

POLS 5377 Scope & Method of Political Science

Week 5 Descriptive Statistics

Measures of Central Tendency

Healey. (2016) *Statistics: A Tool for Social Research*, Chapter 3

Key Question:

- * What is and why central tendency?
- * How to calculate and interpret the three measures of central tendency: the mode, median, and mean?
- * How to choose appropriate measures of central tendency according to level of measurement?
- * What is skewness and how is it identified?

Outline

- * Central tendency – What & Why
- * Mode
- * Median
- * Percentiles, Deciles, and Quartiles
- * Mean
- * Three characteristics of the mean
- * Choosing a measure of central tendency

Central Tendency

- * Summarize information about the most typical, central, or average score of a variable.
- * Univariate descriptive statistics
- * Three measures
 - * Mode: the most common score
 - * Median: the score of the middle case
 - * Mean: the average score
- * The three measures provide different information and need to take the level of measurement into consideration

Case	Days
1	32
2	41
3	70
4	50
5	63
6	50
7	50
8	124
9	50
10	63

The Mode

- * The mode common score
- * Can be used with variables at all levels of measurement
- * Most often used with nominal level variables
- * Finding the mode
 - * Count the number of times each score occurred
 - * The score that occurs most often is the mode

2 2 3 5 7 7 7 8 9

The mode is 7.

The Mode

- * Limitation of the mode
 - * Some distributions have no mode

2 3 5 9 10 11 13 17 19 20
 - * Some distributions have multiple modes

2 2 5 9 10 11 13 **19 19** 20
 - * The mode of an ordinal or interval ratio level variable may not be central to the whole distribution

Test score	Frequency
95	5
68	3
65	4
62	2
60	4
	N= 18

Mode

The Median

- * Exact center of distribution of scores
- * The score of the middle case
- * Can be used with variables measured at the ordinal or interval-ratio levels
- * Cannot be used for nominal level variables
- * Steps:
 - * Array the cases from low to high (or from high to low)
 - * Locate the middle case
 - * If N is odd: the median is the score of the middle case (the middle case: $\frac{N+1}{2}$)
 - * If N is even: the median is the average of the scores of the two middle cases (first middle case: $\frac{N}{2}$; second middle case: $\frac{N}{2} + 1$)

The Median

- * Examples

0 10 10 12 13 15 18 19 39

The median is 7. $(9+1)/2=5^{\text{th}}$ case

0 8 10 10 12 13 15 18 19 39

The median is $(12+13)/2=12.5$ First: $10/2=5^{\text{th}}$ case
Second: $5+1=6^{\text{th}}$ case

- * The median can reduce the effect of outliers in the data

Other Measures of Position

- * Percentiles: point below which a specific percentage of cases fall
- * Deciles: divides distribution into tenths (10%, 20%, 30%...90%)
- * Quartiles: divides distribution into quarters (25%, 50%, 75%)
- * Example: the 6th decile (or 60th percentile) for the sample of 10 test grades is 85, which tells us that 60% of students scored below 85 on the test.

60 65 72 78 80 **85** 88 90 90 92

The 6th decile is 85

- * Median: the 50th percentile, or the 5th decile or the 2nd quartile

Other Measures of Position

- * How to calculate a percentile:
 - * Arrange scores in order from low to high
 - * Multiply the number of cases (N) by the proportional value of the percentile (for example: the 75th percentile would be 0.75)
 - * The resultant value marks the number of the case that falls at the percentile
- * Examples:
 - In a sample of 70 test grades we want to find the 4th decile (or 40th percentile), $70 \times 0.40 = 28$, so the 28th case is the 40th percentile
 - In a sample of 70 test grades we want to find the 3rd quartile (or 75th percentile), $70 \times 0.75 = 52.5$, rounding to 53 the 53rd case is the 75th percentile

The Mean

- * The average score
- * Requires variables measured at the interval-ratio level but is often used with ordinal-level variables
- * Cannot be used for nominal-level variables
- * The mean or arithmetic average, is by far the most commonly used measure of central tendency

The Mean

- * To calculate: add all of the scores and then divide by the number of scores (**N**)

The mathematical formula for the mean is

$$\bar{X} = \frac{\sum(X_i)}{N}$$

where \bar{X} = the mean

$\sum(X_i)$ = the summation of the scores

N = the number of cases

Scores: 1 3 4 5 5 6 8 9

$$\text{Mean } (\bar{X}) = \frac{1+3+4+5+5+6+8+9}{8} = \frac{41}{8} = 5.125$$

Characteristics of the Mean

- * The mean “balances” out all of the scores in a distribution; all scores “cancel out” around the mean

$$\sum(X_i - \bar{X}) = 0$$

- * The mean is the point of minimized variation of the scores, “least squares principle”

$$\sum(X_i - \bar{X})^2 = \text{minimum}$$

- * The mean is affected by all scores; all scores are used in the calculation of the mean

Characteristics of the Mean

- * First, the Mean “balances out” all scores

X_i	$X_i - \bar{X}$
65	$65 - 78 = -13$
73	$73 - 78 = -5$
77	$77 - 78 = -1$
85	$85 - 78 = 7$
90	$90 - 78 = 12$
$\Sigma(X) = 390$	$\Sigma(X_i - \bar{X}) = 0$
$\bar{X} = 390/5 = 78$	

Characteristics of the Mean

* Second, the Mean is the point of "Minimized" variation

1	2	3	4
X_i	$X_i - \bar{X}$	$(X_i - \bar{X})^2$	$(X_i - 77)^2$
65	$65 - 78 = -13$	$(-13)^2 = 169$	$65 - 77 = (-12)^2 = 144$
73	$73 - 78 = -5$	$(-5)^2 = 25$	$73 - 77 = (-4)^2 = 16$
77	$77 - 78 = -1$	$(-1)^2 = 1$	$77 - 77 = (0)^2 = 0$
85	$85 - 78 = 7$	$(7)^2 = 49$	$85 - 77 = (8)^2 = 64$
90	$90 - 78 = 12$	$(12)^2 = 144$	$90 - 77 = (13)^2 = 169$
$\Sigma(X_i) = 390$	$\Sigma(X_i - \bar{X}) = 0$	$\Sigma(X_i - \bar{X})^2 = 388$	$\Sigma(X_i - 77)^2 = 393$

Characteristics of the Mean

* Third, the Mean is affected by all scores

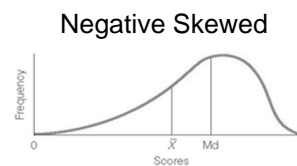
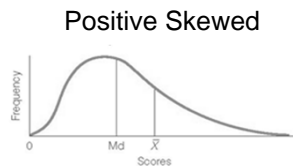
1	2	3	4	5	6
<i>Scores</i>	<i>Measures of Central Tendency</i>	<i>Scores</i>	<i>Measures of Central Tendency</i>	<i>Scores</i>	<i>Measures of Central Tendency</i>
15	Mean = 25	15	Mean = 718	0	Mean = 22
20	Median = 25	20	Median = 25	20	Median = 25
25		25		25	
30		30		30	
35		3500		35	

Characteristics of the Mean

- * Strength
 - * The mean uses all the available information from the variable
- * Weaknesses
 - * The mean is affected by every score
 - * If there are some very high or low scores (as with skewed distributions), the mean may be misleading
 - * Example: in 1984, the University of Virginia announced that its Department of Communication graduates had a mean starting salary of \$55,000. They failed to report that one of the graduates was NBA player, Ralph Sampson.

Means, Medians, and Skew

- * When a distribution has a few very high or low scores, the mean will be pulled in the direction of the extreme scores



- * When an interval-ratio level variable has a pronounced skew, the **median** may be the more trustworthy measure of central tendency

Choosing a Measure of Central Tendency

* Level of measurement

Measure of Central Tendency	Level of Measurement		
	<i>Nominal</i>	<i>Ordinal</i>	<i>Interval-Ratio</i>
Mode	YES	Yes	Yes
Median	No	YES	Yes
Mean	No	Yes (?)	YES

Choosing a Measure of Central Tendency

* Information you aim to report

Use the mode when:	<ol style="list-style-type: none"> 1 The variable is measured at the nominal level. 2 You want a quick and easy measure for ordinal and interval-ratio variables. 3 You want to report the most common score.
Use the median when:	<ol style="list-style-type: none"> 1 The variable is measured at the ordinal level. 2 An interval-ratio variable is badly skewed. 3 You want to report the central score. The median always lies at the exact center of the distribution.
Use the mean when:	<ol style="list-style-type: none"> 1 The variable is measured at the interval-ratio level (except when the variable is badly skewed). 2 You want to report the typical score. The mean is the "fulcrum that exactly balances all of the scores." 3 You anticipate additional statistical analysis.

Choosing a Measure of Central Tendency

- * Example: you conducted a survey on 50 people and ask them three questions:
 - * What state are you from? (nominal)
 - * What is your highest education level? Category: no high school, high school, college, graduate school (ordinal)
 - * How old are you? (interval)
- * Report:
 - * State: the largest number of people were from Texas (the Mode)
 - * Education: the largest number of people were high school graduates; half the people have received college degree (the Mode; the Median)
 - * Age: the largest number of people were 30; half the people are below the age of 32; the mean age is 35.6 years old. (the Mode; the Median; the Mean)

After this lecture:

You should be able to:

- * know the three measures of central tendency: the mode, the median, and the mean
- * calculate and interpret the three measures
- * identify the characteristics of the mean
- * indicate which measure of central tendency is appropriate for each level of measurement.
- * identify skewness of the data distribution.