Measures of Variability

Ray Block, Jr.
PS 585 (Research Methods)
Fall 2003

Today's Blueprint

Last Class

- Univariate Data Analysis (Part 1)
  - Statistical Models
  - Measures of Central Tendency

Today's Class

- Univariate Data Analysis (Part 2)
  - Statistical Models (A Recap)
  - Measures of Variability

Statistical Models (A Recap)

Recap

What Models Are:

- Symbolic representations of social phenomena
  - Statistical models use mathematical/statistical symbols

What Purpose Do [Statistical] Models Serve:

- Discuss significant relationships among *concepts*
- Enable researchers to form testable propositions between *variables*
- Summarize *data*

The Goal of Statistical Modeling:

- To build a model that best represents the real-world phenomena of interest
  - The degree to which a statistical model represents the data collected is known as the *fit* of the model to the data

How Do You Build Statistical Models

- Observe some facts about the world
- Speculate about the process(es) that produced those facts
- Collect data that represent the process(es)
- Reduce the process(es) to a statistical model using the data you collected

Statistical Models Fall Into 2 Categories

- Models measuring Central Tendencies
- Models measuring Variability

Recap

Central Tendencies (The 4 "Ms")

1) The Midpoint/Midrange

- *Description:* Picking the middle slice of bread
- *Level of measurement:* Ordinal, Interval, Ratio
- *Shape of Distribution:* N/A
- *Research Objective:* Crude measure of central tendency
- *Note:* Seldom used in social science

2) The Mode (Mo)

- *Description:* Maximum Frequency
- *Level of Measurement:* Nominal

- *Shape of Distribution:* Most appropriate for bimodal or multimodal
- *Research Objective:* Fast, simple, but rough measure of central tendency

3) The Median (Mdn)
- *Description:* Middlemost Value
- *Level of Measurement:* Ordinal, Interval, or Ratio
- *Shape of Distribution:* Most appropriate for highly skewed
- *Research Objective:* Precise measure of central tendency
- *Note:* Sometimes used to split distributions into categories (i.e. high vs. low)

4) The [Arithmetic] Mean (X-Bar)
- *Description:* Center of Gravity
- *Level of Measurement:* Interval or Ratio
- *Shape of Distribution:* Most appropriate for unimodal symmetrical
- *Research Objective:* Precise measure of central tendency
- *Note:* Most commonly-used central measure. Used for hypothesis tests and other statistical operations

Recap

Finding the mode, median, and mean:
- Arrange scores from highest to lowest
- The mode is the most frequent score
- The Median is the middlemost value in the ordered list of scores
  - If there is an odd number of scores, then median is in the exact middle of the list
  - If there is an even number of scores, then the median is halfway between the two middlemost scores
- Determine the sum of the scores
- Calculate the mean by dividing the sum by the number of scores

Measures of Variability

AKA: Measures of "Spread" "Width" or "Dispersion"

Measures of Variability

- In data analysis, the purpose of calculating measures of dispersion is to discover the extent to which scores differ, cluster, or scatter around a measure of central tendency

Some Measures of Spread:
- The Range
- The Mean Deviation
- The Variance
- The Standard Deviation
- Standard Error

Measures of Variability

- The Range is the difference between the highest and the lowest score: R = H – L
  - Where:
    - R = Range
    - H = Highest score in a distribution
    - L = Lowest score in a distribution
- Advantages:
  - Quick and easy to calculate
- Disadvantage:
  - Crude measure of variability
  - Why? Because it depends only on lowest and highest values in distribution

Measures of Variability

- Deviation = The distance between any given raw score and its mean (X_i – X-Bar)
- Mean Deviation = The average distance between the raw scores and the mean

$$MD = \frac{\Sigma\left|X_i - \overline{X}\right|}{N}$$

Where:

- MD = Mean Deviation
- $\Sigma|X_i$ - X-Bar| = Sum of absolute deviations (disregarding plus or minus signs)
- N = Total number of scores

Step-by-Step Illustration:

- Take the following list of numbers (arranged from highest to lowest):

9
8
6
4
2
1

- Step 1: Find the mean of the distribution

$$\overline{X} = \frac{\sum X_i}{N} = \frac{30}{6} = 5$$

- Step 2: Subtract the mean from each raw score
  - Take the absolute values (ignore the signs)
  - Add up these absolute deviations

$$\sum\left|X_i - \overline{X}\right| = 16$$

- Step 3: To get MD, Divide $\Sigma|X$ - X-bar| by N to adjust for the cases involved

$$MD = \frac{\sum\left|X_i - \overline{X}\right|}{N} = \frac{16}{6} \approx 2.67$$

- Note: Mean deviations are no longer widely used in social sciences. However, calculating MD is not a complete waste of time
- …Here's why…
  - Recall that we took the absolute values to avoid getting minus signs: $\Sigma|X_i$ - X-bar|
  - We use absolute values so that the different signs of values in $\Sigma(X$ - X-bar) do not cancel themselves out

$X_i - \overline{X}$
-4
+3
+1
-1
-3
-4
$\Sigma = 0$

- We can also get around this sign canceling issue by squaring $\Sigma(X_i$ - X-bar)

$(X_i - \overline{X})^2$
16
9
1
1
9
16
$\Sigma = 52$

- Therefore, the variance = The mean of the squared deviations

$$S^2 = \frac{\sum (X_i - \overline{X})^2}{N}$$

- Where:
  - S^2 = Variance
  - $\Sigma(X$ - X-bar$)^2$ = Sum of squared deviations from mean
  - N = Total number of observations

- Variance = The average difference between the mean and the observations made
- Caveat:
  - Squaring the deviations alters the units of measurement
  - We need to bring the units back to their original non-squared values
  - The simplest way to do this is to take the square root everything

$$S = \sqrt{\frac{\sum (X_i - \overline{X})^2}{N}}$$

- Standard deviation = square root of the variance
  - Squared values are not standard (doesn't make sense to talk in terms of things squared)
  - Standard deviations restate variance in standard units

$$s = \sqrt{\frac{\sum\left(X_i - \overline{X}\right)^2}{N}} = \sqrt{\frac{52}{6}} = \sqrt{8.67} \approx 2.94$$

References

- FYI:
  - Levin, Jack and James Alan Fox. 2003. Elementary Statistics in Social Research, 9th Edition. Boston, MA: Pearson Education Group, Inc.
  - Salkind, Neil. 2003. Exploring Research, 5th Edition. Upper Saddle River, NJ: Prentice Hall.