Table 2.13 & Figure 2.9

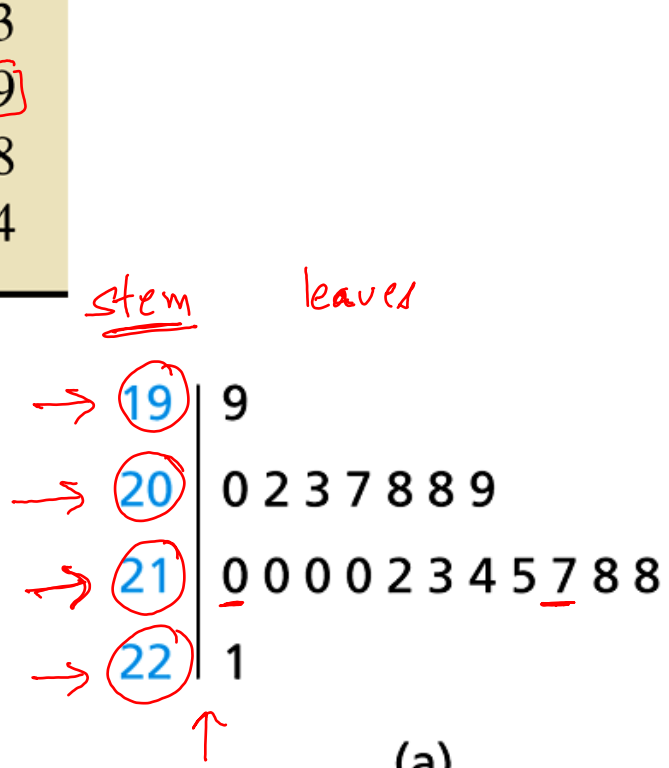
Cholesterol levels
for 20 high-level patients

210	209	212	208
217	207	210	203
208	210	210	199
215	221	213	218
202	218	200	214

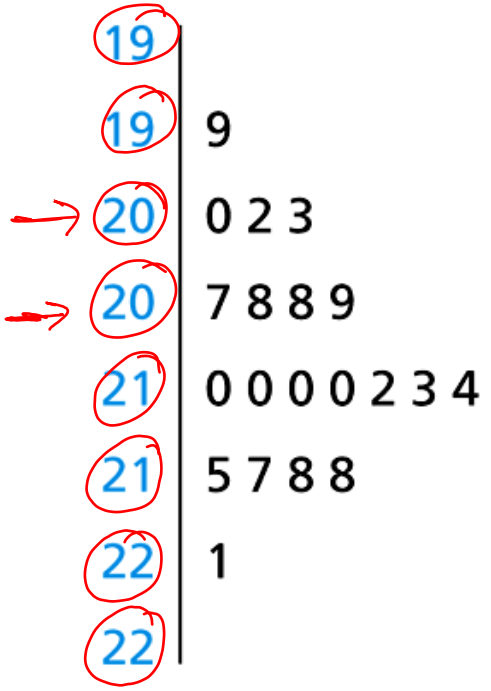
MIN 199

MAX 221

Stem-and-leaf diagram for cholesterol levels:
(a) one line per stem; (b) two lines per stem



(a)



(b)

KEY: 20|8

Read it as = 208 (leaf, stem)

Ex 65,000 70,000, 85,000 - - -

65, 70, 85, - -

65 stem leaf KEY: 6|5 read it as 65,000

Definition 2.12

Distribution of a Data Set

The **distribution of a data set** is a table, graph, or formula that provides the values of the observations and how often they occur.

Ex 7.4 6.5, 3.5, 4.5, - - KEY: 6|5 read it as 6.5

stem leaf

X

Ex 10045 21561, 81613, 99999, - -

Min Max

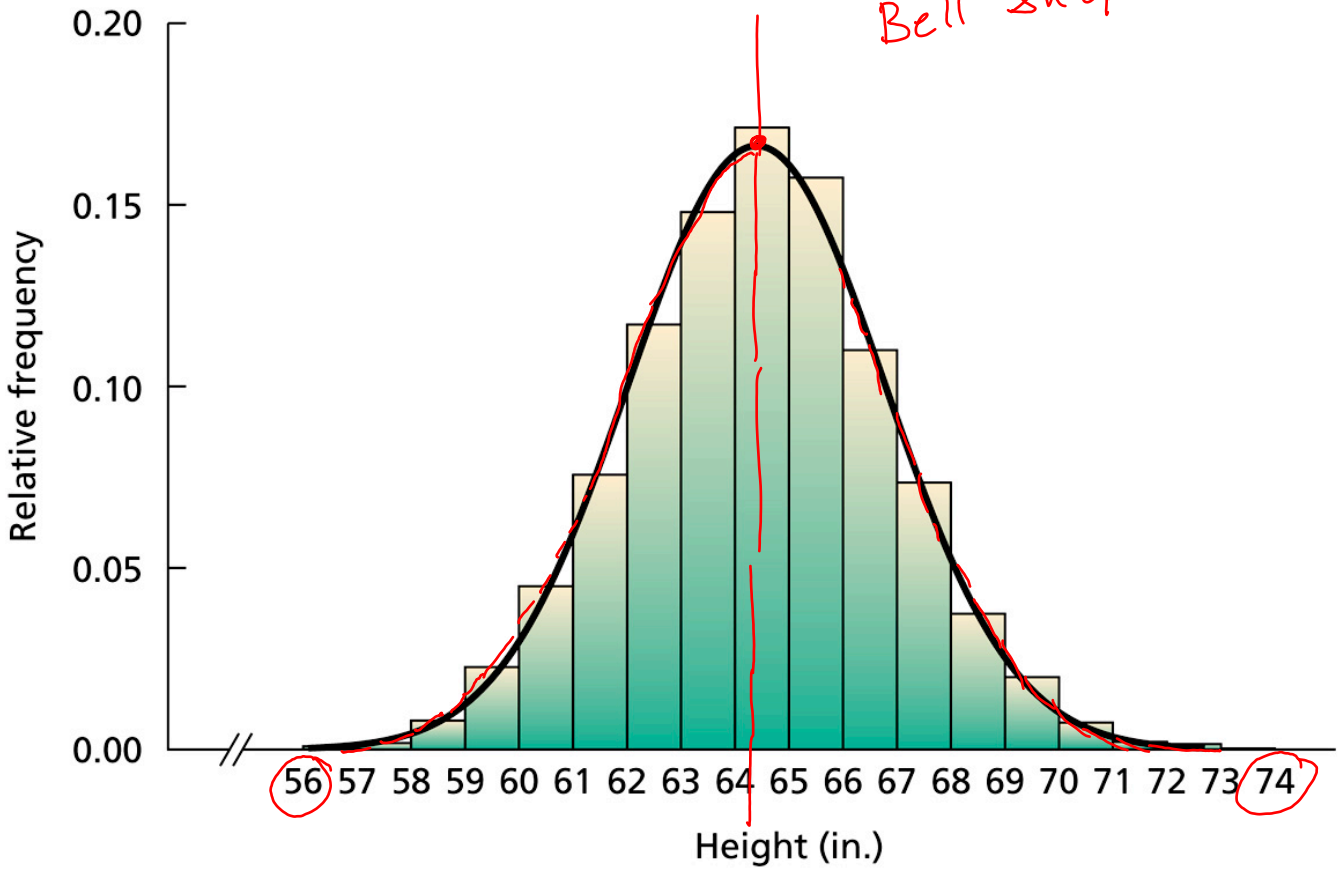
stem	leaf
1004	
9999	

10045 stem leaf

9999 | 9 stem leaf

Figure 2.10

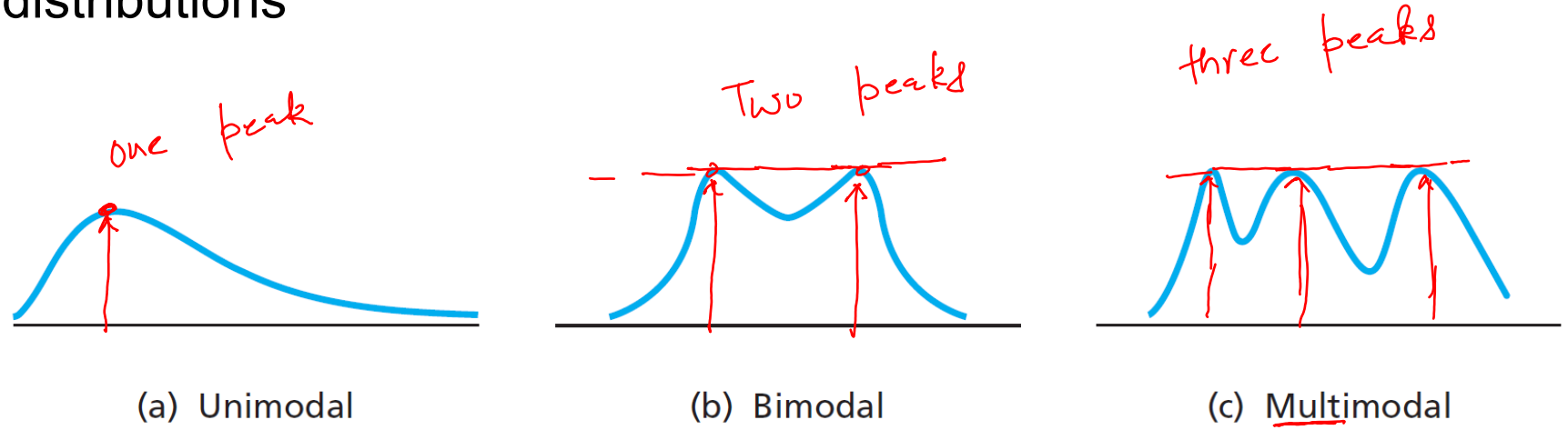
Relative-frequency histogram and approximating smooth curve for the distribution of heights



① MODALITY

Figure 2.11

Examples of (a) unimodal, (b) bimodal, and (c) multimodal distributions



(a) Unimodal (b) Bimodal (c) Multimodal

three or more peaks

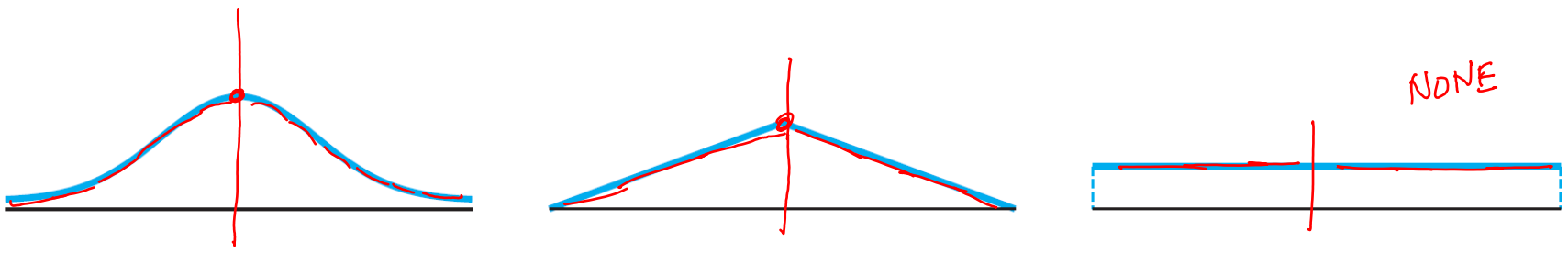
otherwise,

NONE .

(2) SYMMETRY

Figure 2.12

Examples of symmetric distributions: (a) bell shaped, (b) triangular, and (c) uniform

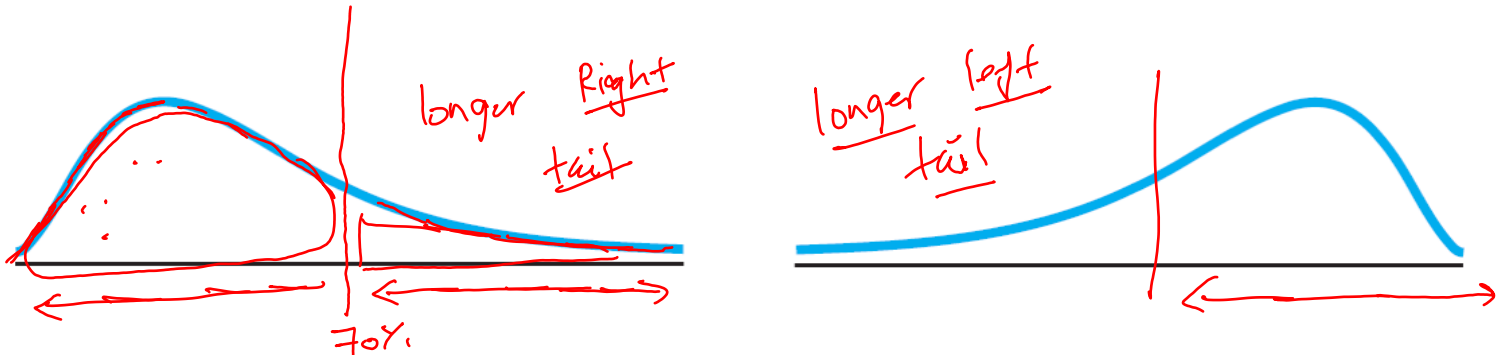


(a) Bell shaped (b) Triangular (c) Uniform (or rectangular)

(3) SKEWNESS

Figure 2.13

Generic skewed distributions: (a) right skewed (b) left skewed

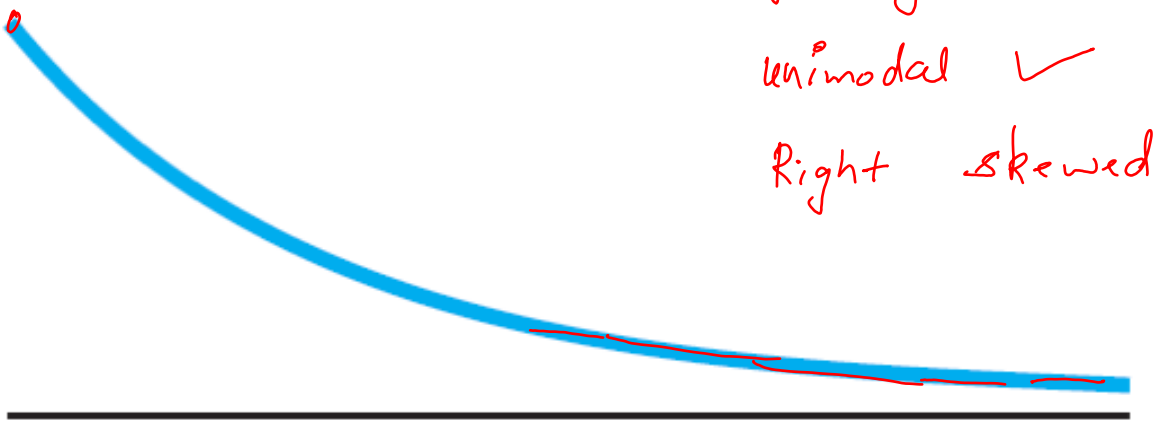


(a) Right skewed

(b) Left skewed

Figure 2.14

Reverse-J-shaped distribution



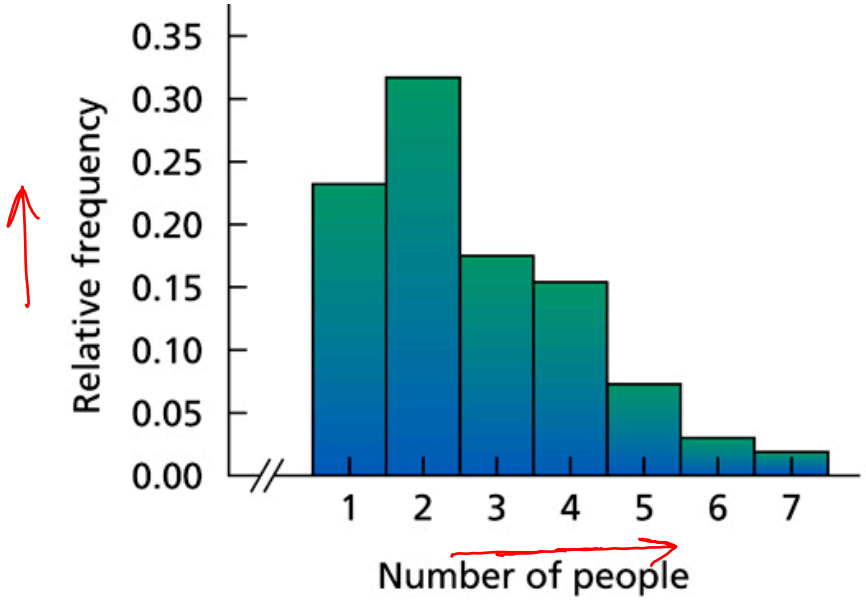
Symmetry X

unimodal ✓

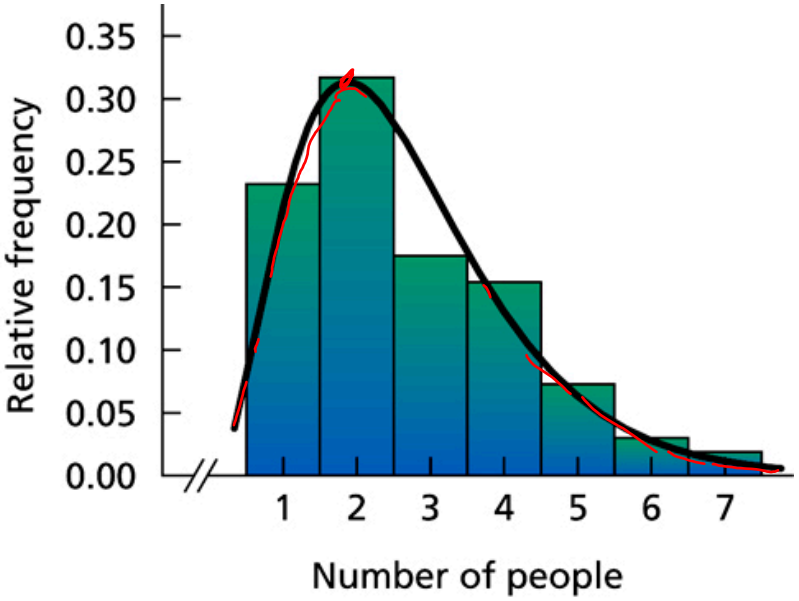
Right skewed ✓

Figure 2.15

Relative-frequency histogram for household size



(a)



(b)

Symmetry X

unimodal ✓

Right skewed ✓

Definition 2.13

Population and Sample Data

Population data: The values of a variable for the entire population.

Sample data: The values of a variable for a sample of the population.

Definition 2.14

Population and Sample Distributions; Distribution of a Variable

The distribution of population data is called the **population distribution,** or the **distribution of the variable.**

The distribution of sample data is called a **sample distribution.**

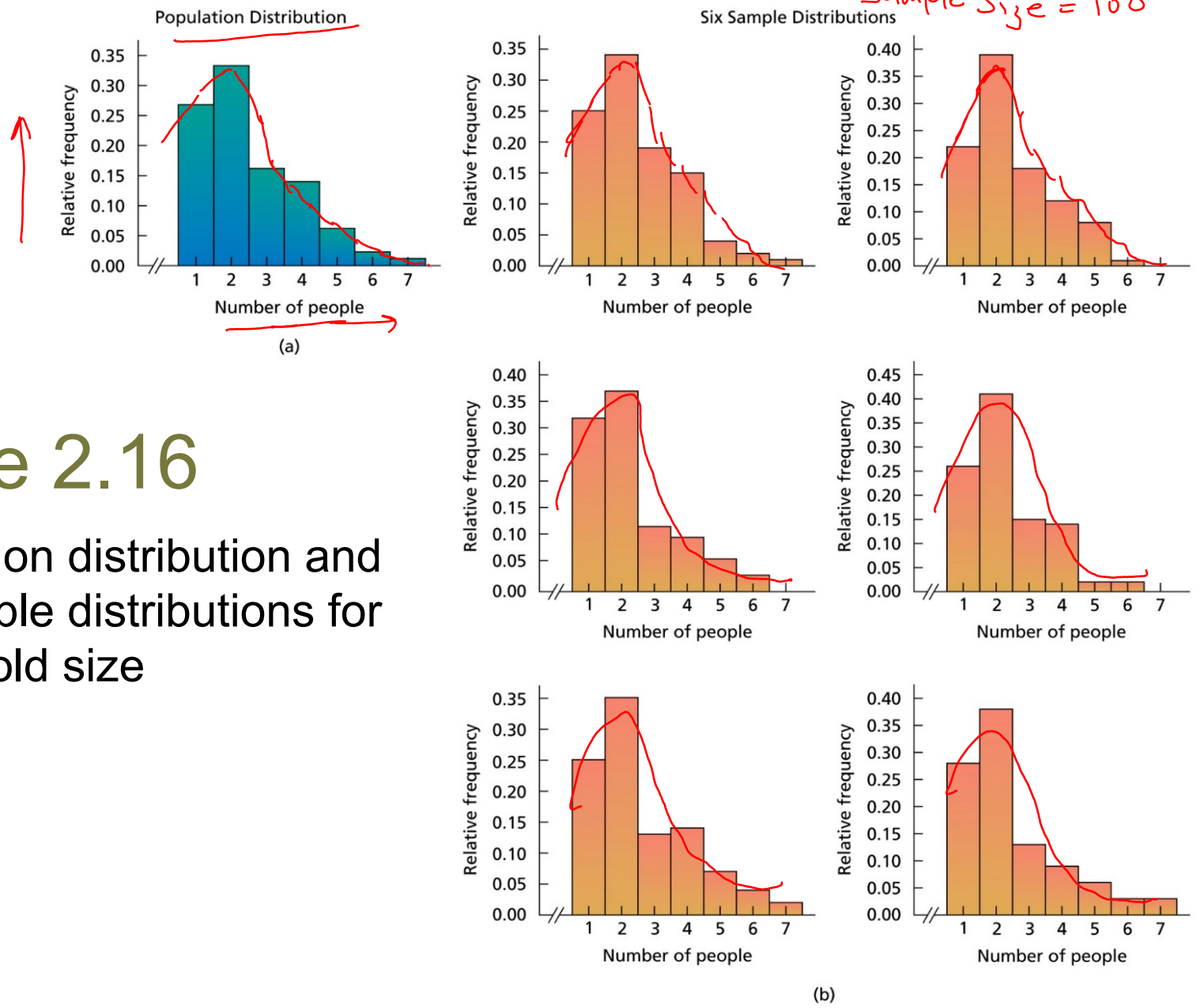


Figure 2.16

Population distribution and six sample distributions for household size

Key Fact 2.1

Population and Sample Distributions

For a simple random sample, the sample distribution approximates the population distribution (i.e., the distribution of the variable under consideration). The larger the sample size, the better the approximation tends to be.

Misleading Graphs

Figure 2.17

Unemployment rates: (a) truncated graph; (b) nontruncated graph

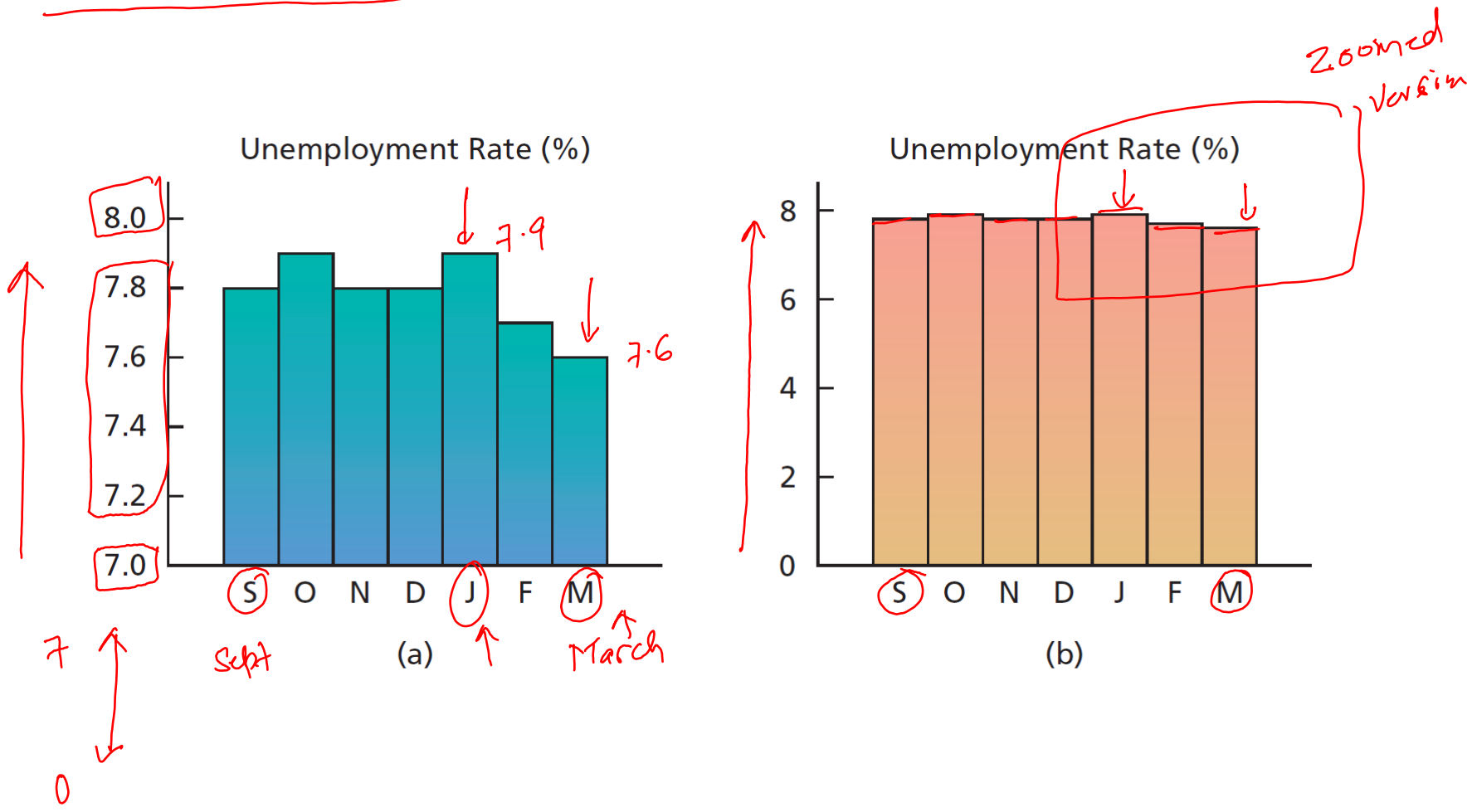
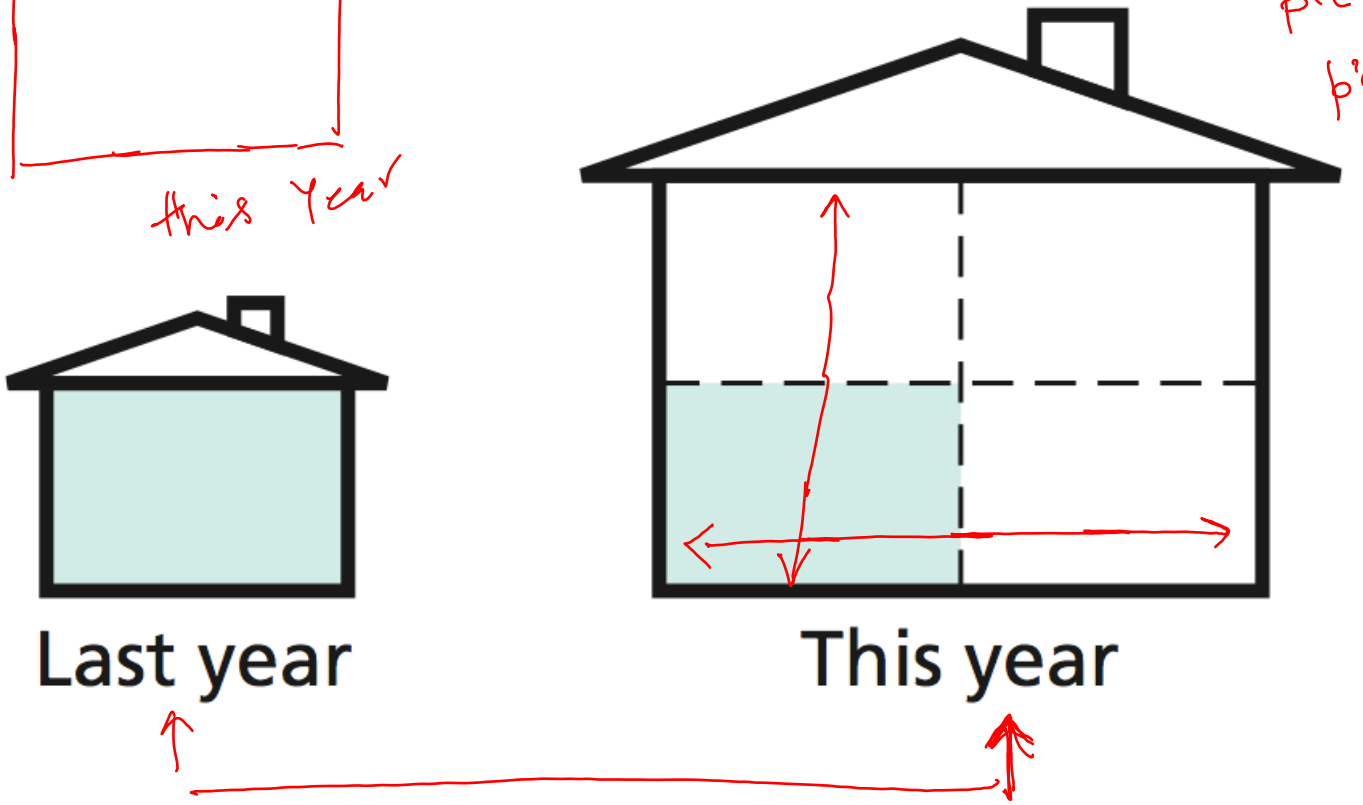


Figure 2.19

Improper scaling: Number of homes this year will be double last year, so the developer doubled the width and height, which makes it look like four times the number of homes will be built.



Chapter 3

Descriptive Measures

Section 3.1

Measures of Center

Definition 3.1

Mean of a Data Set

The **mean** of a data set is the sum of the observations divided by the number of observations.

Definition 3.2

Median of a Data Set

Arrange the data in increasing order.

- If the number of observations is odd, then the **median** is the observation exactly in the middle of the ordered list.
- If the number of observations is even, then the **median** is the mean of the two middle observations in the ordered list.

In both cases, if we let n denote the number of observations, then the median is at position $(n + 1) / 2$ in the ordered list.

Definition 3.3

Mode of a Data Set

Find the frequency of each value in the data set.

- If no value occurs more than once, then the data set has *no mode.*
- Otherwise, any value that occurs with the greatest frequency is a **mode** of the data set.

Tables 3.1, 3.2 & 3.4

Data Set I

$300	300	300	940	300
300	400	300	400	
450	800	450	1050	

Data Set II

$300	300	940	450	400
400	300	300	1050	300

Means, medians, and modes of salaries in Data Set I and Data Set II

Measure of center	Definition	Data Set I	Data Set II
Mean	$\frac{\text{Sum of observations}}{\text{Number of observations}}$	$483.85	$474.00
Median	Middle value in ordered list	$400.00	$350.00
Mode	Most frequent value	$300.00	$300.00