Median

Given data $x_1, x_2, \ldots, x_n$, let $y_1, y_2, \ldots, y_n$ be the same data in nondecreasing order:

$$y_1 \le y_2 \le \cdots \le y_n.$$

If n is odd, the **median** M is the middle value:

$$M = y_{(n+1)/2}.$$

If n is even, any number between the middle two values $y_{n/2}$ and $y_{n/2+1}$ could be considered "a" **median** of the data. When n is even, typically "the" **median** is taken to be the midpoint between, or average of, the middle two values:

$$M = \frac{y_{n/2} + y_{n/2+1}}{2}.$$

The Median as an Error Minimizer

PROBLEM: Prove that the sum of the absolute deviations of the data from x is minimized when x is a median of the data. In other words, prove that

$$\sum_{j=1}^{n} |x - x_j|$$

is minimized when $x = M$.

Hint

Show that for $n > 2$, if $y \in [y_1, y_n]$ then

$$\sum_{j=1}^{n} |y - y_j| = \sum_{j=2}^{n-1} |y - y_j| + y_n - y_1$$

while if $y \notin [y_1, y_n]$ then

$$\sum_{j=1}^{n} |y - y_j| > \sum_{j=2}^{n-1} |y - y_j| + y_n - y_1.$$

Now use mathematical induction to prove the proposition.