Connecting Position and Dispersion

The mean of a data set, $\bar{x} = \frac{1}{n}\sum_{k=1}^{n} x_k$, is connected to the variance of the same set by this interesting fact: define, for any real number m, the quantity

$$v(m) = \frac{1}{n}\sum_{k=1}^{n} (x_k - m)^2$$

Proposition 1. *The function $v(m)$ has a minimum for $m = \bar{x}$*

The proof is not hard: expanding the square, and regrouping the terms, we end up with a quadratic function of m. We can now verify that its vertex occurs at the mean. The main difficulty in the general proof consists in being familiar with summation notation (or deal with cumbersome expressions):

$$\frac{1}{n}\sum_{k=1}^{n} \left(x_k^2 - 2x_k m + m^2\right) = \frac{1}{n}\sum_{k=1}^{n} x_k^2 - 2\frac{m}{n}\sum_{k=1}^{n} x_k + m^2 =$$
$$m^2 - 2m\bar{x} - \frac{1}{n}\sum_{k=1}^{n} x_k^2$$

Now, the third term is irrelevant for the position of the vertex in terms of m. The formula for the vertex horizontal coordinate is, from algebra, given for a generic quadratic function $az^2 + bz + c$, $-\frac{b}{2a}$. Here, this means

$$m = -\frac{-2\bar{x}}{2} = \bar{x}$$

As for the median, introducing the *average deviation* from a reference number

$$d(m) = \frac{1}{n}\sum_{k=1}^{n} \left| x_k - m \right|$$

one can prove that

Proposition 2. *The function $d(m)$ has a minimum when m equals the median*

An intuitive argument for the case of an odd number of points is as follows. Suppose you have an odd number of points on the real line, call it $2k+1$, so that the k-th point is the median M. Looking at $d(M)$, and now choosing a point m smaller than M, we will notice that we decrease the sum of distances from points smaller than M, increase, by the same amount, the sum of distances from points larger than M, but also add to the total the now non-zero distance from M, increasing the average deviation. Similarly, if we choose a point larger than M. The argument for an even number of points $2k$ is similar, and it also shows that any values between the k-th and $(k+1)$-th will do the job.