NATIONAL BUREAU OF STANDARDS REPORT

6263

Draft of

Part IV (Miscellaneous Topics)

for

Manual on Experimental Statistics
for Ordnance Engineers

A REPORT TO
OFFICE OF ORDNANCE RESEARCH
DEPARTMENT OF THE ARMY

U. S. DEPARTMENT OF COMMERCE
NATIONAL BUREAU OF STANDARDS

THE NATIONAL BUREAU OF STANDARDS

Functions and Activities

The functions of the National Bureau of Standards are set forth in the Act of Congress, March 3, 1901, as amended by Congress in Public Law 619, 1950. These include the development and maintenance of the national standards of measurement and the provision of means and methods for making measurements consistent with these standards; the determination of physical constants and properties of materials; the development of methods and instruments for testing materials, devices, and structures; advisory services to Government Agencies on scientific and technical problems; invention and development of devices to serve special needs of the Government; and the development of standard practices, codes, and specifications. The work includes basic and applied research, development, engineering, instrumentation, testing, evaluation, calibration services, and various consultation and information services. A major portion of the Bureau's work is performed for other Government Agencies, particularly the Department of Defense and the Atomic Energy Commission. The scope of activities is suggested by the listing of divisions and sections on the inside of the back cover.

Reports and Publications

The results of the Bureau's work take the form of either actual equipment and devices or published papers and reports. Reports are issued to the sponsoring agency of a particular project or program. Published papers appear either in the Bureau's own series of publications or in the journals of professional and scientific societies. The Bureau itself publishes three monthly periodicals, available from the Government Printing Office: The Journal of Research, which presents complete papers reporting technical investigations; the Technical News Bulletin, which presents summary and preliminary reports on work in progress; and Basic Radio Propagation Predictions, which provides data for determining the best frequencies to use for radio communications throughout the world. There are also five series of nonperiodical publications: The Applied Mathematics Series, Circulars, Handbooks, Building Materials and Structures Reports, and Miscellaneous Publications.

Information on the Bureau's publications can be found in NBS Circular 460, Publications of the National Bureau of Standards ($1.25) and its Supplement ($0.75), available from the Superintendent of Documents, Government Printing Office, Washington 25, D. C.

Inquiries regarding the Bureau's reports should be addressed to the Office of Technical Information, National Bureau of Standards, Washington 25, D. C.

NATIONAL BUREAU OF STANDARDS REPORT

NBS PROJECT		NBS REPORT
1103-40-5146	13 January 1959	6263

Draft of

Part IV (Miscellaneous Topics)

for

Manual on Experimental Statistics
for Ordnance Engineers

Prepared by

Statistical Engineering Laboratory

A Report to

Office of Ordnance Research
Department of the Army

NBS

U. S. DEPARTMENT OF COMMERCE
NATIONAL BUREAU OF STANDARDS

NOTICE

This report is a preliminary draft of Part IV (Miscellaneous Topics) for the Manual on Experimental Statistics for Ordnance Engineers.

No known inaccuracies exist in the present draft, but improvements in arrangement and exposition of some of the material may be made at a later date.

Table of Contents

1. Rejection of Observations

Every experimenter has at some time obtained a set of observations, purportedly taken under the same conditions, in which one observation was widely different, or an outlier from the rest.

The problem that confronts the experimenter is whether he should keep the suspect observation in computation, or whether he should discard it as being a faulty measurement. The word reject will mean reject in computation, since every observation should be recorded. A careful experimenter will want to make a record of his "rejected" observations and where possible detect and analyze carefully their cause(s).

It should be emphasized that we are not discussing the case where we *know* that the observation differs because of an assignable cause, i.e. a dirty test-tube, or a change in operating conditions. We are dealing with the situation where, as far as we are able to ascertain, all the observations are on approximately the same footing. One observation is "suspicious" however, in that it seems to be set apart from the others. We wonder whether it is not so far from the others that we can reject it as being caused by some assignable but thus far unascertainable cause.

Whether a measurement far-removed from the great majority of a set of measurements of a quantity, and thus possibly reflecting a gross error, should have a full vote, a diminished vote, or no vote in the final average - and in the determination of precision - is a very difficult question to answer completely in general terms. Common sense dictates that, if on investigation, a trustworthy explanation of the discrepancy is found, the value concerned should be excluded from the final average and from the estimate of precision, since these presumably are intended to apply to the unadulterated system. If, on the other hand, no explanation for the apparent anomalousness is found, then common sense would seem to indicate that it should be included in computing the final average and the estimate of precision. Experienced investigators differ in this matter. Some, e.g., J.W. Bessel, would always include it. Others would be inclined to exclude it, on the grounds that it is better to exclude a possibly 'good' measurement than to include a possibly 'bad' one. The argument for exclusion is that when a 'good' measurement is excluded we simply lose some of the relevant information, with consequent decrease in precision and the introduction of some bias (both being theoretically computable); whereas when a truly anomalous measurement is included it vitiates our results,

biasing both the final average and the estimate of precision by unknown, and generally unknowable, amounts.

There have been many criteria proposed for guiding the rejection of observations. For an excellent summary and critical review of the classical rejection procedures, and some more modern ones, see ref. [1]. One of the more famous classical rejection rules is "Chauvenet's criterion", which is not recommended. This criterion is based on the normal distribution and advises rejection of an extreme observation if the probability of occurrence of such deviation from the mean of the n measurements is less than 1/2n. Obviously for small n, such a criterion rejects too easily.

A review of the history of rejection criterion, and the fact that new criteria are still being proposed, leads one to realize that no completely satisfactory rule can be devised for any and all situations. One cannot devise a criterion that will not reject a predictable amount from endless arrays of perfectly good data. The amount of data rejected of course depends on the rule. This is the price one pays for using any rule for rejection of data. There is actually nothing better to use than the judgment of an experienced investigator who is thoroughly familiar with his measurement process. For an excellent discussion of this point, see ref. [2]. Statistical rules are given largely for the benefit of

inexperienced investigators, those working with a new process, or those who simply want justification for what they would have done anyway.

Whatever rule is used, it must bear some resemblance to the experimenter's feelings about the nature and possible frequency of errors. For an extreme example - if the experimenter feels that he may make one blunder in twenty, and he uses a rejection rule that throws out 3 in 10, obviously his reported data will be "clean". The one sure way, and the only sure way, to avoid publishing any "bad" results is to throw away all results.

With the foregoing reservations, the next two sections give some suggested procedures for judging outliers. In general the rules to be applied to a single experiment reject only what would be rejected by an experienced investigator anyway. (See section 1.2)

1.1. Rejection of observations in routine experimental work.

Probably the best technique for detection of "errors" (e.g., systematic errors, gross errors) in routine work is the use of control charts for the mean and range. Detailed methods of application of control charts have been described in many places and will not be repeated here - see section 2 for appropriate references.

1.2 Rejection of observations in a single experiment.

We shall assume that we have n observations ranked in order from lowest to highest ($X_1 \leq X_2 \dots \leq X_n$). All of the observations are from a population which has a normal distribution with mean μ and standard deviation σ. We shall discuss four cases.

Case (1) - We do not know anything whatever about the population from which the data were taken, and must make our decision on the basis of the n observations.

Case (2) - We do not know μ or σ, but have an estimate s of σ which is independent of our present sample. The value of s is estimated with f degrees of freedom.

Case (3) - We are willing to assume a value for σ, the standard deviation of the population from which the observations were taken.

Case (4) - The population from which the observations have been taken is well known, and in particular we know the values of μ and σ.

CASE (1): No prior information about the mean and standard deviation

i) Choose α, the probability or risk we are willing to take of rejecting an observation which really belongs in the group.

ii) Look up $r_{1-\alpha/2}$ in Table XI. (NOTE: If the procedure is such that we would consider rejection of observations which were extreme in one direction only -

e.g., where we would reject values which were too large but never those which were too small - then we should replace $\alpha/2$ by α. Our choice as to whether to use $\alpha/2$ or α must be made on a priori grounds and never on the basis of the data to be analyzed).

iii) If $3 \leq n \leq 7$ Compute r_{10},
$8 \leq n \geq 10$ Compute r_{11},
$11 \leq n \leq 13$ Compute r_{21},
$14 \leq n \leq 25$ Compute r_{22},

where r_{ij} is computed as follows:

r_{ij}	X_n is suspect	X_1 is suspect
r_{10}	$(X_n - X_{n-1})/(X_n - X_1)$	$(X_2-X_1)/(X_n-X_1)$
r_{11}	$(X_n-X_{n-1})/(X_n-X_2)$	$(X_2-X_1)/(X_{n-1}-X_1)$
r_{21}	$(X_n-X_{n-2})/(X_n-X_2)$	$(X_3-X_1)/(X_{n-1}-X_1)$
r_{22}	$(X_n-X_{n-2})/(X_n-X_3)$	$(X_3-X_1)/(X_{n-2}-X_1)$

iv) If $r_{ij} > r_{1-\alpha/2}$, reject the suspect observation, otherwise retain it.

CASE (2): Mean unknown, estimate of σ available from previous sample

i) Choose α, the probability or risk we are willing to take of rejecting an observation which really belongs in the group.

ii) Look up $q_{1-\alpha}(n,f)$ in Table IV. (See Note in step ii of Case (1)).

iii) Compute $w = q_{1-\alpha}s$.

iv) If $w < X_n - X_1$, reject the observation which is suspect, otherwise retain it.

CASE (3): Mean unknown, σ known

i) Choose α, the probability or risk we are willing to take of rejecting an observation which really belongs in the group.

ii) Look up $q_{1-\alpha}(n,\infty)$ in Table IV. (See Note in step ii of Case (1)).

iii) Compute $w = q_{1-\alpha}\sigma$.

iv) If $w < X_n - X_1$, reject the observation which is suspect, otherwise retain it.

CASE (4): Mean known, σ known

i) Choose α, the probability or risk we are willing to take of rejecting an observation which really belongs in the group.

ii) Compute $\alpha' = (1-\alpha/2)^{1-n}$. (See note in step ii of Case (1)).

iii) Look up $z_{1-\alpha'}$ in Table Ib.

iv) Compute $a = \mu + \sigma z_{1-\alpha'}$

$$b = \mu - \sigma z_{1-\alpha'}$$

v) Reject any observation which does not lie in the interval from a to b.

References

[1] Paul R. Rider. "Criteria for Rejection of Observations". Washington University Studies - New Series, Science and Technology - No. 8, October 1933.

[2] E.B. Wilson, Jr. "An Introduction to Scientific Research", McGraw-Hill Book Co., Inc., New York, 1952.

Additional Reading

[3] W.J. Dixon. "Processing Data for Outliers", Biometrics, Vol. 9 (1953) pp. 74-89.

[4] A. Hold. "Statistical Theory with Engineering Applications", John Wiley and Sons, 1952, pp. 333-336.

[5] F. Proschan. "Testing Suspected Observations", Industrial Quality Control, Vol. XIII, No. 7, January 1957.

2. The Place of Control Charts in Experimental Work

2.1 Primary objective of control charts

Control charts have very important functions in experimental work, although their use in laboratory situations has been discussed only briefly by most textbooks. Control charts can be used as a form of statistical test in which the primary objective is to test whether or not the process is in "statistical control". The process is in "statistical control" when repeated samples from the process behave as random samples from a stable probability distribution; thus the underlying conditions of a process "in control" are such that it is possible to make predictions in the probability sense.

The control limits are usually computed by using formulae which utilize the information from the samples themselves. These limits are placed on the specific chart and the decision is made that the process was "in control" if all points fall within the line. If all points are not within the limits then the decision is made that the process is not "in control".

The basic assumption underlying most statistical techniques is that the data are a random sample from a stable probability distribution, which is another way of saying that the process is in "statistical control". It is the validity of this basic assumption which the control chart is designed to test. The control chart is used to demonstrate the existence of statistical control, and to monitor a controlled process.

As a monitor, a given control chart indicates a particular type of departure from control.

2.2. Information provided by control charts

Control charts provide a running graphical record of small subgroups of data taken from a repetitive process. Control charts may be kept on any of various characteristics of each small subgroup -- e.g. the average, standard deviation, range, proportion defective. The chart for each particular characteristic is designed to detect certain specified departures in the process from the conditions assumed. The process may be a measurement process as well as a production process. The order of groups is usually with respect to time, but not necessarily so. The grouping is such that the members of a group are more likely to be alike than are different groups.

Primarily control charts can be used to demonstrate whether or not the process is in statistical control. When the charts show lack of control, they indicate where or when the trouble occurred. Often they indicate the nature of the trouble, e.g. trends or runs, sudden shifts in the mean, increased variability, etc.

The control charts, besides serving as a method of testing for control, also provide additional and useful information in the form of estimates of the characteristics of a controlled

process. This is information which is altogether too frequently overlooked. For example, one very important piece of information which can be obtained from a control chart for the range or standard deviation is an estimate of the variability (σ) of a routine measurement or production process. It will be remembered that many of the techniques of Part I are given in parallel for known σ and unknown σ. Most experimental scientists have very good knowledge of the variability of their measurements, but hesitate to assume "known σ" without additional justification. Control charts can be used to provide the justification.

Finally, as was pointed out in section 1, a control chart is the most satisfactory criterion for rejection of observations in a routine laboratory operation.

2.3 References

The basic references for the use of control charts are:

1. For the underlying philosophy of statistical control - Shewhart, W.A., "Economic Control of Quality of Manufactured Product", Van Nostrand Inc., New York, N.Y., 1931. (The first and still a very significant book in the field).

2. For the details of application - "ASTM Manual on Quality Control of Materials", available from American Society for Testing Materials, 1916 Race Street, Philadelphia 3, Pa., 1951. (This manual gives examples of various kinds of charts).

Some recommended general texts on quality control are:

3. Cowden, D.J., "Statistical Methods in Quality Control", Prentice-Hall, Inc., Englewood Cliffs, N.J., 1957.

4. Duncan, A.J. "Quality Control and Industrial Statistics", Richard D. Irwin, Inc., Chicago, 1952.

5. Grant, E.L. "Statistical Quality Control", Second Edition, McGraw-Hill Book Co., New York, 1952.

6. Actual examples of laboratory applications in the chemical field can be found in a series of comprehensive bibliographies published in Analytical Chemistry:

a) Wernimont, G., "Statistics Applied to Analysis", Anal. Chem. 21, 115, (1949).

b) Hader, R.J. and Youden, W.J., "Experimental Statistics", Anal. Chem. 24, 120, (1952).

c) Mandel, J. and Linnig, F.J., "Statistical Methods in Chemistry", Anal. Chem. 28, 770 (1956).

d) Mandel, J. and Linnig, F.J., "Statistical Methods in Chemistry", Anal. Chem. 30, 739, (1958).

These 4 articles are excellent review articles, successively bringing up to data the recent developments in statistical theory and statistical applications which are of interest in chemistry. They contain extensive bibliographies, divided by subject matter, and thus provide means for locating articles on control charts in the laboratory. They are not limited to control chart applications, however.

e) For a special technique with ordnance examples, see: Grubbs, F.E., "The Difference Control Chart with Example of its Use", Industrial Quality Control Vol. III, No. 1, July 1946.

Industrial Quality Control, the monthly journal of the American Society for Quality Control, is the most comprehensive publication in this field.

3. Statistical Techniques for Analyzing Extreme-Value Data*

Classical applications of statistical methods, which frequently concern average values and other quantities following the symmetrical normal distribution, are inadequate when the quantity of interest is the largest or the smallest in a set of magnitudes. This is the situation in a number of fields, in many of which applications of methods for dealing with extremes have already been made. Meteorological phenomena that involve extreme pressures, temperatures, rainfalls, wind velocities, etc., may be treated by extreme-value techniques. The techniques are also applicable in the study of floods and droughts.

Other examples occur in the fracturing of metals, textiles, and other materials under applied force, and in fatigue phenomena [1,5,6]. In these instances, the observed strength of a specimen often differs from the calculated strength, and depends, among other things, upon the length and volume. An explanation is to be found in the existence of weakening flaws assumed to be distributed at random in the body and assumed not to influence one another in any way. The observed strength is determined by that of the weakest region - just as no chain is stronger than its weakest link. It is thus apparent that whenever extreme observations are encountered it will pay to consider the use of extreme-value techniques.

---

*) This section is adapted, with permission, from [10] and [14].

Use of Extreme-Value Technique - Largest Values

A simplified account is given here. For the detailed theory and methods, consult the primary references, [1], [2], and [3]. References [4] thru [15] give some selected applications. More extensive lists of references can be found in [1], [3], and [9].

Figure 1 illustrates the frequency form of a typical curve for the distribution of largest observations.

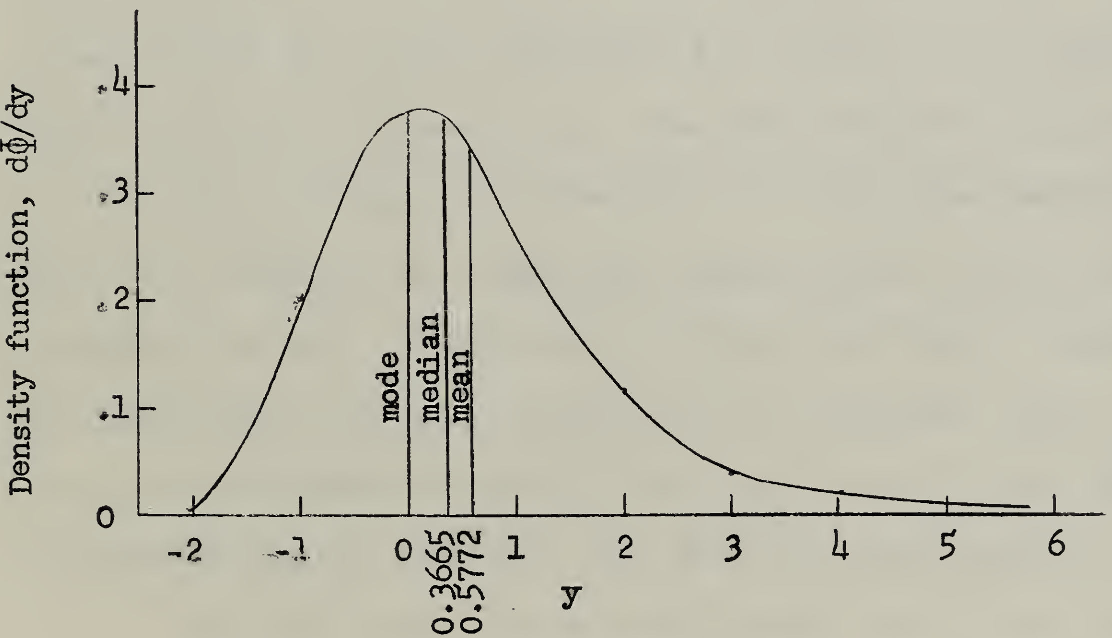

Fig. 1--Theoretical distribution of largest values.

This curve is the derivative of the function

$$\Phi(y) = \exp\left[-\exp(-y)\right] .$$

Unlike the normal distribution this curve is skewed - rising to a maximum to one side of the mean. The chief feature of such a distribution is that very large values have a much greater chance of occurring than very small values. This agrees with common experience. Very low maximum values are most unusual,

while very high ones do occasionally occur. Theoretical considerations lead to a curve of this nature, called the distribution of largest values or the extreme-value distribution.

In using the extreme-value method, all the observed maxima, such as the largest wind velocity observed in each year during a fifty-year period, are first ranked in order of size from the smallest to the largest. These values are given ranks 1 to n and are then transformed into cumulative frequencies P_i by the relation $P_i = i/(n + 1)$, where i is the rank of the ith observation counting from the smallest. Thus a plotting position is obtained for each observation. The data are plotted on a special graph paper, called extreme-value probability paper, designed so that the "ideal" extreme-value distribution will plot exactly as a straight line. Consequently, the closeness of the plotted points to a straight line is an indication of how well the data fit the theory.

Extreme-value probability paper has a uniform scale along one axis, usually the vertical, which is used for the observed values, as in Figure 2. The horizontal axis then serves as the probability scale and is marked according to the doubly-exponential formula. Thus, in Figure 2, the space between 0.01 and 0.5 is much less than the space between 0.5 and 0.99. The limiting values 0 and 1 are never reached, as

is true of any probability paper designed for an unlimited variate.

An extreme-value plot of the maximum atmospheric pressures in Bergen, Norway for the period between 1857 and 1926 (Figure 2) showed by inspection that the observed data satisfactorily fitted the theory. Fitting the line by eye may be sufficient. Details of fitting a computed line can be found in [1]. From the fitted straight line it is possible to predict, for example, that a pressure of 793mm corresponds to a probability of 0.994, that is, pressures of this magnitude have less than 1 chance in 100 of being exceeded in any particular year.

In studies of the normal acceleration increments experienced by an airplane flying through gusty air [9, page 394], an instrument was employed that indicated only the maximum shocks. Thus, only one maximum value was obtained from a single flight. A plot representing 26 flights of the same aircraft indicated that the probability that the largest recorded gust will not be exceeded in any other flight was 0.96, i.e., a chance of four in 100 of encountering a more severe gust than any recorded. A more recent study [13] presents refinements especially adapted to very small samples of extreme data and also to larger samples where it is necessary to obtain the

greatest amount of information from a limited set of costly data.

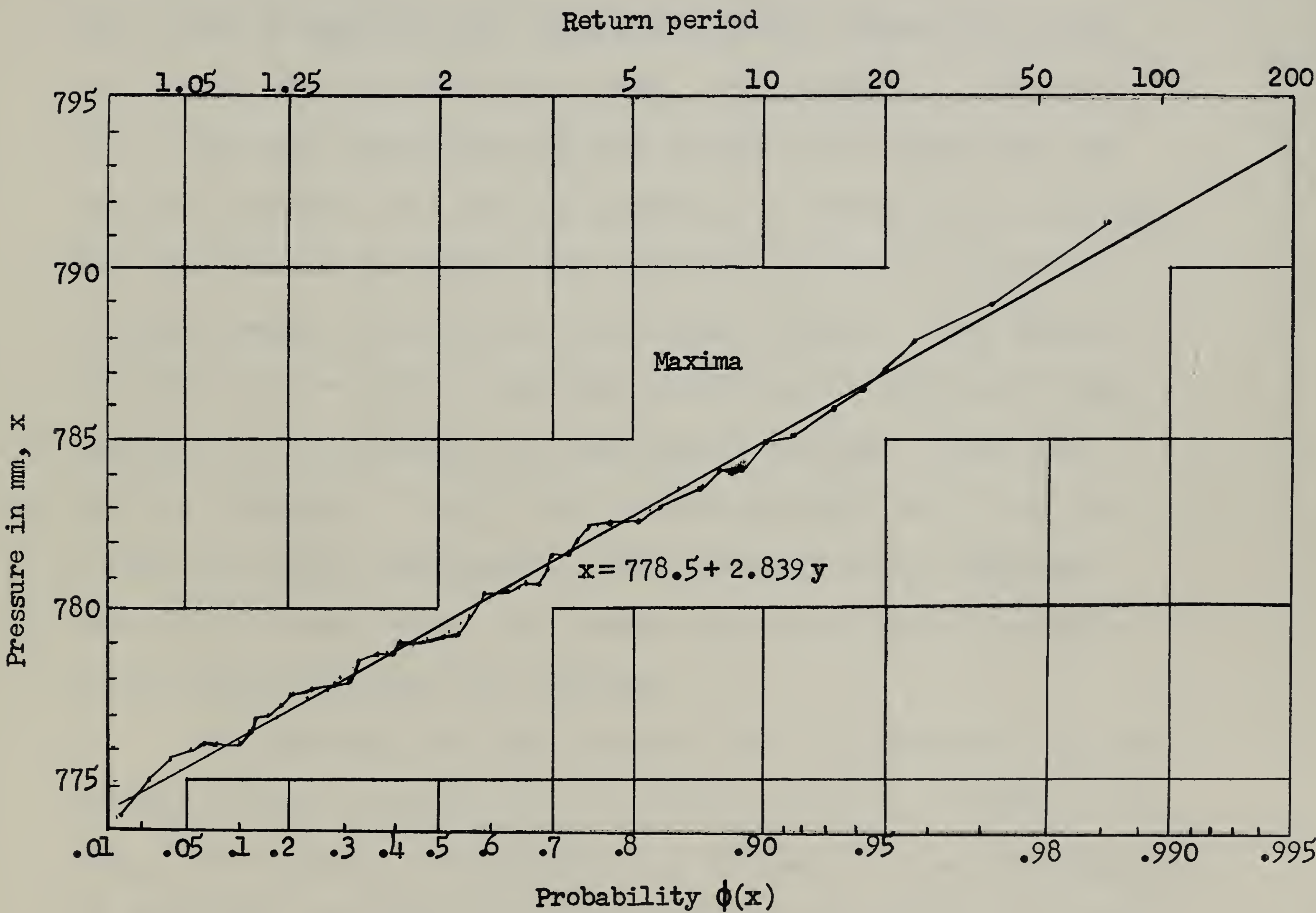

Fig. 2--Annual maxima of atmospheric pressures, Bergen, Norway, 1857-1926.

Smallest Values.

Extreme-value theory can also be used to study the smallest observations, since the corresponding limiting distribution is simply related to the one previously given. The steps in applying the "smallest-value" theory are very similar to the largest-value case. For example, engineers have long been interested in the problem of predicting the tensile strength of a bar or specimen of homogeneous material. One approach is to regard the specimen as being composed of a large number of pieces of very short length. The tensile strength of the entire specimen is obviously limited by the strength of the weakest of these small pieces. Thus the tensile strength at which the entire specimen will fail is a smallest-value phenomenon. The smallest-value approach can be used even though the number and individual strengths of the "small pieces" are unknown.

This method has been applied with considerable success by Shigeo Kase in studying the tensile testing of rubber [11]. Using 200 specimens obtained so as to assure as much homogeneity as possible, he found that the observed distribution of their tensile strengths could be fitted remarkably well by the extreme-value distribution for smallest values. The fitted curve given by the data indicates that one-half of a test group of specimens

may be expected to break under a tensile stress of 105 kg./cm.^{2} or more, while only 1 in a thousand will survive a stress exceeding 126 kg./cm.^{2}.

Missing Observations.

It has been found that fatigue life of specimens under fixed stress can be treated in the same manner as tensile strength -- by using the theory of smallest values. An extensive application of this method is given in [15].

In such cases and in other situations, tests must be stopped before all specimens have failed. This results in a sample with missing values -- a "censored" sample. Methods for handling such data are included in [15].

References

Primary references.

[1] E.J. Gumbel, "Statistical Theory of Extreme Values and Some Practical Applications", National Bureau of Standards Applied Mathematics Series 33, 1954, (U.S. Government Printing Office, Washington, D.C.).

[2] National Bureau of Standards, "Probability Tables for the Analysis of Extreme-Value Data", Applied Mathematics Series 22, 1953. (U.S. Government Printing Office, Washington, D.C.).

[3] E.J. Gumbel, "Statistics of Extremes", Columbia University Press, New York, 1958.

Selected references on specific applications.

[4] American Society for Testing Materials, "Symposium on Fatigue with Emphasis on Statistical Approach - II," ASTM publication 137, Philadelphia, 1952. Discussions by E.J. Gumbel on pp. 22 and 56.

[5] B. Epstein and H. Brooks, "The Theory of Extreme Values and Its Implications in the Study of the Dielectric Strength of Paper Capacitors", J. Applied Physics, Vol. 19, pp. 544-550 (1948).

[6] A.M. Freudenthal and E.J. Gumbel, "On the Statistical Interpretation of Fatigue Tests", Proceedings of the Royal Society A, 216, pp. 309-332 (1953).

[7] A.M. Freudenthal and E.J. Gumbel, "Minimum Life in Fatigue", Journal of the Am. Stat. Assoc., Vol. 49, pp. 575-597, (September 1954).

[8] Y.C. Fung, "Statistical Aspects of Dynamic Loads", Journal of the Aeronautical Sciences, Vol. 20, pp. 317-330, (May 1954).

[9] E.J. Gumbel and P.G. Carlson, "Extreme Values in Aeronautics", Journal of the Aeronautical Sciences, Vol. 21, No. 6, pp. 389-398 (June 1954).

[10] E.J. Gumbel and J. Lieblein, "Some Applications of Extreme-Value Methods", The American Statistician, Vol. 8, No. 4, December 1954.

[11] Shigeo Kase, "A Theoretical Analysis of the Distribution of Tensile Strength of Vulcanized Rubber", Journal of Polymer Science, Vol. 11, No. 5, pp. 425-431, November 1953.

[12] Julius Lieblein, "On the Exact Evaluation of the Variances and Covariances of Order Statistics in Samples from the Extreme-Value Distribution", Annals of Mathematical Statistics, 24, No. 2, pp. 282-287, June 1953.

[13] Julius Lieblein, "A New Method of Analyzing Extreme-Value Data," NACA Technical Note 3053, National Advisory Committee for Aeronautics, January 1954.

[14] Anonymous: "Extreme-Value Methods for Engineering Problems", National Bureau of Standards Technical News Bulletin, 38, No. 2, pp. 29-31, February 1954.

[15] Julius Lieblein and Marvin Zelen, "Statistical Investigation of the Fatigue Life of Deep Groove Ball Bearings", Journal of Research of the National Bureau of Standards, 57, No. 5, November 1956.

4. Transformations

Statistical techniques have to be based on some assumptions. There is often the assumption of normality; in some techniques there is the assumption of equal variance among groups; etc. Analysis of variance tests (in Part III) depend on assumptions of normality, additivity, independence and equality of variance. Real-life data does not always fit the assumptions required for commonly used methods of analysis. Where this is the case, a transformation (change of scale) applied to the raw data may put the data in such form that a conventional analysis can be validly performed.

Many physical measurement processes produce normally-distributed data. Some do not. If the data are definitely known to be non-normal, there are distribution-free techniques available (see Part I, section 2.7). If, however, a simple transformation is known to normalize the particular kind of data, one might prefer to use normal techniques on the transformed data, rather than use distribution-free methods. Certain kinds of data, for example, are pretty definitely known to be approximately normal in logs, and the use of a log transformation in these cases may become routine.

Transformations are also used to achieve equality of variance among groups, and quite often both equality of variance and approximate normality are achieved simultaneously. Transformations are not recommended for uncritical use by the amateur. From small amounts of data it is often not clear which transformation to use and it is not always clear whether or not the transformation has helped. Graphical methods can be used to examine the effect of transformation --- see, for example [2]. The table following, taken from a larger table of Bartlett [1], shows several frequently used variance stabilizing transformations.

[1] M.S. Bartlett, "The Use of Transformations", Biometrics, Vol. 3, No. 1, 1947 (pp. 39-52).

[2] A. Hald, "Statistical Theory with Engineering Applications" John Wiley and Sons, New York, 1952.

Table of Transformations

Transformation	Appropriate Situation	Variance on Transformed Scale	Notes
$\sqrt{x}$	Mean equal to or proportional to variance Example: Poisson distribution Var = mean Other distributions Var = λ^2(mean)	 0.25 $0.25\,\lambda^2$	Use $\sqrt{x+\frac{1}{2}}$ for small integers
$\log_e x$ or $\log_{10} x$	Mean proportional to standard deviation Var = $\lambda^2(\text{mean})^2$ Example: Log normal distribution	For $\log_e x$, For $\log_{10} x$, $.189\lambda^2$	Use $\log_e(x+1)$ or $\log_{10}(x+1)$ to avoid difficulties with zeros
$\sin^{-1}\sqrt{x}$ (radians) $\sin^{-1}\sqrt{x}$ (degrees)	Variance = $\frac{p(1-p)}{n}$ Example: Binomial distribution	For radians, .25/n For degrees, 821/n	
Probit	"All or nothing" response		See Part II, section

5. Notes on Statistical Computations.

5.1 Coding in statistical computations. Coding is the term applied to arithmetical operations on the original data intended to make the numbers easier to handle in computation. The possible coding operations are:

(a) multiplication (or its inverse, division) to change the order of magnitude of the recorded numbers for computing purposes.

(b) addition (or its inverse, subtraction) of a constant - applied to recorded numbers which are nearly equal, to reduce the number of figures which need be carried in computation.

Particularly when the recorded results contain non-significant zeros, (e.g., numbers like .000121 or like 11,100) coding is clearly desirable. There is obviously no point in copying these zeros a large number of times, or in adding additional useless zeros when squaring, etc. Of course, these results could have been given as 121×10^{-4} or 11.1×10^{3} in which case coding for order of magnitude would not be necessary.

The purpose of coding is to save labor in computation. On the other hand, the process of coding and decoding the results introduces more opportunities for error in computation. The decision of whether to code or not must be carefully

considered, weighing the advantage of saved labor against the disadvantage of more likely mistakes. With this in mind the following rules are given for coding and decoding:

1. The whole set of observed results must be treated alike.

2. The possible coding operations are the two general types of arithmetic operations - (a) addition (or subtraction) and (b) multiplication (or division). Either a or b or both together may be used as necessary to make the original numbers more tractable.

3. Needless to say, keep careful note of how the data have been coded.

4. The desired computation is carried out on the coded data.

5. The process of decoding a computed result depends on the computation that has been carried out, and therefore is shown below separately for several common computations.

   a. The mean is affected by every coding operation. Therefore, apply the inverse operation and reverse the order of coding to put the coded mean back into original units. For example, if the data have been coded by first multiplying by 10,000 and then subtracting 120, decode mean by adding 120 and then dividing by 10,000

Example

Obs. results	Coded
.0121	1
.0130	10
.0125	5
Mean .0125	coded mean 5

Decoding:

$$\text{Mean} = \frac{\text{Coded mean} + 120}{10{,}000} = \frac{125}{10{,}000}$$

$$= .0125$$

b. A standard deviation computed on coded data is affected only by multiplication or division. The standard deviation is a measure of dispersion, like the range, and is not affected by adding or subtracting a constant to the whole set of data. Therefore, if the data have been coded by addition or subtraction only, no adjustment is needed in the computed standard deviation. If the coding has involved multiplication (or division) the inverse operation must be applied to the computed standard deviation to bring it back to original units.

c. A variance computed on coded data will have to be multiplied by the square of the coding factor if division has been used in coding; divided by the square of the coding factor if multiplication was used in coding.

d. "Coding" which involves loss of significant figures: Notice that the kind of coding discussed so far has not involved any loss in significant figures. There is a process, however, called "coding" that involves both coding and rounding. This usually arises where the data are considered to be too-finely recorded for the purpose. It is sometimes a desirable pro-

cedure but one should realize that there is more here involved than just arithmetic. This type of coding involves a loss of information and therefore its application requires more judgment.

For example suppose the data consisted of weights of shipments (in pounds) of some bulk material:

7,123
10,056
100,310
5,097
543
⋮
etc.

If the interest is in the average, and the range is large, as here, one may decide to save trouble by working with weights to the nearest hundred pounds. The "coded" data would then be:

71
101
1003
51
5
⋮
etc.

Whether this is sufficient information will depend on the range of the data and the uses to which the results will be put. The only caution is that a higher order of judgment is required than was required in the previous examples. The grouping of data that is done in making a frequency distribution is coding of this nature and does involve rounding.

5.2 Rounding in statistical computations

5.2.1 Rounding of numbers: Rounded numbers are inherent to the process of reading and recording data. The readings of an experimenter are rounded numbers to start with since all measuring equipment is of limited accuracy. Often he records the results to even less accuracy than is attainable with the available equipment simply because they are completely adequate for his immediate purpose. Computers are often required to round numbers either to simplify the arithmetic calculations or because it cannot be avoided, as when 3.1416 is used for π or 1.414 is used for $\sqrt{2}$. When a number is to be rounded to a specific number of significant figures, the rounding procedure should be carried out in accordance with the following rule:

(1) When the figure next beyond the last place to be retained is less than 5, the figure in the last place retained should be kept unchanged.

(2) When the figure next beyond the last figure or place to be retained is greater than 5, the figure in the last place retained should be increased by 1.

(3) When the figure next beyond the last figure to be retained is 5, and

  (a) there are no figures or only zeros beyond this 5,an odd figure in the last place to be retained should be increased by 1; an even figure should be kept unchanged;

  (b) if the 5 is followed by any figures other than zero, the figure in the last place to be retained should be increased by 1 whether odd or even.

The final-recorded result should be rounded off in one step to the most precise value available and not in two or more steps of successive roundings.

5.2.2 Rounding the results of single arithmetic operations: Nearly all numerical calculations arising in the problems of everyday life are in some way approximate. The aim of the

computer should be to obtain results consistent with the data with a minimum of labor. He can be guided in the various arithmetical operations by some basic rules regarding significant figures and the rounding of data.

1. Addition: When several approximate numbers are to be added, the sum should be rounded to the number of decimal places (not significant figures) no greater than in the addend which has the smallest number of decimal places.

   Although the result is determined by the least accurate of the numbers entering the operation, one more decimal place in the more accurate numbers should be retained thus eliminating inherent errors in the numbers.

2. Subtraction: When one approximate number is to be subtracted from another, they must both be rounded off to the same place before subtracting.

   Errors arising from the subtraction of nearly equal approximate numbers are frequent and troublesome often making the computation practically worthless. They can sometimes be avoided when the two nearly equal numbers can be approximated to more significant digits.

3. <u>Multiplication</u>: If the less accurate of two approximate numbers contains n significant digits, their product can be relied upon for n digits only, and should not be written with more.

> As a practical working plan, carry intermediate computations out in full and round off the final result in accordance with the rule stated above.

4. <u>Division</u>: If the less accurate of either the dividend or the divisor contains n significant digits, their quotient can be relied upon for n figures only and should not be written with more.

> Carry intermediate computations out in full and round off the final result in accordance with the rule stated above.

5. <u>Powers and Roots</u>: If an approximate number contains n significant digits, its power can be relied upon for n digits only; its root can be relied upon for at least n digits.

6. <u>Logarithms</u>: If the mantissa of the logarithm in an n-place log table is not in error by more than two units in the last significant figure, the antilog is correct to n-1 significant figures.

The foregoing statements are rules only. Fuller expla-

nations of the rules, tests for determining the accuracy of computations when it is desirable to attain or to state a certain degree of precision, and the effects of rounding on statistical analyses of large numbers of observations are ably handled by Scarborough [1] and Eisenhart [2].

5.2.3 Rounding the results of a series of arithmetic operations: Most engineers and physical scientists well know the rules for reporting a result to the proper number of significant figures. From a computational point of view, they know these rules too well. It is perfectly true, for example, that a product of two numbers should be reported to the same number of significant figures as the least accurate of the two numbers. It is not so true that the two numbers should be rounded to the same number of significant figures before multiplication. A better rule is to round the more accurate number to one more figure than the less accurate one and then round the product to the same number as the less accurate one. The great emphasis against reporting more figures than are reliable has led to a prejudice against carrying enough figures in computation.

Assuming that the reader is familiar with the rules of the preceding section regarding significant figures in a single arithmetical operation, this section will stress the less well-

known difficulties which arise in a computation consisting of a long series of different arithmetic operations. In such a computation, strict adherence to the rules at each stage can wipe out all meaning from the final results.

For example in computing the slope of a straight line fitted to data with 3 significant figures, one would surely not report the slope to 7 significant figures, but if one rounds to 3 significant figures, after each necessary step in the computation, one might end up with no significant figures in the slope.

It is easily demonstrated by carrying out a few computations of this nature that there is real danger of losing all significance by too-strict adherence to the rules, devised for use at the final stage. The greatest trouble of this kind comes where we must subtract two nearly equal numbers, and many statistical computations involve such a step.

The rules generally given for rounding off were given in a period when the average was the only property of interest in a set of data. Reasonable rounding does little damage to the average. Almost always now, however, we calculate the standard deviation, and this statistic does suffer from too-strict rounding. Suppose we have a set of numbers:

	3.1
	3.2
	3.3
Avg. =	3.2

If the three numbers are rounded off to 1 significant figure, they are then all the same. The average of the rounded figures is the same as the rounded average of the original figures, but all information about the variation in the original numbers is lost.

The generally recommended procedure is to carry two or three extra figures throughout the computation and then to round off the final reported answer (e.g., standard deviation, slope of a line, etc.) to a number of significant figures consistent with the original data.

[1] J.B. Scarborough, Chapter I - Numerical Mathematical Analysis, the Johns Hopkins Press, Baltimore (1930).

[2] C. Eisenhart, Chapter 4 - Techniques of Statistical Analysis, McGraw-Hill Book Co., New York (1947).

U. S. DEPARTMENT OF COMMERCE
Lewis L. Strauss, *Secretary*

NATIONAL BUREAU OF STANDARDS
A. V. Astin, *Director*

THE NATIONAL BUREAU OF STANDARDS

The scope of activities of the National Bureau of Standards at its headquarters in Washington, D. C., and its major laboratories in Boulder, Colo., is suggested in the following listing of the divisions and sections engaged in technical work. In general, each section carries out specialized research, development, and engineering in the field indicated by its title. A brief description of the activities, and of the resultant publications, appears on the inside front cover.

WASHINGTON, D. C.

Electricity and Electronics. Resistance and Reactance. Electron Devices. Electrical Instruments. Magnetic Measurements. Dielectrics. Engineering Electronics. Electronic Instrumentation. Electrochemistry.

Optics and Metrology. Photometry and Colorimetry. Optical Instruments. Photographic Technology. Length. Engineering Metrology.

Heat. Temperature Physics. Thermodynamics. Cryogenic Physics. Rheology. Engine Fuels. Free Radicals Research.

Atomic and Radiation Physics. Spectroscopy. Radiometry. Mass Spectrometry. Solid State Physics. Electron Physics. Atomic Physics. Neutron Physics. Radiation Theory. Radioactivity. X-rays. High Energy Radiation. Nucleonic Instrumentation. Radiological Equipment.

Chemistry. Organic Coatings. Surface Chemistry. Organic Chemistry. Analytical Chemistry. Inorganic Chemistry. Electrodeposition. Molecular Structure and Properties of Gases. Physical Chemistry. Thermochemistry. Spectrochemistry. Pure Substances.

Mechanics. Sound. Mechanical Instruments. Fluid Mechanics. Engineering Mechanics. Mass and Scale. Capacity, Density, and Fluid Meters. Combustion Controls.

Organic and Fibrous Materials. Rubber. Textiles. Paper. Leather. Testing and Specifications. Polymer Structure. Plastics. Dental Research.

Metallurgy. Thermal Metallurgy. Chemical Metallurgy. Mechanical Metallurgy. Corrosion. Metal Physics.

Mineral Products. Engineering Ceramics. Glass. Refractories. Enameled Metals. Concreting Materials. Constitution and Microstructure.

Building Technology. Structural Engineering. Fire Protection. Air Conditioning, Heating, and Refrigeration. Floor, Roof, and Wall Coverings. Codes and Safety Standards. Heat Transfer.

Applied Mathematics. Numerical Analysis. Computation. Statistical Engineering. Mathematical Physics.

Data Processing Systems. SEAC Engineering Group. Components and Techniques. Digital Circuitry. Digital Systems. Anolog Systems. Application Engineering.

- Office of Basic Instrumentation.
- Office of Weights and Measures.

BOULDER, COLORADO

Cryogenic Engineering. Cryogenic Equipment. Cryogenic Processes. Properties of Materials. Gas Liquefaction.

Radio Propagation Physics. Upper Atmosphere Research. Ionospheric Research. Regular Propagation Services. Sun-Earth Relationships. VHF Research. Ionospheric Communication Systems.

Radio Propagation Engineering. Data Reduction Instrumentation. Modulation Systems. Navigation Systems. Radio Noise. Tropospheric Measurements. Tropospheric Analysis. Radio Systems Application Engineering. Radio-Meteorology.

Radio Standards. High Frequency Electrical Standards. Radio Broadcast Service. High Frequency Impedance Standards. Electronic Calibration Center. Microwave Physics. Microwave Circuit Standards.

NBS