



Describing Data: Numerical Measures



Chapter 3

Learning Objectives

- LO3-1 Compute and interpret the mean, the median, and the mode
- LO3-2 Compute a weighted mean
- LO3-3 Compute and interpret the geometric mean
- LO3-4 Compute and interpret the range, variance, and standard deviation
- LO3-5 Explain and apply Chebyshev's theorem and the Empirical Rule
- LO3-6 Compute the mean and standard deviation of grouped data

Measures of Location

- ▶ A measure of location is a value used to describe the central tendency of a set of data
- ▶ Common measures of location
 - ▶ Mean
 - ▶ Median
 - ▶ Mode
- ▶ The arithmetic mean is the most widely reported measure of location

Population Mean

POPULATION MEAN

$$\mu = \frac{\sum x}{N}$$

[3-1]

where:

μ represents the population mean. It is the Greek lowercase letter “mu.”

N is the number of values in the population.

x represents any particular value.

Σ is the Greek capital letter “sigma” and indicates the operation of adding.

Σx is the sum of the x values in the population.

- ▶ A measurable characteristic of a population is a parameter

PARAMETER A characteristic of a population.

Example: Population Mean

There are 42 exits on I-75 through the state of Kentucky. Listed below are the distances between exits (in miles).

11	4	10	4	9	3	8	10	3	14	1	10	3	5
2	2	5	6	1	2	2	3	7	1	3	7	8	10
1	4	7	5	2	2	5	1	1	3	3	1	2	1

1. Why is this information a population?
2. What is the mean number of miles between exits?

Example: Population Mean Continued

There are 42 exits on I-75 through the state of Kentucky. Listed below are the distances between exits (in miles).

11	4	10	4	9	3	8	10	3	14	1	10	3	5
2	2	5	6	1	2	2	3	7	1	3	7	8	10
1	4	7	5	2	2	5	1	1	3	3	1	2	1

1. Why is this information a population?

This is a population because we are considering all of the exits in Kentucky.

2. What is the mean number of miles between exits?

$$\mu = \frac{\sum X}{N} = \frac{11 + 4 + 10 + \cdots + 1}{42} = \frac{192}{42} = 4.57$$

Sample Mean

SAMPLE MEAN

$$\bar{x} = \frac{\sum x}{n}$$

[3-2]

where:

$\bar{x}$ represents the sample mean. It is read “x bar.”

n is the number of values in the sample.

x represents any particular value.

Σ is the Greek capital letter “sigma” and indicates the operation of adding.

Σx is the sum of the x values in the sample.

- ▶ A measurable characteristic of a sample is a statistic

STATISTIC A characteristic of a sample.

Example: Sample Mean

Verizon is studying the number of minutes used by clients in a particular cell phone rate plan. A random sample of 12 clients showed the following number of minutes used last month.

90	77	94	89	119	112
91	110	92	100	113	83

What is the arithmetic mean number of minutes used?

$$\begin{aligned}\text{Sample mean} &= \frac{\text{Sum of all values in the sample}}{\text{Number of values in the sample}} \\ \bar{x} &= \frac{\sum x}{n} = \frac{90 + 77 + \cdots + 83}{12} = \frac{1,170}{12} = 97.5\end{aligned}$$

Properties of the Arithmetic Mean

- ▶ Interval or ratio scale of measurement is required
- ▶ All the data values are used in the calculation
- ▶ It is unique
- ▶ The sum of the deviations from the mean equals zero
- ▶ A weakness of the mean is that it is affected by extreme values

Exercise

- ▶ Compute the mean of the following population values:
7, 5, 7, 3, 7, 4.
- ▶ Compute the mean of the following sample values:
5, 9, 4, 10
 - ▶ Show that $\sum (X - \bar{X}) = 0$
- ▶ Suppose you go to the grocery store and spend \$61.85 for the purchase of 14 items. What is the mean price per item?

Exercise

- ▶ There are 10 salespeople employed by Mercedes. The number of new cars sold last month by the respective salespeople were: 15, 23, 4, 19, 18, 10, 10, 8, 28, 19.
 - ▶ Indicate whether it is a statistic or a parameter.
- ▶ EDL selected a random sample of 20 residential customers. Following are the amounts, to the nearest dollar, the customers were charged for electrical service last month:
 - ▶ 54 48 58 50 25 47 75 46 60 70
 - ▶ 67 68 39 35 56 66 33 62 65 67

The Median

MEDIAN The midpoint of the values after they have been ordered from the minimum to the maximum values.

Prices Ordered from Minimum to Maximum		Prices Ordered from Maximum to Minimum
\$ 60,000		\$ 275,000
65,000		80,000
70,000	← Median →	70,000
80,000		65,000
275,000		60,000

Characteristics of the Median

- ▶ The median is the value in the middle of a set of ordered data
- ▶ At least the ordinal scale of measurement is required
- ▶ It is not influenced by extreme values
- ▶ Fifty percent of the observations are larger than the median
- ▶ It is unique to a set of data

Finding the Median

- ▶ To find the median for an even numbered data set
- ▶ Sort the observations and calculate the average of the two middle values

The number of hours a sample of 10 adults used Facebook last month:

3 5 7 5 9 1 3 9 17 10

Arranging the data in ascending order gives:

1 3 3 5 5 7 9 9 10 17

Thus, the median is 6.

The Mode

MODE The value of the observation that occurs most frequently.

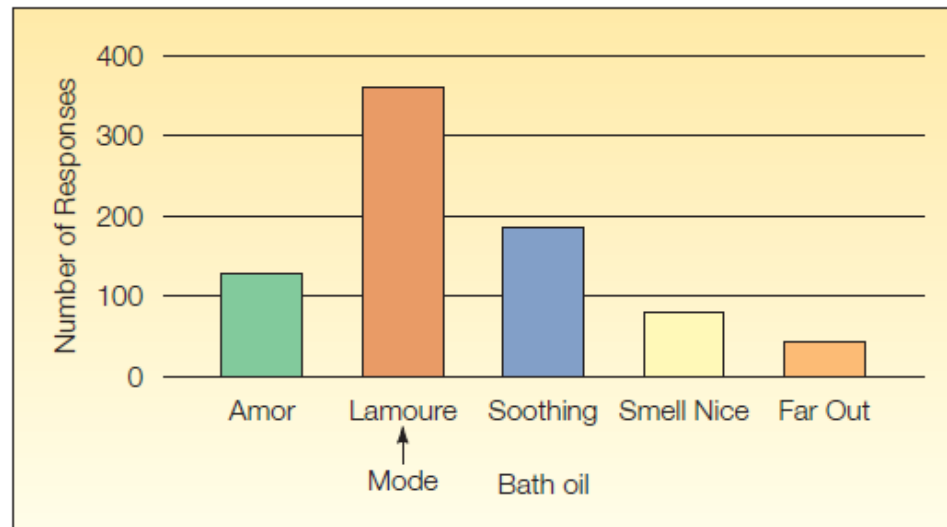


CHART 3-1 Number of Respondents Favoring Various Bath Oils

- ▶ Major Characteristics of the mode:
 - ▶ The mode can be found for nominal level data
 - ▶ A set of data can have more than one mode

The Mode

Recall the data regarding the distance in miles between exits on I-75 through Kentucky. The information is repeated below.

11	4	10	4	9	3	8	10	3	14	1	10	3	5
2	2	5	6	1	2	2	3	7	1	3	7	8	10
1	4	7	5	2	2	5	1	1	3	3	1	2	1

What is the modal distance?

The first step is to organize the distances into a frequency table. This will help us determine the distance that occurs most frequently.

Distance in Miles between Exits	Frequency
1	8
2	7
3	7
4	3
5	4
6	1
7	3
8	2
9	1
10	4
11	1
14	1
Total	42

The distance that occurs most often is one mile. This happens eight times—that is, there are eight exits that are one mile apart. So the modal distance between exits is one mile.

Which of the three measures of location (mean, median, or mode) best represents the central location of this data?



Relative Positions of Mean, Median, and Mode

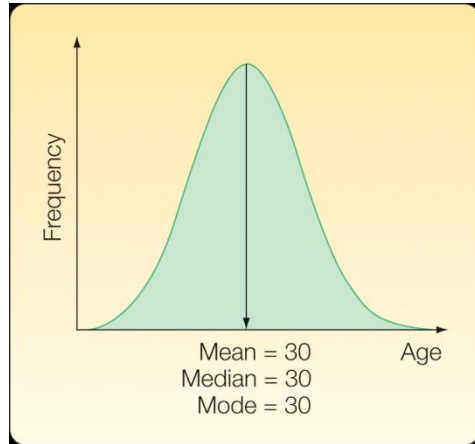


Chart 3-2 A Symmetric Distribution

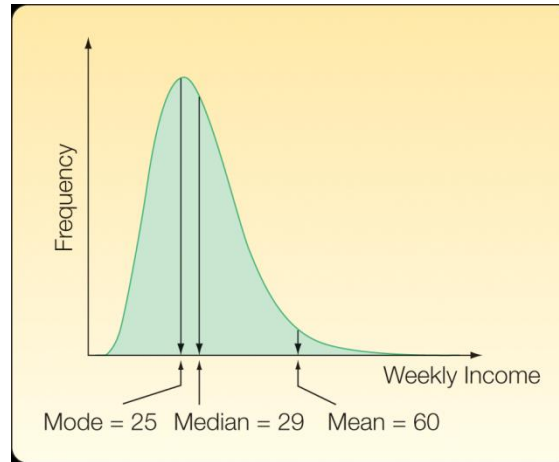


Chart 3-3 A Positively Skewed Distribution

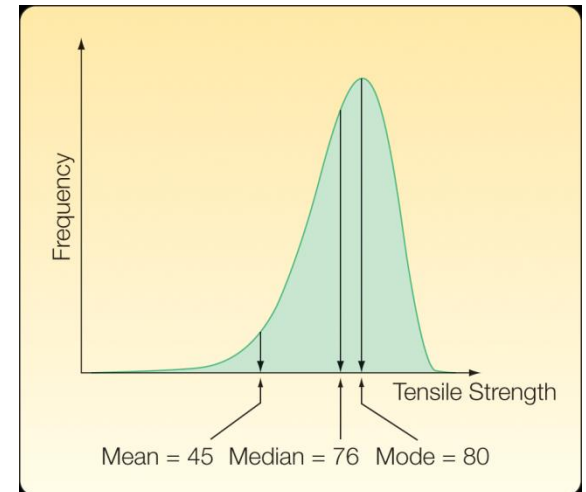


Chart 3-4 A Negatively Skewed Distribution

The Weighted Mean

- ▶ The weighted mean is found by multiplying each observation, x , by its corresponding weight, w

WEIGHTED MEAN

$$\bar{x}_w = \frac{w_1x_1 + w_2x_2 + w_3x_3 + \cdots + w_nx_n}{w_1 + w_2 + w_3 + \cdots + w_n}$$

[3-3]

- ▶ The Carter Construction Company pays its hourly employees \$16.50, \$19.00, or \$25.00 per hour. There are 26 hourly employees: 14 are paid at the \$16.50 rate, 10 at the \$19.00 rate, and 2 at the \$25.00 rate.
- ▶ What is the mean hourly rate paid for the 26 employees?

$$\bar{x}_w = \frac{14(\$16.50) + 10(\$19.00) + 2(\$25.00)}{14 + 10 + 2} = \frac{\$471.00}{26} = \$18.1154$$

The Geometric Mean

- ▶ The geometric mean is the n th root of the product of n positive values
- ▶ The formula for geometric mean is

GEOMETRIC MEAN

$$GM = \sqrt[n]{(x_1)(x_2) \cdots (x_n)}$$

[3-4]

- ▶ The geometric mean is also used to find the rate of change from one period to another

**RATE OF INCREASE
OVER TIME**

$$GM = \sqrt[n]{\frac{\text{Value at end of period}}{\text{Value at start of period}}} - 1$$

[3-5]

Exercise (Weighted Mean)

- ▶ In June, an investor purchased 300 shares of Oracle (an information technology company) stock at \$20 per share. In August, she purchased an additional 400 shares at \$25 per share. In November, she purchased an additional 400 shares, but the stock declined to \$23 per share. What is the weighted mean price per share?
- ▶ Andrews and Associates specialize in corporate law. They charge \$100 an hour for researching a case, \$75 an hour for consultations, and \$200 an hour for writing a brief. Last week one of the associates spent 10 hours consulting with her client, 10 hours researching the case, and 20 hours writing the brief. What was the weighted mean hourly charge for her legal services?

Exercise

17. What would you report as the modal value for a set of observations if there were a total of:

- a. 10 observations and no two values were the same?
- b. 6 observations and they were all the same?
- c. 6 observations and the values were 1, 2, 3, 3, 4, and 4?

18. The following is the number of oil changes for the last 7 days at the Jiffy Lube located at the corner of Elm Street and Pennsylvania Avenue.

determine the (a) mean, (b) median, and (c) mode

41	15	39	54	31	15	33
----	----	----	----	----	----	----

20. The following are the ages of the 10 people in the video arcade at the Southwyck Shopping Mall at 10 A.M. determine the (a) mean, (b) median, and (c) mode

12	8	17	6	11	14	8	17	10	8
----	---	----	---	----	----	---	----	----	---

Geometric mean Exercise (1/2)

The return on investment earned by Atkins Construction Company for four successive years was: 30 percent, 20 percent, −40 percent, and 200 percent. What is the geometric mean rate of return on investment?

The number 1.3 represents the 30 percent return on investment, which is the “original” investment of 1.0 plus the “return” of 0.3. The number 0.6 represents the loss of 40 percent, which is the original investment of 1.0 less the loss of 0.4. This calculation assumes the total return each period is reinvested or becomes the base for the next period. In other words, the base for the second period is 1.3 and the base for the third period is (1.3)(1.2) and so forth.

Then the geometric mean rate of return is 29.4 percent, found by

$$GM = \sqrt[n]{(X_1)(X_2) \cdots (X_n)} = \sqrt[4]{(1.3)(1.2)(0.6)(3.0)} = \sqrt[4]{2.808} = 1.294$$

The geometric mean is the fourth root of 2.808. So, the average rate of return (compound annual growth rate) is 29.4 percent.

Notice also that if you compute the arithmetic mean $[(30 + 20 - 40 + 200)/4 = 52.5]$, you would have a much larger number, which would overstate the true rate of return!

29. Listed below is the percent increase in sales for the MG Corporation over the last 5 years. Determine the geometric mean percent increase in sales over the period.

9.4	13.8	11.7	11.9	14.7
-----	------	------	------	------

Geometric mean Exercise (2/2)

During the decade of the 1990s, and into the 2000s, Las Vegas, Nevada, was the fastest-growing city in the United States. The population increased from 258,295 in 1990 to 607,876 in 2009. This is an increase of 349,581 people, or a 135.3 percent increase over the period. The population has more than doubled. What is the average *annual* percent increase?

There are 19 years between 1990 and 2009, so $n = 19$. Then formula (3–5) for the geometric mean as applied to this problem is:

$$GM = \sqrt[19]{\frac{\text{Value at end of period}}{\text{Value at start of period}}} - 1.0 = \sqrt[19]{\frac{607,876}{258,295}} - 1.0 = 1.0461 - 1.0 = .0461$$

The value of .0461 indicates that the average annual growth over the period was 4.61 percent. To put it another way, the population of Las Vegas increased at a rate of 4.61 percent per year from 1990 to 2009.

Why Study Dispersion?

- ▶ The dispersion is the variation or spread in a set of data
- ▶ The range is the difference between the maximum and minimum values in a set of data
- ▶ The formula for range is

RANGE

Range = Maximum value – Minimum value

[3–6]

- ▶ The major characteristics of the range are
 - ▶ Only two values are used in its calculation
 - ▶ It is influenced by extreme values
 - ▶ It is easy to compute and to understand

Population Variance

POPULATION VARIANCE

$$\sigma^2 = \frac{\sum(x - \mu)^2}{N}$$

[3-7]

where:

σ^2 is the population variance (σ is the lowercase Greek letter sigma). It is read as "sigma squared."

x is the value of a particular observation in the population.

μ is the arithmetic mean of the population.

N is the number of observations in the population.

- ▶ Major characteristics of the variance are:
 - ▶ All observations are used in the calculation
 - ▶ The units are somewhat difficult to work with, they are the original units squared

Sample Variance

SAMPLE VARIANCE

$$s^2 = \frac{\sum(x - \bar{x})^2}{n - 1}$$

[3-9]

where:

s^2 is the sample variance.

x is the value of each observation in the sample.

$\bar{x}$ is the mean of the sample.

n is the number of observations in the sample.

Example: Sample Variance

The hourly wages for a sample of 5 part-time employees at Home Depot are: \$12, \$20, \$16, \$18, and \$19.

The sample mean is \$17.

What is the sample variance?

Hourly Wage (x)	$x - \bar{x}$	$(x - \bar{x})^2$
\$12	-\$5	25
20	3	9
16	-1	1
18	1	1
19	2	4
<u>\$85</u>	<u>0</u>	<u>40</u>

$$s^2 = \frac{\Sigma(x - \bar{x})^2}{n - 1} = \frac{40}{5 - 1}$$

= 10 in dollars squared

Standard Deviation

SAMPLE STANDARD DEVIATION

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

[3-10]

where:

s is the sample standard deviation

x is the value of each observation in the sample

$\bar{x}$ is the mean of the sample

n is the number of observations in the sample



Standard Deviation

- ▶ The major characteristics of the standard deviation are:
 - ▶ It is in the same units as the original data
 - ▶ It is the square root of the average squared distance from the mean
 - ▶ It cannot be negative
 - ▶ It is the most widely used measure of dispersion

Interpretations and Uses of the Standard Deviation

CHEBYSHEV'S THEOREM For any set of observations (sample or population), the proportion of the values that lie within k standard deviations of the mean is at least $1 - 1/k^2$, where k is any value greater than 1.

The arithmetic mean biweekly amount contributed by the Dupree Paint employees to the company's profit-sharing plan is \$51.54, and the standard deviation is \$7.51. At least what percent of the contributions lie within plus 3.5 standard deviations and minus 3.5 standard deviations of the mean?

About 92%, found by

$$1 - \frac{1}{k^2} = 1 - \frac{1}{(3.5)^2} = 1 - \frac{1}{12.25} = 0.92$$

Interpretations and Uses of the Standard Deviation

THE EMPIRICAL RULE For a symmetrical, bell-shaped frequency distribution, approximately 68% of the observations will lie within plus and minus one standard deviation of the mean, about 95% of the observations will lie within plus or minus 2 standard deviations of the mean, and practically all (99.7%) will lie within 3 standard deviations of the mean.

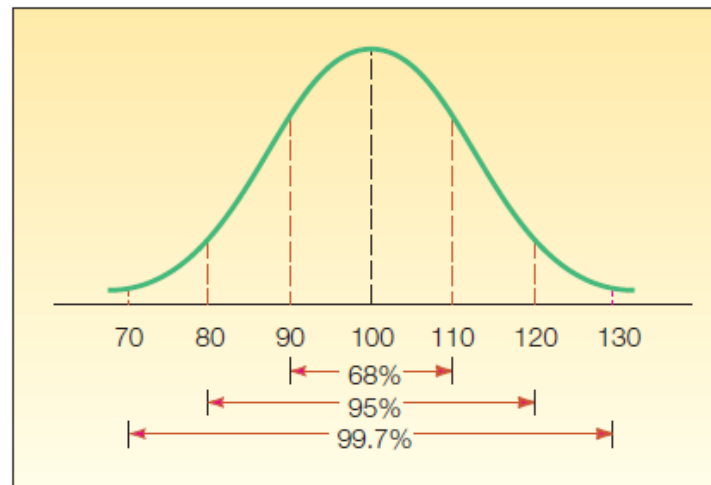


CHART 3-7 A Symmetrical, Bell-Shaped Curve Showing the Relationships between the Standard Deviation and the Percentage of Observations

Sample Mean of Grouped Data

ARITHMETIC MEAN OF GROUPED DATA

$$\bar{x} = \frac{\Sigma fM}{n}$$

[3-11]

where:

$\bar{x}$ is the designation for the sample mean.

M is the midpoint of each class.

f is the frequency in each class.

fM is the frequency in each class times the midpoint of the class.

ΣfM is the sum of these products.

n is the total number of frequencies.



Calculating the Standard Deviation of Grouped Data

- ▶ Applewood Auto Group Frequency Distribution
Compute the standard deviation of the vehicle profits.

Profit	Frequency (<i>f</i>)	Midpoint (<i>M</i>)	<i>fM</i>	$(M - \bar{x})$	$(M - \bar{x})^2$	$f(M - \bar{x})^2$
\$ 200 up to \$ 600	8	400	3,200	−1,451	2,105,401	16,843,208
600 up to 1,000	11	800	8,800	−1,051	1,104,601	12,150,611
1,000 up to 1,400	23	1,200	27,600	−651	423,801	9,747,423
1,400 up to 1,800	38	1,600	60,800	−251	63,001	2,394,038
1,800 up to 2,200	45	2,000	90,000	149	22,201	999,045
2,200 up to 2,600	32	2,400	76,800	549	301,401	9,644,832
2,600 up to 3,000	19	2,800	53,200	949	900,601	17,111,419
3,000 up to 3,400	4	3,200	12,800	1,349	1,819,801	7,279,204
Total	180		333,200			76,169,780

$$s = \sqrt{\frac{\sum f(M - \bar{x})^2}{n - 1}} = \sqrt{\frac{76,169,780}{180 - 1}} = 652.33$$

Ethics and Reporting Results

- ▶ Useful to know the advantages and disadvantages of mean, median, and mode as we report statistics and as we use statistics to make decisions
- ▶ Important to maintain an independent and principled point of view
- ▶ Statistical reporting requires objective and honest communication of any results

Exercise

- ▶ A large pharmaceutical company hires business administration graduates to sell its products. The company is growing rapidly and dedicates only one day of sales training for new salespeople. The company's goal for new salespeople is \$10,000 per month. The goal is based on the current mean sales for the entire company, which is \$10,000 per month. After reviewing the retention rates of new employees, the company finds that only 1 in 10 new employees stays longer than three months. Comment on using the current mean sales per month as a sales goal for new employees. Why do new employees leave the company?

Exercise

- ▶ For below calculate the (a) range, (b) arithmetic mean, (c) mean deviation, and (d) interpret the values.
 - ▶ Ex35. There were five customer service representatives on duty at the Electronic Store during last weekend's sale. The numbers of HDTVs these representatives sold are: 5, 8, 4, 10, and 3.
 - ▶ Ex39. Ten randomly selected young adults living in California rated the taste of a newly developed sushi pizza topped with tuna, rice, and kelp on a scale of 1 to 50, with 1 indicating they did not like the taste and 50 that they did. The ratings were:

34	39	40	46	33	31	34	14	15	45
----	----	----	----	----	----	----	----	----	----

- ▶ In a parallel study, 10 randomly selected young adults in Iowa rated the taste of the same pizza. The ratings were:

28	25	35	16	25	29	24	26	17	20
----	----	----	----	----	----	----	----	----	----

Exercise

Ex 43: The annual report of Dennis Industries cited these primary earnings per common share for the past 5 years: \$2.68, \$1.03, \$2.26, \$4.30, and \$3.58. If we assume these are population values, what is:

- a. The arithmetic mean primary earnings per share of common stock?
- b. The variance?

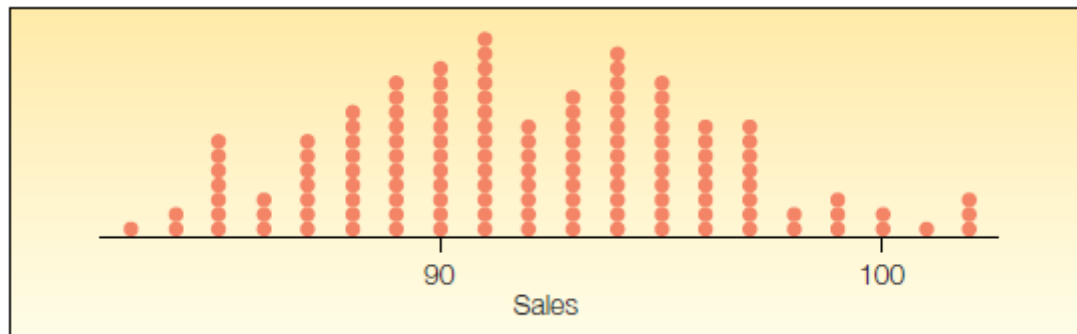
Ex 45: Plywood Inc. reported these returns on stockholder equity for the past 5 years: 4.3, 4.9, 7.2, 6.7, and 11.6. Consider these as population values.

- a. Compute the range, the arithmetic mean, the variance, and the standard deviation.

Ex 53: According to Chebyshev's theorem, at least what percent of any set of observations will be within 1.8 standard deviations of the mean?

Exercise

56. The following graph portrays the distribution of the number of Biggie-sized soft drinks sold at a nearby Wendy's for the last 141 days. The mean number of drinks sold per day is 91.9 and the standard deviation is 4.67.



If we use the Empirical Rule, sales will be between what two values on 68 percent of the days? Sales will be between what two values on 95 percent of the days?