

Chapter 4: What is a measure of Central Tendency?

Numbers that describe what is typical of the distribution

- You can think of this value as where the middle of a distribution lies (the median).
- or
- The value within a distribution of values that has the most cases (mode)
- or
- The mathematical average (mean)

Chapter 4 - 1

Measure of Central Tendency:

The Mode

- The category with the largest frequency (or percentage) in the distribution.

Chapter 4 - 2

The Mode: An Example

- Which of the three candidates represents the "mode" for these candidates?
- Variable=Candidates
Candidate A - 11,769 votes
Candidate B - 39,443 votes
Candidate C - 78,331 votes

Level of measurement = _____

The Mode= _____

Chapter 4 - 3

The Mode: An Example

- Which of the three candidates represents the "mode" for these candidates
- Variable=Candidates
Candidate A - 11,769 votes
Candidate B - 39,443 votes
Candidate C - 78,331 votes

Level of measurement = nominal (why?)

The Mode= Candidate C (why?)

Chapter 4 - 4

The mode can be calculated for variables within all levels of measurement that are: nominal, ordinal, or interval-ratio.

Chapter 4 - 5

Measure of Central Tendency: The Median

- The score that divides the distribution into two equal parts, so that half the units (cases) are above it and half below it.
- The median is the middle score in a distribution.
- The median is appropriate for ordinal or interval-ratio data.

Chapter 4 - 6

Finding the Median for an Ordinal Variable

Job Satisfaction (I am very satisfied with my job)

| Values | Freq | Cummulative Frequency |
|----------------|------|-----------------------|
| Agree Strongly | 5 | 5 |
| Agree | 10 | 15 |
| Undecided | 3 | 18 |
| Disagree | 7 | 25 |
| Dis. Strongly | 3 | 28 |

Total Cases: 28

Steps to Determine Median for Ordinal Var

1. divide total # of cases by 2: $28/2 = 14$
2. determine/calculate the cumulative frequencies
3. locate the value (category) that holds the middle case (unit): "agree" contains the 14th case

Chapter 4 - 7

Finding the Median for an Interval/Ratio Variable

What is the interval/ratio variable below?

What is the median # of hate crimes?

What is the "unit of analysis"?

Number of Hate Crimes in State

NC = 39
Penn = 141
TX = 287
Ohio = 255
Fla = 240

Chapter 4 - 8

Finding Median for Interval/Ratio Variable

of hate crimes by state

Cases

NC = 39
Penn = 141
TX = 287
Ohio = 255
Fla = 240

States ordered low to high

NC = 39
Penn = 141
Fla = 240
Ohio = 255
TX = 287

of cases (or units) = 5

Steps to Determine:

1. Order the variable (hate crimes) from highest to lowest or vice versa
2. Add 1 to the total # of units (states) if there is an odd # of units (e.g. $1+5=6$)
3. divide resulting number by 2 ($6/2 = 3$)
4. Count down that many units (cases) to identify the middle or median (Fla)

Chapter 4 - 9

Percentile

- A score at or below which a specific percentage of the distribution falls.
- For example, the 75th percentile is a score for which 75% of the cases are at or below it.

Chapter 4 - 10

Percentile

Table 1: Satisfaction with Health

| | Freq | Cum Freq | % | Cum % |
|-----------|------|----------|----|-------|
| Very Low | 5 | 5 | 18 | 18 |
| Low | 7 | 12 | 25 | 43 |
| Moderate | 6 | 18 | 21 | 64 |
| High | 7 | 25 | 25 | 89 |
| Very High | 3 | 28 | 11 | 100 |

Total N: 28

Steps to Determine Percentile: determine cumulative percentages and then locate the percentile of interest. The 75th percentile would be which category: _____

Chapter 4 - 11

The Mean

The arithmetic average obtained by adding up all the scores and dividing by the total number of scores.

The mean is used with interval-ratio data.

Can be used with ordinal data but is not very accurate/precise.

Chapter 4 - 12

Formula for the Mean

$$\bar{Y} = \frac{\sum Y}{N}$$

"Y bar" (\bar{Y}) equals the average or the sum of all the scores, Y, divided by the number of scores, N

(for example add up the # of hate crimes for the states and then divide by "N" or the number of states).

Chapter 4 - 13

Calculating the mean with frequency distributions (ordinal variable):

Steps to Determine:

Satisfaction with Health

| | Freq | Category x Freq |
|---------------|------|--------------------|
| 1 - Very High | 5 | 5 |
| 2 - High | 7 | 14 |
| 3 - Moderate | 6 | 18 |
| 4 - Low | 7 | 28 |
| 5 - Very Low | 3 | 15 |

| | | |
|----------|----|----|
| Total N: | 28 | 80 |
|----------|----|----|

$$\bar{Y} = \frac{\sum fY}{N}$$

- multiply each category by its frequency (category x frequency)
- sum all the "category x freq" scores to determine total (80)
- divide total (80) by total number of cases (total N or 28) to get average score (2.86)

Chapter 4 - 14

In-Class Exercise:

Calculate the mode, median, and mean for the grouped frequency below.

Satisfaction with Parking

| Level of Satisfaction | Frequency |
|-------------------------|-----------|
| 1 Very Satisfied | 190 |
| 2 Satisfied | 316 |
| 3 Somewhat Satisfied | 54 |
| 4 Somewhat Dissatisfied | 17 |
| 5 Dissatisfied | 2 |
| 6 Very dissatisfied | 2 |
| TOTAL | 581 |

Chapter 4 - 15

Ordinal (Grouped) Data: Mode

Category with the most cases or "Satisfied (#2)"

Satisfaction with Parking

| Level of Satisfaction | Frequency |
|-------------------------|-----------|
| 1 Very Satisfied | 190 |
| 2 Satisfied | 316 |
| 3 Somewhat Satisfied | 54 |
| 4 Somewhat Dissatisfied | 17 |
| 5 Dissatisfied | 2 |
| 6 Very dissatisfied | 2 |
| TOTAL | 581 |

Chapter 4 - 16

Ordinal (Grouped) Data: Median

- Make sure values are ordered
- Add one to total frequency (if an odd #): $581 + 1 = 582$
- Divide by 2: $582/2 = 291$
- Calculate cumulative frequency and determine which category contains the 291st person (answer is "Satisfied" or #2)

| Level of Satisfaction | Frequency | Cumulative Freq |
|-----------------------|-----------|-----------------|
| 1 V. Satisfied | 190 | 190 |
| 2 Satisfied | 316 | 506 |
| 3 Somewhat Sat. | 54 | 560 |
| 4 Somewhat Dis. | 17 | 577 |
| 5 Dissatisfied | 2 | 579 |
| 6 V. Dissatisfied | 2 | 581 |
| TOTAL | 581 | |

Chapter 4 - 17

Ordinal (Grouped) Data: Mean

- Multiply frequency (# of people) times category
- Sum the scores obtained: 1,074
- Divide by total frequency 1074/581 to obtain mean category (mean=1.85 people per household)

| Level of Satisfaction | Frequency | Category x Frequency |
|-------------------------|-----------|----------------------|
| 1 Very Satisfied | 190 | 190 |
| 2 Satisfied | 316 | 632 |
| 3 Somewhat Satisfied | 54 | 162 |
| 4 Somewhat Dissatisfied | 17 | 68 |
| 5 Dissatisfied | 2 | 10 |
| 6 Very Dissatisfied | 2 | 12 |
| TOTAL | 581 | 1,074 |

Chapter 4 - 18

Considerations for Choosing a Measure of Central Tendency

- For a nominal variable, the mode is the only measure that can be used.
- For ordinal variables, the mode and the median may be used. The median provides more information (taking into account the ranking of categories). Can also use interval/ratio but not precise.
- For interval-ratio variables, the mode, median, and mean may all be calculated. The mean provides the most information about the distribution, but the median is preferred if the distribution is skewed.

Chapter 4 – 19

When choosing the appropriate measure of central tendency for a distribution, what should you consider?

the level of measurement of the variables

(e.g., mode for nominal level)

Chapter 4 – 20

What is usually the appropriate measure of central tendency for interval-ratio level?

the mean

Chapter 4 – 21

What is the primary "weakness" of the mean?

the mean is highly influenced by extreme scores in one direction

(e.g., the mean may not "represent" the true distribution of the cases very well)

Chapter 4 – 22

Example of mean "unrepresentative" of sample

| Sample 1: Score for <u>Five Women</u> | Sample 2: Score for <u>Five Women</u> |
|---|---|
| 100 | 100 |
| 110 | 110 |
| 125 | 125 |
| 125 | 125 |
| 135 | 450 |

What is the mode: _____

What is the median: _____

What is the mean: _____

Chapter 4 – 23

Example of mean "unrepresentative" of sample

| Sample 1: Score for <u>Five Women</u> | Sample 2: Score for <u>Five Women</u> |
|---|---|
| 100 | 100 |
| 110 | 110 |
| 125 | 125 |
| 125 | 125 |
| 135 | 450 |

What is the mode: 125 and 125

What is the median: _____

What is the mean: _____

Chapter 4 – 24

Example of mean "unrepresentative" of sample

| Sample 1: Score for <u>Five Women</u> | Sample 2: Score for <u>Five Women</u> |
|---|---|
| 100 | 100 |
| 110 | 110 |
| 125 | 125 |
| 125 | 125 |
| 135 | 450 |

What is the mode: 125 and 125
 What is the median: 125 and 125
 What is the mean: _____

Example of mean "unrepresentative" of sample

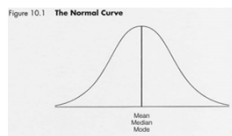
| Sample 1: Score for <u>Five Women</u> | Sample 2: Score for <u>Five Women</u> |
|---|---|
| 100 | 100 |
| 110 | 110 |
| 125 | 125 |
| 125 | 125 |
| 135 | 450 |

What is the mode: 125 and 125
 What is the median: 125 and 125
 What is the mean: 119 and 182

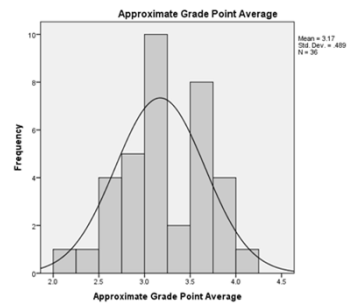
Normal Distributions (also called normal curve)

• Normal Distribution

- Used with **linear** variables
- A bell-shaped and symmetrical theoretical distribution (a theoretical distribution of cases is not an actual distribution of cases),
- with the mean, the median, and the mode all coinciding at its peak and
- with frequencies gradually decreasing at both ends of the curve.



Normal Distributions



Normal Distributions

• Normal Distribution

- What happens when we have a few cases that are far above or below the other cases?
- **Negatively Skewed**: a few extremely low values
- **Positively Skewed**: a few extremely high values

