

how spread out the data really is, measured four different ways

What dispersion measures *

- the mean alone hides how spread out data is. Two datasets can share the same mean and look completely different
- dispersion measures answer one question: how far, on average, do the observations sit from some central value
- the four measures here, range, mean deviation, variance and standard deviation, each answer that question a different way

Range

biggest value minus smallest value in the dataset $R = x_{\max} - x_{\min}$



range measured directly: the biggest and smallest values, and the distance between them

range is quick to compute but throws away every value except the two extremes. One outlier can swing it completely

Four measures, four levels of detail

Range

uses only the two extreme values, ignores everything in between

Mean deviation

uses every value's absolute distance from the mean

Variance

uses every value's SQUARED distance from the mean

Standard deviation

the square root of variance, back in the original units

Mean deviation about the mean

ungrouped data, n observations

$$MD = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|$$

grouped data, $N = \sum f_i$

$$MD = \frac{1}{N} \sum f_i |x_i - \bar{x}|$$

the vertical bars matter: every deviation is made positive before averaging, or the positive and negative deviations about the mean would cancel out to exactly 0

Mean deviation about the median *

- the same idea, but measured from the median M instead of the mean: $MD = \frac{1}{n} \sum |x_i - M|$
- mean deviation about the median is always the SMALLEST possible mean deviation from any single value, a property unique to the median
- mean deviation is rarely asked for grouped data with a shortcut method. Direct substitution into the formula above is usually fastest

Worked: mean deviation about the mean

Q. find the mean deviation about the mean for the data 2, 4, 6, 8, 10

- mean: $\bar{x} = \frac{2+4+6+8+10}{5} = \frac{30}{5} = 6$
- absolute deviations from 6: $|2-6|=4$, $|4-6|=2$, $|6-6|=0$, $|8-6|=2$, $|10-6|=4$
- sum of absolute deviations: $4+2+0+2+4=12$
- $MD = \frac{12}{5} = 2.4$

Answer

mean deviation about the mean is 2.4