MONT 105N – Analyzing Environmental Data
Thinking about central measures–"think, pair, share"
February 1, 2019

Background

Recall that we have introduced two ways to estimate a "middle" or center of a collection of numerical data: $x_1, x_2, \ldots, x_N$. The first is the *mean* or numerical average. We will use the notation

$$\overline{x} = \frac{x_1 + x_2 + \cdots + x_N}{N} = \frac{1}{N} \cdot x_1 + \cdots + \frac{1}{N} \cdot x_N$$

for this. The second is the *median*. The easiest way to describe this is to give the usual method for computing it. To find the median (by hand), we would first sort the numbers into increasing (or non-decreasing) order:

$$x_1 \leq x_2 \leq \cdots \leq x_N$$

(note that this might involve *renumbering* the list if it is not sorted to begin with). Then

$$\text{median} = \begin{cases} x_{M+1} & \text{if } N = 2M+1 \text{ is an odd number} \\ \frac{x_M + x_{M+1}}{2} & \text{if } N = 2M \text{ is an even number} \end{cases}$$

Questions

(1) Let $N = 8$ and

$$x_1 = 8.4, x_2 = 3.5, x_3 = 5.7, x_4 = 3.5, x_5 = 5.7, x_6 = 6.2, x_7 = 3.5, x_8 = 9.3$$

What are $\overline{x}$ and the median for this data set? (Note that the x_i do not need to be all different numbers!) Why are the mean and median different?

(2) What happens to the mean and the median in (1) if x_1 is changed to 10.2?

(3) The x_i in part (1) above can be thought of as a collection of 5 different numbers, with some "repeats." If you group the repeated numbers together, explain how $\overline{x}$ could also be viewed as a *weighted average* of the distinct number values, with weights proportional to the number of times each number appears. (Ask me if the idea of a weighted average is not clear!)

(4) Give examples where each of these two methods for estimating a middle might be superior to the other. That is, try to think of situations where the mean of a collection of numbers might be more informative than the median, and other situations where the opposite would be true.