A NOTE ON STANDARD DEVIATION AND RMS

R.E.Deakin
D.G.Kildea
RMIT University
GPO Box 2476V
MELBOURNE VIC 3001
AUSTRALIA

INTRODUCTION

The aim of this paper is to clarify some confusing nomenclature of some statistical terms frequently used in the surveying literature. These terms are Standard Deviation and Variance, Root Mean Square (RMS), Mean Square Error and Standard Error.

Much confusion arises from a lack of appreciation of the differences between population values and estimates of these population values based on samples. Moreover, there is confusion as to how the statistical concept of unbiasedness when associated with the usual divisor of $n-1$ in the definition of the sample variance leads to a biased estimate of population standard deviation.

The results are not new and can be found in any elementary textbook (or rather any collection of several elementary textbooks), but we demonstrate from examples drawn from the surveying literature that the confusion is definitely real.

POPULATION QUANTITIES AND MATHEMATICAL EXPECTATION

Population is a potential set of quantities that we want to make inferences about based on a sample from that population. Consider a finite population, such as the examination marks m_i of a group of N students in a single subject. In this case calculating statistics presents no problems since we have complete information about the population. The mean μ, variance σ^2 and standard deviation σ of the **finite population** are,

$$\mu = \frac{\sum_i^N m_i}{N} \tag{1}$$

$$\sigma^2 = \frac{\sum_i^N (m_i - \mu)^2}{N} \tag{2}$$

$$\sigma = \sqrt{\frac{\sum_i^N (m_i - \mu)^2}{N}} \tag{3}$$

Note that the variance σ^2 is the average squared difference of a member from the population mean μ.

For large finite populations (such as all students in a country), we may wish to estimate the population average based on the sample average with some idea of the accuracy of estimation.

Now consider surveying measurements, drawn from **infinite populations** with the attendant difficulties of estimation since population averages can never be known. To deal with this, probability density functions have been introduced to model infinite populations. In surveying, **Normal** (Gaussian) probability density functions are the usual model.

A probability density function is a non-negative function where the area under the curve is one. For $f(x) \geq 0$ and $\int_{-\infty}^{+\infty} f(x)\,dx = 1$ the values of $f(x)$ are not probabilities. The probability a member of the population lies in the interval a to b is $\int_a^b f(x)\,dx$. The population mean μ is defined as (Kreyszig 1970, p.77)

$$\mu = \int_{-\infty}^{+\infty} x\, f(x)\,dx \tag{4}$$

where $f(x)$ replaces $\frac{1}{N}$ and the integral replaces the summation in (1).

By analogy with (2), the population variance σ^2 is defined as the mean squared difference from μ (Kreyszig 1970, p.80)

$$\sigma^2 = \int_{-\infty}^{+\infty} (x-\mu)^2\, f(x)\,dx \tag{5}$$

where $f(x)$ replaces $\frac{1}{N}$ and the integral replaces the summation in (2). Similarly to (3), the population standard deviation σ is defined as

$$\sigma = \sqrt{\int_{-\infty}^{+\infty} (x-\mu)^2\, f(x)\,dx} \tag{6}$$

For the family of Normal probability density functions $f(x;\, c, d) = \frac{1}{d\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-c}{d}\right)^2} \quad (d > 0)$ the population mean μ given by (4) is c, the population variance given by (5) is d^2 and the standard deviation σ is d. This explains the usual presentation in statistics texts of the family of Normal distributions as $\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ (e.g. Kreyszig 1970, p.107).

If we denote a potential member of the infinite population as X, another name for the population mean μ is the expectation of X and is usually written as $E\{X\}$.

$$\mu = E\{X\} \tag{7}$$

We can consider more general expectations $E\{g(X)\}$ where (Kreyszig 1970, p.83)

$$E\{g(X)\} = \int_{-\infty}^{+\infty} g(x)\, f(x)\,dx \tag{8}$$

In other words $g(x)\, f(x)$ is substituted for $x\, f(x)$ in the right hand side of (4). In statistics texts X is called a random variable. Hence, using the notation of expectation, the population variance σ^2 is

$$\sigma^2 = \int_{-\infty}^{+\infty} (x-\mu)^2\, f(x)\,dx = E\left\{(X-\mu)^2\right\} \tag{9}$$

ESTIMATES OF POPULATION QUANTITIES AND THE IDEAS OF UNBIASEDNESS AND DEGREES OF FREEDOM

Suppose that we have a sample of n potential members of a Normal population $X_1, X_2, \cdots X_n$. Throughout, we will assume that these members will be obtained independently of one another. Furthermore, we have a rule for estimating one of the three population quantities μ, σ^2 or σ, which we will denote generically as κ. We write this rule as,

$$T = h(X_1, X_2, \cdots X_n) \tag{10}$$

Then T is also a random variable with mean $\mu_T = E\{T\}$ and variance $\sigma_T^2 = E\left\{(T - E\{T\})^2\right\}$. An overall measure of the accuracy of T is its **mean square error** $m^2 = E\left\{(T-\kappa)^2\right\}$. It turns out that (see Appendix)

$$m^2 = E\left\{(T-\kappa)^2\right\} = E\left\{(T - E\{T\})^2\right\} + (\kappa - E\{T\})^2 \tag{11a}$$

This is the equation given by Gauss (1821-28, p. 11) for what he termed m the *mean error* or *mean error to be feared*.

The term $\kappa - E\{T\}$ in (11a) is the difference between the population quantity κ and the average value of the estimating rule T and is termed the **bias** of T. Hence, (11a) can be expressed in words as

$$\text{mean square error} = \text{variance} + (\text{bias})^2 \tag{11b}$$

When the bias is zero, T is said to be **unbiased** and mean square error equals variance.

The usual rule for estimating μ is the sample mean $\overline{X}$,

$$T = \overline{X} = \frac{1}{n}\sum_i^n X_i \tag{12}$$

The usual rule for estimating σ^2 is the sample variance s^2,

$$T = s^2 = \frac{1}{n-1}\sum_i^n \left(X_i - \overline{X}\right)^2 \tag{13}$$

The usual rule for estimating σ is the sample standard deviation s,

$$T = s = \sqrt{\frac{1}{n-1}\sum_i^n \left(X_i - \overline{X}\right)^2} \tag{14}$$

Other rules are possible; e.g. we could use as a rule for estimating σ^2,

$$T = e^2 = \frac{1}{n}\sum_i^n \left(X_i - \overline{X}\right)^2 \tag{15}$$

It was Sir Ronald Aylmer Fisher who demonstrated that s^2 should be preferred to e^2 for estimating σ^2 (Box 1978, pp. 72-3), and this lead him to the idea of $(n-1)$ **degrees of freedom** for estimating σ^2 in a sample of n independent observations.

A simple illustration may explain the importance of the divisor $n-1$ rather than n. Consider a population of $N=1$ only, called m_1. By definition, the population mean is just m_1 [see (1)], and by (2) and (3) σ^2 and σ are both zero since there is no variation in the population. However, if we have a sample of size one, from an infinite population i.e. $n=1$, we must use this single observation to estimate μ. Clearly, we have no further information in this sample to estimate variation in the population. That is we have used all our degrees of freedom to estimate μ, hence, our estimate of σ^2 (or σ) must be undefined. However, if we use e^2 in (15) our estimate is zero but if we use s^2 in (13) **the answer is undefined** (division by zero!) as we should expect. See Kendall and Buckland (1971) for an excellent discussion of the concept of degrees of freedom.

It can be shown that for the rule $T=\overline{X}$, T is **unbiased** for μ and its variance is $\dfrac{\sigma^2}{n}$ (Kreyszig 1970, p. 175), hence the standard deviation of $\overline{X}$ is $\dfrac{\sigma}{\sqrt{n}}$ and is referred to as the **standard error** of $\overline{X}$.

The usual rule for estimating the standard error of $\overline{X}$ is,

$$\text{standard error} = \frac{s}{\sqrt{n}} \tag{16}$$

This formula shows why, intuitively, we think that the larger the sample size the more precise our determinations are. A simple way to remember this is to note that to double the precision of measurement we must quadruple the number of measurements taken.

It also can be shown that s^2 is an <u>unbiased</u> estimator of σ^2 but that s is a <u>biased</u> estimator of σ. This was known as early as 1920 by Fisher (1920, p. 760), long before the concept of unbiasedness was formally introduced as a term in theoretical statistics. Referring to the formulae above, the action of "taking a square root" changes the property of unbiasedness. This is more an accident of mathematics rather than a cause of faulty estimation but it is not well appreciated in general. This result could in fact be used as an argument for using biased estimators. However, it can be shown that the appropriate divisor is c_n for unbiased estimation of σ giving

$$s^* = \sqrt{\frac{1}{c_n}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2} \tag{17}$$

where

$$c_n = \left\{\frac{\sqrt{2}\,\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}\right\}^2 \tag{18}$$

and $\Gamma(x)$ is a gamma function (Spiegel 1968). A proof of this equation is given in the Appendix and it is interesting to note that c_n is undefined when $n=1$ thus s^* is also undefined for a sample of one. Values of c_n are given in Table 1.

n	2	3	4	5	10	15	20	30	90
$n-1$	1	2	3	4	9	14	19	29	89
c_n	0.64	1.57	2.55	3.53	8.51	13.51	18.51	28.50	88.50

Table 1. Values of the divisor c_n for unbiased estimation of σ

So far, the emphasis has been on defining rules for estimating population quantities. When we substitute measured values into these rules for **estimators,** we obtain a single number called an **estimate**. We are interested in three distinct definitions for the population quantities. For example, the mean can have three different connotations. The first is the theoretical definition of μ given by (4); the second is the rule we apply to estimate μ given by (12). The third is the number, or estimate, we obtain when we substitute a particular sample in that rule. This distinction should be borne in mind when using the terms mean and average to describe the same quantity or process. The same subtle differences apply to the usage of standard deviation and mean square error.

SOME EXAMPLES FROM SURVEYING LITERATURE

Confusion of terminology and confusion of useage appear quite frequently in surveying literature some examples follow:

[1] STANDARD ERROR. The Geodetic Glossary (NGS 1986, p.77) says:

"**error, standard** Equivalent, for the most part to *standard deviation*. However, standard error is also used to mean a number of things such as the standard deviation of the mean and the standard deviation calculated from large samples."

It is evident from the foregoing section that standard error is in no way equivalent to standard deviation in any sense. We have shown that standard error has the population sense of $\dfrac{\sigma}{\sqrt{n}}$ which is the standard deviation of the rule

$\bar{x}$ for estimating μ. If we wished to estimate $\frac{\sigma}{\sqrt{n}}$ (standard error) we would use the rule $\frac{s}{\sqrt{n}}$ and would obtain a single number for any particular sample, which we could also call standard error. The importance of standard error is in finding a **confidence interval**.

[2] ROOT MEAN SQUARE (RMS). The Geodetic Glossary (NGS 1986, p.77) says:

"**error, root-mean-square** A quantity measuring deviation of a random variable from some standard or accepted value; its value is determined by

$$s = \sqrt{\sum \left(x_n - \bar{x}_n\right)^2 / N}$$

where $\{x_n\}$ is the set of N random variables, and $\{\bar{x}_n\}$ the corresponding set of accepted values."

This at first glance looks like the rule s above. Confusion springs from the term $\{\bar{x}_n\}$ since $\bar{x}$ is almost universally accepted as a mean value. In fact in this definition, it refers to a set of accepted values and the terms within the summation are in fact differences. The equation would be better expressed as

$$\mathrm{RMS} = \sqrt{\frac{1}{n}\sum_i^n \left(x_i - a_i\right)^2} \tag{19}$$

where a_i refers to accepted value. This is in fact how RMS is used in geoid modelling where the accepted value is often zero (Featherstone et al 1997, Table 1). When the accepted value in any sample is a (a constant) and the mean of the sample is $\bar{x}$, then (19) becomes (see Appendix)

$$(\mathrm{RMS})^2 = \left\{\frac{1}{n}\sum_i^n \left(x_i - \bar{x}\right)^2\right\} + \left(\bar{x} - a\right)^2 \tag{20a}$$

or in words

$$(\mathrm{RMS})^2 = \text{estimate of variance} + (\text{estimate of bias})^2 \tag{20b}$$

and this is the sample analogue of (11b) above.

To illustrate these three formulae the first row of Table 1 in Featherstone et al has a mean of 1.237, a standard deviation of 2.616 and rms of 2.894 based on a sample of 59. The accepted, or target value is zero. In the right-hand-side of (20a)

$$\left(\bar{x} - a\right)^2 = \left(\bar{x} - 0\right)^2 = 1.237^2 = 1.5302$$

and

$$\frac{1}{n}\sum_i^n \left(x_i - \bar{x}\right)^2 = \frac{n-1}{n}s^2 = \frac{58}{59}\,2.616^2 = 6.7275$$

The left-hand-side of (20a) is $2.894^2 = 8.3752$ which is actually $1.237^2 + 2.616^2$. This means that their column, headed st.d, does not correspond to the usual definition of standard deviation but rather to our e, given by (15). Clearly, the difference is of no practical importance to the interpretation of their results.

[3] AN EXAMPLE OF CLEARLY DEFINED ROOT MEAN SQUARE (RMS)

Rapp (1997) in a paper on geoid modelling presents data in Table 2. There are two columns labelled RMS, in each case relating to differences. It is clear that our formula (19) is used with the accepted values a_i equal to zero. Another column in the table gives the means of the difference between the two models. The squares of these means correspond to the bias-squared term in our equations (20a) and (20b).

[4] EXAMPLES OF POORLY DEFINED ROOT MEAN SQUARE (RMS)

Grejner-Brzezinska et al (1998) in a paper on GPS error modelling states:

> *"Examples of estimated standard deviations (Root-Mean-Square, RMS) for position, velocity, and orientation are plotted in Figs. 13-15."*

Clearly, estimated standard deviations are meant and there are no accepted values present.

Langley (1991) in a paper on the mathematics of GPS states:

> *"We can determine σ experimentally by making a large number of observations and calculating the sum of the squares of the errors in the observations divided by one less than the number of observations made. It is this method of computation that gives σ its alias of root-mean-square error (rms) error."*

This corresponds to our formula (19) with accepted values a_i all equal to zero and the divisor $n-1$ instead of n. It will only be a good estimator of σ when the bias is zero. He has used a hybrid of s given in (14) and RMS given in (20a) and (20b).

[5] SCIENTIFIC CALCULATOR FUNCTION KEYS.

Many calculators, e.g. the Hewlett-Packard HP32SII, have two function keys for calculating standard deviation, labelled s and σ. The σ key corresponds to the σ of (3) and is therefore applicable where we have complete information on a finite population. Since it is rare that the entire population could be sampled in any surveying exercise, this value would be inappropriate. The other function key s corresponds to the s of (14), which is an estimator of σ in an infinite population defined by (6).

The difference in the two values that may arise is simply due to the divisor, either n or $n-1$.

Clearly, this difference will only be important in small samples. Traditionally a sample size $n \leq 30$ is regarded as small.

CONCLUSION

We have highlighted the distinction between population values and sample values of statistical parameters and re-stated well-known rules for the calculation of means and variances or their estimators. In the process, we have demonstrated that statisticians regularly use a biased estimator of σ.

Fisher (1922, p. 366) in his summary states:

> *"During the rapid development of practical statistics in the last few decades, the theoretical foundations of the subject have been involved in great obscurity. Adequate distinction has seldom been drawn between the sample recorded and the hypothetical population from which it has been drawn."*

We have shown that this lack of distinction between population values and sample values remains a source of confusion.

Our paper should also clarify the distinction between Root Mean Square (RMS) and standard deviation. RMS values include components of variance and bias and we have demonstrated that the concept of RMS has a population definition and a corresponding estimate in a sample from that population. Using sample RMS and means can indicate bias in mathematical models as is clearly demonstrated in the papers by Rapp (1997) and Featherstone et al (1997).

APPENDIX

Proof that $\text{mean square error} = \text{variance} + (\text{bias})^2$

Mean square error is defined as the average squared difference of a member T of a random population from some population quantity κ. Using the notation of expectation we may write

$$\text{mean square error} = E\left\{(T-\kappa)^2\right\} \tag{A1}$$

We may write

$$\begin{aligned} E\left\{(T-\kappa)^2\right\} &= E\left\{\left[(T-E\{T\})+(E\{T\}-\kappa)\right]^2\right\} \\ &= E\left\{(T-E\{T\})^2+2(T-E\{T\})(E\{T\}-\kappa)+(E\{T\}-\kappa)^2\right\} \\ &= E\left\{(T-E\{T\})^2\right\}+E\left\{2(T-E\{T\})(E\{T\}-\kappa)\right\}+E\left\{(E\{T\}-\kappa)^2\right\} \end{aligned}$$

Now, since $E\{T-E\{T\}\}=0$ and $E\{\text{constant}\}=\text{constant}$, we may write

$$\begin{aligned} E\left\{(T-\kappa)^2\right\} &= E\left\{(T-E\{T\})^2\right\}+E\left\{(E\{T\}-\kappa)^2\right\} \\ &= E\left\{(T-E\{T\})^2\right\}+(\kappa-E\{T\})^2 \\ &= \text{variance} + (\text{bias})^2 \end{aligned} \tag{A2}$$

Proof that s^* is an unbiased estimator of the population standard deviation σ.

For ν independent normal random variables $X_1, X_2, \cdots X_\nu$ each with with mean 0 and variance 1 the sum of the squares, usually denoted by χ^2_ν (chi-squared), has a probability density function

$$f(x) = \frac{1}{2^{\left(\frac{\nu}{2}\right)}\Gamma\left(\frac{\nu}{2}\right)} x^{\left(\frac{\nu-2}{2}\right)} e^{-\left(\frac{x}{2}\right)} \tag{A3}$$

where $f(x)=0$ when $x<0$ and ν is a positive integer known as the degrees of freedom.

If $X_1, X_2, \cdots X_n$ are independent normal random variables with mean μ and variance σ^2 then (Kreyszig 1970, p. 181)

$$T = \frac{(n-1)s^2}{\sigma^2} \tag{A4}$$

has a χ^2_{n-1} distribution with a probability density function

$$f(t) = \frac{1}{2^{\left(\frac{n-1}{2}\right)}\Gamma\left(\frac{n-1}{2}\right)} t^{\left(\frac{n-3}{2}\right)} e^{-\left(\frac{t}{2}\right)} \tag{A5}$$

where $n-1$ has replaced ν in (A3). Now from (A4) we may write s equal to some function of T or

$$s = g(T) = \frac{\sigma}{\sqrt{n-1}}\sqrt{\frac{(n-1)s^2}{\sigma^2}}$$

and by the general definition of expectation, we may write the expected value of s as

$$E\{s\} = E\{g(T)\} = \int_0^\infty g(t)\, f(t)\, dt = \int_0^\infty \frac{\sigma}{n-1}\sqrt{\frac{(n-1)s^2}{\sigma^2}}\, f(t)\, dt \tag{A6}$$

but $\sqrt{\frac{(n-1)s^2}{\sigma^2}} = t^{\frac{1}{2}}$ which when substituted into (A6) with (A5) simplifies to

$$E\{s\} = \frac{\sigma}{n-1} \frac{1}{2^{\left(\frac{n-1}{2}\right)} \Gamma\left(\frac{n-1}{2}\right)} \int_0^\infty t^{\left(\frac{n-2}{2}\right)} e^{-\left(\frac{t}{2}\right)} dt$$

$$= \frac{\sigma}{n-1} \frac{1}{2^{\left(\frac{n-1}{2}\right)} \Gamma\left(\frac{n-1}{2}\right)} 2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right) \int_0^\infty \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} t^{\left(\frac{n-2}{2}\right)} e^{-\left(\frac{t}{2}\right)} dt \qquad \text{(A7)}$$

and the integral in (A7) evaluates to unity. This simplifies to

$$E\{s\} = \frac{\sqrt{2}\,\Gamma\left(\frac{n}{2}\right)}{\sqrt{n-1}\,\Gamma\left(\frac{n-1}{2}\right)} \sigma = k\sigma$$

Since k is a constant, the rules of expectation can be used to give

$$E\{s^*\} = E\left\{\frac{1}{k}s\right\} = \sigma$$

and s^* is an unbiased estimator of the population standard deviation σ.

Proof that $(\mathrm{RMS})^2 = \text{estimate of variance} + (\text{bias})^2$ when accepted value a_i is a constant

From (19) with $a_i = a$ (a constant) we have

$$(\mathrm{RMS})^2 = \frac{1}{n}\sum_i^n (x_i - a)^2 \qquad \text{(A8)}$$

We may write

$$(\mathrm{RMS})^2 = \frac{1}{n}\sum_{i=1}^n \left[(x_i - \bar{x}) + (\bar{x} - a)\right]^2$$

$$= \frac{1}{n}\sum_{i=1}^n \left[(x_i - \bar{x})^2 + 2(\bar{x} - a)(x_i - \bar{x}) + (\bar{x} - a)^2\right]$$

$$= \frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2 + \frac{1}{n}\sum_{i=1}^n 2(\bar{x} - a)(x_i - \bar{x}) + \frac{1}{n}\sum_{i=1}^n (\bar{x} - a)^2$$

Now, since $\sum_{i=1}^n (x_i - \bar{x}) = 0$ and

$\sum_{i=1}^n (\text{constant}) = n(\text{constant})$ we may write

$$(\mathrm{RMS})^2 = \left\{\frac{1}{n}\sum_i^n (x_i - \bar{x})^2\right\} + (\bar{x} - a)^2 \qquad \text{(A9)}$$

REFERENCES

Box, Joan Fisher., 1978, *R.A. Fisher, The Life of a Scientist,* John Wiley & Sons, New York.

Featherstone, W.E., Gilliland, J.R. and Kearsley, A.H.W., 1997, 'Data preparations for a new Australian gravimetric geoid', *The Australian Surveyor,* Vol. 42, No. 1, March 1997, pp. 33-44.

Fisher, R.A., 1920, 'A mathematical examination of the methods of determining the accuracy of an observation by the mean error and by the mean square error', *Monthly Notices of the Royal Astronomical Society,* 80, pp. 758-70.

Fisher, R.A., 1922, 'On the mathematical foundations of theoretical statistics', *Philosophical Transactions of the Royal Society of London,* Series A, Vol. 222, pp. 309-68.

Gauss, C.F., 1821–28, *Theory of the Combination of Observations Least Subject to Errors: Part One, Part Two, Supplement.* A translation of *Theoria Combinationis* by G.W.Stewart, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 1995.

Grejner-Brzezinska, D.A., Da, R. and Toth, C., 1998, 'GPS error modelling and OTF ambiguity resolution for high-accuracy GPS/INS integrated system', *Journal of Geodesy,* Vol. 72, No. 11, November 1998, pp. 626-38.

Kendall, Maurice G. and William R. Buckland, 1971, *A Dictionary of Statistical Terms,* 3rd edn, Oliver & Boyd, Edinburgh.

Kreyszig, Erwin, 1970, *Introductory Mathematical Statistics,* John Wiley & Sons, New York.

Langley, R.B., 1991, 'The mathematics of GPS', *GPS WORLD,* Vol. 2, No. 7, July/August 1991, pp. 45-50.

NGS, 1986, *Geodetic Glossary,* National Geodetic Survey, Rockville, MD, September 1986.

Rapp, R.H., 1997, 'Use of potential coefficient models for geoid undulation determinations using a spherical harmonic representation of the height anomaly/geoid undulation difference', *Journal of Geodesy,* Vol. 71, No. 5, April 1997, pp. 282-89.

Spiegel, Murray R., 1968, *Mathematical Handbook,* Schaum's Outline Series in Mathematics, McGraw-Hill, New York.