. Cowden, Prentice-Hall, Inc., New York, 1940.

ENGINEER'S MANUAL OF STATISTICAL METHODS, by Col. L. E. Simon, United States Ordnance, John Wiley & Sons, New York, 1941.

ECONOMIC CONTROL OF QUALITY OF MANUFACTURED PRODUCT - by W. A. Shewhart (D. Van Nostrand Co., New York, 1939).

QUALITY CONTROL OF MUNITIONS - G. D. Edwards, War Department, Washington, D.C. "Army Ordnance," 1942.

Ottawa, Canada,
February 19th, 1943.
HHF:PES:GHB.