Simplified Statistics for Small Numbers of Observations

R. B. Dean, and W. J. Dixon

Anal. Chem., **1951**, 23 (4), 636-638• DOI: 10.1021/ac60052a025 • Publication Date (Web): 01 May 2002

Simplified Statistics for Small Numbers of Observations

R. B. DEAN AND W. J. DIXON
University of Oregon, Eugene, Ore.

Several short-cut statistical methods can be used with considerable saving of time and labor. This paper presents some useful techniques applicable to a small number of observations. Advantages of the median as a substitute for the average are discussed. The median is especially useful when a quick decision must be made and gross errors are suspected. The range may be used to estimate the standard deviation and confidence intervals with little loss in precision for small numbers of observations. A convenient criterion for rejecting gross errors is presented. A table gives efficiencies and conversion factors for two to ten observations. The statistical methods presented are sufficiently simple to find use by those who feel that "statistics are too much trouble." These methods are usually more exact than the statistics of large numbers applied directly to small numbers of observations.

ALTHOUGH all scientific workers are aware that their numerical observations may be analyzed by statistical methods, many have expressed the opinion that the classical methods of statistics are unnecessarily complicated. It is not generally realized that there are available a number of short-cut statistics which will yield a large fraction of the available information with less computation. This paper calls attention to some of these statistics which are applicable to ten or fewer independent observations. These methods are especially valuable when it is necessary to make quick decisions.

SCATTERING OF DATA

In general, repeated observations of the "same" quantity will not be identical. We normally recognize this tendency toward scattering of the data by calculating an average value which is an estimate of the "true" value (the average of the entire population of measurements). For this purpose it is customary to calculate the arithmetic mean, $\bar{x}$. In order to follow general usage, the term "average" is used synonymously with "arithmetic mean" in this paper and is denoted by the symbol $\bar{x}$.

It is also generally recognized that the extent of the scattering is inversely related to the reliability of the measurements. The measures of this scattering which are frequently used are the standard deviation, s, and the average deviation.

Whenever a large number of independent fluctuations contribute to the spread in the values of a population, the values will approximate a normal bell-shaped error curve. This distribution of the population is the one for which most statistical tables have been prepared and is fortunately the one observed frequently in practice. In this paper all conclusions are based on a normally distributed population.

The tabulated values presented apply only if independent random samples are taken from the original population. For example, if five samples of an ore are individually dissolved and made up to volume and the iron in these samples is then determined in duplicate on aliquots of each sample, there would be only five independent observations representing the five averages of the aliquots.

Subject to the conditions that the observations are independent and normally distributed, the best statistics known are the arithmetic mean, $\bar{x}$, for the average, and the standard deviation, s, or the variance, s^2, for the spread.

The term "best" applied to $\bar{x}$ and s^2 is based on the fact that these estimates are the most efficient known estimates for the mean and variance, respectively, of the population when the observations follow the normal (or Gaussian) law of errors. The term "most efficient" indicates a minimum variance or dispersion for the sampling distributions of these statistics—i.e., least amount of spread for repeated estimates of the population mean or population variance.

The classical statistical theory for large numbers of observations is presented, in part, in most textbooks of analytical chemistry (*11, 16*). Rarely do these books recognize the fact that parts of this theory are inaccurate for small numbers of observations. For example, we can no longer expect the interval $\bar{x} \pm 0.6745s/\sqrt{n}$ to include the population mean 50% of the time when the $\bar{x}$ and s are calculated from n observations. Tables for the correct multiplier (Student's t) to replace 0.6745 for various numbers of observations and for various probabilities of covering the population mean are available (*8, 14*). When n is greater than 20 to 30, Student's t is closely approximated by the normal curve of error. To make use of the exact statistical methods, most chemists will have to use tables for the chosen number of observations, for very few analyses are run on as many as 20 samples or replicates. The less efficient statistics presented in this paper also require the use of tables, but the computations are less tedious and the efficiency is adequate for many purposes. Tables of useful functions are presented for two to ten observations (*5*). Accurate values of some of these functions are not yet available for larger numbers of observations.

For convenience a series of observations will be arranged in ascending order of magnitude and assigned the symbols:

$$x_1, x_2, x_3, \ldots\ldots x_{n-1}, x_n$$

The best estimate of the central value is the average, $\bar{x}$

$$\bar{x} = \frac{(x_1 + x_2 + \ldots + \ldots x_{n-1} + x_n)}{n} = \frac{(\Sigma x_i)}{n} \qquad (1)$$

The spread of the values is measured most efficiently by the variance, s^2. The computation is as follows:

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots (x_n - \bar{x})^2}{n - 1} = \frac{\Sigma(x_i - \bar{x})^2}{n - 1} \qquad (2)$$

and

$$s = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n - 1}} \qquad (3)$$

It is essential to use $n - 1$ instead of n in the denominator to avoid a biased estimate of σ^2, the variance of the population. Instead of calculating the average, $\bar{x}$, we can use the median value, M, as a measure of the central value. M is the middle value or the average of the two values nearest the middle. If $n = 5$, then $M = x_3$; if $n = 6$, then $M = (x_3 + x_4)/2$. M is an estimate of the central value and has the advantage that it is not influenced markedly by extraneous values. The efficiency (defined here as the ratio of the variances of the sampling distributions of these two estimates of the "true" mean) of M, E_M, is given in column 2 of Table I (*4, 10*). It varies from 1.00 for only two observations (where the median is identical with the average, $\bar{x}$) down to 0.64 for large numbers of observations. It can be shown (*4, 10*) that the numerical value of the efficiency implies, for example, that the median, M, from 100 observations (where the efficiency is 0.64) conveys as much information about

Table I. Efficiencies and Conversion Factors for Two to Ten Observations

No. of Observations, n	Efficiency: Of median, E_M	Efficiency: Of range, E_W	Deviation Factor, K_W	Confidence Factors: Student's $t_{0.95}$	Confidence Factors: Student's $t_{0.99}$	Confidence Factors: Range $t_{w\,0.95}$	Confidence Factors: Range $t_{w\,0.99}$	Rejection Quotient $Q_{0.90}$
2	1.00	1.00	0.89	12.7	64	6.4	31.83	...
3	0.74	0.99	0.59	4.3	10	1.3	3.01	0.94
4	0.84	0.98	0.49	3.2	5.8	0.72	1.32	0.76
5	0.69	0.96	0.43	2.8	4.6	0.51	0.84	0.64
6	0.78	0.93	0.40	2.6	4.0	0.40	0.63	0.56
7	0.67	0.91	0.37	2.5	3.7	0.33	0.51	0.51
8	0.74	0.89	0.35	2.4	3.5	0.29	0.43	0.47
9	0.65	0.87	0.34	2.3	3.4	0.26	0.37	0.44
10	0.71	0.85	0.33	2.26	3.25	0.23	0.33	0.41
∞	0.64	0.00	0.00	1.960	2.576	0.00	0.00	0.00

the central value of the population as the average, $\bar{x}$, calculated from 64 observations.

The median of a small number of observations can usually be determined by inspection. Although it is less efficient than the average if the population is normally distributed, M may be more efficient than the average, $\bar{x}$, if gross errors are present. A chemist is frequently faced with the problem of deciding whether or not to reject an observation that deviates greatly from the rest of the data. The median is obviously less influenced by a gross error than is the average. It may be desirable to use the median in order to avoid deciding whether a gross error is present. It has been shown that for three observations from a normal population the median is better than the "best two out of three"—i.e., average of the two closest observations (*18*).

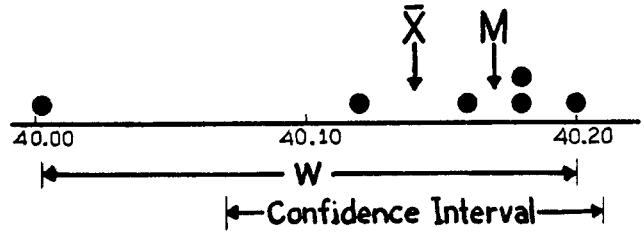

Figure 1. Graphical Presentation of Data Chosen in Example

No attempt is made to give a complete treatment of the problem of gross errors here, but an approach based on a recent fairly complete summary of this problem (*2*, *3*) is given in the last section of this paper.

The following series of observations represents calculated percentages of sodium oxide in soda ash. The data have been arranged in order of magnitude, and are presented graphically in Figures 1 and 2.

40.02	x_1
40.12	x_2
40.16	x_3
40.18	x_4
40.18	x_5
40.20	x_6
$\bar{x}$ = 40.14	M = 40.17

The first value may be considered doubtful. The average is 40.14 and the median is 40.17. We may wish to place more confidence in the median, 40.17, than in the average, 40.14. (If the series were obtained in the order given, we might justifiably rule out the first observation as showing lack of experience or other "starting up" errors.)

The range of the observations, w, is the difference between the greatest and least value; $w = x_n - x_1$. The range, w, is a convenient measure of the dispersion. It is highly efficient for ten or fewer observations, as is evident from column 3 of Table I (*5*, *17*). This high relative efficiency arises in part from the fact that the standard deviation is a poor estimate of the dispersion for a small number of observations, even though it is the best known estimate for a given set of data. The range is also more efficient than the average deviation for less than eight observations. To convert the range to a measure of dispersion independent of the number of observations we must multiply by the factor, K_w, which is tabulated in column 4 (*5*, *17*). This factor adjusts the range, w, so that on the average we estimate the standard deviation of the population. The product $wK_w = s_w$ is therefore an estimate of the standard deviation which can be obtained from the range. In the series presented above, the range is 0.18. From the table we find that K_w for six observations is 0.40, so $s_w = 0.072$. The standard deviation, s, calculated according to Equation 3, equals 0.066.

As n increases the efficiency of the range decreases. If the data are randomly presented, such as in order of production rather than in order of size, the average of the ranges of successive subgroups of 6 or 8 is more efficient than a single range. The same table of multipliers is used, the appropriate K_w being determined by the subgroup size. This factor is the reciprocal of d_2, the divisor used—e.g., in quality control work—to convert the range to an estimate of the standard deviation. Tables of d_2 are given for n up to 100 (*1*, *9*), but Lord has shown in a recent publication (*13*, *15*) that the most efficient subgroup size for estimating the standard deviation from the average range is from 6 to 8.

CONFIDENCE LIMITS

Although s and s_w are useful measures of the dispersion of the original data, we are usually more interested in the confidence interval or confidence limits. By the confidence interval we mean the distance on either side of $\bar{x}$ in which we would expect to find, with a given probability, the "true" central value. For example, we would expect the true average to be covered by the 95% confidence limits 95% of the time. By taking wider confidence limits, say 99% limits, we can increase our chances of catching the "true" average but the interval will necessarily be longer. The shortest interval for a given probability corresponds to the "t test" of Student (*8*, *14*). $\bar{x} \pm ts/\sqrt{n}$ is the confidence interval. t varies with the number of observations and the degree of confidence desired. For convenience t is tabulated for 95 and 99% confidence values (columns 5 and 6).

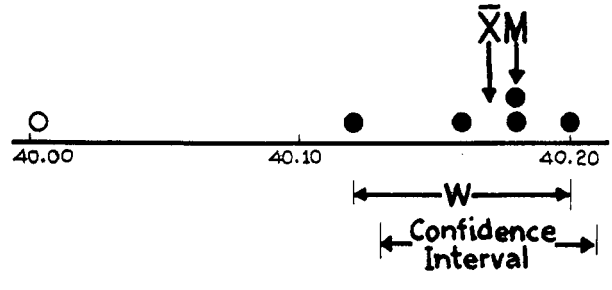

Figure 2. Graphical Presentation of Charges Produced by Rejection of a Questionable Observation

Confidence limits might be calculated in a similar manner, using s_w obtained from the range and a corresponding but different table for t. However, it is more convenient to calculate the limits directly from the range as $\bar{x} \pm wt_w$. The factor for converting w to s_w has been included in the quantity t_w, which is tabulated in column 7 for 95% confidence and in column 8 for 99% confidence (*6*, *12*).

For a set of six observations t_w is 0.40 and the range of the set previously listed is 0.18, so wt_w equals 0.072. Hence, we can

report as a 95% confidence interval, 40.14 ± 0.072. If we calculate a confidence interval using the tables of t and the calculated value of s we find $t = 2.6$ at the 95% level and $ts\sqrt{6} = 0.078$. We would report 40.14 ± 0.078, a result substantially the same as that obtained from the range of this particular sample.

If the standard deviation of a given population is known or assumed from previous data, we can use the normal curve to calculate the confidence limits. This situation may arise when a given analysis has been used for fifty or so sets of analyses of similar samples. The standard deviation of the population may be estimated by averaging the variance, s^2, of the sets of observations or from K_w times the average range of the sets of observations. The interval $\bar{x} \pm 1.96s/\sqrt{m}$, where $\bar{x}$ is computed from a new set of m observations, can be expected to include the population average 95% of the time.

EXTRANEOUS VALUES

Simplified statistics have been presented, which enable one to obtain estimates of a central value and to set confidence limits on the result. Let us consider the problem of extraneous values. The use of the median eliminates a large part of the effect of extraneous values on the estimate of the central value, but the range obviously gives unnecessary weight to an extraneous value in an estimate of the dispersion. On the other hand, we may wish to eliminate extraneous values which fail to pass a screening test.

One very simple test, the Q test, is as follows:

Calculate the distance of a doubtful observation from its nearest neighbor, then divide this distance by the range. The ratio is Q where

$$Q = (x_2 - x_1)/w$$

or

$$Q = (x_n - x_{n-1})/w$$

If Q exceeds the tabulated values, the questionable observation may be rejected with 90% confidence (*2*, *3*, *7*). In the example cited 40.02 is the questionable value and

$$Q = (40.12 - 40.02)/(40.20 - 40.02) = 0.56$$

This just equals the tabulated value of 0.56 for 90% confidence, so we may wish to reject the value.

If we had decided to reject an extreme low value if Q is as large or larger than would occur 90% of the time in sets of observations from a normal population, we would now reject the observation 40.02. In other words, a deviation this great or greater would occur by chance only 10% of the time at one or the other end of a set of observations from a normally distributed population.

By rejecting the first value we increase the median from 40.17 to 40.18 and the average from 40.14 to 40.17 (see Figure 2). The standard deviation, s, falls from 0.067 to 0.030 (it might be greater than 0.030 if we have erred in rejecting the value 40.02) and s_w falls from 0.072 to 0.034. The 95% confidence interval corresponding to the t test is now 40.17 ± 0.038 and from the median and range is 40.18 ± 0.040, a reduction of about one half in the length of the interval. (The 95% is only approximate, as we have performed an intermediate statistical test.)

LITERATURE CITED

(1) Am. Soc. Testing Materials, "Manual on Presentation of Data," 1945.
(2) Dixon, W. J., *Ann. Math. Stat.*, **21**, 488 (1950).
(3) *Ibid.*, in press.
(4) Dixon, W. J., and Massey, F. J., "Introduction to Statistical Analysis," p. 238, New York, McGraw-Hill Book Co., 1951.
(5) *Ibid.*, p. 239.
(6) *Ibid.*, p. 242.
(7) *Ibid.*, p. 319.
(8) Fisher, R. A., and Yates, Frank, "Statistical Tables for Biological, Agricultural, and Medical Research," Edinburgh, Oliver and Boyd, 1943.
(9) Grant, E. L., "Statistical Quality Control," p. 536, New York, McGraw-Hill Book Co., 1946.
(10) Kendall, M. G., "Advanced Theory of Statistics," Vol. II, p. 6, London, Charles Griffin & Co., 1946.
(11) Kolthoff, I. M., and Sandell, E. B., "Textbook of Quantitative Inorganic Analysis," p. 270, New York, Macmillan Co., 1948.
(12) Lord, E., *Biometrika*, **34**, 41 (1947).
(13) *Ibid.*, **37**, 64 (1950).
(14) Merrington, Maxine, *Ibid.*, **32**, 300 (1942).
(15) Pearson, E. S., *Ibid.*, **37**, 88 (1950).
(16) Pierce, W. C., and Haenisch, E. L., "Quantitative Analysis," p. 48, New York, John Wiley & Sons, 1948.
(17) Tippett, L. H. C., *Biometrika*, **17**, 364 (1925).
(18) Youden, W., unpublished data.

RECEIVED May 27, 1950. Presented before the Division of Analytical Chemistry at the 117th Meeting of the AMERICAN CHEMICAL SOCIETY, Houston, Tex.

Determination of Intrinsic Low Stress Properties of Rubber Compounds

USE OF INCLINED PLANE TESTER

W. B. DUNLAP, JR., C. J. GLASER, JR., AND A. H. NELLEN
Lee Rubber & Tire Corp., Conshohocken, Pa.

PHYSICAL test methods for vulcanized natural and synthetic rubber compounds as commonly employed today do not provide an accurate basis for evaluating service performance, especially in the case of tire compounds. Poor correlation between ordinary laboratory physical tests and observed performance in tire service on the road is the rule rather than the exception.

Substantial improvement is needed in developing tests which are less subject to the human variable, tests in which the criteria examined are more in line with the actual service conditions, and tests which are mechanically simple and easy to perform. One significant advance has been made in this direction by the development of the National Bureau of Standards' comparatively new strain test (*3*). This paper reports progress made by the Lee laboratories in developing one phase of an evaluation system which more accurately predicts compound performance than will the methods previously employed.

As a starting point for this work, it was decided first to determine the important physical characteristics of a tire compound and, secondly, to devise a new test or alter an existing test to give the desired information. It is assumed that an accurate appraisal of the physical characteristics listed in Table I is required.