| 1 | Biology |
| :--- | :--- |
| 1 | Chemistry |
| 1 | Earth Science, Environmental Science, Physical Science, Biology <br> II, Chemistry II, <br> Physics, Physics II, or Physics for Technology (one unit) <br> American History |
| 1 | World History, Western Civilization, or World Geography 1 <br> Civics and Free Enterprise (one unit combined) or Civics (one <br> unit) |
| Fine Arts Survey (or substitute two units of performance courses <br> in music, dance and/or theater; or substitute two units of visual <br> art; or substitute two units of studio art; or substitute one unit of <br> an elective from among the other subjects listed in this core <br> curriculum) <br> Foreign Language (two units in the same language) |  |
| $1 / 2$ | Computer Science, Computer Literacy, or Business Computer <br> Applications (or substitute at least one-half unit of an elective <br> course related to computers approved by the State Board of <br> Elementary and Secondary Education or one-half unit as an <br> elective from among the other subjects listed in this core <br> curriculum) |

## Visualizing the Sample Standard Deviation

Jyotirmoy Sarkar

Mamunur Rashid
The standard deviation (SD) of a random sample is defined as the square-root of the sample variance, which is the 'mean' squared deviation of the sample observations from the sample mean. Here, we interpret the sample SD as the square-root of twice the mean square of all pairwise balf deviations between any two sample observations. This interpretation leads to a geometric visualization of the sample SD, and a more

## elementary explanation as to why the denominator in the sample variance

 is one less than the sample size.To summarize the values of a variable measured on a set of units randomly selected from a population, we oftentimes record the sample mean as a measure of center, and the sample standard deviation (SD) as a measure of spread. The sample mean is used to estimate the population mean and the sample SD is used to the population SD . Needless to say, the mean and the SD are the two most frequently used summary measures in probability and statistics (Lesser, Wagler, \& Abormegah, 2014). Now-a-days these quantities are introduced to students as early as in middle school (National Governors' Association, 2010).

For the readers' benefit, we recall the definitions of these quantities. Suppose that we have a sample of $n$ numbers $x_{1}, x_{2}, \ldots, x_{n}$. The sample mean (denoted by $\bar{x}$ ) is defined as the sum of the values divided by the number of values; that is,

$$
\begin{equation*}
\bar{x}=\frac{1}{n} \sum_{i \neq 1}^{n} x_{i} \tag{1}
\end{equation*}
$$

The sample variance ${ }^{i}{ }^{i} \mathrm{Z}$ denoted by $\mathrm{s}^{2}$ ) is the 'mean' ${ }^{1}$ squared deviation of the sample observations from the sample mean $\bar{x}$, and is defined by

$$
\begin{equation*}
s^{2}=\frac{1}{n-1} \sum_{i=1}^{n}\left(x_{i}-\bar{x}\right)^{2} \tag{2}
\end{equation*}
$$

which simplifies to

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

[^0]The sample SD (denoted by $s$ ) is the positive square-root of the sample variance given by

$$
\begin{equation*}
s=\sqrt{\frac{1}{n-1} \sum_{i=1}^{n}\left(x_{i}-\bar{x}\right)^{2}}=\sqrt{\frac{1}{n-1}\left[\sum_{i=1}^{n} x_{i}^{2}-n \bar{x}^{2}\right]} \tag{3}
\end{equation*}
$$

One interpretation of the sample mean $\bar{X}$ as the center of gravity is well known: If the entire dataset is viewed as a dot plot, with the dots representing heavy balls on a weightless number line, then the mean can be thought of as the location of a fulcrum that would keep the balls in balance. An alternative geometric interpretation of the mean, as a vertical line that equalizes the areas of two regions formed by the empirical cumulative distribution function of the sample, is given in Sarkar and Rashid (2015a).

We found some visual representations of the sample SD in the literature. See, for example, Maverick (1932), and Embse and Engebretsen (1996). However, these are only depictions of the sample SD after it has been calculated numerically. There is no obvious way to verify the correctness of these depictions except by computing the sample SD numerically once again!

The interpretation of the sample SD $\boldsymbol{S}$ is not as straight-forward as that of the sample mean $\bar{x}$. Some authors skip the interpretation of the sample SD, and only discuss its usefulness. Other authors, perhaps in haste, have interpreted the sample SD as a 'typical distance' of each observation from the sample mean, causing conflict with the notion of the mean (absolute) deviation. ${ }^{2}$ The correct interpretation of the

[^1]sample SD , as the square-root of the 'mean' squared deviation (RMSD) from the sample mean leaves unanswered why the denominator in $(2)$ is $(n-1)$ instead of the naturally anticipated $n$. Typical explanations involve the concepts of "unbiased estimation" (see, for example, Ugarte, Militino, \& Arnholt, 2008, pp. 247), or "degrees of freedom" (df) (see Martin, 2003). In other words, we choose the denominator in (2) to be ( $n-1$ ) so that the bias involved in estimating the population variance is eliminated (and the bias involved in estimating the population SD is reduced) from what they would be if the denominator were chosen to be $n$. Alternatively, since one df is used up to estimate the population mean there remain $(n-1) \mathrm{df}$ to estimate the population variance. However, more often than not, students are not familiar with the concepts of estimation and df.

In Section 2, we provide an alternative, but equivalent, expression for the sample SD. Utilizing this result, in Section 3, we construct a geometric object that represents the sample SD based on all pairwise deviations, and hence can be visualized. Suggestions for other geometric objects which help visualize the sample SD, based on deviations from the mean, can be found in Sarkar and Rashid (2015b). Furthermore, our equivalent expression for the sample SD yields an elementary explanation as to why the denominator in (2) is $(n-1)$.

## An Equivalent Expression for the Sample Variance

Let us begin with the notion of deviation ${ }^{3}$ between two numbers $a$ and $b$ given by $|a-b|$. Likewise, the deviation

$$
\mathrm{MD}=\frac{1}{n} \sum_{i=1}^{n}\left|x_{i}-\bar{x}\right| \cdot
$$

It measures the mean distance of each observation from the sample mean.
${ }^{3}$ We use the word 'deviation' to mean 'absolute difference.'
of each number $a$ and $b$ from their average $(a+b) / 2$, is given by the balf-deviation between them, or $|a-b| / 2$. Given a random sample of size $n>2$, how can we combine all

$$
\binom{n}{2}=n(n-1) / 2
$$

pairwise deviations (PD) (or pairwise half-deviations, PHD) into one single measure of spread?

We can calculate either the mean or the mean square of these pairwise deviations (or half-deviations). For example, the mean of pairwise deviations (MPD), the mean of pairwise half-deviations (MPHD), the mean square of pairwise deviations (MS-PD), and the mean square of pairwise half-deviations (MS-PHD) are given below.

$$
\begin{array}{ll}
\text { MPD }=\sum_{i=1}^{n}\left|x_{i}-x_{j}\right| /\binom{n}{2}, & \text { MPHD }=\sum_{i=1}^{n}\left(\frac{\left|x_{i}-x_{j}\right|}{2}\right) /\binom{n}{2}, \\
\text { MS-PD }=\sum_{i=1}^{n}\left(\left|x_{i}-x_{j}\right|\right)^{2} /\binom{n}{2}, & \text { MS-PHD }=\sum_{i=1}^{n}\left(\frac{\left|x_{i}-x_{j}\right|}{2}\right)^{2} /\binom{n}{2} \tag{4}
\end{array}
$$

It is easy to verify that MPHD $=$ MPD/2 and MSPHD $=$ MS-PD/4. It turns out that MS-PD is exactly twice the sample variance! Hence, the square root of MS-PD (RMS-PD) is $\sqrt{2}$ times the sample SD; that is,

$$
\text { RMS-PD }=\sqrt{2} s .
$$

Equivalently, MS-PHD is half the sample variance, which we state in Proposition 1 below. Interested readers may see a proof in the Appendix. Here, we only give an example to illustrate the result.

Proposition 1. The MS-PHD given in (4) equals half the sample variance given in (2); that is,

$$
\begin{equation*}
\sum_{1 \leq i<j \leq n} \sum\left(\frac{\left|x_{i}-x_{j}\right|}{2}\right)^{2} /\binom{n}{2}=\frac{s^{2}}{2} \tag{5}
\end{equation*}
$$

Example 1. Suppose that our data consist of $n=4$ values (sorted from the smallest to the largest): 11, 20, 22, 35 . Then the sample mean is $\bar{x}=(11+20+22+35) / 4=22$. The deviations of the $x$-values from $\bar{x}$ are: $11,2,0,13$; and so the sample variance is $\left(11^{2}+2^{2}+0^{2}+13^{2}\right) / 3=294 / 3=98$. The $\binom{4}{2}=6$ pairwise deviations are: $9,11,24,2,15,13$. These PD's may be sorted as: $2,9,11,13,15,24$. Hence, we have MS-PD $=$ $\left(2^{2}+9^{2}+11^{2}+13^{2}+15^{2}+24^{2}\right) / 6=1176 / 6=196$, which is twice the sample variance. Consequently, MS-PHD $=$ MS-PD $/ 4=49$, or half the sample variance, as Proposition 1 states.

In view of Proposition 1, the sample variance is twice the MS-PHD, and the sample SD is the square-root of twice the MS-PHD. Furthermore, there is no ambiguity about the division by $\binom{n}{2}$ on the left hand side (LHS) of (5), since there are $\binom{n}{2}=n(n-1) / 2$ possible pairwise half-deviations (though not all are distinct). The proof of Proposition 1 reveals that the numerator on the LHS of (5) also has a factor $n$, which cancels with that in the denominator, leaving the other factor $(n-1)$ in the denominator. This explains why the denominator in (2) is $(n-1)$.

## Geometric Visualization of the Sample SD

Proposition 1 lends itself to the construction of a geometric object that represents the sample SD, thereby making it possible to visualize the sample SD. We will describe the construction in five steps. But first, let us give a geometric meaning to a typical term on the LHS of (5). Suppose that we
have a right isosceles triangle ${ }^{4}$ (RIT) whose hypotenuse equals any typical PD, say $h=\left|x_{i}-x_{j}\right|$. Then the area of such an RIT is

$$
\left(\frac{h}{2}\right)^{2}=\left(\frac{\left|x_{i}-x_{j}\right|}{2}\right)^{2}
$$

See Fig. 1(a). When this RIT is translated in the threedimensional space orthogonally to itself through a distance of $w=1 /\binom{n}{2}$, it generates a right prism whose volume equals

$$
\left(\frac{h}{2}\right)^{2} w=\left(\frac{\left|x_{i}-x_{j}\right|}{2}\right)^{2} /\binom{n}{2}
$$

See Fig. 1(b). Moreover, note that the bottom face of the right prism so generated is a rectangle of size $h \times w=\left|x_{i}-x_{j}\right| \times\left[1 /\binom{n}{2}\right]$, and so its area is $h w=\left|x_{i}-x_{j}\right| /\binom{n}{2}$.

[^2]

Figure 1. Area of a right isosceles triangle (RIT) and volume of a right prism.

## Steps to construct the sample SD geometrically

The method of construction of a geometric object that helps visualize the sample SD is described below and illustrated in Figures 2-4. Let us first assume that the given sample consists of $n$ distinct values. See Fig. 2, which uses the data in Example 1, and also Fig. 3, which uses another data set. Thereafter, Remark 1 explains how to handle ties, which is illustrated in Fig. 4, using a third data set.

Step 0 (Preliminary preparation). Given the dot plot of $x$, showing the random sample of $n$ distinct values, draw a vertical reference line $\ell$ given by $x=x_{\min }$, and a horizontal reference line (just below the dot plot of $x$ ) to the right of $\ell$ with all other $x$-values marked on it and surrounded by tiny circles. From each of these marked points draw a line with slope 1 (or the so-called $45^{\circ}$ line) that intersects $\ell$ at some point, from which draw a horizontal line to the right of $\ell$. Mark and surround with a tiny circle every point of intersection so generated which are not on $\ell$. Then counting row by row, there will be
exactly $(n-1)+(n-2)+\cdots+2+1=\binom{n}{2}$ marked points, each representing a PD in terms of its distance from $\ell$.

Step 1 (Dot plot of PD). Projecting the marked points (surrounded by tiny circles) vertically up, we construct the dot plot of PD, which we depict just above the given dot plot of $x$. Note that even though the marked points are distinct, some of them may be vertically aligned (as in Fig. 3). Hence, these $\binom{n}{2}$ PD's need not be distinct.

Step 2 (CDF of PD). Suppose that the distinct values of the PD's are $0<d_{1}<d_{2}<\ldots<d_{K}$ with associated frequencies $f_{1}, f_{2}, \ldots, f_{K}$. Clearly then, $f_{1}+f_{2}+\ldots+f_{K}=\binom{n}{2}$. Draw the cumulative distribution function (CDF) of the PD's: It is a step function of the form $y=G(d)$, which begins at height $y=0$ for $d \leq 0$, has jumps of magnitude $f_{k} /\binom{n}{2}$ occurring at $d_{k}$ for $k=1,2, \ldots, K$; and ends at height $y=1$ for $d \geq d_{K}$.

Step 3 (Erect right prisms). If we superimpose on the graph of the CDF of PD horizontal lines $y=t /\binom{n}{2}$ for $t=1,2, \ldots,\binom{n}{2}$, stretching from $\ell$ to the CDF, we see $\binom{n}{2}$ rectangles corresponding to the $\binom{n}{2}$ PD's. Each rectangle has width given by a PD and height $1 /\binom{n}{2}$. Using each of these $\binom{n}{2}$ rectangles as the $x y$-face, we erect a right prism of $y$-thickness $1 /\binom{n}{2}$ and $x \approx$-cross-section given by an RIT whose hypotenuse equals the width of the rectangle (or the corresponding PD). Prisms with identical $x \approx$-crosssections may be blended together, as done in Fig. 3. Then the total volume $V_{+}$of all $\binom{n}{2}$ such right prisms is precisely the LHS expression in (5), or the MS-PHD. Therefore, in view of Proposition 1, the total volume $V_{+}$also equals half the sample variance.

Step 4 (Search for a composite right prism). Consider a single composite right prism having $y$-thickness 1 and a $x \approx$-cross-section in the shape of an RIT. Allow the size of this RIT to vary by changing the $x$-size until we find a suitable size such that the volume of the single composite right prism equals $V_{+}$. Then each leg of that RIT equals the sample SD, and the hypotenuse of that RIT is $\sqrt{2}$ times the sample SD, or the RMS-PD.

Remark 1. When the $x$-values are not all distinct, as in Fig. 4, additional care must be taken to determine the multiplicity of each intersection point, described in Step 1 above, representing a possible value of PD: (0) We must also consider the intersection points on $\ell$ and circle them if their multiplicity is non-zero; (1) The intersection point on $\ell$, corresponding to a given observation $u$ with frequency $f_{u}$, must be assigned a multiplicity of $\binom{f_{u}}{2}$; and (2) The intersection point not on $\ell$, corresponding to two given observations $u$ and $v$ with respective frequencies $f_{u}$ and $f_{v}$, must be assigned a multiplicity of $f_{u} f_{v}$.

Remark 2. In Step 2, we can further draw a vertical line that equalizes the areas of the shaded regions (see Fig. 2-4. See below). This line depicts the MPD. See details in Sarkar and Rashid (2015a).

Remark 3. From Fig. 2-4, we note that MPD $\leq$ RMS-PD $=$ $\sqrt{2} s$. This is, in fact, a general result which follows from the celebrated Cauchy-Swartz inequality. Interested readers may see Bellman (1997, pp. 126). For our purpose in this paper, this inequality provides a nice check in our visualization of the sample SD.


Figure 2. A geometric construction of the sample SD when the given numbers are distinct and the PD's are also distinct.



Figure 4. A geometric construction of the sample SD when the given numbers (and hence the PD's) are tied. The frequency of each distinct value of a PD is determined as explained in Remark 1
above. In particular, the frequency of $\mathrm{PD}=0$ is $\sum_{x}\binom{f_{x}}{2}$, where $\mathrm{f}_{x}$ is the frequency of each distinct $x$-value.

## Discussion

In this paper we have interpreted the sample variance as twice the mean square of all pairwise half-deviations. Thereby we provided a geometric way to visualize the sample SD. We also found a direct explanation of why the denominator in the definition of sample variance is one less than the sample size, without referring to unbiasedness or df. However, we must admit that this alternative interpretation is not the recommended way to compute the sample variance, since instead of the $n$ deviations we must now deal with $\binom{n}{2}$ pairwise half-deviations.

In fact, one can visualize the RMSD based on the all deviations from the sample mean in a manner analogous to that based on all pairwise half-deviations. Recall that the RMSD is defined by

$$
\begin{equation*}
\tilde{S}^{2}=\frac{1}{n} \sum_{i=1}^{n}\left(x_{i}-\bar{x}\right)^{2}=\frac{1}{n}\left[\sum_{i=1}^{n} x_{i}^{2}-n \bar{x}^{2}\right] \tag{6}
\end{equation*}
$$

Note that $\tilde{s}^{2}$ is the correct mean square deviation from the sample mean, where both uses of the word "mean" involve division by $n$.

To visualize the RMSD based on the $n$ deviations, of course, Step 0 is not needed. Working with the $n$ deviations, we follow Steps 1-4. Then the hypotenuse of the composite RIT, obtained in Step 4, equals the RMSD. Finally, we must implement one more step to obtain the sample SD as follows.

Step 5 (Modify the composite right prism). Replace the composite right prism having volume $V_{+}, y$-thickness 1 and a $x \approx$-cross-section in the shape of an RIT by another composite right prism having the same volume $V_{+}$, but a
smaller $y$-thickness $(1-1 / n)$ and a $x \approx$-cross-section in the shape of an RIT. Then the base of this new composite RIT equals the sample SD.

The method is demonstrated in Fig. 5 using the data of Example 1. Recall that the sample variance is 98 , and the MSD is $\left(11^{2}+2^{2}+0^{2}+13^{2}\right) / 4=73.5$.


Figure 5. A geometric construction of the RMSD $\widetilde{s}$ and the sample SD $s$ based on the deviations from the sample mean.

We contend that our geometric way of viewing the sample SD (and also the RMSD) will help readers develop a better intuition about these concepts. We plan to test this hypothesis in a statistically designed experiment.

## References

Bellman, R. (1997). Introduction to matrix analysis (2nd ed.). Santa Monica, CA: The RAND Corporation.
Embse, C. \& Engebretsen, A. (1996). Visual Representations of Mean and Standard Deviation. The Mathematics Teacher, 89(8), 688-692.
Lesser, L.M., Wagler, A.E., \& Abormegah, P. (2014). Finding a happy median: Another balance representation for measures of center. Journal of Statistics Education, 22(3), 1$27 . \quad$ Retrieved from www.amstat.org/publications/jse/v22n3 /lesser.pdf.
Maverick, L.A. (1932). Graphic presentation of standard deviation. Journal of the American Statistical Association, 27(179), 287-297.
Martin, M. (2003). "It's like...you know": The use of analogies and heuristics in teaching introductory statistical methods. Journal of Statistics Education, 11(2). Retrieved from http://www.amstat.org/publications/ jse/v11n2/martin.html.
National Governors Association and Council of Chief State School Officers (2010). Common Core state standards (College- and career-readiness standards and K-12 standards in English language arts and math). Washington, DC: National Governors Association Center for Best Practices and Council of Chief State School Officers. Retrieved from http://www.corestandards.org.
Sarkar, J. \& Rashid, M. (2015a). A geometric view of the mean of a set of numbers. Teaching Statistics, 38(3), 77-82.

Sarkar, J. \& Rashid, M. (2015b). Visualizing mean, median, mean deviation and standard deviation of a set of numbers. The American Statistician, 70(3), 304-312.
Ugarte, M.D, Militino, A.F., \& Arnholt, A.T. (2008). Probability and statistics with R. Boca Raton, FL: CRC Press.


[^0]:    ${ }^{1}$ Whereas mean involves a division by the sample size, we write 'mean' when the division is by one less than the sample size.

[^1]:    ${ }^{2}$ Mean (absolute) deviation is defined by

[^2]:    ${ }^{4}$ A right isosceles triangle is a three-sided plane figure which has one right angle ( $90^{\circ}$ ) and two legs (sides adjacent to the right angle) that are of equal length.

