Introductory STATISTICS

9TH EDITION

Neil WEISS

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.

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

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

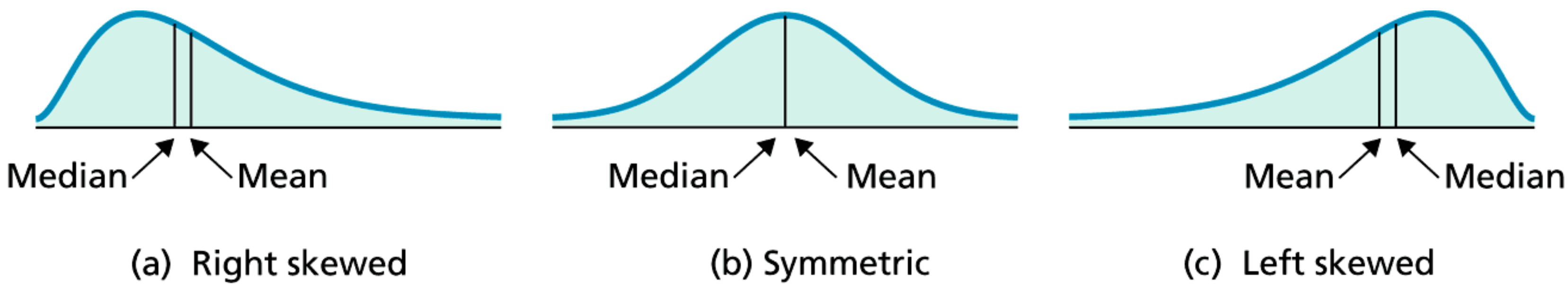
Mode of a Data Set

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.

Figure 3.1

Relative positions of the mean and median for (a) right-skewed, (b) symmetric, and (c) left-skewed distributions



Definition 3.4

Sample Mean

For a variable x, the mean of the observations for a sample is called a **sample mean** and is denoted $\bar{\mathbf{x}}$. Symbolically,

$$\bar{x} = \frac{\Sigma x_i}{n},$$

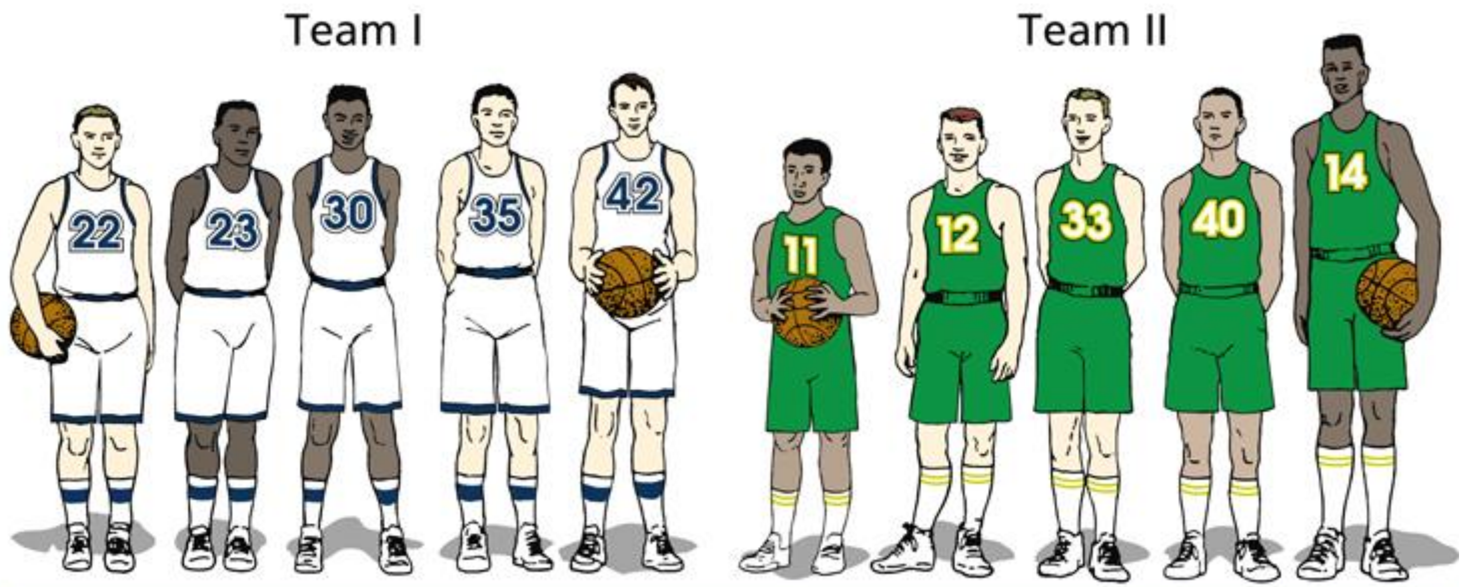
where n is the sample size.

Section 3.2

Measures of Variation

Figure 3.2

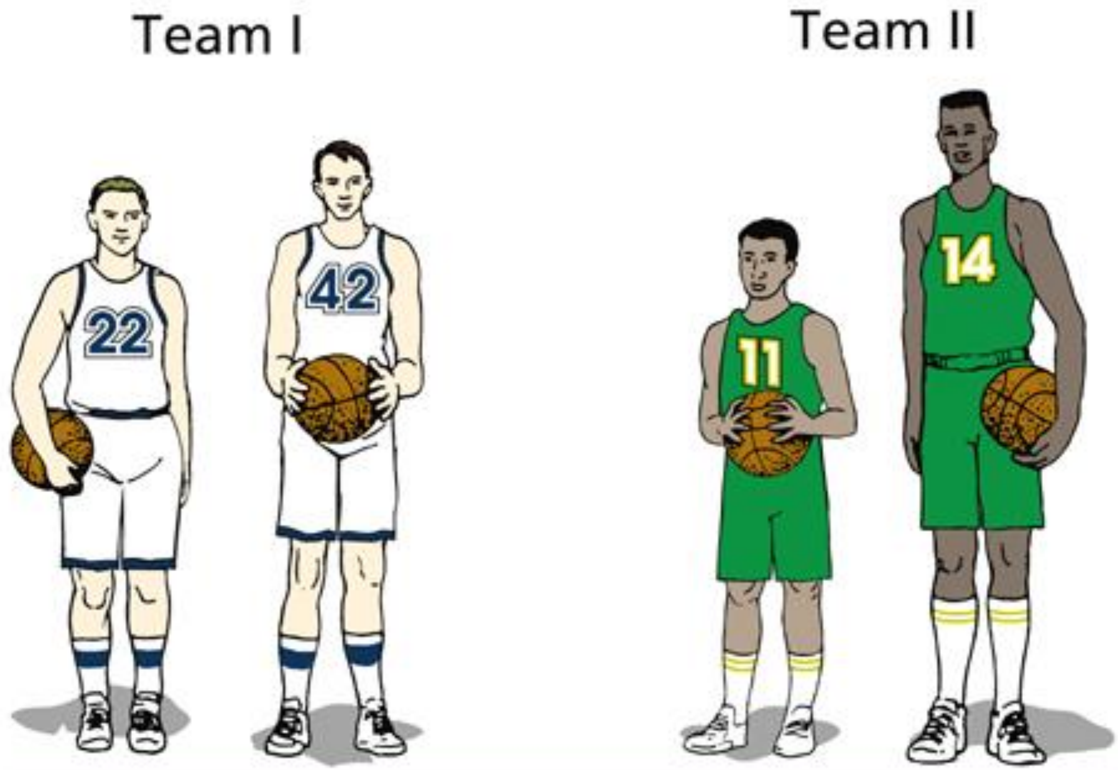
Five starting players on two basketball teams



	Team I					Team II				
Feet and inches	6'	6'1"	6'4"	6'4"	6'6"	5'7"	6'	6'4"	6'4"	7'
Inches	72	73	76	76	78	67	72	76	76	84

Figure 3.3

Shortest and tallest starting players on the teams



	Team I		Team II	
Feet and inches	6'	6'6"	5'7"	7'
Inches	72	78	67	84

Definition 3.5

Range of a Data Set

The **range** of a data set is given by the formula

$$\text{Range} = \text{Max} - \text{Min},$$

where Max and Min denote the maximum and minimum observations, respectively.

Definition 3.6

Sample Standard Deviation

For a variable x, the standard deviation of the observations for a sample is called a **sample standard deviation.** It is denoted $\mathbf{s_x}$ or, when no confusion will arise, simply **s**. We have

$$s = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1}},$$

where n is the sample size and $\bar{x}$ is the sample mean.

Formula 3.1

Computing Formula for a Sample Standard Deviation

A sample standard deviation can be computed using the formula

$$s = \sqrt{\frac{\Sigma x_i^2 - (\Sigma x_i)^2/n}{n-1}},$$

where n is the sample size.

Tables 3.10 & 3.11

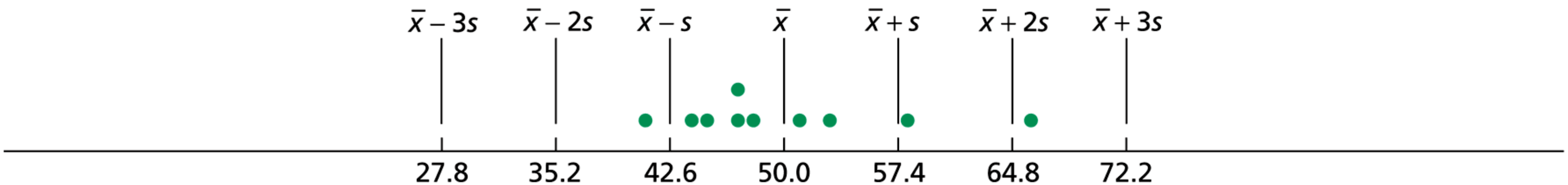
Data sets that have different variation

Data Set I	41	44	45	47	47	48	51	53	58	66
Data Set II	20	37	48	48	49	50	53	61	64	70

Means and standard deviations of the data sets in Table 3.10

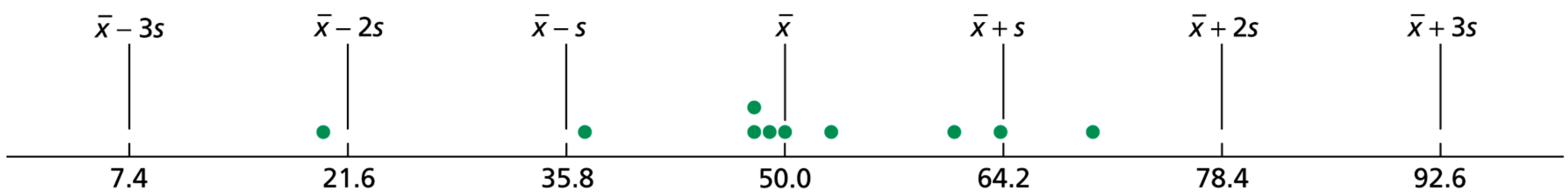
Data Set I	**Data Set II**
$\bar{x} = 50.0$ $s = 7.4$	$\bar{x} = 50.0$ $s = 14.2$

Figure 3.5



$\bar{x} - 3s$
$\bar{x} - 2s$
$\bar{x} - s$
$\bar{x}$
$\bar{x} + s$
$\bar{x} + 2s$
$\bar{x} + 3s$
27.8
35.2
42.6
50.0
57.4
64.8
72.2

Figure 3.6



$\bar{x} - 3s$
$\bar{x} - 2s$
$\bar{x} - s$
$\bar{x}$
$\bar{x} + s$
$\bar{x} + 2s$
$\bar{x} + 3s$
7.4
21.6
35.8
50.0
64.2
78.4
92.6

Section 3.3

The Five-Number Summary; Boxplots

Definition 3.7

Quartiles

Arrange the data in increasing order and determine the median.

- The **first quartile** is the median of the part of the entire data set that lies at or below the median of the entire data set.
- The **second quartile** is the median of the entire data set.
- The **third quartile** is the median of the part of the entire data set that lies at or above the median of the entire data set.

Definition 3.8

Interquartile Range

The **interquartile range,** or **IQR,** is the difference between the first and third quartiles; that is, IQR = $Q_3 - Q_1$.

Definition 3.9

Five-Number Summary

The **five-number summary** of a data set is
Min, Q_1, Q_2, Q_3, Max.

Definition 3.10

Lower and Upper Limits

The **lower limit** and **upper limit** of a data set are

$$\text{Lower limit} = Q_1 - 1.5 \cdot IQR;$$
$$\text{Upper limit} = Q_3 + 1.5 \cdot IQR.$$

Procedure 3.1

To Construct a Boxplot

Step 1 Determine the quartiles.

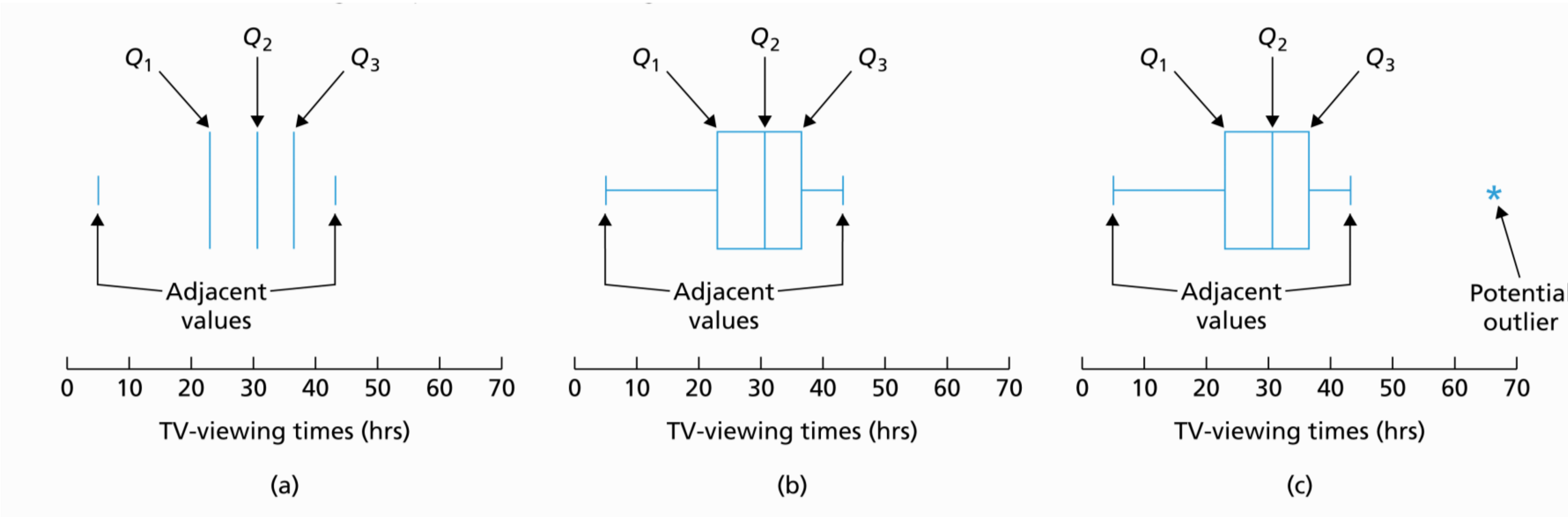
Step 2 Determine potential outliers and the adjacent values.

Step 3 Draw a horizontal axis on which the numbers obtained in Steps 1 and 2 can be located. Above this axis, mark the quartiles and the adjacent values with vertical lines.

Step 4 Connect the quartiles to make a box, and then connect the box to the adjacent values with lines.

Step 5 Plot each potential outlier with an asterisk.

Figure 3.9



Q_1
Q_2
Q_3
Adjacent values
0 10 20 30 40 50 60 70
TV-viewing times (hrs)
(a)
Q_1
Q_2
Q_3
Adjacent values
0 10 20 30 40 50 60 70
TV-viewing times (hrs)
(b)
Q_1
Q_2
Q_3
Adjacent values
Potential outlier
0 10 20 30 40 50 60 70
TV-viewing times (hrs)
(c)

Section 3.4

Descriptive Measures for Populations; Use of Samples

Definition 3.11

Population Mean (Mean of a Variable)

For a variable x, the mean of all possible observations for the entire population is called the **population mean** or **mean of the variable x.** It is denoted μ_x or, when no confusion will arise, simply μ. For a finite population,

$$\mu = \frac{\Sigma x_i}{N},$$

where N is the population size.

Definition 3.12

Population Standard Deviation (Standard Deviation of a Variable)

For a variable x, the standard deviation of all possible observations for the entire population is called the **population standard deviation** or **standard deviation of the variable x.** It is denoted σ_x or, when no confusion will arise, simply σ. For a finite population, the defining formula is

$$\sigma = \sqrt{\frac{\Sigma(x_i - \mu)^2}{N}},$$

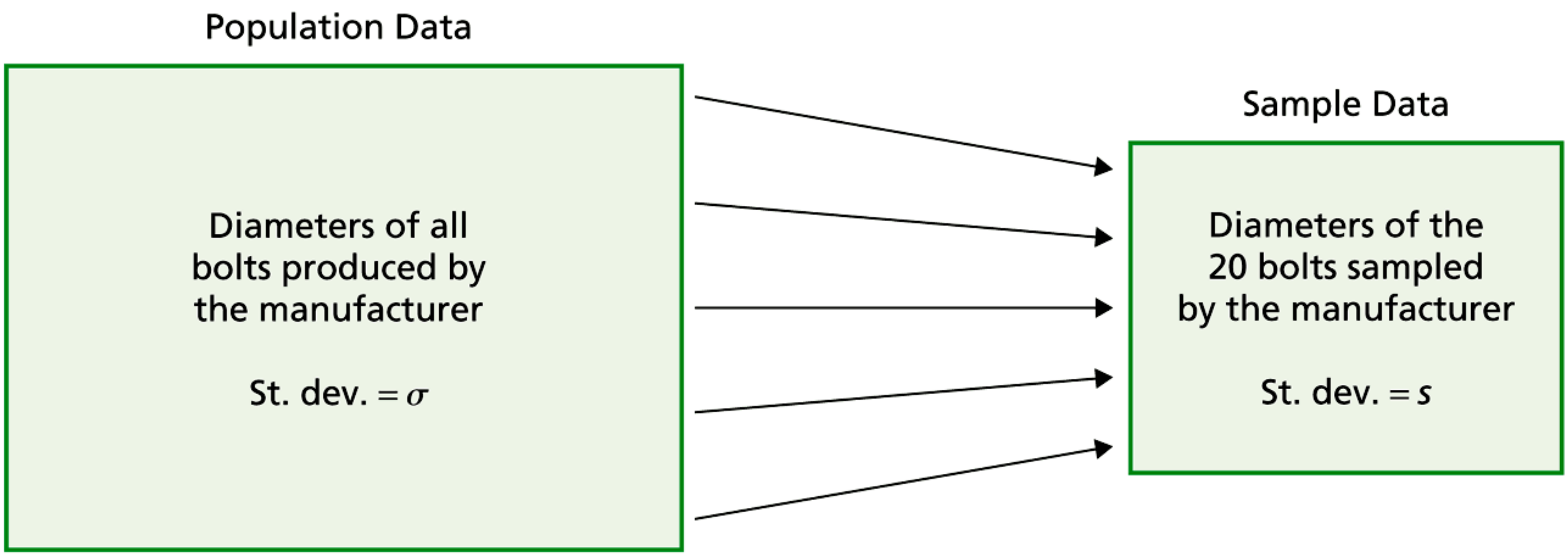
where N is the population size.

The population standard deviation can also be found from the computing formula

$$\sigma = \sqrt{\frac{\Sigma x_i^2}{N} - \mu^2}.$$

Figure 3.13 & Definition 3.13

Population and sample for bolt diameters



Parameter and Statistic

Parameter: A descriptive measure for a population.

Statistic: A descriptive measure for a sample.

Definition 3.14 & 3.15

Standardized Variable

For a variable x, the variable

$$z = \frac{x - \mu}{\sigma}$$

is called the **standardized version** of x or the **standardized variable** corresponding to the variable x.

z-Score

For an observed value of a variable x, the corresponding value of the standardized variable z is called the ***z-score*** of the observation. The term **standard score** is often used instead of *z-score.*