TAKE IT TO THE MAT

A NEWSLETTER ADDRESSING THE FINER POINTS OF MATHEMATICS INSTRUCTION

Southern Nevada Regional Professional Development Program
April/May 2005 — High School Edition

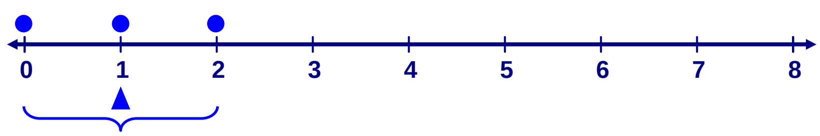
www.rpdp.net

In this final issue of *Take It to the MAT* for the 2004–05 school year, we'll look at a common data analysis question with which students have difficulty, but is actually fairly easy to understand. The question is about what happens to summary measures—mean, median, range, etc.—if the data from which they are computed are transformed in some consistent way.

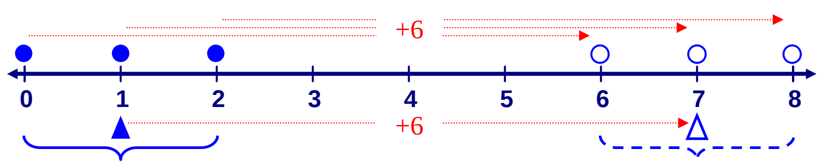
For example: *Given a set of data, what would happen to its mean if every data value were increased by 6? What would happen to its range?*

There are two approaches to this question. The first, would be to simply create a data set, find the mean and range, add 6 to each data value, and compute the new mean and range. For instance, take the set {0, 1, 2}. It has a mean of $\frac{0+1+2}{3}=1$, and a range of 2 – 0 = 2. Adding six to each datum—datum is singular, data is plural—gives us the new set {6, 7, 8} which has a mean of $\frac{6+7+8}{3}=7$, and a range of 8 – 6 = 2. The mean increased by 6, but the range stayed the same. This, however, is a brute force approach that only serves us in this one case. What follows is a second approach that will lead us to a more general conclusion.

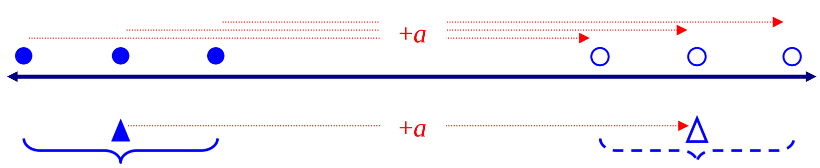
We'll start with our chosen data set {0, 1, 2}. Mark those values on a number line with dots, their mean with a triangle, and the range with a horizontal brace.



Now, we'll move each point six units to the right.



As we can see, the mean also shifts 6 units to the right and the range stays the same. Of course, this is true of any set of data, and any shift. If one adds (or subtracts) the same number, say a, to all data, the mean will change in the same way, but the range stays the same.



Actually, the usual measures of center (mean, median, mode, etc.) shift in the same way as the data, measures of location (minimum, maximum, quartiles, percentiles, etc.), shift in the same way, and measures of spread (range, interquartile range, standard deviation) do not change.

Think of it as having several people in a single file line and asking them to take a step forward. The center moves as they move, their locations change in the same way, and each is in the same relative location to the others and center, but they maintain the same distances apart.

Depending on the level of the students, one could take several approaches to "prove" this. One method would be to do "brute force" on several sets of data and recognize a pattern. Not a proof to be sure, but it certainly would get the point across. A "graphical" representation, like that on page one, might be more meaningful to learners, or at least provide a more general justification as to what happens to measures of center and spread when adding (or subtracting) a constant.

In algebra classes, it should be no problem for students to do a proof that goes something like this: We have a data set of n values, $\{x_1, x_2, x_3, \ldots, x_n\}$, where $x_1 \le x_2 \le x_3 \le \ldots \le x_n$. What happens to the mean and range of the data if the constant a is added to each datum?

The mean of the data is $\dfrac{x_1+x_2+x_3+\ldots+x_n}{n}$ and the range is $x_n - x_1$. When a is added to each datum, the set $\{x_1 + a, x_2 + a, x_3 + a, \ldots, x_n + a\}$ has a mean of

$$\frac{(x_1+a)+(x_2+a)+(x_3+a)+\ldots+(x_n+a)}{n} = \frac{(x_1+x_2+x_3+\ldots+x_n)+na}{n} = \frac{x_1+x_2+x_3+\ldots+x_n}{n}+a$$

and a range of $(x_n+a)-(x_1+a) = x_n + a - x_1 - a = x_n - x_1$.

Turning things up a notch, precalculus students should be able to approach this using summation notation. The mean of the orginal data is $\dfrac{1}{n}\sum_{i=1}^{n} x_i$ and the mean of the transformed set is

$$\frac{1}{n}\sum_{i=1}^{n}(x_i+a) = \frac{1}{n}\left[\sum_{i=1}^{n}x_i + \sum_{i=1}^{n}a\right] = \frac{1}{n}\left[\sum_{i=1}^{n}x_i + na\right] = \frac{1}{n}\sum_{i=1}^{n}x_i + a.$$

In calculus, we might ask students to compare the mean of a probability density function $f(x)$, which is $\int_{-\infty}^{\infty} x\, f(x)\,dx$, to the transformed mean when we add a to x, $\int_{-\infty}^{\infty} (x+a)\, f(x+a)\,dx$.

So what's the big deal with adding or subtracting a constant? It could make our life easier. Take the case of the data {34, 38, 39, 40, 43, 45, 48}. A quick glance at the data would tell us that the mean is probably about 40. Let's subtract 40 from each datum: {–6, –2, –1, 0, 3, 5, 8}. That's a lot nicer group of numbers to add up. Negative one and –2 cancel the three, –6 and 5 add up to –1, plus 8 equals 7. The sum of 7 divided by 7 data values gives us a mean of 1. Since our new data is 40 less than the orginal set, so must the mean. Thus, the orginal mean is 41. This isn't a big deal if one has a calculator, but on the *Nevada High School Proficiency Exam* ….

A final question: what if we multiplied (or divided) each datum by some constant? What would happen to measures of center, location, and spread then? The answer to that is left to the reader.