VARIABILITY

CHAPTER 4

OVERVIEW

- VARIABILITY provides a quantitative measure of the degree to which scores in a distribution are spread out or clustered together.
- A good measure of variability serves two purposes:
  - It tells how much distance to expect between one score and another or how much distance to expect between an individual score and the mean.
  - It measures how well an individual score represents the entire distribution. Moreover, it provides information about how much error to expect if you are using a sample to represent a population.
- There are three different measures of variability: range, the interquartile range and standard deviation

THE RANGE

- The range refers to the distance between the largest score (Xmax) and the smallest score in the distribution (Xmin).
- **IMPORTANT!!!** You should take into account the real limits of the maximum and minimum X values
- It is typically used with interval and ratio scale measurements of a continuous variable.
- However, the range can be used with discrete variables that are measured on interval and ratio scale

THE RANGE

- The range is considered to be an **unreliable** measure of variability. Because it is **completely determined by the two extreme values** and **ignores the other scores** in the distribution.

THE INTERQUARTILE RANGE

- It measures the range covered by the middle 50% of the distribution.
- To find the interquartile range

1) Locate the boundary that seperates the lowest 25% from the rest of the distribution. It is called first quartile and identified with Q1
2) Locate the boundary that seperates the top 25% from the rest of the distribution. It is called third quartile and identified with Q3
3) **Interquartile range is then defined as the distance between Q1 and Q3**

THE INTERQUARTILE RANGE

- When the interquartile range is used to measure variability, it is commonly transformed into the semi-interquartile range
- Semi-interquartile range measures the distance from the middle of the distribution to the boundaries that define the middle 50%

- **Semi-interquartile range= Q3-Q1/2**

THE INTERQUARTILE RANGE

- The semi-interquartile range gives a better and stable variability because it is derived from the middle 50% of the distribution and it is less likely to be influenced by extreme scores.

- However, it does not provide a complete picture of the variability because it only considers the middle 50% but ignores the other 50%

STANDARD DEVIATION AND VARIANCE FOR A POPULATION

- Standard deviation is the most common and important measure of variability.

- It uses the mean of a distribution as a reference point and **measures variability by considering the distance between each score and mean**

STANDARD DEVIATION AND VARIANCE FOR A POPULATION

- In order to calculate the standard deviation
- STEP 1: We determine the deviation which refers to the distance from the mean for each individual score
  - There are two parts to a deviation score:
    - The sign (+ or -) that tells the direction from the mean
    - The number that gives the actual distance from the mean

- Deviation=$X - \mu$

STANDARD DEVIATION AND VARIANCE FOR A POPULATION

- STEP 2: To calculate the mean of deviation scores
- NOTE that **the sum of the deviation scores is always zero.** So the **mean of deviation scores is always zero**. Therefore mean of deviation scores is no value as a measure of variability.

STANDARD DEVIATION AND VARIANCE FOR A POPULATION

- STEP 3: In order to solve the problem in step 2, we square the each deviation score. Using the squared values, we compute the mean squared deviation which is called **variance**
- However, the concept of squared distance (variance) is not easy to understand. Therefore, we make a correction.
- STEP 4: We take **the square root of the variance which is called standard deviation**.

STANDARD DEVIATION AND VARIANCE FOR A POPULATION

- Standard deviation and variance are used only with numerical scores that are obtained from interval or ratio scales.

STANDARD DEVIATION AND VARIANCE FOR A SAMPLE

- Samples consistently tend to be less variable than their populations.
- A few extreme scores in the population tend to make population variability relatively large.
- However, these extreme values are unlikely to be obtained when we are selecting a sample.
- Therefore, sample variability is relatively small.
- Sample variability gives a biased estimate of population variability. It underestimates the population parameter.
- In order to obtain an accurate and unbiased sample variance, we make a correction in the calculation of sample variance.

TRANSFORMATION OF SCALE

- Adding a constant to every score does not affect any distances in the distribution and therefore does not change the standard deviation
- Multiplying each score by a constant causes the standard deviation to be multiplied by the same constant.

VARIANCE AND INFERENTIAL STATISTICS

- Variability plays an important role in inferential statistics, the variability in the data influences how easy it is to see patterns.
  - Low variability means the existing patterns can be seen easily
  - High variability tends to obscure any patterns that might exist.
- **Error Variance** refers to the variance that exists in a set of sample. Error variance occurs because there is unexplained and uncontrolled differences between scores.
- When the error variance increases, it is difficult to see any systematic differences between sample mean difference

COMPARING MEASURES OF VARIABILITY

- Extreme Scores:
  - Range is most affected.
  - Standard Deviation and variance are also influenced by extreme scores
  - Semi-interquartile range is less affected
- Sample size
  - When the sample size increases, the range is unaffected
  - Standard Deviation, variance and semi-interquartile range are relatively affected.
- Open-Ended Distributions
  - You can't compute the range, SD and Variance
  - Only measure is semi-interquartile range