

4 Chapter 2

CHAPTER 2. MEASURING AND DESCRIBING VARIABLES

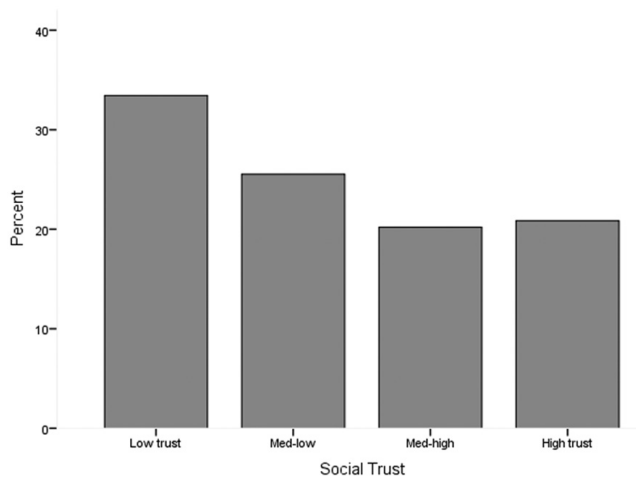
1. A. Age: name/interval; military dictatorship: value/nominal; strongly oppose: value/ordinal; election year: name/interval; 62 percent: value/interval; Asian: value/nominal; class rank: value/ordinal; commute distance: value/interval.

2. A. Social trust.

(i)

| Social trust | Frequency | Percent | Cumulative percent |
|--------------|-----------|---------|--------------------|
| Low | 437 | 33.4 | 33.4 |
| Medium-low | 334 | 25.5 | 59.0 |
| Medium-high | 264 | 20.2 | 79.2 |
| High | 273 | 20.8 | 100.0 |
| | 1,308 | 100.0 | |

(ii)



(iii) Mode is low trust. (iv) Median is medium-low trust. (v) High dispersion.

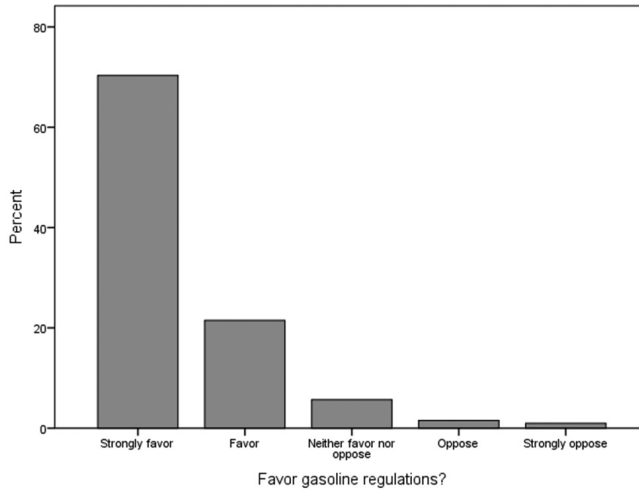
(vi) The median and mode are different values, indicating high dispersion. Also, the cases are spread out across the categories.

B. Gas regulations.

(i)

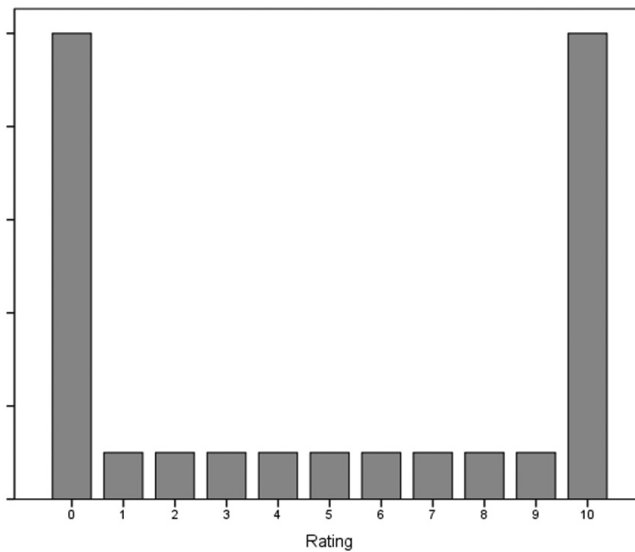
| Favor gas regulations? | Frequency | Percent | Cumulative percent |
|--------------------------|-----------|---------|--------------------|
| Strongly favor | 645 | 70.3 | 70.3 |
| Favor | 197 | 21.5 | 91.8 |
| Neither favor nor oppose | 52 | 5.7 | 97.5 |
| Oppose | 14 | 1.5 | 99.0 |
| Strongly oppose | 9 | 1.0 | 100.0 |
| Total | 917 | 100.0 | |

(ii)



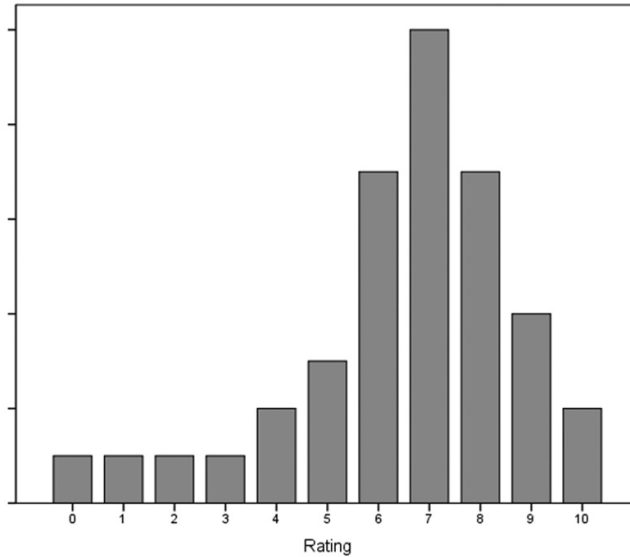
(iii) Mode is strongly favor. (iv) Median is strongly favor. (v) Low dispersion. (vi) The median and mode are the same value, indicating low dispersion. Also, the cases are heavily concentrated in the modal (median) category, strongly favor.

3. A. Students should sketch a bar chart showing large percentages of respondents grouped at the extreme values of the approval rating scale, such as the following:

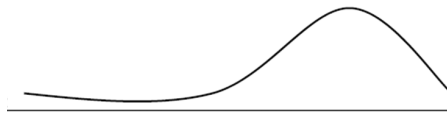


B. Set 2: median, 5, two modes (bimodal), 2 and 7. Senator Foghorn claims that the distribution of opinion is bimodal, creating a mean and median at or near the center of the scale (a score of around 5). Polarization implies a bimodal distribution, with cases clustered near each extreme, such as around scores of 2 and 7.

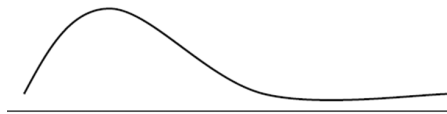
- C. Students should sketch a bar chart showing a clustering of respondents toward higher values of the scale. Plausible values would be high and fairly uniform. For example: mode, 7; median, 6.



4. A. Negative skew:



- B. Positive skew:



- C. No skew:



5. A. *The Barack Obama feeling thermometer has a positive skew.* (i) False. (ii) The thermometer's mean is below its median, suggesting a negative skew. The mode is 0, telling us that extremely low values are pulling the mean down, also indicating a negative skew.
- B. *The mean of the Hillary Clinton thermometer provides an accurate measure of central tendency.* (i) True. (ii) This is a judgment call, but the mean (58.8) and the median (60) are close to each other, suggesting that the distribution is not badly skewed. Therefore, the mean may be used to measure central tendency.

- C. *Hillary Clinton's ratings have a greater amount of variation than Barack Obama's ratings.*
- (i) False. (ii) For the Clinton thermometer, the interquartile range falls between 40 and 85. For the Obama thermometer, the interquartile range falls between 30 and 85. The larger interquartile range for the Obama thermometer indicates that his ratings are more spread out; that is, that the Obama scale has greater variation.

Chapter Two: Measuring and Describing Variables

Learning Objectives

1. How to recognize the essential features of a variable
2. How to determine a variable's level of measurement
3. How to describe the central tendency of a variable
4. How to describe the amount of dispersion in a variable

Chapter Summary

When conducting research in political science it is important that a researcher be able to **recognize the essential features of a variable**. A variable is a point of data that may have different values or meanings across a collection of data. For example, if a researcher were studying the political attitudes of individuals s/he would have at least two variables to consider. The first is the political attitude of the people in the study, which could range from extremely liberal to extremely conservative. A second variable the researcher might study would be the gender of the individual, which is usually categorized as male or female. Having those two variables would allow a researcher to compare the attitudes of men and women on various political issues.

There are a few more things a researcher should know about his variables before beginning a project, including the **level of measurement** of the variable. In quantitative research variables are typically measured at one of three levels: **nominal, ordinal, or interval**. A **nominal** level variable is one in which the variable represents a difference between the units of analysis based upon a specific characteristic being measured, such as gender. A second level of measurement is **ordinal variables**. Ordinal variables are those that can be ordered according to the degree to which a characteristic is present, such as an individual's level of education. The third level of measurement is the **interval variable**. An interval level variable is one that measures the presence of a characteristic along a continuous spectrum, such as a person's age.

Finally, two additional elements that researchers need to know are how to describe a variable's **central tendency** and the amount of **dispersion** in a variable. Central tendency is simply a description of where the center of the variable is found and may be described several different ways. For example, a researcher may describe the average (or **mean**) age of respondents to a survey. S/he may also state the **median** age, which is the point where half the respondents are younger than a particular age and half are older than that age. The level of **dispersion** in a variable tells us how tightly clustered the values of a variable are around the central tendency. A variable with low dispersion is tightly clustered around a few categories while a variable with high dispersion has values spread across many categories.

Discussion Questions

1. Suppose a researcher wanted to learn about the relationship between party affiliation and individual views on gun control. What variables would the researcher need to collect and at what level would they be measured?
2. Discuss the advantages and disadvantages of using the mean and the median to describe the central tendency of a variable? Is either of them more reliable than the other?
3. Suppose a researcher wanted to learn about the views of Democrats and Republicans on President Obama's job performance and were asked to rate him on a scale from 0-10 (0 being total disapproval to 10 being complete approval). Would the results reflect a high level of dispersion or a low level of dispersion? Why?

Measuring Variables

- Two essential elements for every variable are the variable name and the variable values. Every variable must have at least two values.
 - For example, party affiliation is the name of a variable while Republican, Democrat, Independent, Libertarian, etc. are the values of the variable.

Levels of Measurement

- Researchers typically distinguish between three different levels at which a variable may be measured—nominal, ordinal, and interval.

Additive Indexes

- An index is an additive combination of ordinal variables that are identically coded and measure the same concept.
 - An example could be an index designed to measure political liberalism by asking respondents a series of questions related to political issues such as abortion, gun control, and same sex marriage. Respondents could be asked to rate their support for each issue on a scale from 1 to 5, with 1 being no support and 5 being high support. The numeric value for each question is then added up to measure the concept of political liberalism. In this example the maximum score would be 15, which would indicate a high level of liberalism while the minimum score would be a 3, which would indicate a low level of liberalism.

Describing Variables

- The most frequent way to describe variables in society is by the use of numbers.
 - For example, colleges and universities use a number, the grade point

average, to describe a student's academic performance.

- Variables are often described using one of three measures of **central tendency**—the mode, the median, or the mean.
 - The **mode** describes the most frequently occurring value in a variable, such as the number of runs scored in a baseball game.
 - Say your favorite team plays ten games and scores five runs in six of those games. The mode of runs scored is five. This is said to be **unimodal** because five is the most frequent number of runs. However, if the team scored four runs on three occasions and five runs on three occasions then the mode would be both four and five since each number occurred with the same frequency. This referred to as a **bimodal** distribution.
 - The **median** is another measure of central tendency that provides the value at which half the values fall below that value and half fall above it.
 - Think about the median income in a community. Suppose it is \$70,000. That means half the families make less than \$70,000 and half the families make more than \$70,000.
 - The **mean** is also used to describe a measure's central tendency. The mean is the arithmetic average of a set of numbers.
 - Using the example above we might find the mean income of families in a community to be \$100,000. It should be noted the mean may be skewed by outliers (extremely high or low values). For example, suppose in our fictional community there are several families making over \$1 million a year. This would likely push the average income higher and cause the mean to be less informative than the median or mode.

Nominal Variables

- Nominal level is used to describe categories within a variable, such as home ownership. One either owns a home or does not own a home
- Gender is another good example to use as it is typically a two category nominal level variable consisting of males and females. An individual does not possess more of less male gender or female gender. It is simply a category, which leads some researchers to classify nominal level variables as categorical variables.

Ordinal Variables

- Variables measured at the ordinal level may be **rank-ordered**. That is, someone or something may have more or less of the characteristic being measured by the variable.
 - Think about the level of education individuals have completed. One may have completed an advanced graduate degree while another

may have dropped out of high school. The one with the graduate degree has more education than the one who dropped out of high school.

Interval Variables

- Interval variables provide precise measurements of a characteristic, such as age or income or any other characteristic, which can be measured along a continuous spectrum.
 - A frequently used interval measurement in political research is the **feeling thermometer**. Individuals are asked to rate how they feel towards a particular individual on a scale from zero (ice cold) to 100 (extremely warm).
 - Used in conjunction with another variable a researcher would be able to learn how different groups of people feel about someone or something such as a presidential candidate or climate change.

Web Resources

1. Watch the video about levels of measurement at <https://youtu.be/A5zlhbmBghI>
2. Watch this fascinating video about measures of central tendency at <https://youtu.be/5C9LBF3b65s>
3. Dispersion is easily explained by this video: <https://youtu.be/9mniDp6tg-4>

Exercises

1. Collect the month of birth for all the students in the classroom and create a frequency distribution table for them. What is the mode(s) for birth month?
2. Collect the ages of all students in the classroom and calculate both the average age of students and the median age of students. Are they the same or different? If they are different, why?
3. Construct an additive index measuring school spirit. What variables will be used and how will they be measured?



The
ESSENTIALS of
**Political
Analysis**

PHILIP H. POLLOCK III

5e

**Chapter 2:
Measuring and
Describing Variables**

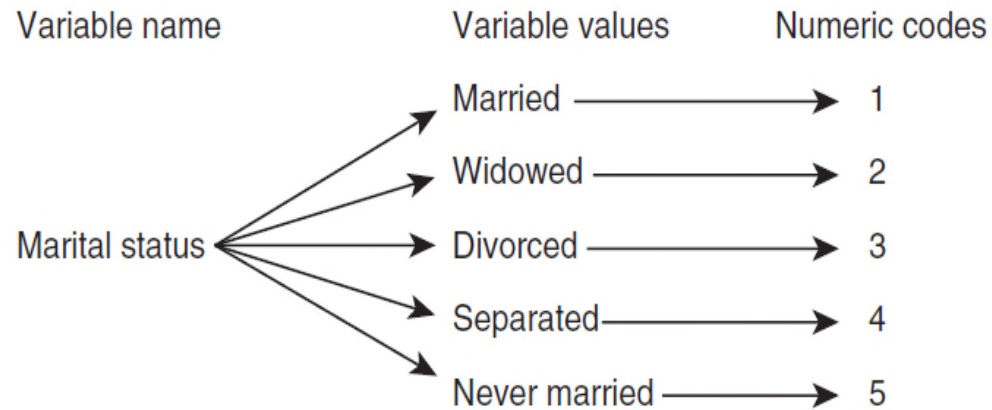
Topics: Chapter 2

- Measuring Variables
 - *Levels of Measurement*
 - *Additive Indexes*
- Describing Variables
 - *Nominal Variables*
 - *Ordinal Variables*
 - *Interval Variables*

Variables

- A **variable** is an empirical measurement of a characteristic.
- Every variable has one name and at least two possible values.
 - Name: region. Values: South, Northeast, Midwest, West
 - Name: income. Values: \$42,000; \$42,001; \$16,152

Figure 2-1 Anatomy of a Variable



Name Vs. Value?

- Think of a unit of analysis
- Ask: “What is this unit’s _____?”
 - The word that fills in the blank is always the name of a variable.
- The question is always answered by one of the variable’s values.
 - You would say, “What is this person’s marital status?” (variable name)
 - You wouldn’t say, “What is this person’s single?” (variable value)

3 Levels of Precision

- Three levels of measurement or “degrees of precision”
 - 1. Interval level variables have values that are the most precise (eg. Income in dollars).
 - 2. Ordinal-level variables have values that are somewhat less precise (eg. approval of disapproval same-sex marriage).
 - 3. Nominal-level variables have the least precise values (eg., gender: male and female) .

Interval Level Variables

- Have values that communicate the exact amount of the characteristic being measured
- Use a widely-recognized scale
- Comparing two units of analysis
 - classify into different values of the characteristic
 - rank by relative amount of the characteristic
 - compute the exact difference between units
- For interval variables, the numeric codes convey the exact quantities of the measured characteristic.
 - Person A is 21 years old. Person B is 22. The values are the ages, in years. The difference in ages is exactly 1 year.

Ordinal Level Variables

- Have values that communicate the relative amount of the characteristic being measured.
 - may use a scale but the scale measures relative amount, not absolute amount
- Comparing units of analysis
 - classify into different values on the characteristic
 - rank by relative amount of the characteristic
- For ordinal variables, the numeric codes represent relative amounts, not absolute amounts.
 - Person A “approves” of same-sex marriage (numeric code 1). Person B “disapproves” (numeric code 2). Person B has *relatively* more disapproval than Person A. But the difference is not 1 unit of approval.

Nominal Level Variables

- Have values that communicate differences
 - may be coded with numbers but the numbers just differentiate between units
- Comparing units
 - Classify into different values on the characteristic
 - Person A is male (numeric code 1). Person B is female (numeric code 2).
- The codes simply represent differences. The numbers have no inherent meaning.
 - not exact amounts of gender
 - not relative amounts of gender
 - only differences in gender

Box 2-1 Nominal Means “Name,” Not “Number”

Students sometimes mistakenly equate the word *nominal* with *number*. Given this mistaken assumption, for example, one might misidentify an interval-level variable, such as age—measured by a number, number of years—as a nominal variable. This would be incorrect. The word nominal means “in name only.” Thus, variables whose values are names, or whose numeric codes only represent names, are nominal variables.

Ordinal scales

- A common practice: summing ordinal items into an *additive* scale
 - Likert scales the most common
 - all items in the scale must measure the same concept
- The similarity of the questions ensures that only one concept is measured.
- The greater the number of questions, the more reliable the scale.



The ESSENTIALS of Political Analysis

Likert example: Egalitarianism

Table 2-1 Items in a Likert Scale

1. Our society should do whatever is necessary to make sure that everyone has an equal opportunity to succeed.
 2. We have gone too far in pushing equal rights in this country.
 3. One of the big problems in this country is that we don't give everyone an equal chance.
 4. This country would be better off if we worried less about how equal people are.
 5. It is not really that big a problem if some people have more of a chance in life than others.
 6. If people were treated more equally in this country, we would have many fewer problems.
-

Source: 2012 American National Election Study. Variables `egal_equal`, `egal_toofar`, `egal_bigprob`, `egal_worryless`, `egal_notbigprob`, and `egal_fewerprobs`.

Note: After looking at each statement, respondents are asked: "Do you agree strongly, agree somewhat, neither agree nor disagree, disagree somewhat, or disagree strongly with this statement?"

Describing variables

- Any variable can be described by its:
 1. Central tendency
 - “average” value
 - value that best typifies the variable
 2. Dispersion
 - extent to which the units of analysis are distributed across the values of the variable

Measures of Central Tendency

- **Mode:** The “common-most” value
 - may be used with any level of measurement
 - for nominal variables, the only measure that may be used
- **Median:** The “middle-most” value
 - 50% of cases fall above / 50% fall below the median
 - the same as the 50th percentile
 - for ordinal or interval variables
- **Mean:** The arithmetic average
 - add up all the values and divide by the number of cases
 - for interval variables only

Dispersion

- a neglected aspect of description
- In political discussion, we use dispersion to describe variables.
 - “polarization,” “consensus,” “equality”
- A variable has **maximum dispersion** if the cases are spread evenly across all values of the variable.
- A variable has **minimum dispersion** if the cases all fall into one value of the variable.

Frequency distributions

- 3 (sometimes 4) columns
- Column 1: labels of the variable's values
- Column 2: raw number of cases in each value
- Column 3: percentage of cases in each value
- Column 4: cumulative percentages showing the percentage of cases in or below each value
 - Ordinal and interval variables only

Graphs: Bar charts

- The variable's values appear on the horizontal axis.
- The percentage of cases in each value are recorded on the vertical axis.
- for variables at all levels of measurement

Examples

- **Nominal**
 - region of residence
- **Ordinal**
 - religious attendance
 - ideological self-placement
- **Interval**
 - TV hours
 - support for spending on government programs
 - feeling thermometers

Region of residence (Nominal)

are visually pleasing and elegant. The variable's values are labeled along the horizontal axis

Table 2-2 Region of Residence (tabular)

| Region | Frequency | Percentage |
|--------------------|-----------|------------|
| New England | 101 | 5.1 |
| Middle Atlantic | 237 | 12.0 |
| East North Central | 326 | 16.5 |
| West North Central | 116 | 5.9 |
| South Atlantic | 423 | 21.4 |
| East South Central | 115 | 5.8 |
| West South Central | 201 | 10.2 |
| Mountain | 148 | 7.5 |
| Pacific | 307 | 15.5 |
| Total | 1,975 | 100.0 |

Source: 2012 General Social Survey.

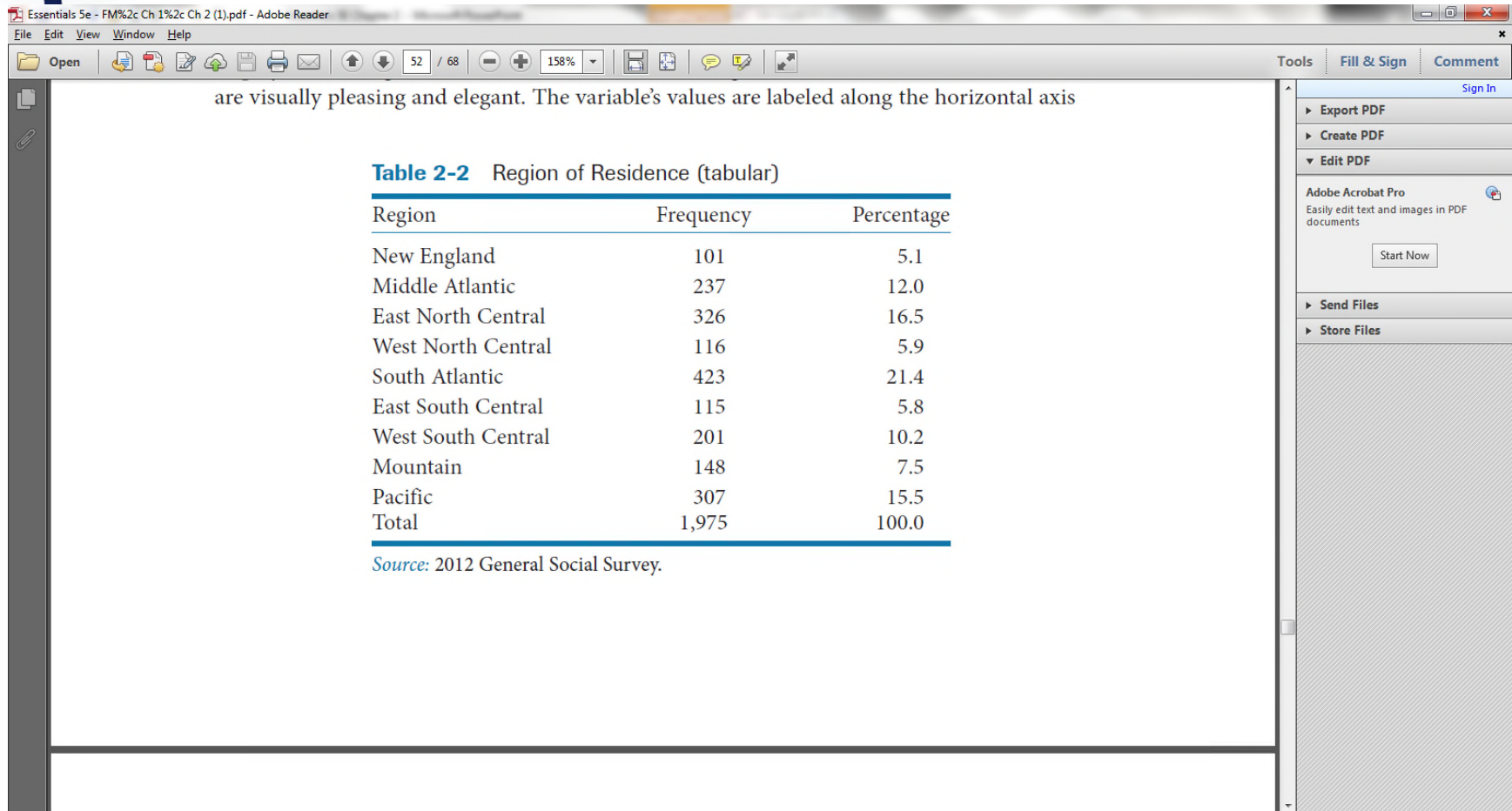
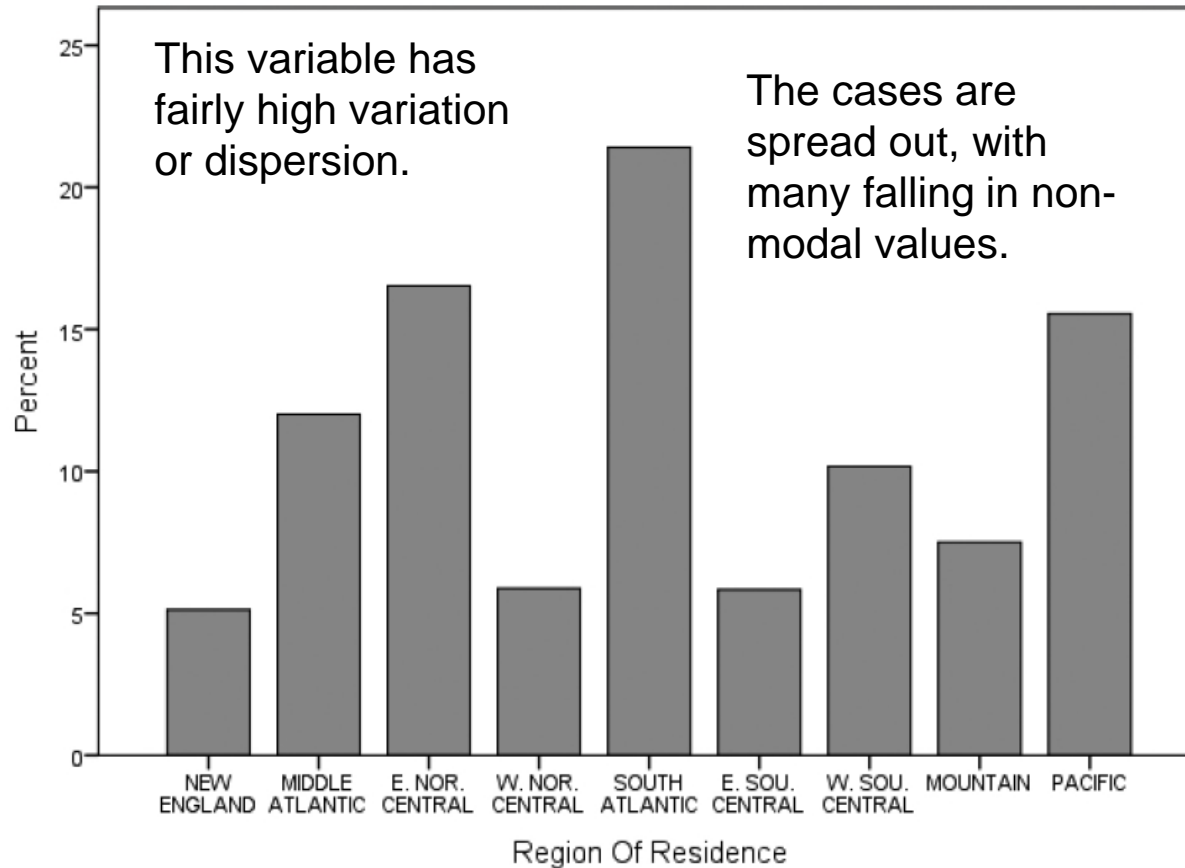


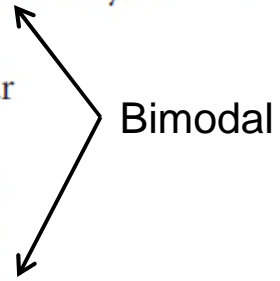
Figure 2-2 Region of Residence (graphic)



Religious attendance (Ordinal)

Table 2-3 Attendance at Religious Services (tabular)

| Attendance | Frequency | Valid percentage | Cumulative percentage |
|--------------------------------|-----------|------------------|-----------------------|
| Never or less than once a year | 595 | 30.3 | 30.3 |
| Once a year | 256 | 13.0 | 43.3 |
| Several times a year | 213 | 10.8 | 54.1 |
| Once a month | 133 | 6.8 | 60.9 |
| 2-3 times a month | 174 | 8.9 | 69.7 |
| Nearly every week | 79 | 4.0 | 73.8 |
| Every week or more | 516 | 26.2 | 100.0 |
| Total | 1,967 | 100.0 | |



Source: 2012 General Social Survey.

Note: Question: “How often do you attend religious services?”

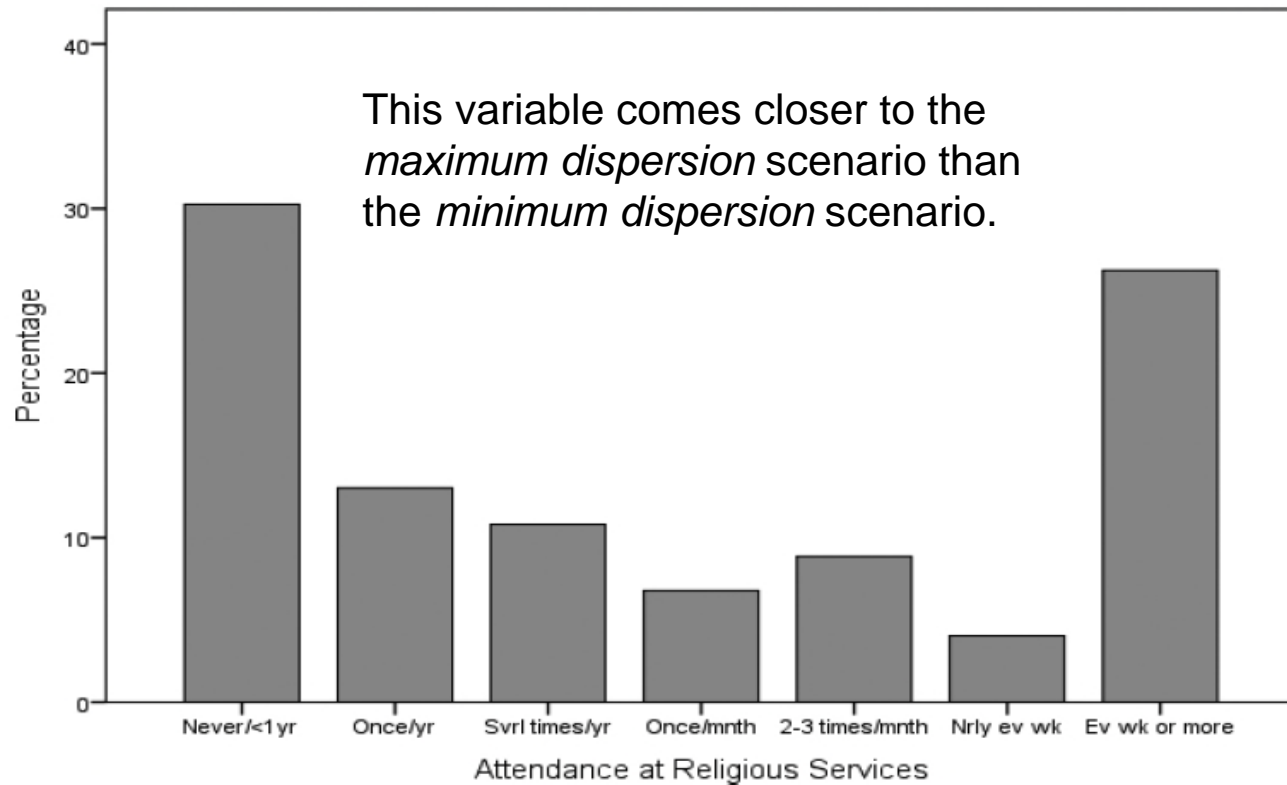
Table 2-3 Attendance at Religious Services (tabular)

| Attendance | Frequency | Valid percentage | Cumulative percentage |
|--------------------------------|-----------|------------------|-----------------------|
| Never or less than once a year | 595 | 30.3 | 30.3 |
| Once a year | 256 | 13.0 | 43.3 |
| Several times a year ← Median | 213 | 10.8 | 54.1 |
| Once a month | 133 | 6.8 | 60.9 |
| 2–3 times a month | 174 | 8.9 | 69.7 |
| Nearly every week | 79 | 4.0 | 73.8 |
| Every week or more | 516 | 26.2 | 100.0 |
| Total | 1,967 | 100.0 | |

Source: 2012 General Social Survey.

Note: Question: “How often do you attend religious services?”

Figure 2-3 Attendance at Religious Services (graphic)



Source: 2012 General Social Survey.

Note: Question: "How often do you attend religious services?"

Ideological Self-Placement (Ordinal)

Table 2-4 Ideological Self-Placement (tabular)

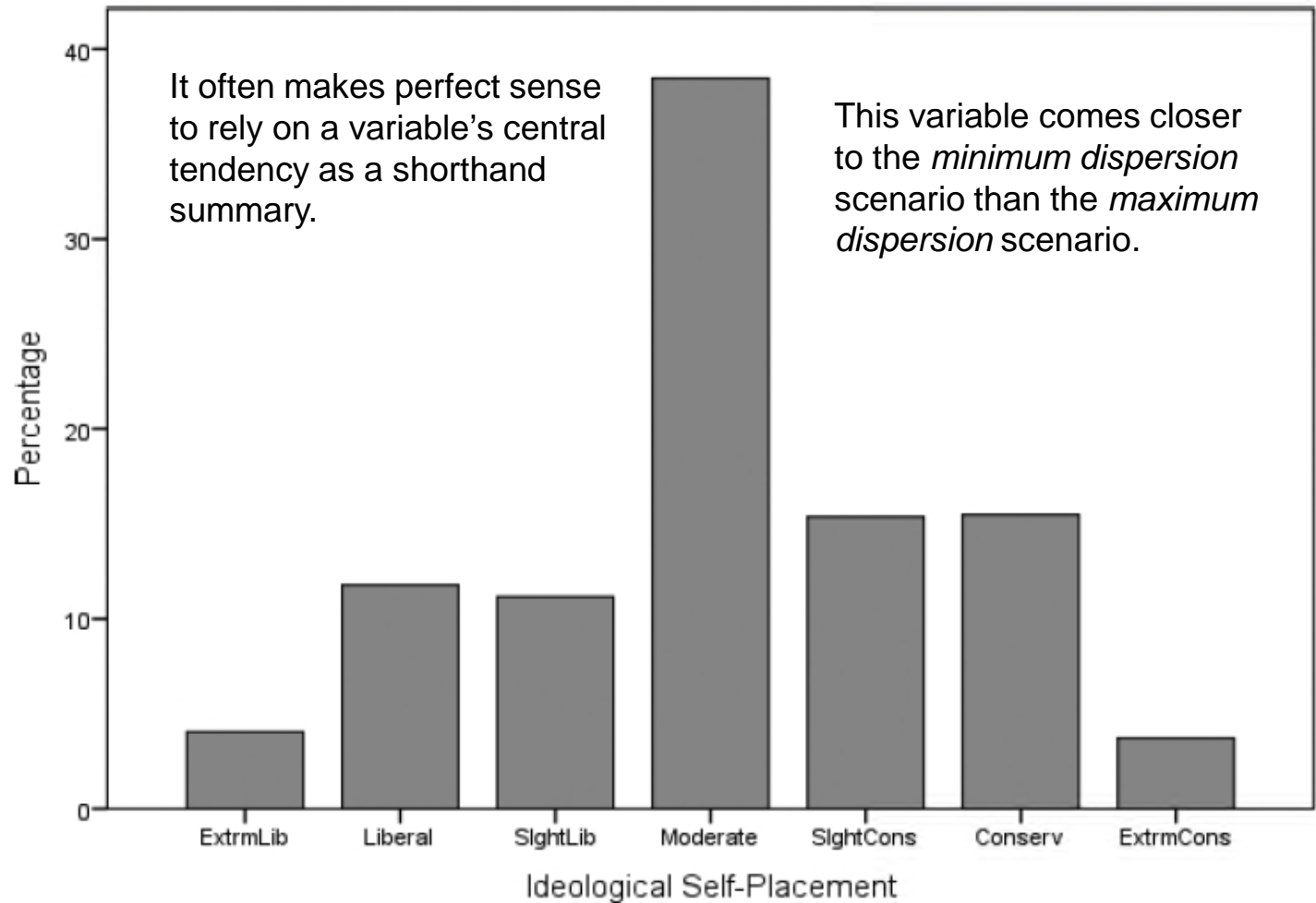
| Ideology | Frequency | Percentage | Cumulative percentage |
|---|-----------|------------|-----------------------|
| Extremely liberal | 76 | 4.0 | 4.0 |
| Liberal | 220 | 11.8 | 15.8 |
| Slightly liberal | 209 | 11.2 | 27.0 |
| Moderate ← Mode = "Moderate" Median = "Moderate" | 720 | 38.5 | 65.4 |
| Slightly conservative | 288 | 15.4 | 80.8 |
| Conservative | 290 | 15.5 | 96.3 |
| Extremely conservative | 70 | 3.7 | 100.0 |
| Total | 1,873 | 100.0 | |

Source: 2012 General Social Survey.

Note: Question: "We hear a lot of talk these days about liberals and conservatives. I'm going to show you a 7-point scale on which the political views that people might hold are arranged from extremely liberal (point 1) to extremely conservative (point 7). Where would you place yourself on this scale?"

The ESSENTIALS of Political Analysis

Figure 2-4 Ideological Self-Placement (graphic)



TV hours (Interval)

Table 2-5 Hours Watching TV (tabular)

| Hours | Frequency | Percentage | Cumulative percentage |
|--------------|--------------|--------------|-----------------------|
| 0 | 86 | 6.6 | 6.6 |
| 1 | 291 | 22.2 | 28.8 |
| 2 | 346 | 26.4 | 55.2 |
| 3 | 222 | 17.0 | 72.2 |
| 4 | 164 | 12.5 | 84.7 |
| 5 | 56 | 4.3 | 89.0 |
| 6 | 55 | 4.2 | 93.2 |
| 7 | 16 | 1.2 | 94.5 |
| 8 | 31 | 2.4 | 96.8 |
| 9 | 3 | .2 | 97.0 |
| 10 | 14 | 1.0 | 98.0 |
| 12 | 9 | .7 | 98.7 |
| 13 | 1 | .1 | 98.8 |
| 14 | 4 | .3 | 99.1 |
| 15 | 4 | .3 | 99.5 |
| 16 | 1 | .1 | 99.5 |
| 18 | 0 | .0 | 99.6 |
| 20 | 1 | .1 | 99.7 |
| 22 | 1 | .1 | 99.7 |
| 24 | 4 | .3 | 100.0 |
| Total | 1,309 | 100.0 | |

← Mode = 2
Median = 2

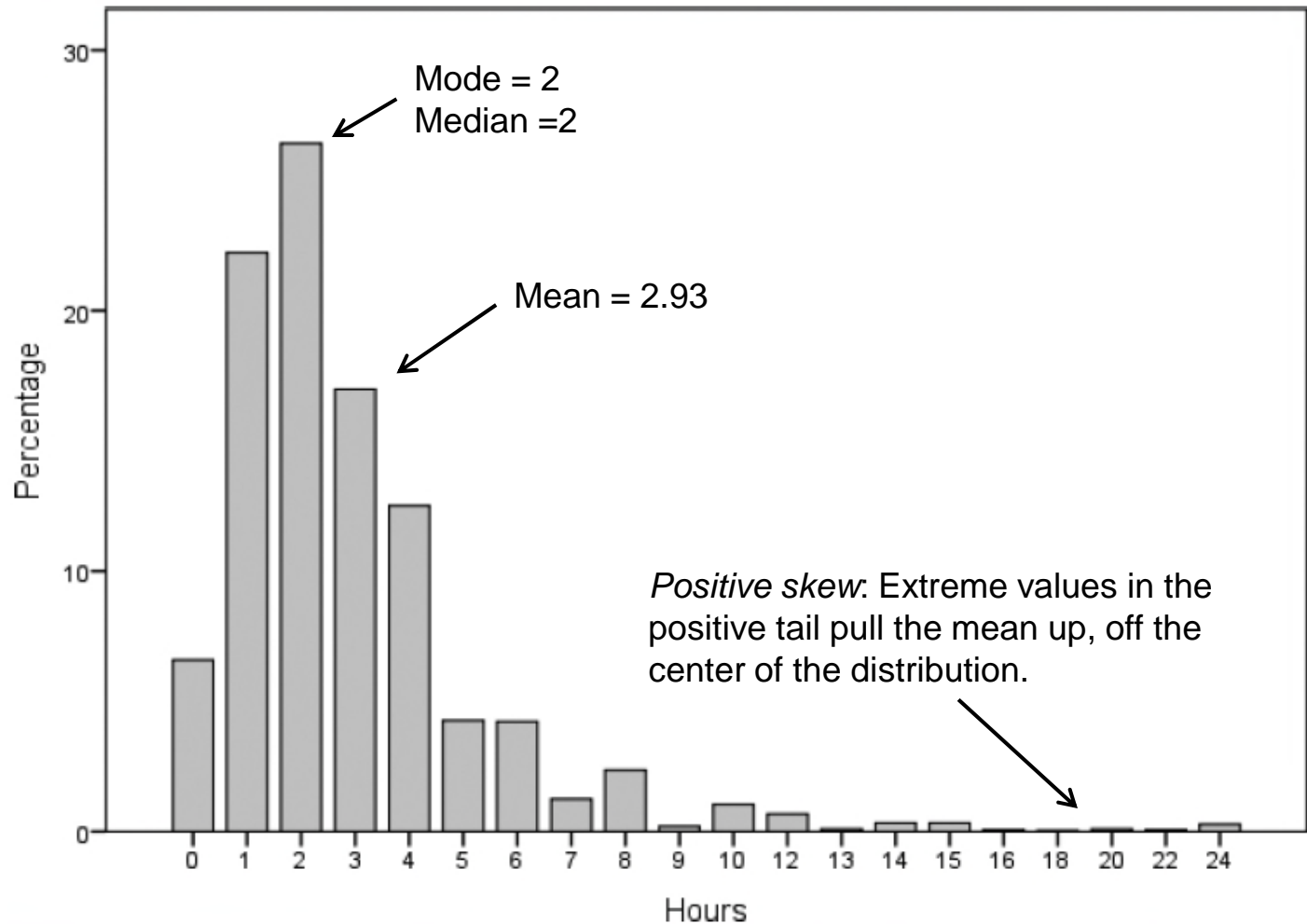
Source: 2012 General Social Survey.

Note: Question: "On the average day, about how many hours do you personally watch television?"

Pollock, The Essentials of Political Analysis. © 2016,
SAGE Publications

The ESSENTIALS of Political Analysis

Figure 2-5 Hours Watching TV (graphic)



Source: 2012 General Social Survey.

Support for spending on 17 programs (Interval)

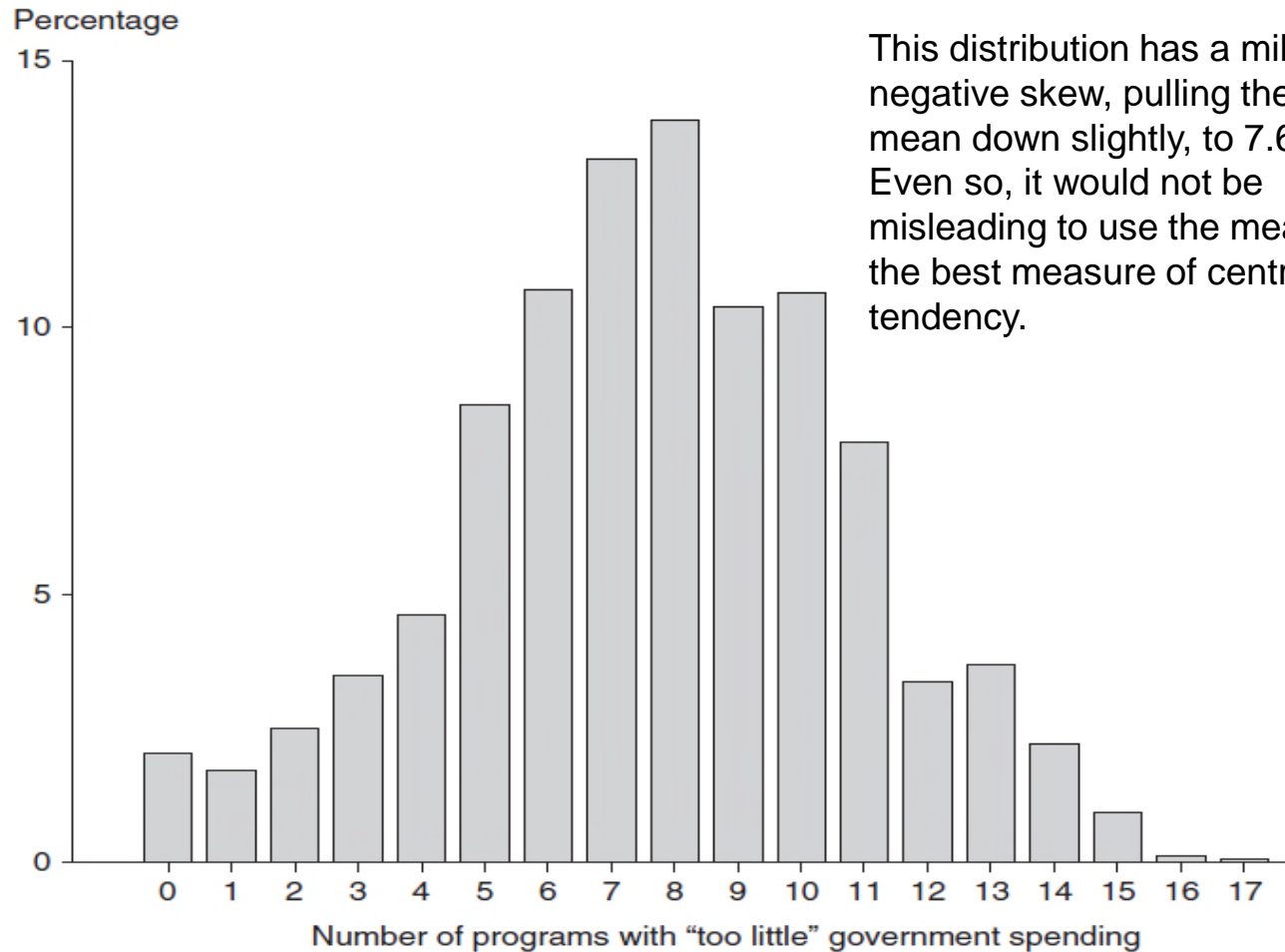
Table 2-6 Number of Programs with “Too Little” Government Spending (tabular)

| Number of programs | Frequency | Percentage | Cumulative percentage |
|--------------------|-----------|------------|-----------------------|
| 0 | 14 | 2.0 | 2.0 |
| 1 | 11 | 1.7 | 3.7 |
| 2 | 17 | 2.5 | 6.2 |
| 3 | 23 | 3.5 | 9.7 |
| 4 | 31 | 4.6 | 14.4 |
| 5 | 57 | 8.6 | 22.9 |
| 6 | 72 | 10.7 | 33.6 |
| 7 | 88 | 13.2 | 46.8 |
| 8 | 93 | 13.9 | 60.7 |
| 9 | 70 | 10.4 | 71.1 |
| 10 | 72 | 10.7 | 81.8 |
| 11 | 53 | 7.9 | 89.6 |
| 12 | 23 | 3.4 | 93.0 |
| 13 | 25 | 3.7 | 96.7 |
| 14 | 15 | 2.2 | 98.9 |
| 15 | 6 | .9 | 99.8 |
| 16 | 1 | .1 | 99.9 |
| 17 | 1 | .1 | 100.0 |
| Total | 672 | 100.0 | |

Source: 2008 General Social Survey.

The ESSENTIALS of Political Analysis

Figure 2-6 Number of Programs with “Too Little” Government Spending (graphic)



Source: 2008 General Social Survey.

Interquartile range

- non-statistical measure of dispersion for interval (and ordinal) variables
- especially informative when two distributions are being compared.
- Defined as the range of a variable's values that defines the "middle half" of a distribution:
 - between the upper boundary of the lowest quartile (which is the same as the 25th percentile) and the lower boundary of the upper quartile (the 75th percentile)
- graphically depicted by a **box plot**

Feeling thermometers (Interval)

Table 2-7 Summary Information for Two Interval Variables

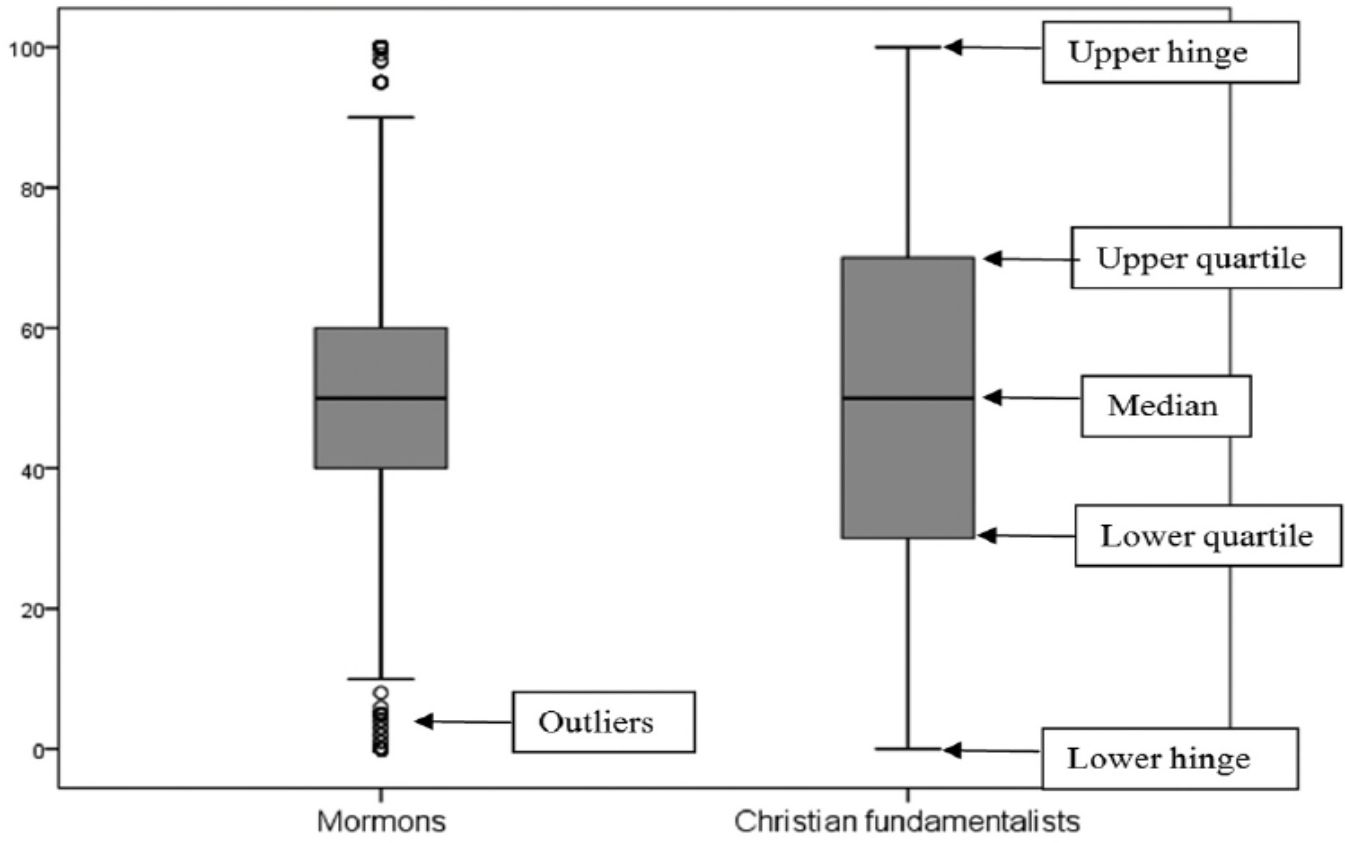
| | Mormons | Christian fundamentalists |
|---------------------|---------|---------------------------|
| Mean | 50.5 | 48.9 |
| Median | 50 | 50 |
| Mode | 50 | 50 |
| <i>Percentiles:</i> | | |
| 25 | 40 | 30 |
| 50 | 50 | 50 |
| 75 | 60 | 70 |

Source: 2012 American National Election Study.

The ESSENTIALS of Political Analysis

Box plots

Figure 2-7 Box Plots of Two Interval Variables



Source: 2012 American National Election Study.

Note: Figure annotations based on Robert I. Kabacoff, *R in Action: Data Analysis and Graphics with R* (Shelter Island, N.Y.: Manning, 2011), 133.

Pollock, The Essentials of Political Analysis. © 2016,

SAGE Publications