

# 1. What are *mean, median, mode, variance, and standard deviation*

## The Mean and Mode

The *sample mean* is the average and is computed as the sum of all the observed outcomes from the sample divided by the total number of events. We use  $\bar{x}$  as the symbol for the sample mean. In math terms,

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x$$

where  $n$  is the sample size and the  $x$  correspond to the observed values.

## Median, and Trimmed Mean

One problem with using the mean, is that it often does not depict the typical outcome. If there is one outcome that is very far from the rest of the data, then the mean will be strongly affected by this outcome. Such an outcome is called an *outlier*. An alternative measure is the median. The *median* is the middle score. If we have an even number of events, we take the average of the two middles. The median is better for describing the typical value. It is often used for income and home prices.

## Variance, Standard Deviation and Coefficient of Variation

The mean, mode, median, and trimmed mean do a nice job in telling where the center of the data set is, but often we are interested in more. For example, a pharmaceutical engineer develops a new drug that regulates iron in the blood. Suppose she finds out that the average sugar content after taking the medication is the optimal level. This does not mean that the drug is effective. There is a possibility that half of the patients have dangerously low sugar content while the other half have dangerously high content. Instead of the drug being an effective regulator, it is a deadly poison. What the pharmacist needs is a measure of how far the data is spread apart. This is what the variance and standard deviation do. First we show the formulas for these measurements. Then we will go through the steps on how to use the formulas.

We define the *variance* to be

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x - \bar{x})^2$$

and the *standard deviation* to be

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x - \bar{x})^2}$$

### 3. What is Gaussian distribution?

normal distribution, also called Gaussian distribution, the most common distribution function for independent, randomly generated variables. Its familiar bell-shaped curve is ubiquitous in statistical reports, from survey analysis and quality control to resource allocation.

The graph of the normal distribution is characterized by two parameters: the mean, or average, which is the maximum of the graph and about which the graph is always symmetric; and the standard deviation, which determines the amount of dispersion away from the mean. A small standard deviation (compared with the mean) produces a steep graph, whereas a large standard deviation (again compared with the mean) produces a flat graph. *See the figure.*