MthStat 465, Spring 2005, Lecture Number 1// Measures of Dispersion

A single number will not indicate how spread out the data is. In order to capture spread, as well as location, there are several possibilities. In this section we discuss univariate data, represented by $(d_1, \ldots, d_N)$.

1. Range

This is perhaps the crudest measure of spread. Simply subtract the smallest observation from the largest. This is very succeptible to outliers.

2. Interquartile Range

This is less sensitive to outliers. The interquartile range is obtained by substracting the first quartile from the third quartile.

3. Five number summary

The five number summary gives, in order, the minimum, the first quartile, the median, the third quartile and the maximum. It can be represented graphically by a box-and-whisker plot.

4. Mean Absolute Deviation

Once one has decided on a middle value, we can quantify how far the data lies from it. If x denotes the middle value, then

$$N^{-1} \sum_{k=1}^{N} |x - d_k|$$

is the **Mean Absolute Deviation** about x. As we have seen, the smallest Mean Absolute Deviation is when x is a median of the data. It is quite common to give the Mean Absolute Deviation about the mean, even though the median will do better.

5. Standard Deviation

We have seen that the mean, $\overline{d}$ is the unique minimizer of

$$S(x) := \sum_{k=1}^{N} (x - d_k)^2.$$

The quantity

$$s := \sqrt{N^{-1} S(\overline{d})}$$

is called the **standard deviation**. The standard deviation is to the mean what the Mean Absolute Deviation about the median is to median.

These minimization ideas will re-occur in the next lecture about Least Squares.