THE MEAN SQUARE SUCCESSIVE DIFFERENCE

By J. von Neumann,[1] R. H. Kent, H. R. Bellinson and B. I. Hart

Aberdeen Proving Ground

1. Introduction. In making measurements, every precaution is generally taken to hold the conditions of the experiment constant, in order that the population, whose parameters are to be estimated from the observations, shall remain fixed throughout the experiment. One wishes each observation to come from the same population, or what is the same thing if normality is assumed, from populations having the same means and standard deviations.

There are cases, however, where the standard deviation may be held constant, but the mean varies from one observation to the next. If no correction is made for such variation of the mean, and the standard deviation is computed from the data in the conventional way, then the estimated standard deviation will tend to be larger than the true population value. When the variation in the mean is gradual, so that a trend (which need not be linear) is shifting the mean of the population, a rather simple method of minimizing the effect of the trend on dispersion is to estimate standard deviation from differences. It is for this purpose that the mean square successive difference

$$\delta^2 = \frac{\sum_{i=1}^{n-1} (x_{i+1} - x_i)^2}{n - 1} \tag{1}$$

is suggested. The subscript i in this expression refers to the temporal order of the observation x_i.

In using δ^2 for estimating standard deviation, the distribution of δ^2 in random samples is of interest, since questions of bias, efficiency, and confidence interval require consideration. δ^2 may be used, in addition, to determine whether a trend actually exists; in this case one must know whether δ^2 differs significantly from

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}, \tag{2}$$

which measures variance independently of the order of the observations, and consequently includes the effect of the trend.

[1] Institute for Advanced Study, Princeton, N. J. Also member of Scientific Advisory Committee of the Ballistic Research Laboratory, Aberdeen Proving Ground.

The distribution of δ^2 is considered in this paper; it is hoped that others will shortly publish methods of estimating the probability that $\delta^2 \leqq ks^2$ as a function of k and the sample size n.

2. History. A somewhat similar procedure is suggested by "Student" [1] and E. S. Pearson [2] who consider the situation in which a shift may occur in the mean of the population, but where pairs of observations may be made with no shift in mean between them; standard deviation may be estimated from the differences between these pairs. The method can be generalized, and

$$s' = \sqrt{\frac{\sum_{i=1}^{n/2} (x_{2i} - x_{2i-1})^2}{n}}$$

is an estimate of the standard deviation. n must, of course, be an even integer. This estimate has the advantage that its properties are fully known: s' is distributed as the standard deviation with $f = n/2$ degrees of freedom. It will be noted that this estimate does not involve the successive differences, but only the alternate ones. Although there are $n - 1$ available successive differences, this estimate uses only the $n/2$ independent differences. The mean square successive difference is based on all $n - 1$ successive differences, and should therefore provide a more efficient estimate of σ than does s'.

There is, of course, nothing new in the concept of estimating the standard deviation from differences. Even as far back as 1870, an interest in the method appears to have existed. Jordan [3] devised methods based on sums of powers of the differences. Helmert [4] gave more careful consideration to the case of the first power, i.e. the sum of the absolute differences. In both these cases, however, all the $n(n - 1)/2$ differences that can be established from a sample of n observations were included in the estimate, so that the estimate was of no value in reducing the effect of a trend. Helmert realized this, for he pointed out that the estimate obtained from the sum of squares of the differences is exactly that obtained by the more conventional procedure of squaring deviations from the mean.

The usefulness of the differences between successive observations only appears to have been realized first by ballisticians, who faced the problem of minimizing effects due to wind variation, heat and wear in measuring the dispersion of the distance traveled by shell. Vallier [5] appears to have been the first to estimate dispersion from successive differences. Cranz and Becker [6] commended the mean successive difference

$$E_d = \frac{\sum_{i=1}^{n-1} | x_{i+1} - x_i |}{n - 1}.$$

To establish the precision of E_d in estimating σ, Cranz and Becker quoted Helmert's paper, and so erred in saying that their method was superior to that of the mean deviation. Helmert's procedure, based on $n(n - 1)/2$ differences, is indeed more precise (for $n > 10$) than the mean deviation

$$M.D. = \frac{\sum_{i=1}^{n} | x_i - \bar{x} |}{n},$$

but the mean successive difference is based on but $n - 1$ differences, and so is not as precise.

Bennett [7] appears to have suggested the use of successive differences independently of the European ballisticians. In recent years, the method of estimation by the mean square successive difference δ^2 was put into practice in the Ballistic Research Laboratory at the Aberdeen Proving Ground, U. S. Army, by L. S. Dederick

3. Bias and efficiency. The moments of δ^2 in samples drawn from a normal population are derived in Section 6 of this paper. The moments are used at this point to establish the estimate of variance, and the efficiency of this estimate.

The mean value of δ^2 in samples taken at random from a normal population is

$$E(\delta^2) = 2\sigma^2. \tag{3}$$

δ^2 consequently offers an unbiased estimate of variance, and this estimate is

$$\frac{\delta^2}{2} = \frac{\sum_{i=1}^{n-1} (x_{i+1} - x_i)^2}{2(n - 1)}. \tag{4}$$

The second moment, i.e., the variance, of δ^2 in samples of size n :

$$\sigma^2_{\delta^2} = \frac{4(3n - 4)}{(n - 1)^2} \sigma^4. \tag{5}$$

As the sample size is increased, the distribution of δ^2 appears to approach the normal. It is therefore appropriate to consider the efficiency as defined by Fisher [8]. Accordingly, the efficiency of δ^2 is

$$\left[\frac{\sigma_{s^2}}{E(s^2)} \Big/ \frac{\sigma_{\delta^2}}{E(\delta^2)}\right]^2.$$

Since

$$\sigma^2_{s^2} = \frac{2(n - 1)}{n^2} \sigma^4,$$

and

$$E(s^2) = \frac{n - 1}{n} \sigma^2,$$

the efficiency of δ^2 in estimating the standard deviation is

(6)
$$\frac{2(n-1)}{3n-4} = \frac{2}{3}\left[1 + \frac{1}{3n-4}\right].$$

The efficiency is unity for $n = 2$, since in this case the two statistics have the same distribution. It therefore appears that the efficiency decreases as the sample size increases, but approaches 2/3 as a limiting value for n very large.

4. Summary of procedure. Having a statistic which estimates a parameter of a population, it is desirable to know the distribution of that statistic as computed from samples taken at random from that population. At present, the distribution of δ^2 in samples of n has not been obtained. The difficulty is in the fact that the successive differences are not independent. The first difference, $d_1 = x_2 - x_1$, and the second difference, $d_2 = x_3 - x_2$, are related in that they both involve x_2. Similar correlation exists between every successive pair of differences between successive observations.

For $n = 2$, and samples taken from a normal population, the distribution of δ^2 is known. Since

$$\delta^2 = (x_2 - x_1)^2 = 2\sum_{i=1}^{2}(x_i - \bar{x})^2 = 4s^2,$$

the distribution of δ^2 is similar to that of s^2 for this sample size.

For $n = 3$, the distribution of δ^2 has been derived analytically. The derivation is indicated in Section 5 of this paper. For $n > 3$, only the moments of the distribution have thus far been obtained. A Pearson type distribution has been fitted to the first three moments to obtain an approximate representation of the true distribution.

5. Distribution of δ^2. In the case of a sample of n taken from a normal population, the probability that the first observation lies between x_1 and $x_1 + dx_1$, while the second lies between x_2 and $x_2 + dx_2$, etc., is

(7)
$$\left[\frac{1}{\sigma\sqrt{2\pi}}\right]^n e^{-(x_1^2+x_2^2+\cdots+x_n^2)/2\sigma^2}\, dx_1 dx_2 \cdots dx_n.$$

If $y_i = x_{i+1} - x_i$, this expression becomes

(8)
$$\left[\frac{1}{\sigma\sqrt{2\pi}}\right]^n e^{-Q(x_1,y_1,y_2,\cdots,y_{n-1})/2\sigma^2}\, dx_1\, dy_1\, dy_2 \cdots dy_{n-1},$$

where Q is a quadratic form in x_1 and the y's. Since

$$\delta^2 = \frac{\sum_{i=1}^{n-1} y_i^2}{n-1},$$

the probability that δ^2 shall be less than some value δ_0^2 is

$$(9)\quad P(\delta^2 < \delta_0^2) = \left[\frac{1}{\sigma\sqrt{2\pi}}\right]^n \underset{\sum_{i=1}^{n-1} y_i^2 < (n-1)\delta_0^2}{\int\int \cdots \int} \int_{-\infty}^{+\infty} e^{-Q(x_1, y_1, \cdots y_{n-1})/2\sigma^2} dx_1\, dy_1 \cdots dy_{n-1}.$$

After the integration with respect to x_1 is carried out, the quadratic form in the exponent may be normalized by a transformation to new coordinates z_i linearly related to the y's. The z's may be so chosen that all the terms z_i^2 in the exponent have the same coefficient, in which case

$$(10)\quad P(\delta^2 < \delta_0^2) = c_1 \int\int \cdots \int e^{-k \sum_{i=1}^{n-1} z_i^2} \frac{\partial(y_1, y_2, \cdots, y_{n-1})}{\partial(z_1, z_2, \cdots, z_{n-1})} dz_1\, dz_2 \cdots dz_{n-1}.$$

As a result of such a transformation, the sphere of integration in (9) becomes an ellipsoid in (10). By changing to polar coordinates, with

$$(11)\quad \begin{aligned} r^2 &= \sum_{i=1}^{n-1} z_i^2, \\ P(\delta^2 < \delta_0^2) &= c_1 \int\int e^{-kr^2} r^{n-2}\, d\Omega\, dr, \end{aligned}$$

in which Ω is the solid angle in the space of $n - 1$ dimensions. The limits of integration with respect to Ω as a function of r must be found; this involves the evaluation of the solid angle subtended by the surface bounded by the intersection of the $(n - 1)$-dimensional sphere and the $(n - 1)$-dimensional ellipsoid. If $\Omega = \phi(r)$,

$$(12)\quad P(\delta^2 < \delta_0^2) = c_2 \int_0^a e^{-kr^2} \phi(r) r^{n-2}\, dr,$$

in which a is the longest semi-axis of the $(n - 1)$-dimensional ellipsoid corresponding to the given value of δ^2.

For $n = 3$, (9) becomes

$$(13)\quad \begin{aligned} P(\delta^2 < \delta_0^2) &= \left[\frac{1}{\sigma\sqrt{2\pi}}\right]^3 \underset{y_1^2+y_2^2<2\delta_0^2}{\int\int} \int_{-\infty}^{+\infty} \exp\left[-\frac{1}{3\sigma^2}(y_1^2 + y_2^2 + y_1 y_2) \right. \\ &\qquad \left. - \frac{3}{2\sigma^2}\left(x_1 + \frac{2y_1 + y_2}{3}\right)^2\right] dx_1\, dy_1\, dy_2 \\ &= \frac{1}{2\sqrt{3}\,\pi\sigma^2} \underset{y_1^2+y_2^2<2\delta_0^2}{\int\int} e^{-(y_1^2+y_1y_2+y_2^2)/3\sigma^2} dy_1\, dy_2. \end{aligned}$$

Normalizing the quadratic form in the exponent,

$$(14)\quad P(\delta^2 < \delta_0^2) = \frac{1}{2\sqrt{3}\,\pi\sigma^2} \underset{z_1^2+z_2^2<2\delta_0^2}{\int\int} e^{-(z_1^2+\frac{1}{3}z_2^2)/2\sigma^2} dz_1\, dz_2,$$

and in polar coordinates

$$(15)\quad \begin{aligned} P(\delta^2 < \delta_0^2) &= \frac{1}{2\sqrt{3}\,\pi\sigma^2}\int_0^{\delta_0\sqrt{2}}\int_0^{2\pi} re^{-r^2[\cos^2\theta + \frac{1}{3}\sin^2\theta]/2\sigma^2}\,d\theta\,dr \\ &= \frac{1}{2\sqrt{3}\,\pi\sigma^2}\int_0^{\delta_0\sqrt{2}} re^{-r^2/2\sigma^2}\left[\int_0^{2\pi} e^{r^2\sin^2\theta/3\sigma^2}\,d\theta\right]dr. \end{aligned}$$

The integral in brackets can be shown to be a Bessel function of zero order; for let

$$r^2/3\sigma^2 = -2iu,$$

$$\phi = \frac{\pi}{2} - 2\theta,$$

then

$$(16)\quad \int_0^{2\pi} e^{r^2\sin^2\theta/3\sigma^2}\,d\theta = e^{-iu}\int_{-\pi}^{\pi} e^{iu\sin\phi}\,d\phi = 2\pi e^{-iu}J_0(u).$$

Consequently, (15) takes the form

$$(17)\quad P(\delta^2 < \delta_0^2) = \frac{1}{\sigma^2\sqrt{3}}\int_0^{\delta_0\sqrt{2}} re^{-r^2/3\sigma^2}J_0\left(\frac{ir^2}{6\sigma^2}\right)dr = F(\delta_0^2).$$

The probability density function

$$(18)\quad \begin{aligned} p(\delta^2) &= \frac{dF(\delta^2)}{d\delta^2} \\ &= \frac{1}{\sigma^2\sqrt{3}}e^{-2\delta^2/3\sigma^2}J_0\left(\frac{i\delta^2}{3\sigma^2}\right) \\ &= \frac{1}{\sigma^2\sqrt{3}}e^{-2\delta^2/3\sigma^2}\left[1 + \frac{1}{2^2}\frac{\delta^4}{3^2\sigma^4} + \frac{1}{2^2 4^2}\frac{\delta^8}{3^4\sigma^8} + \frac{1}{2^2 4^2 6^2}\frac{\delta^{12}}{3^6\sigma^{12}} + \cdots\right]. \end{aligned}$$

6. Moments. The t-th moment of δ^2 about the origin is defined by

$$(19)\quad \mu_t' = E[(\delta^2)^t],$$

or

$$(20)\quad \begin{aligned} (n-1)^t\mu_t' &= E\left(\left[\sum_{i=1}^{n-1}(x_{i+1} - x_i)^2\right]^t\right) \\ &= E\left(\left[2\sum_{i=1}^{n} x_i^2 - (x_1^2 + x_n^2) - 2\sum_{i=1}^{n-1} x_{i+1}x_i\right]^t\right). \end{aligned}$$

For any value of t, the expansion can be performed, and similar terms collected and enumerated. The values of x can be considered as true errors, i.e. as deviations from the true mean, without affecting the conclusions. If the original population from which the samples have been drawn is normal, with standard deviation σ, then:

$$E(x^{2k-1}) = 0 \tag{21}$$
$$E(x^{2k}) = \frac{(2k)!}{2^k k!}\sigma^{2k},$$

and since, in the null case where the mean of the population remains constant, successive observations are independent, then

$$E(x_i^r x_j^s) = E(x^{r+s}), \qquad i = j \tag{22}$$
$$E(x_i^r x_j^s) = E(x^r)E(x^s), \qquad i \neq j.$$

These relations are sufficient for the evaluation of μ_t'. For example, in the case of the second moment, $t = 2$:

$$(n-1)^2\mu_2' = E\left(\left[2\sum_{i=1}^{n} x_i^2 - (x_1^2 + x_n^2) - 2\sum_{i=1}^{n-1} x_{i+1}x_i\right]^2\right). \tag{23}$$

Now:

$$\left[2\sum_{i=1}^{n} x_i^2 - (x_1^2 + x_n^2) - 2\sum_{i=1}^{n-1} x_{i+1}x_i\right]^2$$
$$= 4\left(\sum_{i=1}^{n} x_i^2\right)^2 + (x_1^2 + x_n^2)^2 + 4\left(\sum_{i=1}^{n-1} x_{i+1}x_i\right)^2$$
$$- 4(x_1^2 + x_n^2)\sum_{i=1}^{n} x_i^2 - 8\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n-1} x_{i+1}x_i + 4(x_1^2 + x_n^2)\sum_{i=1}^{n-1} x_{i+1}x_i$$
$$= 4\left[\sum_{i=1}^{n} x_i^4 + \sum_{i,j=1, i\neq j}^{n} x_i^2 x_j^2\right] + [x_1^4 + 2x_1^2 x_n^2 + x_n^4]$$
$$+ 4\left[\sum_{i=1}^{n-1} x_{i+1}^2 x_i^2\right] - 4\left[x_1^4 + x_1^2\sum_{i=2}^{n} x_i^2 + x_n^2\sum_{i=1}^{n-1} x_i^2 + x_n^4\right]$$
$$+ [\text{terms containing odd powers of } x_i].$$

The mean of these terms is found by using (21) and (22), and the number of each type of term present is enumerated:

$$4[n(3\sigma^4) + n(n-1)\sigma^2\sigma^2] + [3\sigma^4 + 2\sigma^2\sigma^2 + 3\sigma^4] + 4[(n-1)\sigma^2\sigma^2]$$
$$- 4[3\sigma^4 + \sigma^2(n-1)\sigma^2 + \sigma^2(n-1)\sigma^2 + 3\sigma^4] = (4n^2 + 4n - 12)\sigma^4.$$

Consequently

$$\mu_2' = \frac{4(n^2 + n - 3)}{(n-1)^2}\sigma^4. \tag{24}$$

The first four moments about the origin were evaluated by this procedure, and from these, the moments about the mean are readily determined. The results are:

$$
(25)\quad
\begin{aligned}
\mu_1' &= 2\sigma^2\\
\mu_2' &= \frac{4(n^2+n-3)}{(n-1)^2}\sigma^4\\
\mu_3' &= \frac{8(n^3+6n^2+2n-21)}{(n-1)^3}\sigma^6\\
\mu_4' &= \frac{16(n^4+14n^3+53n^2-8n-231)}{(n-1)^4}\sigma^8\\
\mu_1 &= 0\\
\mu_2 &= \frac{4(3n-4)}{(n-1)^2}\sigma^4\\
\mu_3 &= \frac{32(5n-8)}{(n-1)^3}\sigma^6\\
\mu_4 &= \frac{48(9n^2+46n-112)}{(n-1)^4}\sigma^8.
\end{aligned}
$$

It should be noted at this point that the above fourth moment is incorrect for $n = 2$. One of the terms in the expansion of the right side of (20), for $t = 4$, is

$$x_1^2x_n^2\sum_{i=1}^{n-1}x_{i+1}^2x_i^2.$$

For $n = 2$, the mean value of this term is

$$E(x_1^2x_2^2x_2^2x_1^2) = E(x_1^4)E(x_2^4) = 9\sigma^8,$$

whereas for $n > 2$, the mean value is

$$E(x_1^4x_2^2x_n^2) + E\left(x_1^2x_n^2\sum_{i=2}^{n-2}x_{i+1}^2x_i^2\right) + E(x_1^2x_{n-1}^2x_n^4) = (n+3)\sigma^8.$$

7. Pearson type fit to distribution of δ^2. From the moments it is found that

$$
(26)\quad
\begin{aligned}
\beta_1 &= \frac{\mu_3^2}{\mu_2^3} = \frac{16(5n-8)^2}{(3n-4)^3},\\
\beta_2 &= \frac{\mu_4}{\mu_2^2} = \frac{3(9n^2+46n-112)}{(3n-4)^2}.
\end{aligned}
$$

As n becomes large, β_1 and β_2 approach 0 and 3 respectively; the distribution therefore appears to approach the normal for large samples. For finite sample sizes, the values of β_1 and β_2 correspond to those of the Pearson Type VI distribution,

$$p\left(\frac{\delta^2}{\sigma^2}\right) = c\left(\frac{\delta^2}{\sigma^2} + a_1\right)^{q_2}\left(\frac{\delta^2}{\sigma^2} + a_2\right)^{-q_1}.$$

The origin of this distribution is at $\delta^2 = -a_1\sigma^2$, but the origin of the true distribution must be at $\delta^2 = 0$. By taking $a_1 = 0$ so that the origin is at $\delta^2 = 0$, we obtain what appears to be a suitable approximation

$$p\left(\frac{\delta^2}{\sigma^2}\right) = c\left(\frac{\delta^2}{\sigma^2}\right)^{q_2}\left(\frac{\delta^2}{\sigma^2} + a_2\right)^{-q_1}. \tag{27}$$

The parameters are determined by equating the 1st, 2nd and 3rd moments of (27) to the corresponding moments of the true distribution, with the result that

$$\begin{aligned} q_2 &= \frac{3n^4 - 10n^3 - 18n^2 + 79n - 60}{8n^3 - 50n + 48}, \\ q_1 &= \frac{4 - \mu_2(q_2 + 1)(q_2 + 3)}{4 - \mu_2(q_2 + 1)}, \\ a_2 &= \frac{2(q_1 - q_2 - 2)}{q_2 + 1}, \\ c &= \frac{a_2^{q_1 - q_2 - 1}}{B(q_2 + 1, q_1 - q_2 - 1)}. \end{aligned} \tag{28}$$

Values of these parameters for selected values of n are given in Table I. The sixth and seventh columns of this table give the values of β_2 for the distribution (27) and for the true distribution, respectively.

TABLE I

(1) n	(2) q_1	(3) q_2	(4) a_2	(5) c	(6) β_2 (27)	(7) β_2 True	(8) Ratio (6)/(7)
5	24.4391	0.6391	26.6000	5.8800×10^{34}	8.807	8.504	1.036
7	31.1286	1.3857	23.2571	4.9285×10^{42}	6.948	6.758	1.028
10	41.2830	2.5079	20.9667	9.4934×10^{54}	5.658	5.538	1.022
15	58.2113	4.3806	19.2659	4.0240×10^{75}	4.718	4.645	1.016
20	75.1210	6.2543	18.4351	1.8063×10^{96}	4.269	4.217	1.012
25	92.0189	8.1285	17.9417	8.1097×10^{116}	4.006	3.965	1.010
50	176.4443	17.5018	16.9651	1.3386×10^{220}	3.494	3.475	1.005

The *Tables of the Incomplete Beta-Function* [9] can be used to evaluate the probability integral of the distribution (27),

$$\begin{aligned} P\left(\frac{\delta^2}{\sigma^2} < \frac{\delta_0^2}{\sigma^2}\right) &= c\int_0^{\delta_0^2/\sigma^2}\left(\frac{\delta^2}{\sigma^2}\right)^{q_2}\left(\frac{\delta^2}{\sigma^2} + a_2\right)^{-q_1} d\left(\frac{\delta^2}{\sigma^2}\right) \\ &= 1 - I_x(q_1 - q_2 - 1, q_2 + 1) \\ x &= \frac{a_2}{a_2 + \delta_0^2/\sigma^2}, \end{aligned} \tag{29}$$

for $n \leqq 14$. For $n > 14$, the probability integral may be determined by quadrature. Some values of the probability integral for $n = 50$ are given in Table II. A comparison with the integral of the normal curve having the same first two moments indicates that a sample of somewhat more than 50 is required before the normal curve becomes a satisfactory approximation to the distribution (27).

TABLE II

$$P\left(\frac{\delta^2}{\sigma^2} < \frac{\delta_0^2}{\sigma^2}\right) \quad \textit{for } n = 50$$

δ_0^2/σ^2	(29)	Normal
.50	.00000	.00118
.75	.00031	.00563
1.00	.00647	.02129
1.25	.04393	.06418

REFERENCES

[1] "Student," *Biometrika*, Vol. 19(1927), p. 158.

[2] E. S. Pearson, *Application of Statistical Methods to Industrial Standardisation and Quality Control*, London, 1935, p. 62.

[3] W. Jordan, *Astronomische Nachrichten*, Vol. 74(1869), pp. 209–226.

[4] F. R. Helmert, *Astronomische Nachrichten*, Vol. 88(1876), pp. 127–132.

[5] E. Vallier, *Balistique Experimentale*, Paris, 1894, p. 166.

[6] C. Cranz and K. Becker, *Exterior Ballistics*, (trans. from 2nd German edition) London, 1921, p. 383.

[7] A. A. Bennett, unpublished report to the Chief of Ordnance, U. S. Army, circa 1918.

[8] R. A. Fisher, *Phil. Trans.* A, Vol. 222(1922), p. 316.

[9] Karl Pearson (Editor), *Tables of the Incomplete Beta-Function*, London: Biometrika Office, 1934.