Central Tendency and Dispersion

POSC 3410 – Quantitative Methods in Political Science

Steven V. Miller

Department of Political Science

Goal for Today

Describe variables by reference to central tendency and dispersion.

Defining and Measuring Variables

Last lecture focused on a typology of variables.

1. Nominal
2. Ordinal
3. Interval

Correct classification will condition how we can *describe* variables.

Central Tendency

The most common description of interest is the **central tendency**.

- This is the variables "typical", or "average" value.
- This takes on different forms contingent on variable type.

Think of what follows as a "tool kit" for researchers.

- More precise variables allow for more precise measures.
- Use the right tool for the job, if you will.

Mode

The **mode** is the most basic central tendency statistic.

- It identifies the most frequently occurring value.

Suppose I have a random sample of 50 students and measured party affiliation.

- Democrats: 26; Republicans: 20; Others: 4

What's the modal category?

Mode

If I randomly grabbed a student from that sample and guessed "Democrat", I would be right 26 times of 50 (on average).

- No other guess, on average, would be as good.

This is the only central tendency statistic for nominal variables.

Median

The **median** is the middlemost value.

- It's the most precise statistic for ordinal variables.
- It's a useful robustness check for interval variables too.

Formally, a median m exists when the following equalities are satisfied.

$$P(X \leq m) \geq \frac{1}{2} \text{ and } P(X \geq m) \geq \frac{1}{2} \tag{1}$$

Finding the Median

Order the observations from lowest to highest and find what value lies in the exact middle.

- The median is the point where half the values lie below and half are above.
- We can do this when our variables have some kind of "order".
- Medians of nominal variables are nonsensical.

Mean

The arithmetic **mean** is used only for interval variables.

- This is to what we refer when we say "average".

Formally, i through n:

$$\frac{1}{n}\Sigma x_i \tag{2}$$

We can always describe interval variables with mode and median.

- We cannot do the same for ordinal or nominal with the mean.

Dispersion

We also need to know variables by reference to its **dispersion**.

- i.e. "how average is 'average'?"
- How far do variables deviate from the typical value?
- If they do, measures of central tendency can be misleading.

The interval variable with no dispersion problem is one in which the mode, median, and mean are the same value.

- This will not happen when there is a significant **skew**, or a **bimodal** distribution.

Frequency Distribution

A **frequency distribution** is a summary of a variable's values.

Table 2-2 Region of Residence (tabular)

Region	Frequency	Percentage
Northeast	344	17.0
Midwest	427	21.1
South	737	36.4
West	515	25.5
Total	2,023	100.0

Source: 2008 General Social Survey.

Cumulative Percentage

A **cumulative percentage** is the percentage of cases at or below a given value.

Table 2-3 Attendance at Religious Services (tabular)

Attendance	Frequency	Percentage	Cumulative percentage
Never or less than once a year	579	28.8	28.8
Once a year	272	13.5	42.3
Several times a year	223	11.1	53.4
Once a month	151	7.5	60.9
2–3 times a month	182	9.1	69.9
Nearly every week	93	4.6	74.5
Every week or more	513	25.5	100.0
Total	2,013	100.1[a]	

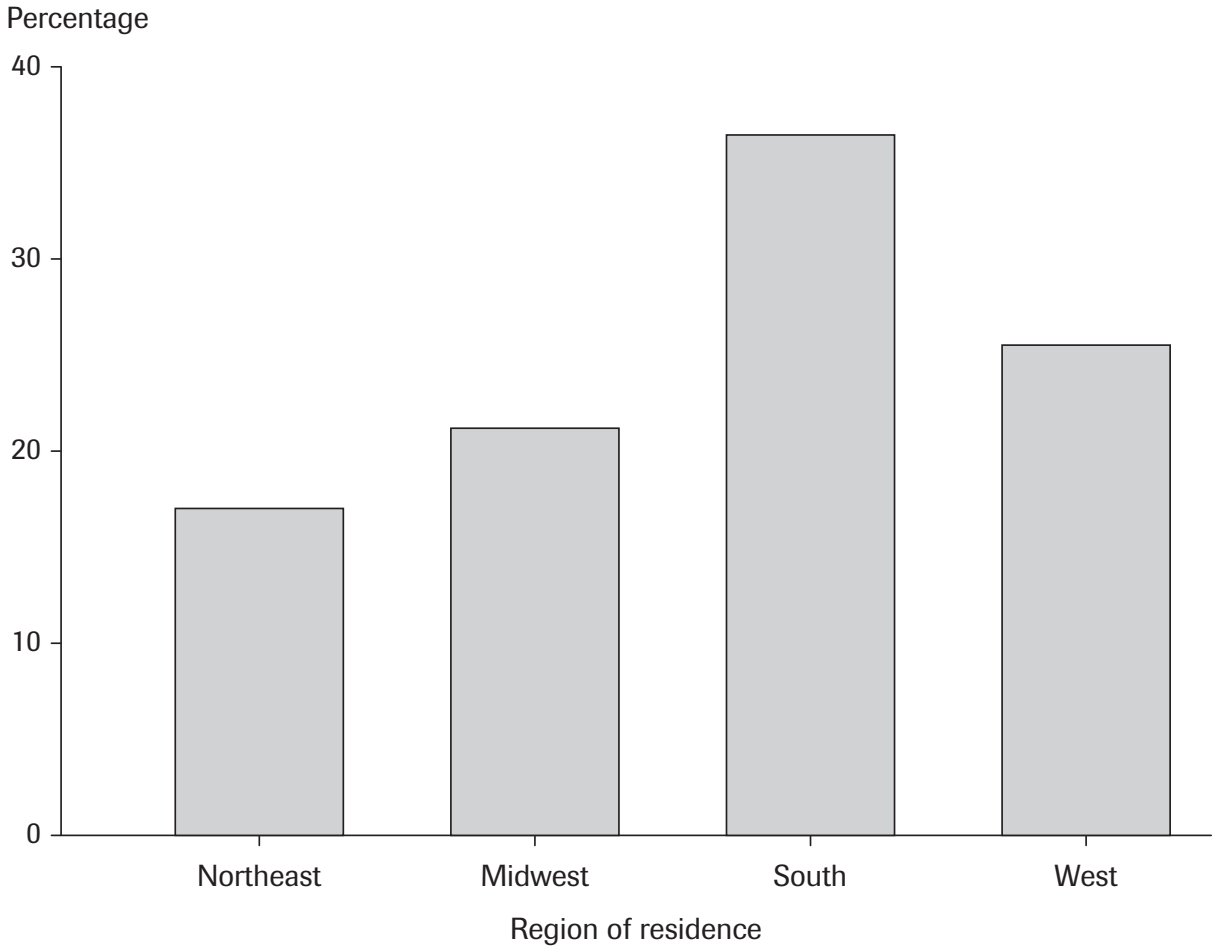
Source: 2008 General Social Survey.

Note: Question: "How often do you attend religious services?" The GSS records nine response categories. Table 2-3 combines "Never" with "Less than once a year" and "Every week" with "Several times a week."

[a]Percentages do not sum to 100.0 percent due to rounding.

Bar Chart

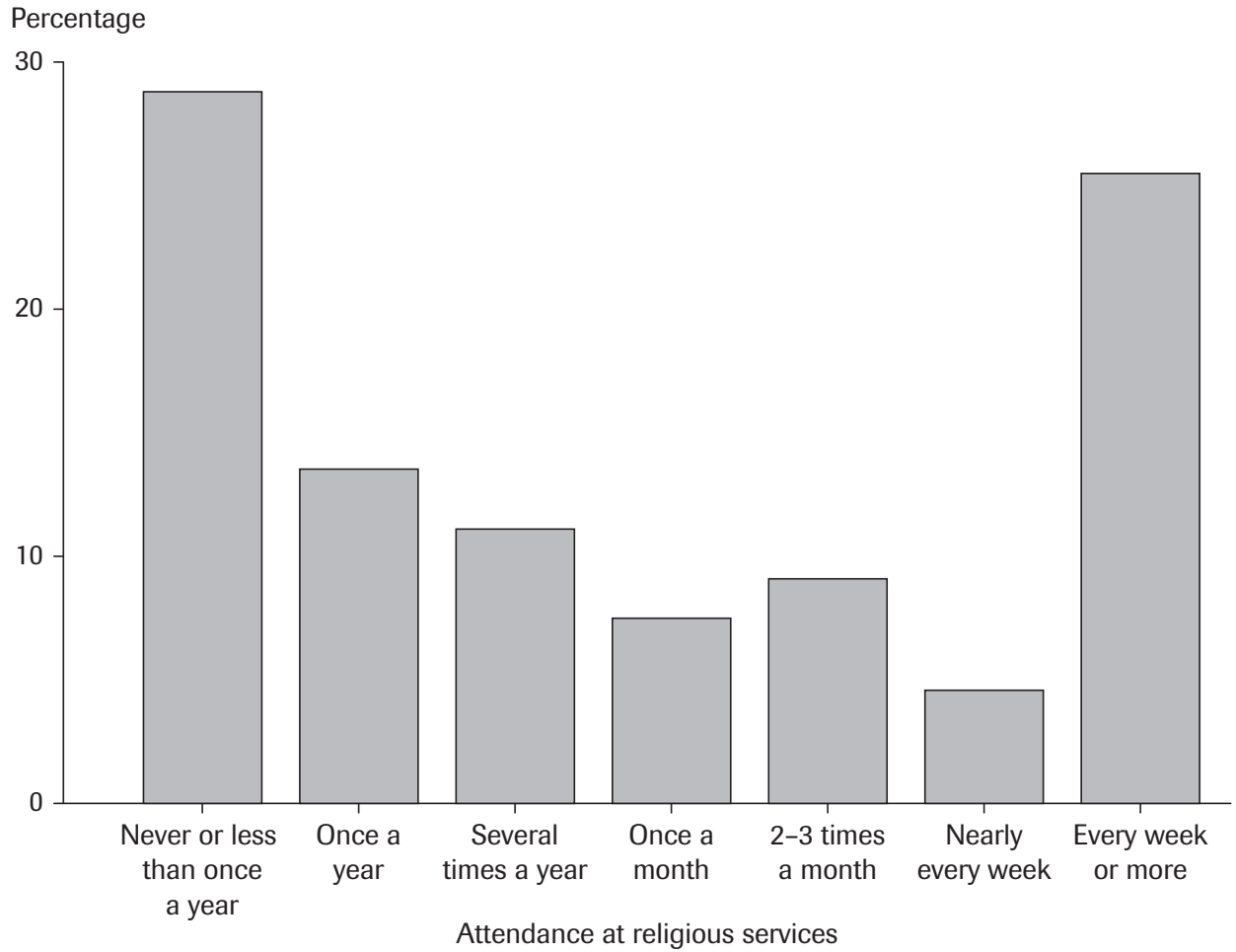
Figure 2-2 Region of Residence (graphic)



Source: 2008 General Social Survey.

Bar Chart

Figure 2-3 Attendance at Religious Services (graphic)

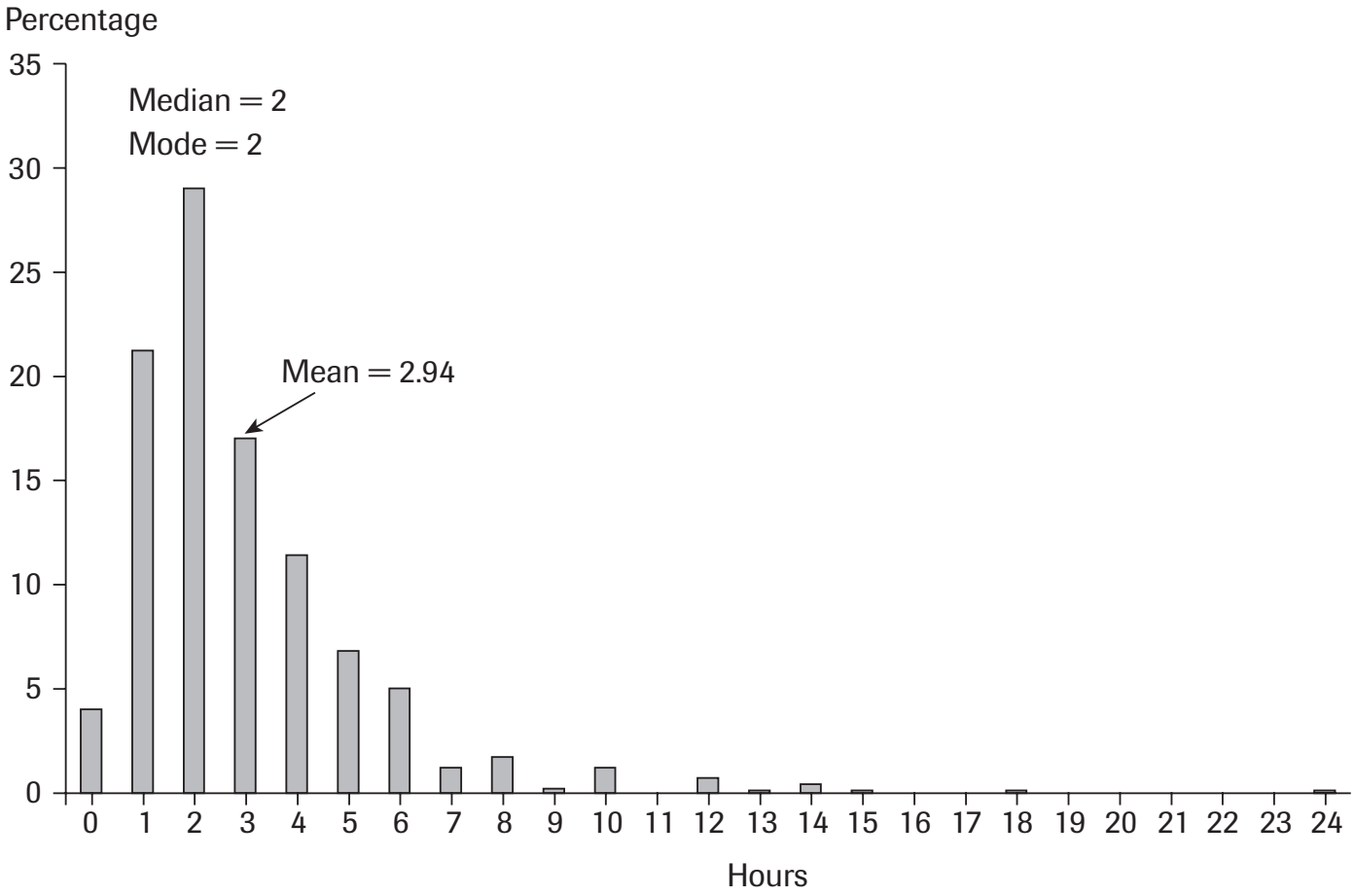


Source: 2008 General Social Survey.

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

Issues of Skew

Figure 2–4 Hours Watching TV (graphic)

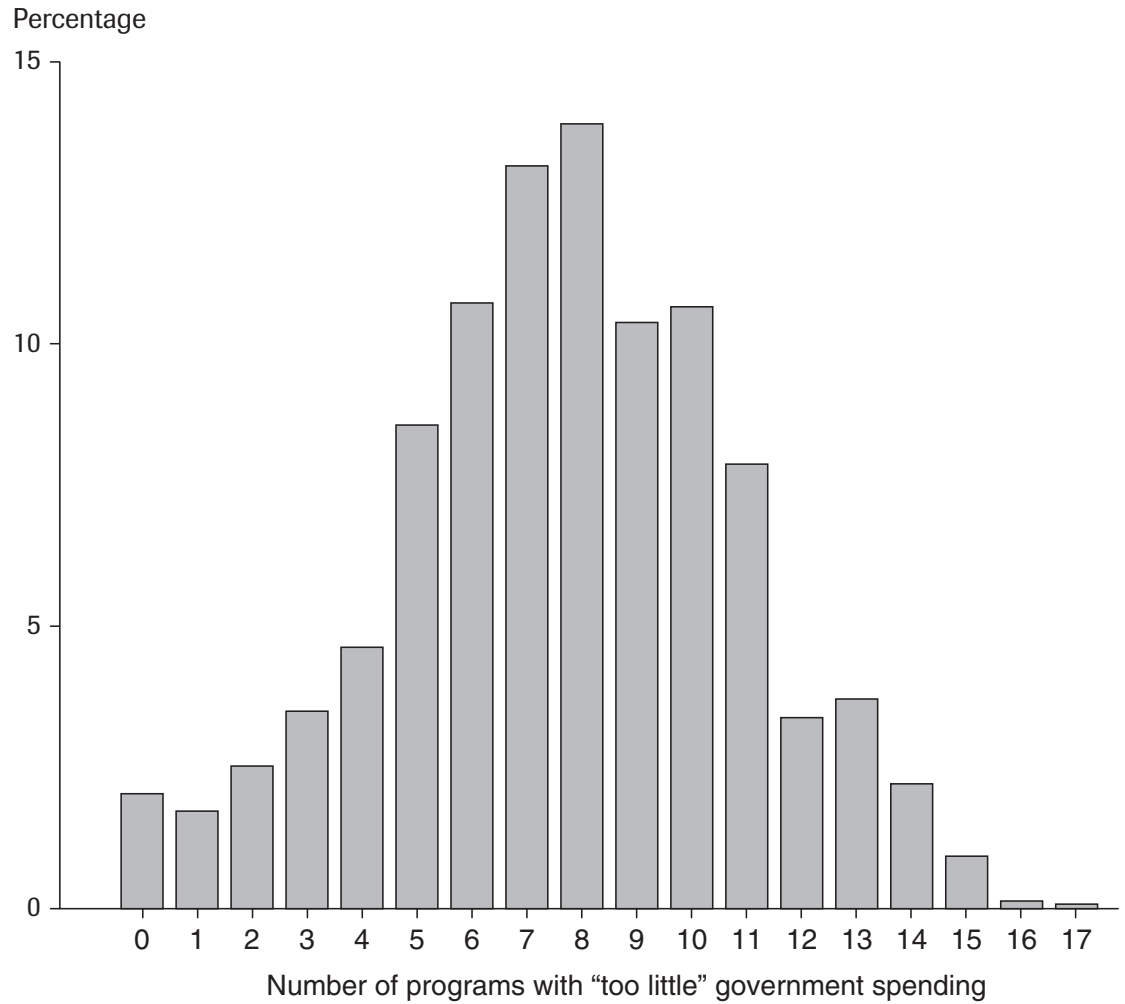


Source: 2006 General Social Survey.

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

Issues of Skew

Figure 2-5 Number of Programs with "Too Little" Government Spending (graphic)



Source: 2008 General Social Survey.

Conclusion

Here are some final thoughts.

- There is a reason we discuss "median income" and not the "average income".
- The mean of a dummy variable communicates the percentage of 1s, divided by 100.
- Skew is mostly a problem of mean variables, and a problem of degree.

Always look carefully at your data!

Table of Contents