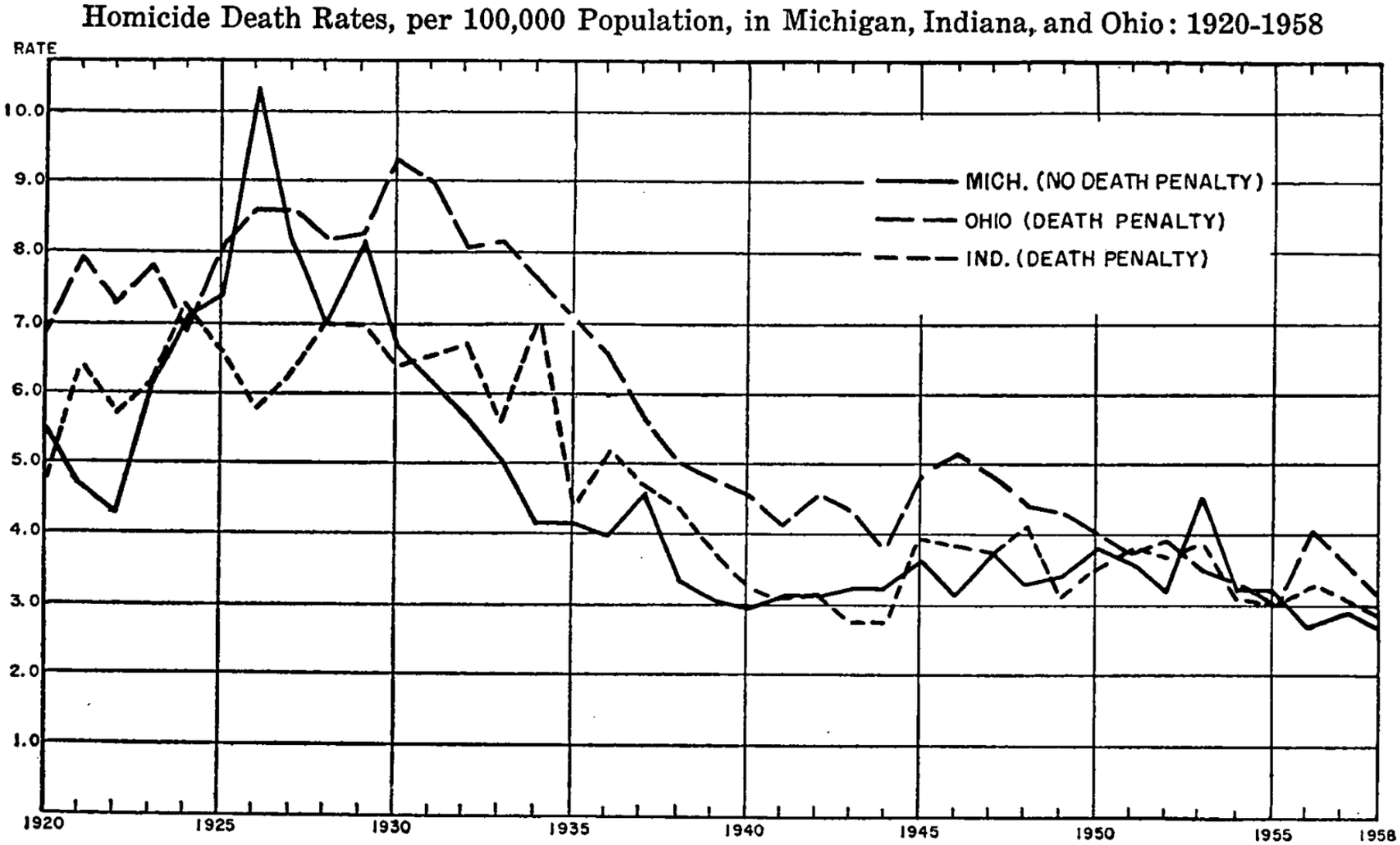
Lesson 9

Thursday 2/22/24

Review

Binary/Dichotomous Variables

New Example: we have 10 side-by-side state pairs where one state has the death penalty and the other one doesn't (Sellin, 1962, Capital Punishment. Federal Probation, 25:3-11 at page 8).



Review

Binary/Dichotomous Variables

- New example: we have 12 side-by-side state pairs where one state has the death penalty and the other one doesn't.
- First batch of data: in 8 comparisons, the death penalty state had a higher homicide rate than the non-death-penalty state.
- Second batch of data: in the other 4 states, the death-penalty state had a lower homicide rate (perhaps along the lines of what deterrence theory would predict).
- Now, suppose we give each of the 4 comparisons that are consistent with what deterrence theory would predict a code of 1 and the other 8 comparisons get a code of 0. What proportion of the comparisons support deterrence theory?

$$p(\text{pro-Deterrence}) = \frac{\text{Sum of Scores}}{\text{\# of Scores}} = \frac{1+1+1+1+0+0+0+0+0+0+0+0}{12} \approx 0.333$$

Review

Base and Comparison Years

- The crude homicide death rate was 5.5/100k population in 2015 and 6.0/100k population in 2016 (data are here).

- One way to compare these 2 numbers is to say that the homicide rate increased by 0.5/100k population from 2015 to 2016.

- The reason the first year (2015) is the base year is simply because it's easier to interpret.

- Consider the alternative: We could say that the homicide rate decreased when we go from 2016 (base year) to 2015 (comparison year). This seems like an odd statement.

- So, researchers usually consider the base year to be the first year and the comparison year to be the second year.

- If it's easier for you to remember, researchers usually subtract the Time 1 measurement from the Time 2 measurement (i.e., Time 2 - Time 1) to document change.

Review

Percent Change v. Absolute Change

- So, we've established that the U.S. homicide rate increased by 0.5/100K population from 2015 to 2016.

- The next problem is that this statement is hard to interpret without knowing what the scale of measurement is.

- The percent change statistic provides a sense of scale to the comparison.

- For our example, we can say the absolute change was 0.5 homicides/100k population or we could say that homicides increased by ((6.0-5.5)/5.5) x 100 = 9.1%.

- Let's work another example: using the same CDC homicide table as above (linked here), notice that in 1990 the homicide rate was 9.9/100K population while in 2010 it was 5.3/100k population. The change was 5.3 - 9.9 = -4.6

- Is 4.6 a big number or a small number?

Review

Percent Change v. Absolute Change

- Is 4.6 a big number or a small number?
- Well, when you consider that the starting point was 9.9, it seems like a pretty big number. Without knowing the starting point it is hard to tell, isn't it?
- We can express the relative magnitude of the change by calculating the percent change statistic:

$$\text{Percent Change} = \frac{\text{Comparison Year} - \text{Base Year}}{\text{Base Year}} \times 100$$

OR

$$\text{Percent Change} = \frac{\text{Time 2 Measurement} - \text{Time 1 Measurement}}{\text{Time 1 Measurement}} \times 100$$

Review

Percent Change v. Absolute Change

- Is 4.6 a big number or a small number?
- Well, when you consider that the starting point was 9.9, it seems like a pretty big number. Without knowing the starting point it is hard to tell, isn't it?
- We can express the relative magnitude of the change by calculating the percent change statistic:

$$\text{Percent Change} = \frac{5.3 - 9.9}{9.9} \times 100 = -46.5\%$$

- Holy Cow! - that's almost a 50% drop, which is huge.
- Notice, the word "drop" corresponds to the negative sign of the percent change statistic.

Example from U.S. Bureau of Labor Statistics

Calculating percent changes

One-month percent changes

Movements of an index from one month to another are usually expressed as percent changes rather than as changes in the index value, because index value changes are affected by the level of the index in relation to its base period, while percent changes are not. To find the percent change, you first subtract the earlier index value from the later one, then divide that difference by the earlier index value, and finally multiply the result by 100. Table 1 shows an example of a CPI one-month change between November 2021 and December 2021 using the CPI-U U.S. city average series for **all items**, not seasonally adjusted. You can find full historical data for this series in our online database.

Table 1, 1-month percent change in CPI-U U.S. city average series for all items

Quantity	Representation	Value
Index value, December 2021		278.802
Index value, November 2021		277.948
Difference in index values	278.802 - 277.948	0.854
Ratio of difference to earlier index value	0.854 / 277.948	0.003
Ratio multiplied by 100	.003 * 100	0.3

Practice Question 6

Age at Release from Prison	N =	# Rearrested
Age ≤ 24	62,700	47,000
Age 25-39	206,000	141,500
Age ≥ 40	140,600	82,400
Total	409,300	270,900

What is the recidivism rate for the 2 oldest age groups combined?

a. 0.472 or 47.2%
b. 0.585 or 58.5%
c. 0.646 or 64.6%
d. 0.821 or 82.1%

$$\text{RR} = \frac{\text{\# of Recidivists}}{\text{\# of People at Risk}} = \frac{141,500 + 82,400}{206,000 + 140,600} = 0.646$$

Practice Question 7

Consider the following 2 cities: Sunset Beach which had 85,000 residents and Crescent Beach which had 107,000 residents (according to last year's Census). Next, we learn from the 2 cities' police departments that Crescent Beach had 103 robberies last year while Sunset Beach had 95 robberies. Which city had the higher robbery rate? What was the robbery rate for the higher of the 2 cities?

a. Sunset Beach / 111.8
b. Crescent Beach / 87.3
c. Sunset Beach / 104.2
d. Crescent Beach / 78.5

$$\text{Sunset Beach Robberies per 100k Population} = \frac{95}{85,000} \times 100,000 = 111.8$$

$$\text{Crescent Beach Robberies per 100k Population} = \frac{103}{107,000} \times 100,000 = 96.3$$

Practice Question 8

Age at Release from Prison	N =	# Rearrested
Age ≤ 24	62,700	47,000
Age 25-39	206,000	141,500
Age ≥ 40	140,600	82,400
Total	409,300	270,900

What is the unit of analysis in the BJS recidivism study above? (short-answer question).

Answer: the unit of analysis is each individual person who was released from prison.

Practice Question 9

Suppose we do a study where we examine America's 100 largest cities and we count the number of these cities where at least one person was killed by the police last year. Next, we calculate the proportion of the cities where at least one person was killed by the police. What is the unit of analysis for this study?

a. individual persons who were killed by the police
b. the police officers who killed citizens
c. each of the cities in the group of America's 100 largest cities
d. the states where the 100 largest cities are located

Practice Question 9 Solution

Suppose we do a study where we examine America's 100 largest cities and we count the number of these cities where at least one person was killed by the police last year. Next, we calculate the proportion of the cities where at least one person was killed by the police. What is the unit of analysis for this study?

a. individual persons who were killed by the police
b. the police officers who killed citizens
c. each of the cities in the group of America's 100 largest cities
d. the states where the 100 largest cities are located

Practice Question 10

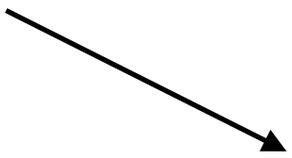
Suppose we do a study where we examine the records of a random sample of juvenile arrestees in a large American city. For each of the arrestees in our sample, we record the person's age at the time they were first arrested (i.e., age of onset). What is the level of measurement for this variable?

a. nominal
b. ordinal
c. interval
d. ratio

Practice Question 10 Solution

Suppose we do a study where we examine the records of a random sample of juvenile arrestees in a large American city. For each of the arrestees in our sample, we record the person's age at the time they were first arrested (i.e., age of onset). What is the level of measurement for this variable?

a. nominal
b. ordinal
c. interval
d. ratio



Continuous & True Zero = Ratio!

Practice Question 11

Using the sample of juvenile arrestees in PQ10, we also record for each juvenile a binary/dichotomous variable measuring whether the child's case was waived up to the adult criminal justice system. Suppose there were 837 kids in the sample and 94 were waived up to the adult system. What is the sample estimate of p(waived) for this sample?

a. 0.038
b. 0.079
c. 0.112
d. 0.187

$$p(\text{Waived}) = \frac{\#\text{ Waived}}{\#\text{ Who Could Have Been Waived}} = \frac{94}{837} = 0.112$$

Practice Question 12

We visited a drug treatment court each Monday for 3 months. Each day, we ask all of the defendants in drug court to participate in a survey of drug usage patterns. This sample would be an example of a:

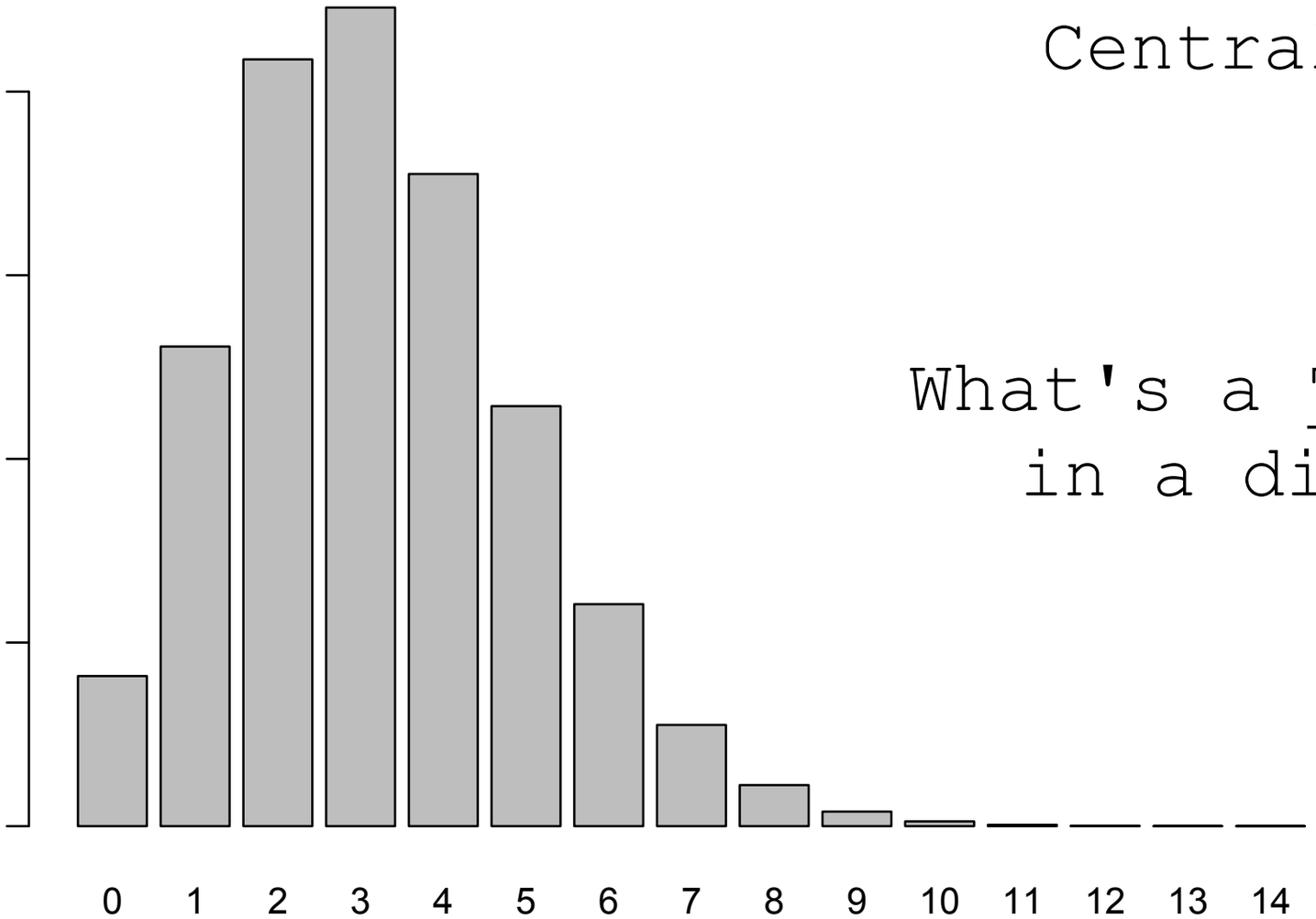
a. random sample
b. systematic sample
c. snowball sample
d. purposive sample

Practice Question 12 Solution

We visited a drug treatment court each Monday for 3 months. Each day, we ask all of the defendants in drug court to participate in a survey of drug usage patterns. This sample would be an example of a:

a. random sample
b. systematic sample
c. snowball sample
d. purposive sample

Chapter 4: Measures of Central Tendency



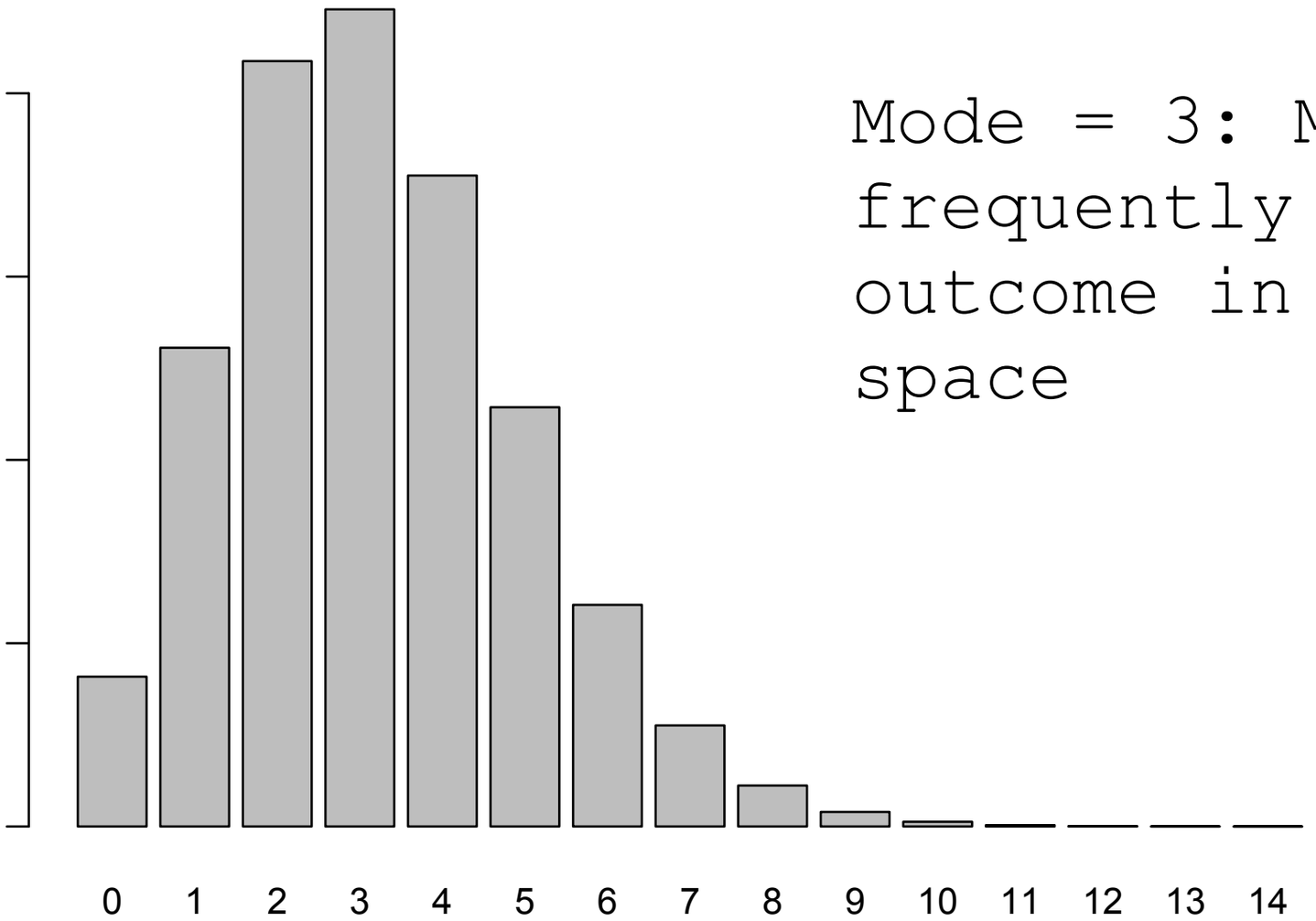
Central Tendency

=

What's a Typical Value
in a distribution

Example variable: people released from prison and followed for 5 years: # of rearrests during that time.

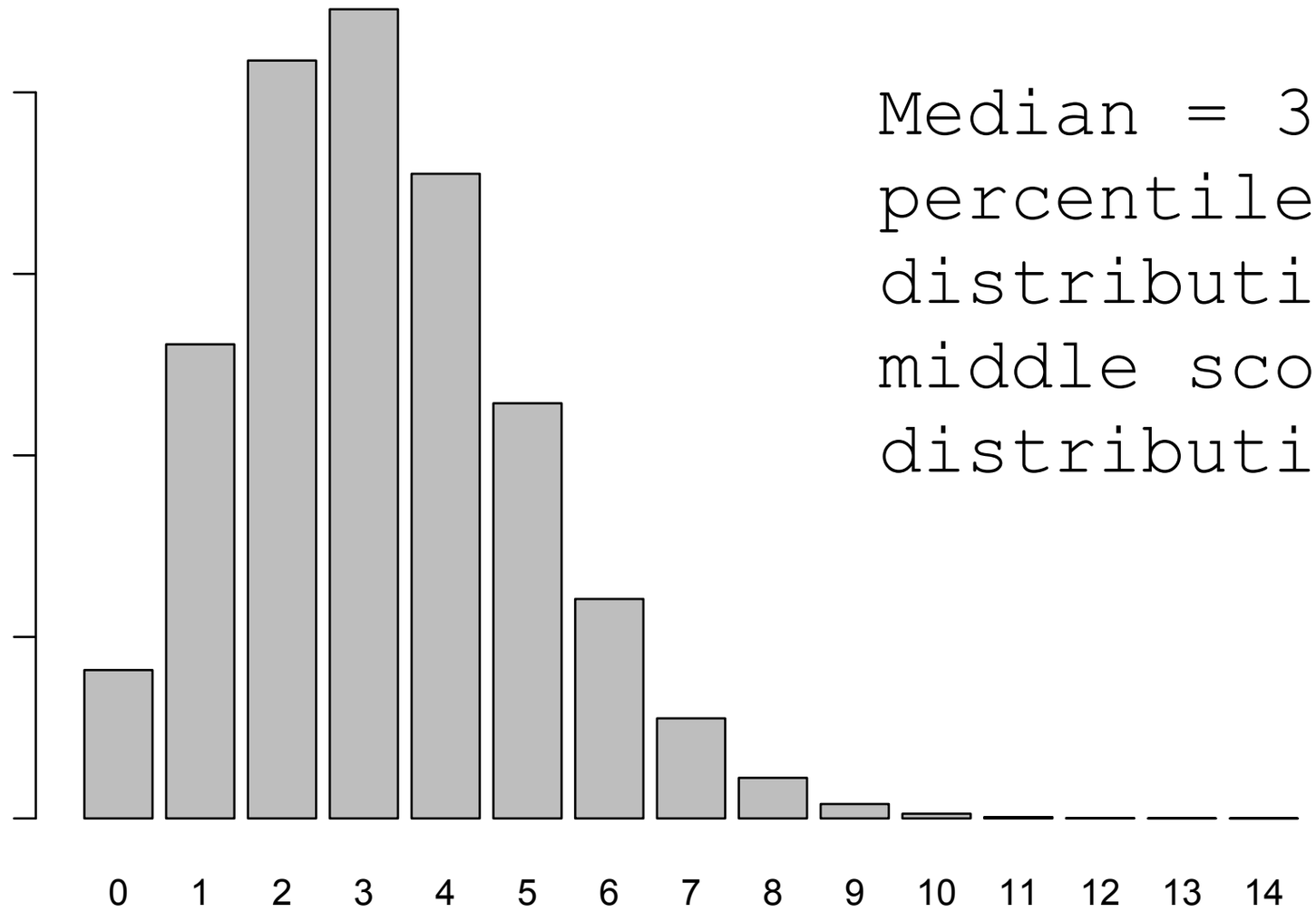
The Sample Mode



Mode = 3: Most frequently occurring outcome in the sample space

Example variable: people released from prison and followed for 5 years: # of rearrests during that time.

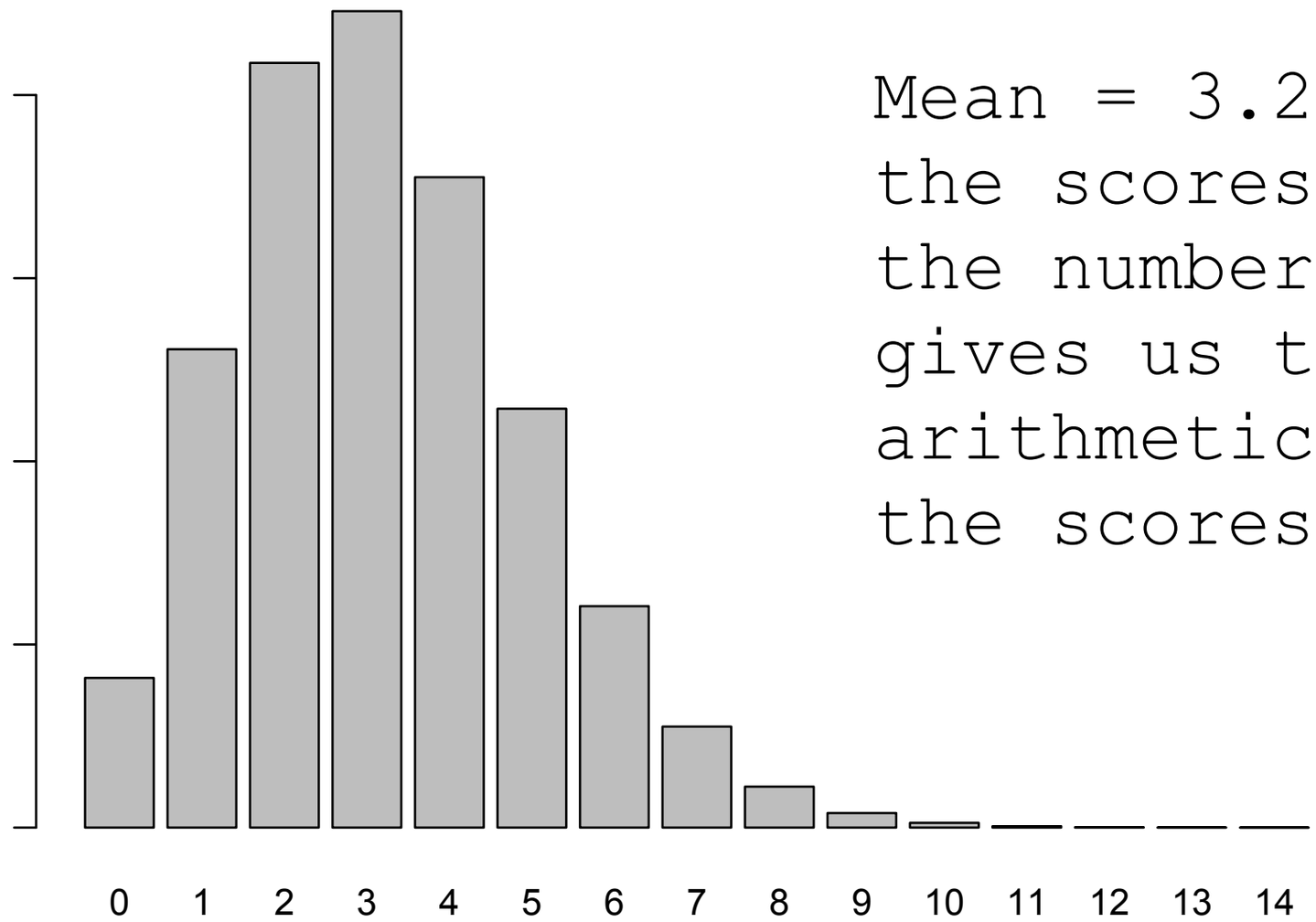
The Sample Median



Median = 3: The 50th percentile of the distribution; the middle score of the distribution.

Example variable: people released from prison and followed for 5 years: # of rearrests during that time.

The Sample Mean



Mean = 3.2: The sum of the scores divided by the number of scores gives us the arithmetic average of the scores.

Example variable: people released from prison and followed for 5 years: # of rearrests during that time.

Practice Question 13

Age at Release from Prison	N =	# Rearrested
Age ≤ 24	62,700	47,000
Age 25-39	206,000	141,500
Age ≥ 40	140,600	82,400
Total	409,300	270,900

What is the modal age at release from prison?

a. age ≤ 24
b. age 25-39
c. age ≥ 40
d. 206,000
e. 62,700

Practice Question 13 Solution

Age at Release from Prison	N =	# Rearrested
Age ≤ 24	62,700	47,000
Age 25-39	206,000	141,500
Age ≥ 40	140,600	82,400
Total	409,300	270,900

What is the modal age at release from prison?

a. age ≤ 24
b. age 25-39
c. age ≥ 40
d. 206,000
e. 62,700

Practice Question 14

This is a list of every person executed in Florida since the death penalty was reinstated in the mid-1970's. Based on the list, I created a dataset of the execution waiting times for the last 12 people who were executed.

Dataset: 26,16,37,30,33,32,23,34,20,24,26,33

What is the mode for this dataset?

a. 26
b. 32
c. 2
d. it is bimodal

Practice Question 14 Solution

This is a list of every person executed in Florida since the death penalty was reinstated in the mid-1970's. Based on the list, I created a dataset of the execution waiting times for the last 12 people who were executed.

Dataset: 26,16,37,30,33,32,23,34,20,24,26,33

What is the mode for this dataset?

a. 26
b. 32
c. 2
d. it is bimodal

Practice Question 15

Dataset: 26,16,37,30,33,32,23,34,20,24,26,33

What is the median for this dataset?

a. 27
b. 28
c. 30
d. 33

Step 1: Sort the observations in ascending order

Sorted Dataset: 16,20,23,24,26,26,30,32,33,33,34,37

Practice Question 15 (Cont'd)

Sorted Dataset: 16,20,23,24,26,26,30,32,33,33,34,37

Step 2: Count the # of observations → the answer is 12

Step 3: Is the # of observations even or odd? → 12 is an even number

Step 4a: If even, divide the number of observations by 2 → 12/2 = 6

Step 4b: If odd, find the middle observation; the score of that observation will be the median. → NA

Practice Question 15 (Cont'd)

Sorted Dataset: 16,20,23,24,26,26,30,32,33,33,34,37

Step 5: If the # of scores is even, calculate the average of the number that is in the position you calculated in step 4a AND the number that is immediately above that position.

→ Since 12/2 = 6, the 6th observation has the score of 26.

The 7th observation in the sorted list has the score of 30.

$$\text{Median} = \frac{26 + 30}{2} = \frac{56}{2} = 28$$

Practice Question 15 Solution

Dataset: 26,16,37,30,33,32,23,34,20,24,26,33

What is the median for this dataset?

a. 27
b. 28
c. 30
d. 33