Measures of Variability

Why use Measures of Variability?

- Although measures of central tendency can be helpful, they only tell part of the story.
- When we use measures of central tendency alone they may be misleading rather than informing.

Measures of Variability

- The measure of variability is a number that describes diversity or variability in the distribution.
- Basically it shows us how much variation and diversity there is within the answers we received.

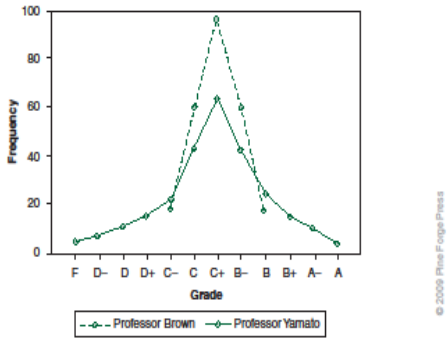
Measures of Variability

- There are five measures of variability that we will discuss in this chapter:
  - The index of qualitative variation
  - The range
  - The interquartile rage
  - The standard deviation
  - The variance

Measures of Variability

- We look at variation and diversity so we can think about the differences between the respondents not just the similarities.
- When we just look at the similarities and not the differences we are not looking at the whole picture.

Figure 5.1 Distribution of Grades for Professors Brown's and Yamato's Statistics Classes



The Index of Qualitative Variation (IQV)

- The index of qualitative variation is a measure of variability for nominal variables like race or ethnicity.
- The index is a number that can range from 0.00 to 1.00.

The Index of Qualitative Variation (IQV)

- When all the cases in the distribution are in one category there is no variation and the IQV is 0.00.
- When all the cases in the distribution are distributed evenly across the categories there is maximum variation and the IQV is 1.00

Two Ways to Calculate IQV

- $IQV = \frac{\text{Total Observed Differences}}{\text{Total Possible Differences}}$
- $IQV = \frac{K(100^2 - \Sigma\%^2)}{100^2\ (K\text{-}1)}$
- $IQV = \frac{K(n^2 - \Sigma f^2)}{n^2\ (K\text{-}1)}$

Index of Qualitative Variation: First Formula

- **IQV** – A measure of variability for **nominal variables**. It is based on the **ratio** of the total number of **differences** in the distribution to the maximum number of **possible differences** within the same distribution.

$$IQV = \frac{K(100^2 - \Sigma\%^2)}{100^2 (K-1)}$$

Where

K= the number of categories

N= *100*= the total number of cases in the distribution

$\Sigma\%^2$ = the sum of all squared percentages

Percentage of Ethnicity

Ethnicity	Percentage
White	90%
African American	6%
Hispanic	4%
Total	100%

Expressing the IQV as a Percentage

- You simply multiply the IQV by 100 to express it as a percentage.
- When you express it as a percentage you are saying that the number of differences are a certain percent.
- What would be the percentage for the last example .28?

Index of Qualitative Variation: Second Formula

- **IQV** – A measure of variability for **nominal variables**. It is based on the **ratio** of the total number of **differences** in the distribution to the maximum number of **possible differences** within the same distribution.

$$IQV = \frac{K(n^2 - \Sigma f^2)}{n^2 (K-1)}$$

Where

K= the number of categories
N= the total number of cases in the distribution
Σf^2= the sum of all squared frequencies

Find the Total Observed Differences

Race	Frequency
White	480
African American	31
Hispanic	23
Total (N)	534

The Range

- The range is the simplest and most straightforward measure of variation.
- The range measures variables at the Ordinal level.
- It is the difference between the highest and lowest scores in the distribution.

The Range

- Range = highest score–lowest score
- In a distribution of age the oldest person included in the study was 89 years old and the youngest person in the study was 10 years old, what is the range?

The Range

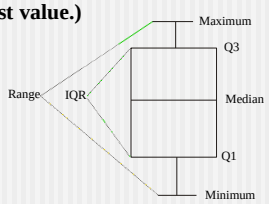
- Although the range is simple and quick to calculate, it is a crude measure because it is based on only the lowest and highest scores.
- These could be extreme and rather atypical, which might make the range a misleading indicator of the variation in a distribution.

The Interquartile Range

- To remedy this limitation we can use an alternative to the range – the interquartile range.
- The IQR is a measure of variation for ordinal variables.
- The IQR uses the middle scores of the distribution, the scores at the 50% mark.

The Box Plot

- **The Box Plot is a graphic device that visually presents the following elements: the range, the inter-quartile range, the median, the quartiles, the minimum (lowest value,) and the maximum (highest value.)**



The Interquartile Range

- The IQR = Q3 – Q1
- The first quartile (Q1) is the 25th percentile, the point where 25% of the cases fall below it and 75% of the cases above it.
- The third quartile (Q3) is the 75th percentile, the point where 75% of the cases fall below it and 25% of the cases above it.

The Interquartile Range

- Like the range, the IQR uses only two scores. Because the two scores are the middle scores and not the extreme scores we do not face the problem of misrepresentation of the distribution.

Calculating the IQR

- First we need to find Q1 and Q3. To do so order the scores on the distribution from highest to lowest or lowest to highest.
- Next, we need to find the first quartile, (Q1). To do this we multiply N by .25

Find the Q1

State	%	State	%	State	%
Nevada	92.5	Wyoming	26.3	Minnesota	13.8
Alaska	91.3	Georgia	25.9	Kentucky	13.5
Hawaii	62.5	Montana	25.6	Illinois	13.3
Arizona	54.9	Texas	24.6	Vermont	13.3
New Mexico	39.7	Connecticut	21.6	North Dakota	13.0
Florida	39.6	Michigan	21.0	West Virginia	12.6
South Carolina	37.1	New Hampshire	20.9	Oklahoma	12.6
Utah	36.9	Ohio	20.0	Massachusetts	12.2
Delaware	35.7	New Jersey	19.2	South Dakota	12.2
California	28.6	Louisiana	15.4	Nebraska	8.3
Oregon	28.5	Wisconsin	15.2	Washington, DC	3.8

The Interquartile Range

- Next, you need to find the Q3.
- To do this you simply multiply N this time by .75

Find the Q3

State	%	State	%	State	%
Nevada	92.5	Wyoming	26.3	Minnesota	13.8
Alaska	91.3	Georgia	25.9	Kentucky	13.5
Hawaii	62.5	Montana	25.6	Illinois	13.3
Arizona	54.9	Texas	24.6	Vermont	13.3
New Mexico	39.7	Connecticut	21.6	North Dakota	13.0
Florida	39.6	Michigan	21.0	West Virginia	12.6
South Carolina	37.1	New Hampshire	20.9	Oklahoma	12.6
Utah	36.9	Ohio	20.0	Massachusetts	12.2
Delaware	35.7	New Jersey	19.2	South Dakota	12.2
California	28.6	Louisiana	15.4	Nebraska	8.3
Oregon	28.5	Wisconsin	15.2	Washington, DC	3.8

Find the IQR

- Now that we know the Q1 and the Q3
- The IQR = Q3 – Q1 =
- What is the IQR?
- What is the range?

The Box Plot

- The box plot visually represents the range and the interquartile range.
- The box plot gives us a way to visually examine the center, the variation, and the shape of distributions of interval-ratio variables.

The Variance and the Standard Deviation

- The variance and standard deviation are two closely related measures of variation.
- The variance is the average of the squared deviations from the mean of the distribution.
- The standard deviation is the square root of the variance
- Both measure variability in interval-ratio variables.

Variance

- **Variance** – A measure of variation for interval-ratio variables; it is the average of the squared deviations from the mean

$$s_Y^2 = \frac{\sum(Y - \overline{Y})^2}{N - 1}$$

Standard Deviation

- **Standard Deviation** – A measure of variation for interval-ratio variables; it is equal to the square root of the variance.

$$s = \sqrt{s_Y^2} = \sqrt{\frac{\sum(Y - \overline{Y})^2}{N - 1}}$$

The Variance

- The variance is a measure of how spread out a distribution is.
- S_y^2 = Variance

$$s_y^2 = \frac{\sum(Y-\overline{Y})^2}{N-1}$$

Variance

- first calculate the mean.
- Subtract the mean from each score to find the deviation.
- Square each deviation
- Sum the squared deviations
- Divide the sum by N-1
- The answer is the variance

Find the Variance

City	Murder rate per 100,00
New York	16.1
Los Angles	24.5
Honolulu	4.3
San Diego	7.9

Standard Deviation

- The standard deviation is the square root of the variance.
- The standard deviation is easier to interpret because the variance is expressed in squared percentages.
- What is the standard deviation for the murder rate?

Considerations for Choosing a Measure of Variability

- **For nominal variables, you can *only* use IQV (index of qualitative variation).**
- **For ordinal variables, you can calculate the IQV or the IQR (inter-quartile range.) Though, the IQR provides more information about the variable.**
- **For interval-ratio variables, you can use IQV, IQR, or variance/standard deviation. The standard deviation (also variance) provides the most information, since it uses all of the values in the distribution in its calculation.**