Nearness Relations Among Measures of Central Tendency and Dispersion: Part 1

Warren Page
V. N. Murty

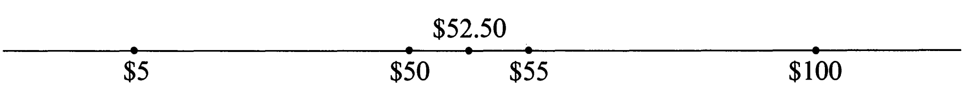
Warren Page, Professor of Mathematics at New York City Technical College, received his Ph.D. from the Polytechnic Institute of Brooklyn in 1971. Since then he has authored more than 25 articles and a text, "Topological Uniform Structures." Dr. Page is editor of TYCMJ's Classroom Capsules column, and he is editor of MAA's recent book, "Two-Year College Mathematics Readings."

Vedula N. Murty is Professor of Mathematics and Statistics at the Capitol Campus of The Pennsylvania State University, Middletown, where he has been employed since 1970. He is the author of several articles. His interests are in the fields of Mathematical Statistics, Sampling Theory and Practice, Experimental Designs, Operations Research, and Analysis.

1. Introduction. If ten winning lottery tickets paid as follows

$$\{\$5, \$50, \$50, \$50, \$50, \$55, \$55, \$60, \$75, \$100\},$$

how much would you be willing to pay for the next assumed winning ticket? Would you pay the mean (i.e., average) amount $\mu = \$55$ won? Or, would you disregard the extreme winning payoffs and pay the median (i.e., middlemost) amount $\mu^* = \$52.50$ won? Perhaps you would be willing to pay the modal (i.e., most likely) amount $\mu_0 = \$50$ won. One thing is clear: These representative-type averages, called *measures of central tendency*, seem rather close together when compared with the range ($R = \$100 - \$5 = \$95$) in amounts won.



Apparently the range is too large a *measure of dispersion* for effectively comparing μ, μ^*, and μ_0. But what about the other common measure of dispersion, the standard deviation σ? Since σ is based on deviations from the mean, none of which are larger than the range, we always have $\sigma \leqslant R/B$ for some constant $B > 0$. On the other hand, $R/A \leqslant \sigma$ for a large enough constant $A > 0$. Thus, $R/A \leqslant \sigma \leqslant R/B$ and it seems only natural to ask if we can use σ to effectively thread μ, μ^*, μ_0, and R together. In the previous set of lottery winnings, for example,

$$\sigma = \sqrt{\frac{\sum_{i=1}^{10}(x_i - 55)^2}{10}} \approx 22.47,$$

and we see that

$$\frac{R}{\sqrt{2(10)}} < \sigma < \frac{R}{2} \quad \text{and} \quad |\mu - \mu^*|, |\mu_0 - \mu^*|, |\mu_0 - \mu| < \sigma.$$

The preceding inequalities hint at several beautiful relationships, among the basic measures of central tendency and dispersion, that are not mentioned in most statistics texts. Our objective is to call attention to these important, yet frequently overlooked, relationships. Accordingly, we offer some elementary derivations and extensions of the following:

The range R, standard deviation σ, mean μ, median μ^*, and mode μ_0 (having frequency m) of a set of real numbers $\{x_1 \leqslant x_2 \leqslant \cdots \leqslant x_n\}$ satisfy

$$\frac{R}{\sqrt{2n}} \leqslant \sigma \leqslant \frac{\sqrt{\left[\frac{n}{4}^2\right]}}{n} \cdot R, \tag{I}$$

$$|\mu - \mu^*| \leqslant \sigma, \tag{II}$$

$$|\mu_0 - \mu| \leqslant \sqrt{\frac{n}{m}} \cdot \sigma, \quad \text{and} \quad |\mu_0 - \mu^*| \leqslant \frac{n}{m} \cdot \sigma. \tag{III}$$

(The symbol $[\alpha]$ denotes the greatest-integer less than or equal to α.)

The above inequalities are particularly noteworthy since they provide students with simple checks as to whether their computations of these measures are of the right order of magnitude. Note, for example, that $[n^2/4] = n^2/4$ for n even, and $[n^2/4] = (n^2 - 1)/4$ for n odd. Thus, $R/\sqrt{2n} \leqslant \sigma \leqslant R/2$ for all n.

Finally, these inequalities have an aesthetic appeal—most notably, the attractive manner in which they combine and interlace such important statistical notions. Other related results will be given—as, for example, a simple proof that

$$\sum_{i=1}^{n} |x_i - \mu^*| \leqslant \sum_{i=1}^{n} |x_i - \xi| \qquad \textit{for every real number } \xi. \tag{IV}$$

In this spirit, we shall also show that:

If $\{x_1 \leqslant x_2 \leqslant \cdots \leqslant x_n\}$ is partitioned into nonempty sets $A = \{x_1 \leqslant \cdots \leqslant x_r\}$ and $B = \{x_{r+1} \leqslant \cdots \leqslant x_n\}$ with respective means $\bar{x}_A$ and $\bar{x}_B$, then

$$\mu - \sigma \cdot \sqrt{\frac{n-r}{r}} \leqslant \bar{x}_A \leqslant x_r \leqslant x_{r+1} \leqslant \bar{x}_B \leqslant \mu + \sigma \cdot \sqrt{\frac{r}{n-r}} . \tag{V}$$

A simple application of (V) leads to another interesting inequality that extends (II).

2. Range Bounds for the Standard Deviation. The key ingredient in this article is the Cauchy–Schwarz inequality

$$\left(\sum_{i=1}^{n} a_i b_i\right)^2 \leqslant \left(\sum_{i=1}^{n} a_i^2\right) \cdot \left(\sum_{i=1}^{n} b_i^2\right) \tag{1}$$

for real numbers $\{a_i, b_i : 1 \leqslant i \leqslant n\}$. A particularly useful form of (1) can be obtained by taking $b_i = x_i - \mu$ and by noting that $\sum_{i=1}^{n}(x_i - \mu)^2 = n\sigma^2$. Substituting and squaring in (1), we get

$$\left\{\sum_{i=1}^{n} a_i(x_i - \mu)\right\}^2 \leqslant \left(\sum_{i=1}^{n} a_i^2\right) \cdot n\sigma^2. \tag{2}$$

If $a_i = 1$ for $x_i \geqslant \mu$ and $a_i = -1$ for $x_i < \mu$, then (2) reduces to

$$\sum_{i=1}^{n} |x_i - \mu| \leqslant n\sigma. \tag{3}$$

The first inequality in (I) follows from (2). Specifically,

$$R = (x_n - \mu) + (\mu - x_1) \leqslant \sqrt{2} \cdot \sqrt{n\sigma^2} .$$

To establish the second inequality in (I), define the real-valued (standard deviation) function

$$\sigma : E^n \longrightarrow R$$

$$t = (t_1, t_2, \ldots, t_n) \rightsquigarrow \sigma(t) = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(t_i - \bar{t}\right)^2} ,$$

where $\bar{t} = (1/n)\sum_{i=1}^{n} t_i$. We abbreviate $\sigma(x)$ for our original set $x = \{x_1 \leqslant x_2 \leqslant \cdots \leqslant x_n\}$ by σ. Using the definition of σ and (1), one can easily verify that

$$\sigma(\lambda u) = |\lambda| \sigma(u) \quad \text{and} \quad \sigma(u \pm v) \leqslant \sigma(u) + \sigma(v) \tag{4}$$

for all scalars $\lambda \in \mathbb{R}$ and vectors $u = (u_1, u_2, \ldots, u_n)$ and $v = (v_1, v_2, \ldots, v_n)$ in E^n. For any constant vector $a = (a, a, \ldots, a)$ in E^n, one also has

$$\sigma(a) = 0 \quad \text{and} \quad \sigma(u \pm a) = \sigma(u). \tag{5}$$

Now let $y = (y_1, y_2, \ldots, y_n)$, where $y_i = (1/R)\cdot(x_i - x_1)$ for $i = 1, 2, \ldots, n$. Then

$$0 = y_1 \leqslant y_2 \leqslant \cdots \leqslant y_{n-1} \leqslant y_n = 1$$

and

$$\sigma = \sigma(y)\cdot R \tag{6}$$

follows from (4) and (5). Our focus is on $\sigma(y)$. Let $e_i (1 \leqslant i \leqslant n-1)$ denote the vector $(0, 0, \ldots, 0, 1, 1, \ldots, 1)$ having zeros for the first i coordinates and ones for the remaining $n - i$ coordinates. Then

$$y = \sum_{i=1}^{n-1} (y_{i+1} - y_i) e_i$$

and, since $\sum_{i=1}^{n-1} (y_{i+1} - y_i) = 1$,

$$\sigma(y) \leqslant \max_{1 \leqslant i \leqslant n-1} \sigma(e_i) \tag{7}$$

follows from (4). The coordinates of e_i have mean $(n-i)/n$, and so

$$\sigma(e_i) = \sqrt{\frac{1}{n}\left\{ i\left(\frac{n-i}{n}\right)^2 + (n-i)\left(\frac{i}{n}\right)^2 \right\}} = \frac{\sqrt{(n-i)i}}{n}.$$

Therefore, we can obtain $\max_{1 \leqslant i \leqslant n-1} \sigma(e_i)$ by maximizing the discrete function

$$F(i) = (n-i)i$$

for $i = 1, 2, \ldots, n-1$. A few simple calculations reveal that

$$\max_{1 \leqslant i \leqslant n-1} F(i) = \left\{ \begin{array}{ll} \dfrac{n^2}{4}, & \text{when } n \text{ is even} \\ \dfrac{n^2-1}{4}, & \text{when } n \text{ is odd} \end{array} \right\}.$$

Therefore,

$$\max_{1 \leqslant i \leqslant n-1} \sigma(e_i) = \frac{1}{n}\cdot\sqrt{\max_{1 \leqslant i \leqslant n-1} F(i)} = \frac{\sqrt{[n^2/4]}}{n}. \tag{8}$$

Substitution of (8) into (7), and then (7) into (6), yields the second inequality in (I).

Remark: Sher, using (2), recently showed in [6] that $\sigma \leqslant R/2$. Our second inequality in (I) is sharper than $\sigma \leqslant R/2$ since $[n^2/4] = n^2/4$ when n is even, and $[n^2/4] = (n^2-1)/4 < n^2/4$ when n is odd. It is interesting to note that $R/2$ *is the best approximate upper bound for* σ *since* $\dfrac{\sqrt{[n^2/4]}}{n}\cdot R = \sqrt{1-(1/n^2)}\cdot\dfrac{R}{2}$ *increases to* $\dfrac{R}{2}$ *as n becomes large.*

The inequalities in (I) do more than provide range bounds for σ; they also provide information on how dispersed the scores $\{x_1 \leqslant x_2 \leqslant \cdots \leqslant x_n\}$ are. A good way to see this is to consider the following form of Tchebycheff's Theorem:

For any value $\xi > 0$, there are at most $M = n/\xi^2$ terms satisfying $|x_i - \mu| \geqslant \xi\sigma$.

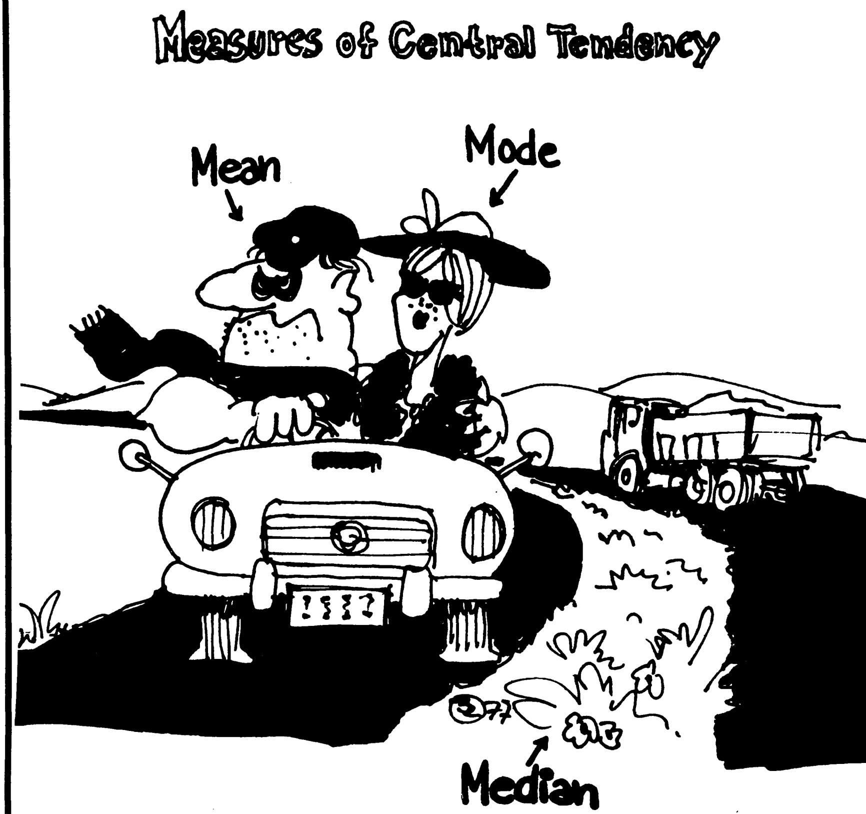
(This can also be stated and verified by replacing "$\geqslant$" and "at most" by "$>$" and "less than.") To verify this, assume that M terms satisfy $|x_i - \mu| \geqslant \xi\sigma$. Then

$$M\xi^2\sigma^2 \leqslant \sum_{M}(x_i - \mu)^2 \leqslant \sum_{j=1}^{n}(x_j - \mu)^2 = n\sigma^2 \quad \text{and} \quad M \leqslant \frac{n}{\xi^2}.$$

The following example illustrates how (I) refines Tchebycheff's Theorem.

Example 1. Assume that $n \geqslant 2$ and let $\xi = 1$. Then there are at most n terms (and possibly none!) that satisfy $|x_i - \mu| \geqslant \sigma$. According to (I), however, at least one of the data's endpoints meets this condition. (Otherwise, $R = |x_n - \mu| + |x_1 - \mu| < 2\sigma$ and $R/2 < \sigma$ contradicts $\sigma \leqslant R/2$.) Thus, at least one endpoint is not within σ distance of μ.

How Many Standard Deviations Apart are the Mean, Median, and Mode?



Cartoon by Geoffrey Akst & Jan Zanenberg; Courtesy J. CUNY Math Discussion Group 7 (1979–80).

If we take $\xi = \sqrt{n}$, then there are no terms satisfying $|x_i - \mu| > \sqrt{n}\,\sigma$. Therefore, Tchebycheff's Theorem shows that $R = |x_n - \mu| + |x_1 - \mu| \leqslant 2\sqrt{n}\,\sigma$ and $R/(2\sqrt{n}) \leqslant \sigma$. (This also follows since $n\sigma^2 = \sum_{i=1}^{n}(x_i - \mu)^2$ implies that $|x_i - \mu| \leqslant \sqrt{n}\,\sigma$ for each i.) But $R/(\sqrt{2n}) \leqslant \sigma$ from (I), and this lower bound is better than $R/(2\sqrt{n})$ by a factor of $\sqrt{2}$.

Finally, consider $\xi = \sqrt{n/2}$. Then at most one term (necessarily x_1 or x_n) satisfies $|x_i - \mu| > \sqrt{n/2}\,\sigma$. This is the same as stating that at least $n - 1$ terms (one of which must be x_1 or x_n) satisfy $|x_i - \mu| \leqslant \sqrt{n/2}\,\sigma$. But all this follows directly from (I) as well, since, $|x_1 - \mu| > \sqrt{n/2}\,\sigma$ and $|x_n - \mu| > \sqrt{n/2}\,\sigma$ together yield the contradiction that $R > \sqrt{2n}\,\sigma$ and $\sigma < R/\sqrt{2n}$.

3. Here we shall again use (3) in order to establish that

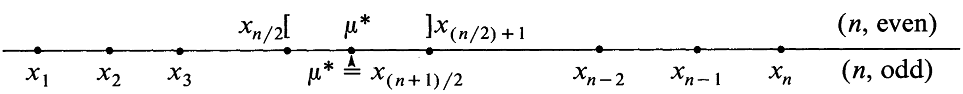
$$|\mu - \mu^*| \leqslant \sigma, \tag{II}$$

$$|\mu_0 - \mu| \leqslant \sqrt{\frac{n}{m}}\cdot\sigma, \quad \text{and} \quad |\mu_0 - \mu^*| \leqslant \frac{n}{m}\cdot\sigma \tag{III}$$

for the ordered set of real numbers $\{x_1 \leqslant x_2 \leqslant \cdots \leqslant x_n\}$. But first, for the sake of completeness, let us formally define the median. If $\#\{x_i < \zeta\}$ denotes the number of terms in $\{x_1 \leqslant x_2 \leqslant \cdots \leqslant x_n\}$ that are less than the real number ζ (and $\#\{x_i \leqslant \zeta\}$ is defined similarly), then the *median* of $\{x_1 \leqslant x_2 \leqslant \cdots \leqslant x_n\}$ is any real value μ^* that satisfies

$$\#\{x_i < \mu^*\} \leqslant \frac{n}{2} \leqslant \#\{x_i \leqslant \mu^*\}.$$

Example 2. If n is even, any value $\mu^* \in [x_{n/2}, x_{(n/2)+1}]$ can be taken as the median (although it is customary to take μ^* as the midpoint $\frac{1}{2}(x_{n/2} + x_{(n/2)+1})$ of this interval). If n is odd, then μ^* is uniquely defined as $x_{(n+1)/2}$.



The preceding example provides a visual proof of a rather interesting fact. Consider any real value ξ (mark such a point on the line above), and observe that for each $i = 1, 2, \ldots, [(n+1)/2]$:

$$|x_i - \mu^*| + |x_{n-i+1} - \mu^*| \leqslant |x_i - \xi| + |x_{n-i+1} - \xi|, \tag{9}$$

with equality holding if and only if $x_1 \leqslant \xi \leqslant x_{n-i+1}$. If we sum (9) over all the data, there follows the important result that

$$\sum_{i=1}^{n} |x_i - \mu^*| \leqslant \sum_{i=1}^{n} |x_i - \xi| \qquad \textit{for every real number } \xi. \tag{IV}$$

It is a simple matter now to prove that $|\mu - \mu^*| \leqslant \sigma$. Indeed,

$$\sum_{i=1}^{n} |x_i - \mu| \leqslant n\sigma$$

by (3), and an application of (IV) yields

$$|\mu - \mu^*| = \frac{1}{n}\left|\sum_{i=1}^{n} x_i - n\mu^*\right| = \frac{1}{n}\left|\sum_{i=1}^{n}(x_i - \mu^*)\right| \leqslant \frac{1}{n}\sum_{i=1}^{n}|x_i - \mu| \leqslant \sigma.$$

Remark: Evidently $R/\sqrt{2n} \nleqslant |\mu - \mu^*|$ for any set of data having $\mu = \mu^*$ and $R \neq 0$. The other, perhaps enticing, conjecture that $|\mu - \mu^*| \leqslant R/\sqrt{2n}$ may or may not be true. For instance, it is easily verified that $|\mu - \mu^*| \leqslant R/\sqrt{2n}$ for $1 \leqslant n \leqslant 4$. On the other hand, $\mu - \mu^* = \frac{2}{5} \nleqslant 1/\sqrt{10} = R/\sqrt{2n}$ for the set $\{0,0,0,1,1\}$.

Our next point of interest concerns the mode of $\{x_1 \leqslant x_2 \leqslant \cdots \leqslant x_n\}$. Suppose that μ_0 exists and has frequency m. The first inequality of

$$|\mu_0 - \mu| \leqslant \sqrt{\frac{n}{m}} \cdot \sigma \quad \text{and} \quad |\mu_0 - \mu^*| \leqslant \frac{n}{m} \cdot \sigma \tag{III}$$

is clear since $m|\mu_0 - \mu|^2 \leqslant \sum_{i=1}^{n}(x_i - \mu)^2 = n\sigma^2$; the second inequality in (III) follows from (IV) and (3). Specifically, $m|\mu_0 - \mu^*| \leqslant \sum_{i=1}^{n}|x_i - \mu^*| \leqslant \sum_{i=1}^{n}|x_i - \mu| \leqslant n\sigma$.

Remark: It should be noted that $|\mu_0 - \mu^*| \leqslant |\mu_0 - \mu| + |\mu - \mu^*| \leqslant (1 + \sqrt{n/m})\sigma$, and $1 + \sqrt{n/m} < n/m$ for $n/m > (3 + \sqrt{5})/2$.

The inequalities in (III) provide the following observation: The mean appears to be "more strongly attracted" to the mode than does the median.

4. Other Useful Inequalities. Suppose we partition the set $\{x_1 \leqslant x_2 \leqslant \cdots \leqslant x_n\}$ into nonempty subsets

$$A = \{x_1 \leqslant x_2 \leqslant \cdots \leqslant x_r\} \quad \text{and} \quad B = \{x_{r+1} \leqslant x_{r+2} \leqslant \cdots \leqslant x_n\}$$

whose means are denoted $\bar{x}_A$ and $\bar{x}_B$, respectively. It may be surprising to observe that

$$\mu - \sigma \cdot \sqrt{\frac{n-r}{r}} \leqslant \bar{x}_A \leqslant x_r \leqslant x_{r+1} \leqslant \bar{x}_B \leqslant \mu + \sigma \cdot \sqrt{\frac{r}{n-r}}\ . \tag{V}$$

We see that $\bar{x}_A \leqslant x_r \leqslant x_{r+1} \leqslant \bar{x}_B$. The other two inequalities are easily obtained from the Cauchy–Schwarz inequality. If we choose

$$a_i = \begin{cases} 1 - r/n, & \text{for } \ 1 \leqslant i \leqslant r \\ -r/n, & \text{for } \ r+1 \leqslant i \leqslant n \end{cases},$$

then

$$\sum_{i=1}^{n} a_i^2 = r\left(1 - \frac{r}{n}\right) \quad \text{and} \quad \sum_{i=1}^{n} a_i(x_i - \mu) = r(\bar{x}_A - \mu).$$

Substituting these expressions into (2) and simplifying, we get

$$(\bar{x}_A - \mu)^2 \leqslant \left(\frac{n-r}{r}\right) \cdot \sigma^2. \tag{10}$$

Since

$$n\mu = \sum_{i=1}^{n} x_i = r\bar{x}_A + (n-r)\bar{x}_B,$$

we can solve for $\bar{x}_A - \mu$. Substituting the right-hand side of

$$\bar{x}_A - \mu = -\left(\frac{n-r}{r}\right) \cdot (\bar{x}_B - \mu)$$

into (10), we obtain

$$(x_B - \mu)^2 \leqslant \left(\frac{r}{n-r}\right) \cdot \sigma^2. \tag{11}$$

Thus, by taking square roots in (10) and (11), we get (V).

Remark: It is interesting to note that when n is even and $r = n/2$, both $|\bar{x}_A - \mu| \leqslant \sigma$ and $|\bar{x}_B - \mu| \leqslant \sigma$ follow from (V). This indicates a rather heavy concentration of values about μ. (Think of the centers of mass, $\bar{x}_A$ and $\bar{x}_B$, being within σ distance of the total center of mass μ.)

A simple application of (V) yields the following for $r = 1, 2, \ldots, \left[\frac{n+1}{2}\right]$:

$$\mu - \sigma \cdot \sqrt{\frac{n-r}{r}} \leqslant x_r \leqslant x_{n-r+1} \leqslant \mu + \sigma \cdot \sqrt{\frac{n-r}{r}}. \tag{VI}$$

The first inequality in (VI) has already been established in (V). Since $r \leqslant (n+1)/2$, we also have $r \leqslant n - r + 1$ and $x_r \leqslant x_{n-r+1}$. Now let $y_i = -x_{n-i+1}$ for each $i = 1, 2, \ldots, n$. Then the first inequality in (VI) can be used to relate the mean $\mu(y)$ and the standard deviation $\sigma(y)$ of $\{y_1 \leqslant y_2 \leqslant \cdots \leqslant y_n\}$ by

$$\mu(y) - \sigma(y) \cdot \sqrt{\frac{n-r}{r}} \leqslant y_r.$$

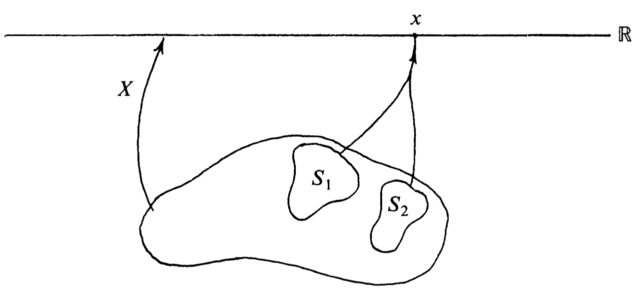
But it is well known (or easily verified!) that $\mu(y) = -\mu$ and $\sigma(y) = \sigma$. And since $y_r = -x_{n-r+1}$, the above inequality can be rewritten as

$$x_{n-r+1} \leqslant \mu + \sigma \cdot \sqrt{\frac{n-r}{r}}.$$

The thrust of (VI) *is that* $|\xi - \mu| \leqslant \sigma \cdot \sqrt{(n-r)/r}$ *for every real value* ξ *in the interval* $[x_r, x_{n-r+1}]$.

Example 3. Consider the important case where $\xi = \mu^*$. Then (VI) yields a nice, crisp proof of (II). To see this, take $r = [(n+1)/2]$. If n is even, $r = n/2$ and $n - r + 1 = (n/2) + 1$. Furthermore, $(n-r)/r = 1$. Therefore, (Example 2) $\mu^* \in [x_r, x_{n-r+1}]$ and $|\mu - \mu^*| \leqslant \sigma$ follows from (VI). If n is odd, then $r = (n+1)/2$ and $(n-r)/r = (n-1)/(n+1)$. Therefore, $\mu^* = x_r$ and it also follows from (VI) that $|\mu - \mu^*| \leqslant \sqrt{(n-1)/(n+1)} \cdot \sigma < \sigma$.

5. Probability Distributions. A *random variable* on a probability space (S, P) is simply a real-valued mapping X on (S, P). Notationwise, $P(X = x)$ denotes the probability of the set $X^{-1}(x) = \{s \in S : X(s) = x\}$.



$P(X = x) = P\{X^{-1}(x)\} = P(S_1 \cup S_2)$

In this section, we shall briefly consider random variables that take on a discrete (i.e., finite or countably infinite) set of values $X(S) = \{x_1 < x_2 < \cdots\}$. It is standard, in this context, to let $P(X = x_i)$ be denoted p_i for each natural number i. A discrete random variable X has *expected value*

$$E(X) = \sum_i x_i p_i \tag{12}$$

and *standard deviation*

$$s(X) = \sqrt{\textstyle\sum_i \{x_i - E(X)\}^2 p_i} = \sqrt{\textstyle\sum_i x_i^2 p_i - \{E(X)\}^2}\,, \tag{13}$$

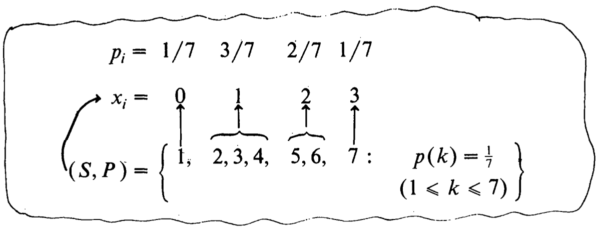
provided that the above sums are absolutely convergent. (Thus, $E(X)$ and $s(X)$ always exist when $X(S)$ is finite.)

The *mode* of X is defined as any value $\mu_0 \in X(S)$ that has the absolute maximal probability p_0. (The mode does not exist when each member of $X(S) = \{x_1 < x_2 < \cdots < x_n\}$ has the same probability since $1/n = \max_{1 \leqslant i \leqslant n} p_i$ is a relative, but not absolute, maximal probability.)

The *median* of X is not always uniquely defined; it is any real value μ^* that satisfies

$$\sum_{x_i < \mu^*} p_i \leqslant \frac{1}{2} \leqslant \sum_{x_i \leqslant \mu^*} p_i\,. \tag{14}$$

Example 4. The set of seven numbers $\{0,1,1,1,2,2,3\}$, also written $\{(x_i, f_i): \sum_{i=1}^{4} f_i = 7\} = \{(0,1),(1,3),(2,2),(3,1)\}$, may be viewed as a probability distribution $\{(x_i, p_i = f_i/7): 1 \leqslant i \leqslant 4\} = \{(0,\frac{1}{7}),(1,\frac{3}{7}),(2,\frac{2}{7}),(3,\frac{1}{7})\}$.



x_i	f_i
0	1
1	3
2	2
3	1

$$\sum_{i=1}^{4} f_i = 7.$$

In general, every finite set of real numbers $\{(x_i, f_i): \sum_{i=1}^{n} f_i = N\}$ can be viewed as a probability distribution $\{(x_i, p_i = f_i/N): 1 \leqslant i \leqslant n\}$. This set and its corresponding distribution have the same range. The mode and the median can also be taken as the same value. Furthermore,

$$E(X) = \sum_{i=1}^{n} x_i p_i = \frac{1}{N} \sum_{i=1}^{n} x_i f_i = \mu \tag{15}$$

and

$$s(x) = \sqrt{\sum_{i=1}^{n} x_i^2 p_i - \{E(X)\}^2} = \sqrt{(1/N)\sum_{i=1}^{n} x_i^2 f_i - \mu^2} = \sigma. \tag{16}$$

Example 5. Consider the distribution $\{(x_i, p_i): 1 \leqslant i \leqslant 6\} = \{(5,.1),(50,.4),(55,.2),(60,.1),(75,.1),(100,.1)\}$, where it is understood that $X(S) = \{x_i : 1 \leqslant i \leqslant 6\}$ for a random variable X on some probability space (S,P). Clearly, $X(S)$ has a range $R = 95$. Furthermore, $\mu_0 = 50$ corresponds to $p_0 = .4$. A few simple calculations reveal that $E(X) = 55$ and $s(X) \approx 22.47$. And, according to (14), the median is any value $\mu^* \in [50, 55)$. Thus our inequalities

$$|E(X) - \mu^*| \leqslant s(X) \quad \text{and} \quad |\mu_0 - \mu^*| \leqslant |\mu_0 - E(X)| \leqslant \frac{1}{\sqrt{p_0}} \cdot s(X)$$

are consistent with (II) and (III). Although $s(X) \leqslant \left(\sqrt{[6^2/4]}\,/6\right) \cdot R$, it is not true that $R/\sqrt{2(6)} \leqslant s(X)$, as might have been anticipated from (I). This seeming anomaly can be cleared up once we observe that $\{p_i : 1 \leqslant i \leqslant 6\}$ is a set of rational numbers having a lowest common denominator (LCD) of $N = 10$. Indeed, our distribution $\{(x_i, p_i) : 1 \leqslant i \leqslant 6\}$ and the corresponding set of ten numbers $\{(x_i, f_i = 10p_i) : 1 \leqslant i \leqslant 6\}$ (our set of winning lottery payoffs described in the introduction) have the same range and the same mode. Furthermore, $E(X) = \mu$ and $s(X) = \sigma$ by (15) and (16). If we take the median of $\{(x_i, p_i) : 1 \leqslant i \leqslant 6\}$ to be the uniquely defined median $\mu^* = 52.50$ of $\{(x_i, f_i) : 1 \leqslant i \leqslant 6\}$, then everything in (I)–(III) can be rewritten for $\{(x_i, p_i) : 1 \leqslant i \leqslant 6\}$ once we replace "n" by "N" and "m/n" by "p_0."

One question still remains here: Do (II) and (III) hold for every possible median $\mu^* \in [50, 55)$ of $\{(x_i, p_i) : 1 \leqslant i \leqslant 6\}$? The answer is affirmative. Suppose ξ is any real number such that $\sum_{x_i < \xi} p_i \leqslant \frac{1}{2} \leqslant \sum_{x_i \leqslant \xi} p_i$. For the set $\{(x_i, f_i = 10p_i) : 1 \leqslant i \leqslant 6\}$, this translates to

$$\sum_{x_i < \xi} f_i \leqslant 5 \leqslant \sum_{x_i \leqslant \xi} f_i . \tag{17}$$

Therefore, $\xi \in [x_5, x_6]$ and it remains only to take $r = 10/2 = 5$ in (VI) in order to conclude (as in Example 3) that $|\mu - \xi| \leqslant \sigma$ (i.e., $|E(X) - \mu^*| \leqslant s(X)$).

Remark: The preceding result holds in general. If "10" is replaced by "N," then (17) becomes $\sum_{x_i < \xi} f_i \leqslant N/2 \leqslant \sum_{x_i \leqslant \xi} f_i$. Thus, $\xi \in [x_{N/2}, x_{(N/2)+1}]$ when N is even, and $\xi = x_{(N+1)/2}$ when N is odd. In either case, we have $|E(X) - \xi| \leqslant s(X)$.

We can summarize our discovery as follows:

Let $\{(x_i, p_i) : 1 \leqslant i \leqslant n\}$ be a distribution of ordered values $\{x_1 < x_2 < \cdots < x_n\}$, where $\{p_i : 1 \leqslant i \leqslant n\}$ is a set of rational numbers, in reduced form, that has an LCD *of N. Then*:

$$\frac{R}{\sqrt{2N}} \leqslant s(X) \leqslant \frac{\sqrt{[N^2/4]}}{N} \cdot R \tag{I$'$}$$

and

$$|E(X) - \mu^*| \leqslant s(X). \tag{II$'$}$$

If $p_0 = \max_{1 \leqslant i \leqslant n} \{p_i > 1/n\}$ exists, then each $\mu_0 \in \{x_1, \ldots, x_n : p_i = p_0\}$ satisfies

$$|\mu_0 - E(X)| \leqslant \frac{1}{\sqrt{p_0}} s(X) \quad \text{and} \quad |\mu_0 - \mu^*| \leqslant \frac{1}{p_0} \cdot s(X). \tag{III$'$}$$

Two important questions arise naturally: What, if anything, can be said for finite distributions $\{(x_i, p_i) : 1 \leqslant i \leqslant n\}$ whose p_i are not all rational? And, what about distributions which are infinite? The following examples begin to hint at the direction in which we are heading.

Example 6. The two-point distribution $\{(x_i, p_i) : i = 1, 2\} = \{(0, 1 - (\sqrt{2}/2)), (1, \sqrt{2}/2)\}$ has $R = 1$, and $\mu_0 = 1$ corresponds to $p_0 = \sqrt{2}/2$. Here $E(X) = \sqrt{2}/2$ and

$s(X) = \sqrt{(\sqrt{2}-1)/2}$. By definition, $\mu^* = 1$. Thus, (II)′ and (III)′ hold, but the first inequality in (I)′ fails. If p_1 and p_2 are interchanged, (II)′ and (III)′ still hold and (I)′ fails again. (It may be instructive to verify that (I)′ holds, for a two-point distribution, if and only if $p_1 = p_2 = \frac{1}{2}$.)

Example 7. The distribution $\{(x_i, p_i) : i = 1, 2, \ldots\} = \{(i, 2^{-i}) : i = 1, 2, \ldots\}$ has $\mu_0 = 1$ corresponding to $p_0 = \frac{1}{2}$, and the median is any value $\mu^* \in [1, 2]$. Furthermore, $E(X) = 2$ and $s(X) = \sqrt{2}$. (For our purposes it suffices to note that $s^2(X) = \sum_{i=1}^{\infty}(i^2/2^i) - 4 > \sum_{i=1}^{7}(i^2/2^i) - 4 > 1$.) Thus, although (I)′ is meaningless (since $X(S)$ is unbounded), both (II)′ and (III)′ remain true.

Example 8. As is well known, the familiar normal distribution (corresponding to a random variable with density function $f(x) = (1/\sqrt{2\pi}) \cdot e^{-x^2/2}$) has $E(X) = \mu^* = 0 = \mu_0$ (with $p_0 = 1/\sqrt{2\pi}$) and $s(X) = 1$. Thus, (II)′ and (III)′ are satisfied for this continuous distribution.

By now the conjecture that (II)′ and (III)′ are true for every probability distribution should have surfaced. This is almost the case; elementary proofs that shed light on these intriguing conjectures will be given in the next issue of TYCMJ.

6. Conclusion . . . or Beginning? Recent Classroom Capsules ([6], [1]) stimulated the research that lead to this article and the generalizations that will follow. Two related notes ([2], [8]), just discovered by the authors, appear to have been lost in the literature. In this vein, the authors thank Stephen Book for bringing [3]–[5] and [9] to our attention. Unfortunately, these articles were written for "specialists," and so they are not readily accessible to a wide audience.

It is our hope that the article will be welcomed by TYCMJ readers as an invitation and a challenge—an invitation to discover (or rediscover!) other useful relationships among the important statistical measures and a challenge to present these discoveries in a manner that can be appreciated by one's colleagues. A good example is the recent proof [7] that:

For any set of real numbers $[x_1 \leqslant x_2 \leqslant \cdots \leqslant x_n\}$, *the set of standard scores* $\{z_i = (x_i - \mu)/\sigma : 1 \leqslant i \leqslant n\}$ *satisfies* $\sqrt{2n} \leqslant \sum_{i=1}^{n}|z_i| \leqslant n$.

(Note, therefore, that $z_1 = -1 = -z_2$ for any two values $\{x_1, x_2 > x_1\}$.)

REFERENCES

1. S. A. Book and L. Sher, How close are the mean and the median?, TYCMJ, 10:3 (1979) 202–204.
2. H. Hotelling and L. M. Solomons, The limits of a measure of skewness, Ann. Math. Stat., 3 (1932) 141–142.
3. K. N. Majindar, Improved bounds on a measure of skewness, Ann. Math. Stat., 33 (1962) 1192–1194.
4. C. L. Mallows and D. Richter, Inequalities of Chebyshev type involving conditional expectations, Ann. Math. Stat., 40 (1969) 1922–1932.
5. S. Moriguti, A modification of Schwarz's inequality with applications to distributions, Ann. Math. Stat., 24 (1953) 107–113.
6. L. Sher, The range of the standard deviation, TYCMJ, 10:1 (1974) 33 and MATYCJ, 11:3 (1977) 197.
7. L. Sher and W. Page, Bounds for the sum of the absolute standard scores, TYCMJ, 10:5 (1979) 351–353.

8. G. W. Thomson, Bounds for the ratio of range to standard deviation, Biometrika, 42 (1955) 268–269.
9. H. Wolkowicz and G. P. H. Styan, Extension of Samuelson's inequality, Amer. Statist., 33:3 (1979) 143–144.

As noted in the last paragraph of Section 5, the concluding installment of this article will appear in the next issue of this journal.