The Normal Distribution

The normal distribution is extremely important in statistics and has some specific characteristics that make it so useful.

Characteristics of the Normal Distribution

- It is symmetrical, meaning that the higher half and the lower half of the distribution are mirror images of each other. Second, the mean, median, and mode are all in the same place, in the center of the distribution (i.e., the peak of the bell curve).
- Because of this second feature, the normal distribution is highest in the middle, so it is **unimodal**, and it curves downward toward the higher values at the right side of the distribution and toward the lower values on the left side of the distribution.
- Finally, the normal distribution is **asymptotic**, meaning that the upper and lower tails of the distribution never actually touch the baseline, also known as the X axis. This is important because it indicates that the probability of a score in a distribution occurring by chance is never zero.

Why Is the Normal Distribution So Important?

- A number of statistics that begin with the assumption that scores are normally distributed.
- When researchers want to know what the exact probability is of something occurring in their sample just due to chance. For example, if the average student in my sample consumes 2,000 calories a day, what are the chances, or probability, of having a student in the sample who consumes 5,000 calories or more a day?
- The three characteristics of the normal distribution are each critical in statistics because they allow us to make good use of **probability** statistics.
- In addition, researchers often want to be able to make inferences about the population based on the data they collect from their sample. To determine whether some phenomenon observed in a sample represents an actual phenomenon in the population from which the sample was drawn, **inferential** statistics are used.
- For example, suppose I begin with an assumption that in the population of men and women there is no difference in the average number of calories consumed in a day. This assumption of no difference is known as a **null hypothesis**.
- Now suppose that I select a sample of men and a sample of women, compare their average daily calorie consumption, and find that the men eat an average of 200 calories more per day than do the women. Given my null hypothesis of no difference, what is the probability of finding a difference this large between my samples ***by chance***?
- To calculate these probabilities, I need to rely on the normal distribution, because the characteristics of the normal distribution allow statisticians to generate exact probability statistics.
- It is important to note that the normal distribution is what is known in statistics as a theoretical distribution.

The Relationship Between the Sampling Method and the Normal Distribution

The relationship between the normal distribution and the sampling methods is as follows. The probabilities generated from the normal distribution depend on (a) the shape of the distribution and (b) the idea that the sample is not somehow systematically different from the population.

- If I select a sample randomly from a population, I know that this sample may not look the same as another sample of equal size selected randomly from the same population. But any differences between my sample and other random samples of the same size selected from the same population would differ from each other randomly, not systematically. In other words, my sampling method was not biased such that I would continually select a sample from one end of my population (e.g., the more wealthy, the better educated, the higher achieving) if I continued using the same method for selecting my sample.
- Contrast this with a convenience sampling method. If I only select schools that are near my home or work, I will continually select schools with similar characteristics. For example, if I live in the "Bible Belt" (an area of the southern United States that is more religious, on average, than other parts of the country), my sample will probably be biased in that my sample will likely hold more fundamentalist religious beliefs than the larger population of schoolchildren.
- Suppose that I live and work in Cambridge, Massachusetts. Cambridge is in a section of the country with an inordinate number of highly educated people because there are a number of high-quality universities in the immediate area (Harvard, MIT, Boston College, Boston University, etc.).
  - If I conduct a study of student achievement using a convenience sample from this area, and try to argue that my sample represents the larger population of students in the United States, probabilities that are based on the normal distribution may not apply.
  - That is because my sample will be more likely than the national average to score at the high end of the distribution.
  - If, based on my sample, I try to predict the average achievement level of students in the United States, or the percentage that score in the bottom quartile, or the score that marks the 75th percentile, all of these predictions will be off, because the probabilities that are generated by the normal distribution assume that the sample is not biased. If this assumption is violated, we cannot trust our results.

Skew and Kurtosis

Skew

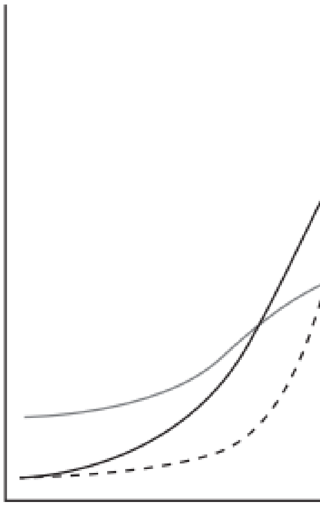
Two characteristics used to describe a distribution of scores are **skew** and **kurtosis**.

- When a sample of scores is not normally distributed (i.e., not the bell shape), there are a variety of shapes it can assume. One way a distribution can deviate from the bell shape is if there is a bunching of scores at one end and a few scores pulling a tail of the distribution out toward the other end.
- If there are a few scores creating an elongated tail at the higher end of the distribution, it is said to be positively skewed.
- If the tail is pulled out toward the lower end of the distribution, the shape is called negatively skewed.
- Skew does not affect the median as much as it affects the mean, however.
  - So a positively skewed distribution will have a higher mean than median, and a negatively skewed distribution will have a lower mean than median.
  - If you recall that the mean and the median are the same in a normal distribution, you can see how the skew affects the mean relative to the median.
- skewed distributions can distort the accuracy of the probabilities based on the normal distribution.
- For example, if most of the scores in a distribution occur at the lower end with a few scores at the higher end (positively skewed distribution), the probabilities that are based on the normal distribution will underestimate the actual number of scores at the lower end of this skewed distribution and overestimate the number of scores at the higher end of the distribution. In a negatively skewed distribution, the opposite pattern of errors in

prediction will occur.

Kurtosis

- Kurtosis refers to the shape of the distribution in terms of height, or flatness.
- When a distribution is symmetrical but has a peak that is higher than that found in a normal distribution, it is called leptokurtic.
- When a distribution is flatter than a normal distribution, it is called platykurtic.
- Because the normal distribution contains a certain percentage of scores in the middle area (i.e., about 68 percent of the scores fall between one standard deviation above and one standard deviation below the mean), a distribution that is either platykurtic or leptokurtic will likely have a different percentage of scores near the mean than will a normal distribution.



- Specifically, a leptokurtic distribution will probably have a greater per