Chapter 3: Measures of Variability

I. Measures of central tendency vs. Measures of variability

   a. Measures of central tendency (e.g., mean, median, mode) provide useful, but limited information. Information is insufficient in regards to the dispersion (i.e., variability) of scores of a distribution.

   b. Three measures of variability that researchers typically examine: range, variance, and standard deviation. Standard deviation is the most informative and widely used of the three.

II. Range

   a. Definition: The range is the difference between the largest (maximum value) score and the smallest score (minimum value) of a distribution

   b. Gives researchers a sense of how spread out the scores of a distribution, but it is not practical and misleading at times.

   c. When it may be used: Researchers may want to know whether all of the response categories on a survey question have been used and/or to have a sense of the overall balance in the distribution.

   d. Interquartile Range (IQR)
      a. Definition: The difference between the 75th percentile (third quartile) and 25th percentile (first quartile) scores in a distribution
      b. When the scores in a distribution are arranged in order, from smallest to largest (or vice-versa), the IQR contains scores in the two middle quartiles.

III. Variance

   a. Definition: The sum of the squared deviations (between the individual scores and the mean of a distribution) divided by the number of cases in the population, or by the number of cases minus one in the sample

   b. Provides a squared statistical average of the amount of dispersion in a distribution of scores. Rarely is variance looked at by itself because it does not use the same scale as the original measure of a variable, because it is squared.
      a. Why have variance? Why not go straight to standard deviation?
         1. We need to calculate the variance before finding the standard deviation. That is because we need to *square* the deviation scores so they will not sum to zero. These squared deviations produce the variance. Then we need to take the *square root* to find the standard deviation.

   c. The fundamental piece of the variance formula, which is the sum of the squared deviations, is used in a number of other statistics, most notably analysis of variance (ANOVA)

IV. Standard Deviation

a. Definition: The average deviation between the individual scores in the distribution and the mean for the distribution.
   1. Note that this is not technically correct, particularly for sample data where the sum of squared deviations is divided by $n - 1$, not N. But the sample estimate of the standard deviation is an estimate of the average deviation (absolute value) between the mean of a distribution and the scores in the distribution.

b. Useful statistic; provides a handy measure of how spread out the scores are in the distribution.

c. When combined, the mean and standard deviation provide a pretty good picture of what the distribution of the scores is like.

V. Sample statistics as estimates of population parameters

a. Researchers are generally concerned with what a sample tells them about the population from which the sample was drawn. Statistics generated from sample data are used to make inferences about the population.

b. The formulas for calculating the variance and standard deviation of sample data are actually designed to make sample statistics better *estimates* of the population parameters (i.e., the population variance and standard deviation)

VI. Formulas for calculating the variance

Sample variance	Sample standard deviation	Population standard deviation
$s^2 = \frac{\Sigma(X - \overline{X})^2}{n-1}$	$s = \sqrt{\frac{\Sigma(X - \overline{X})^2}{n-1}}$	$\sigma = \sqrt{\frac{\Sigma(X - \mu)^2}{N}}$

a. Formulas for calculating the variance and the standard deviation are virtually identical. Square root in standard deviation formula is only difference.

b. Calculating the variance is the same for both sample and population data except the denominator for the sample formula, which is $n - 1$

c. These formulas for calculating the variance are known as deviation score formulas. There are other formulas (e.g., raw score formula).

VII. Differences between the sample and population formulas: Why $n - 1$?

a. If population mean is unknown, use the sample mean as an estimate. But sample mean probably will differ from the population mean

b. Whenever using a number *other than* the sample mean to calculate the variance, a *larger* variance will be found. This will be true regardless of whether the number used in the formula is smaller or larger than the actual mean

c. Because the sample mean usually differs from the population mean, the variance and standard deviation will probably be smaller when calculated using the sample mean than it would have been if we had used the population mean

d. When using the sample mean to generate an *estimate* of the population variance or standard deviation, it will actually *under*estimate the size of the population mean

e. To adjust underestimation:
   a. use $n - 1$ in the denominator in sample formulas

f. Smaller denominators produce larger overall variance and standard deviation statistics, making it a more accurate estimate of the population parameters

VIII. Sample Statistics vs. Population Parameters

a. Researchers usually working with a sample that represents a larger population.

b. Sample statistics are generally considered estimates of population parameters, as is the case of the standard deviation.

c. How much of a difference between using N and $n - 1$ in the denominator depends on size of sample
   a. If sample is large, virtually no difference
   b. If sample is small, relatively large difference between the results produced by the population and sample formulas

IX. Summary

a. Measures of variability provide information about how spread out the scores of a distribution are (i.e., the dispersion).

b. The range, variance, and standard deviation are three common measures of variability.
   a. Of these, standard deviation is the most widely used.

c. Taken together, measures of central tendency and variability can provide useful information quickly about the characteristics of a distribution.