Measures of Variability

James H. Steiger

Overview

- Discuss Common Measures of Variability
  - Range
  - Semi-Interquartile Range
  - Variance
  - Standard Deviation
- Derive Computational Formulas for the Standard Deviation and Variance

Measures of Variability

- Several measures are in common use:
  - The Range
  - The Semi-Interquartile Range
  - The Sample Variance
  - The Sample Standard Deviation

The Range

- The *range* is simply the difference between the highest and lowest score in a set of data
- The *inclusive range* is the difference between the upper and lower real limits for the data

The Semi-Interquartile Range

- This is defined as half the difference between the 75th and 25th percentiles in the data, or

$$Q = \frac{Q_3 - Q_1}{2} = \frac{P_{75} - P_{25}}{2}$$

Variance

- The fundamental notion of variance in a statistical population is *the average squared deviation score.*
- However, when one samples from a population, the average squared deviation score in the sample tends (in the long run) to underestimate the variance in the population.

The Sample Variance

- The *sample variance* is defined as

$$S^2 = \frac{1}{N-1} \sum_{i=1}^{N} \left(X_i - \bar{X}_{\bullet} \right)^2$$

The Sample Standard Deviation

The sample standard deviation is simply the square root of the sample variance, i.e.,

$$S = \sqrt{S^2}$$

Computational Formulas

- Computing the sum of squared deviations by subtracting the mean from each value, then squaring, requires two passes through the numbers – one pass to compute the mean, a second pass to compute the deviation scores, square them, and sum.
- It is possible to compute the variance in one pass through the numbers by accumulating the sum of the raw scores *and the sum of squared raw scores.*

Computational Formulas

- The following formula can be used for data sets of moderate size.

$$S^2 = \frac{1}{N-1}\left(\sum_{i=1}^{N} X_i^2 - \frac{\left(\sum_{i=1}^{N} X_i\right)^2}{N}\right)$$

Calculating the Sample Variance

- Suppose your data are 1,2,3,4,5. You can first calculate the mean, and use the sum of squared deviations, *or* you can use the computational formula. Both methods are illustrated on the following slides.

Calculating the Sample Variance

X	dx	dx^2	X^2
5	2	4	25
4	1	1	16
3	0	0	9
2	–1	1	4
1	–2	4	1
15	0	10	55

Deviation Score Formula

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{N} dx_i^2 = \frac{1}{4}10 = 2.5$$

$$S = \sqrt{2.5} = 1.58$$

Raw Score Computational Formula

$$S^2 = \frac{1}{N-1}\left(\sum_{i=1}^{N} X_i^2 - \frac{\left[\sum_{i=1}^{N} X_i \right]^2}{N} \right)$$

$$= \frac{1}{4}\left(55 - \frac{15^2}{5} \right) = \frac{1}{4}(10) = 2.5$$

Variance of Combined Groups

- The variance of combined groups is a function of both the variability within the groups and the extent to which the groups themselves are separated on the number line. For example, consider the example on the next slide.

Variance of Combined Groups

- Suppose the groups have small variability within group, but the two groups are spread apart, like this

  X X X   Y Y Y

- When you combine the data, you get a much larger variance than in either of the original groups.

Variance of Combined Groups

- If the groups have small variability within group, but the two groups are closer, the combined group would have lower variance, as below

X X X Y Y Y

Variance of Combined Groups

The formula for combined variance of two groups may be written

$$S^2_{combined} = \frac{\sum_{j=1}^{J}\left(N_j - 1\right)S_j^2 + \sum_{j=1}^{J} N_j\left(\bar{X}_{\bullet j} - \bar{X}_{\bullet\bullet}\right)^2}{\left(\sum_{j=1}^{J} N_j\right) - 1}$$

Order Statistic Phenomena, and the Expected Value of the Range

- Glass and Hopkins devote a section of Chapter 5, pages 74-75, to a discussion of the “Expected Value of the Range.” The discussion culminates in Table 5.1 on page 75. As you can see, much of the information in the table is redundant.

Order Statistic Phenomena, and the Expected Value of the Range

- Table 5.1 deals with a phenomenon that can be crucial to understanding certain experimental results. Then phenomenon is this: The larger a sample, the more extreme, all other things being equal, the outstanding cases in the sample will be.

Order Statistic Phenomena, and the Expected Value of the Range

- So, for example, the tallest boy in a class of 1000 will, on average, be much taller than the tallest boy in a class of 50.
- Table 5.1 is really out of place in the textbook. Ideally, it would appear after chapter 6.

Order Statistic Phenomena, and the Expected Value of the Range

- The expected value of the range is the difference between the expected value of the highest observation and the expected value of the lowest observation in samples of a given size.
- Large groups will have more extreme values at both ends of the continuum, and this is reflected in a larger range.