EXAMPLE 9 **Using Chebyshev's Inequality**

Problem: Using the data from University A in Table 7.

(a) Determine the minimum percentage of students who have IQ scores within 3 standard deviations of the mean according to Chebyshev's Inequality.

(b) Determine the minimum percentage of students who have IQ scores between 67.8 and 132.2, according to Chebyshev's Inequality.

(c) Determine the actual percentage of students who have IQ scores between 67.8 and 132.2.

Approach

(a) We use Chebyshev's Inequality with $k = 3$.

(b) We have to determine the number of standard deviations 67.8 and 132.2 are from the mean of 100.0. We then substitute this value of k into Chebyshev's Inequality.

(c) We refer to Table 7 and count the number of observations between 67.8 and 132.2. We divide this result by 100, the number of observations in the data set.

Solution

(a) We use Chebyshev's Inequality with $k = 3$ and determine that at least $\left(1 - \dfrac{1}{3^2}\right)100\% = 88.9\%$ of all students have IQ scores within 3 standard deviations of the mean. Since the mean of the data set is 100.0 and the standard deviation is 16.1, at least 88.9% of the students have IQ scores between $\overline{x} - ks = 100.0 - 3(16.1) = 51.7$ and $\overline{x} + ks = 100 + 3(16.1) = 148.3$.

(b) Since 67.8 is exactly 2 standard deviations below the mean $[100 - 2(16.1) = 67.8]$ and 132.2 is exactly 2 standard deviations above the mean $[100 + 2(16.1) = 132.2]$, we use Chebyshev's Inequality with $k = 2$ to determine that at least $\left(1 - \dfrac{1}{2^2}\right)100\% = 75\%$ of all IQ scores lie between 67.8 and 132.2.

(c) Of the 100 IQ scores listed, 96 or 96% is between 67.8 and 132.2. Notice that Chebyshev's inequality provides a rather conservative result.

Historical Notes

Pafnuty Chebyshev was born on May 16, 1821, in Okatovo, Russia. In 1847, he began teaching mathematics at the University of St. Petersburg. Some of his more famous work was done on prime numbers. In particular, he discovered a way to determine the number of prime numbers less than or equal to a given number. Chebyshev also studied mechanics, including rotary motion. Chebyshev was elected a Fellow of the Royal Society in 1877. He died on November 26, 1894, in St. Petersburg.

Now Work Problem 39

Because the Empirical Rule requires that the distribution be bell shaped, while Chebyshev's Inequality applies to all distributions, the Empirical Rule provides results that are more precise.

3.2 ASSESS YOUR UNDERSTANDING

Concepts and Vocabulary

1. Would it be appropriate to say that a distribution with a standard deviation of 10 centimeters is more dispersed than a distribution with a standard deviation of 5 inches? Support your position.

2. What is meant by the phrase *degrees of freedom* as it pertains to the computation of the sample variance?

3. Are any of the measures of dispersion mentioned in this section resistant? Explain.

4. The sum of the deviations about the mean always equals _____.

5. What does it mean when a statistic is biased?

6. What makes the range less desirable than the standard deviation as a measure of dispersion?

7. In one of Sullivan's statistics sections, the standard deviation of the heights of all students was 3.9 inches. The standard deviation of the heights of males was 3.4 inches and the standard deviation of females was 3.3 inches. Why is the standard deviation of the entire class more than the standard deviation of the males and females considered separately?

8. The standard deviation is used in conjunction with the _____ to numerically describe distributions that are bell shaped. The _____ measures the center of the distribution, while the standard deviation measures the _____ of the distribution.

9. *True or False:* When comparing two populations, the larger the standard deviation, the more dispersion the distribution has, provided that the variable of interest from the two populations has the same unit of measure.

10. *True or False:* Chebyshev's Inequality applies to all distributions regardless of shape, but the Empirical Rule holds only for distributions that are bell shaped.

Skill Building

In Problems 11–16, find the population variance and standard deviation or the sample variance and standard deviation as indicated.

11. Sample: 20, 13, 4, 8, 10

12. Sample: 83, 65, 91, 87, 84

13. Population: 3, 6, 10, 12, 14

14. Population: 1, 19, 25, 15, 12, 16, 28, 13, 6

15. Sample: 6, 52, 13, 49, 35, 25, 31, 29, 31, 29

16. Population: 4, 10, 12, 12, 13, 21

17. Crash Test Results The Insurance Institute for Highway Safety crashed the 2007 Audi A4 four times at 5 miles per hour. The costs of repair for each of the four crashes are as follows:

$976, $2038, $918, $1899

Compute the range, sample variance, and sample standard deviation cost of repair.

18. Cell Phone Use The following data represent the monthly cell phone bill for my wife's phone for six randomly selected months:

$35.34, $42.09, $39.43, $38.93, $43.39, $49.26

Compute the range, sample variance, and sample standard deviation phone bill.

19. Concrete Mix A certain type of concrete mix is designed to withstand 3,000 pounds per square inch (psi) of pressure. The strength of concrete is measured by pouring the mix into casting cylinders 6 inches in diameter and 12 inches tall. The concrete is allowed to set up for 28 days. The concrete's strength is then measured (in psi). The following data represent the strength of nine randomly selected casts:

3960, 4090, 3200, 3100, 2940, 3830, 4090, 4040, 3780

Compute the range, sample variance, and sample standard deviation for the strength of the concrete (in psi).

20. Flight Time The following data represent the flight time (in minutes) of a random sample of seven flights from Las Vegas, Nevada, to Newark, New Jersey, on Continental Airlines.
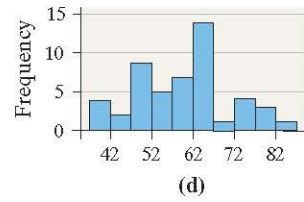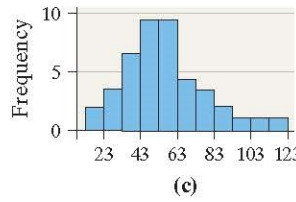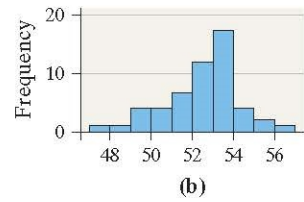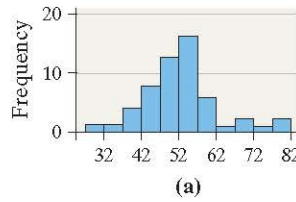
282, 270, 260, 266, 257, 260, 267

Compute the range, sample variance, and sample standard deviation of flight time.

21. Which histogram depicts a higher standard deviation? Justify your answer.

(a)　　　　　(b)
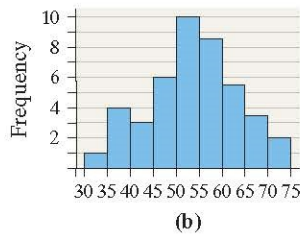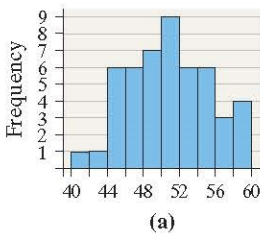
22. Match the histograms to the summary statistics given.

	Mean	Median	Standard Deviation
I	53	53	1.8
II	60	60	11
III	53	53	10
IV	53	53	22

(a)　　　　　(b)

(c)　　　　　(d)

Applying the Concepts

23. pH in Water The acidity or alkalinity of a solution is measured using pH. A pH less than 7 is acidic, while a pH greater than 7 is alkaline. The following data represent the pH in samples of bottled water and tap water.

(a) Which type of water has more dispersion in pH using the range as the measure of dispersion?

(b) Which type of water has more dispersion in pH using the standard deviation as the measure of dispersion?

Tap	7.64	7.45	7.47	7.50	7.68	7.69
	7.45	7.10	7.56	7.47	7.52	7.47
Bottled	5.15	5.09	5.26	5.20	5.02	5.23
	5.28	5.26	5.13	5.26	5.21	5.24

Source: Emily McCarney, student at Joliet Junior College

24. Reaction Time In an experiment conducted online at the University of Mississippi, study participants are asked to react to a stimulus. In one experiment, the participant must press a key upon seeing a blue screen. The time (in seconds) to press the key is measured. The same person is then asked to press a key upon seeing a red screen, again with the time to react measured. The results for six study participants are listed in the table.

(a) Which color has more dispersion using the range as the measure of dispersion?

(b) Which color has more dispersion using the standard deviation as the measure of dispersion?

Participant Number	Reaction Time to Blue	Reaction Time to Red
1	0.582	0.408
2	0.481	0.407
3	0.841	0.542
4	0.267	0.402
5	0.685	0.456
6	0.450	0.533

Source: PsychExperiments at the University of Mississippi (www.olemiss.edu/psychexps)

25. Pulse Rates The following data represent the pulse rates
NW (beats per minute) of nine students enrolled in a section of
Sullivan's course in Introductory Statistics. Treat the nine
students as a population.

Student	Pulse
Perpetual Bempah	76
Megan Brooks	60
Jeff Honeycutt	60
Clarice Jefferson	81
Crystal Kurtenbach	72
Janette Lantka	80
Kevin McCarthy	80
Tammy Ohm	68
Kathy Wojdyla	73

(a) Compute the population standard deviation.
(b) Determine three simple random samples of size 3, and
compute the sample standard deviation of each sample.
(c) Which samples underestimate the population standard
deviation? Which overestimate the population standard
deviation?

26. Travel Time The following data represent the travel time
(in minutes) to school for nine students enrolled in
Sullivan's College Algebra course. Treat the nine students
as a population.

Student	Travel Time	Student	Travel Time
Amanda	39	Scot	45
Amber	21	Erica	11
Tim	9	Tiffany	12
Mike	32	Glenn	39
Nicole	30		

(a) Compute the population standard deviation.
(b) Determine three simple random samples of size 4, and
compute the sample standard deviation of each sample.
(c) Which samples underestimate the population standard
deviation? Which overestimate the population standard
deviation?

27. A Fish Story Ethan and Drew went on a 10-day fishing trip.
The number of smallmouth bass caught and released by the
two boys each day was as follows:

Ethan	9	24	8	9	5	8	9	10	8	10
Drew	15	2	3	18	20	1	17	2	19	3

(a) Find the population mean and the range for the number
of smallmouth bass caught per day by each fisherman. Do
these values indicate any differences between the two
fishermen's catches per day? Explain.
(b) Find the population standard deviation for the number of
smallmouth bass caught per day by each fisherman. Do
these values present a different story about the two fisher-
men's catches per day? Which fisherman has the more
consistent record? Explain.
(c) Discuss limitations of the range as a measure of
dispersion.

28. Soybean Yield The following data represent the number of
pods on a sample of soybean plants for two different plot
types. Which plot type do you think is superior? Why?

Plot Type	Pods								
Liberty	32	31	36	35	44	31	39	37	38
No Till	35	31	32	30	43	33	37	42	40

Source: Andrew Dieter and Brad Schmidgall, students at Joliet Junior
College

29. The Empirical Rule The following data represent the weights
(in grams) of a random sample of 50 M&M plain candies.

0.87	0.88	0.82	0.90	0.90	0.84	0.84
0.91	0.94	0.86	0.86	0.86	0.88	0.87
0.89	0.91	0.86	0.87	0.93	0.88	
0.83	0.95	0.87	0.93	0.91	0.85	
0.91	0.91	0.86	0.89	0.87	0.84	
0.88	0.88	0.89	0.79	0.82	0.83	
0.90	0.88	0.84	0.93	0.81	0.90	
0.88	0.92	0.85	0.84	0.84	0.86	

Source: Michael Sullivan

(a) Determine the sample standard deviation weight. Express
your answer rounded to three decimal places.
(b) On the basis of the histogram drawn in Section 3.1,
Problem 29, comment on the appropriateness of using
the Empirical Rule to make any general statements
about the weights of M&Ms.
(c) Use the Empirical Rule to determine the percentage of
M&Ms with weights between 0.803 and 0.947 gram.
Hint: $\overline{x} = 0.875$.
(d) Determine the actual percentage of M&Ms that weigh
between 0.803 and 0.947 gram, inclusive.
(e) Use the Empirical Rule to determine the percentage of
M&Ms with weights more than 0.911 gram.
(f) Determine the actual percentage of M&Ms that weigh
more than 0.911 gram.

30. The Empirical Rule The following data represent the length
of eruption for a random sample of eruptions at the Old
Faithful geyser in Calistoga, California.

108	108	99	105	103	103	94
102	99	106	90	104	110	110
103	109	109	111	101	101	
110	102	105	110	106	104	
104	100	103	102	120	90	
113	116	95	105	103	101	
100	101	107	110	92	108	

Source: Ladonna Hansen, Park Curator

(a) Determine the sample standard deviation length of
eruption. Express your answer rounded to the nearest
whole number.
(b) On the basis of the histogram drawn in Section 3.1,
Problem 30, comment on the appropriateness of using
the Empirical Rule to make any general statements
about the length of eruptions.

(c) Use the Empirical Rule to determine the percentage of eruptions that last between 92 and 116 seconds. *Hint*: $\bar{x} = 104$.

(d) Determine the actual percentage of eruptions that last between 92 and 116 seconds, inclusive.

(e) Use the Empirical Rule to determine the percentage of eruptions that last less than 98 seconds.

(f) Determine the actual percentage of eruptions that last less than 98 seconds.

31. Which Car Would You Buy? Suppose that you are in the market to purchase a car. With gas prices on the rise, you have narrowed it down to two choices and will let gas mileage be the deciding factor. You decide to conduct a little experiment in which you put 10 gallons of gas in the car and drive it on a closed track until it runs out gas. You conduct this experiment 15 times on each car and record the number of miles driven.

Car 1				
228	223	178	220	220
233	233	271	219	223
217	214	189	236	248

Car 2				
277	164	326	215	259
217	321	263	160	257
239	230	183	217	230

Describe each data set. That is, determine the shape, center, and spread. Which car would you buy and why?

32. Which Investment Is Better? You have received a year-end bonus of $5,000. You decide to invest the money in the stock market and have narrowed your investment options down to two mutual funds. The following data represent the historical quarterly rates of return of each mutual fund for the past 20 quarters (5 years).

Mutual Fund A				
1.3	−0.3	0.6	6.8	5.0
5.2	4.8	2.4	3.0	1.8
7.3	8.6	3.4	3.8	−1.3
6.4	1.9	−0.5	−2.3	3.1

Mutual Fund B				
−5.4	6.7	11.9	4.3	4.3
3.5	10.5	2.9	3.8	5.9
−6.7	1.4	8.9	0.3	−2.4
−4.7	−1.1	3.4	7.7	12.9

Describe each data set. That is, determine the shape, center, and spread. Which mutual fund would you invest in and why?

33. Rates of Return of Stocks Stocks may be categorized by industry. The following data represent the 5-year rates of return (in percent) for a sample of financial stocks and energy stocks ending December 3, 2007.

(a) Compute the mean and the median rate of return for each industry. Which sector has the higher mean rate of return? Which sector has the higher median rate of return?

(b) Compute the standard deviation for each industry. In finance, the standard deviation rate of return is called

Financial Stocks				
16.01	29.66	8.58	61.90	16.27
18.30	47.16	31.40	26.57	12.15
0.86	25.95	16.54	15.21	25.89
14.44	77.82	50.75	15.52	1.05
7.82	8.30	10.01	7.24	13.53

Energy Stocks				
44.39	7.22	18.82	106.05	41.29
42.79	23.10	44.87	34.18	40.67
29.32	39.08	20.35	52.94	42.56
39.89	13.07	71.03	28.68	33.76
24.75	17.39	41.29	26.61	22.26

Source: morningstar.com

risk. Typically, an investor "pays" for a higher return by accepting more risk. Is the investor paying for higher returns in this instance? Do you think the higher returns are worth the cost?

34. Temperatures It is well known that San Diego has milder weather than Chicago, but which city has more dispersion in temperatures over the course of a month? In particular, which city has more dispersion in high temperatures in the month of November? Use the following data, which represent the daily high temperatures for each day in November, 2007. In which city would you rather be a meteorologist? Why?

Daily High Temperature, Chicago (°F)					
54.0	43.0	52.0	39.9	51.8	39.2
57.2	42.8	66.0	44.6	41.0	39.0
55.4	52.0	60.8	42.8	30.9	48.2
53.6	48.0	55.4	61.0	37.9	48.0
53.6	51.1	46.4	61.0	37.9	35.6

Daily High Temperature, San Diego (°F)					
66.0	63.0	64.4	64.9	64.4	71.1
64.9	63.0	70.0	63.0	64.0	71.1
66.2	63.0	80.6	63.0	62.6	73.9
64.4	63.0	84.0	62.6	69.8	73.9
64.9	66.0	84.0	62.1	72.0	71.1

Source: National Climatic Data Center

35. The Empirical Rule One measure of intelligence is the Stanford–Binet Intelligence Quotient (IQ). IQ scores have a bell-shaped distribution with a mean of 100 and a standard deviation of 15.

(a) What percentage of people has an IQ score between 70 and 130?

(b) What percentage of people has an IQ score less than 70 or greater than 130?

(c) What percentage of people has an IQ score greater than 130?

36. The Empirical Rule SAT Math scores have a bell-shaped distribution with a mean of 515 and a standard deviation of 114.

Source: College Board, 2007

(a) What percentage of SAT scores is between 401 and 629?

(b) What percentage of SAT scores is less than 401 or greater than 629?

(c) What percentage of SAT scores is greater than 743?

37. The Empirical Rule The weight, in grams, of the pair of kidneys in adult males between the ages of 40 and 49 have a bell-shaped distribution with a mean of 325 grams and a standard deviation of 30 grams.

(a) About 95% of kidney pairs will be between what weights?

(b) What percentage of kidney pairs weighs between 235 grams and 415 grams?

(c) What percentage of kidney pairs weighs less than 235 grams or more than 415 grams?

(d) What percentage of kidney pairs weighs between 295 grams and 385 grams?

38. The Empirical Rule The distribution of the length of bolts has a bell shape with a mean of 4 inches and a standard deviation of 0.007 inch.

(a) About 68% of bolts manufactured will be between what lengths?

(b) What percentage of bolts will be between 3.986 inches and 4.014 inches?

(c) If the company discards any bolts less than 3.986 inches or greater than 4.014 inches, what percentage of bolts manufactured will be discarded?

(d) What percentage of bolts manufactured will be between 4.007 inches and 4.021 inches?

39. Chebyshev's Inequality In December 2007, the average price
NW of regular unleaded gasoline excluding taxes in the United States was $3.06 per gallon, according to the Energy Information Administration. Assume that the standard deviation price per gallon is $0.06 per gallon to answer the following.

(a) What minimum percentage of gasoline stations had prices within 3 standard deviations of the mean?

(b) What minimum percentage of gasoline stations had prices within 2.5 standard deviations of the mean? What are the gasoline prices that are within 2.5 standard deviations of the mean?

(c) What is the minimum percentage of gasoline stations that had prices between $2.94 and $3.18?

40. Chebyshev's Inequality According to the U.S. Census Bureau, the mean of the commute time to work for a resident of Boston, Massachusetts, is 27.3 minutes. Assume that the standard deviation of the commute time is 8.1 minutes to answer the following:

(a) What minimum percentage of commuters in Boston has a commute time within 2 standard deviations of the mean?

(b) What minimum percentage of commuters in Boston has a commute time within 1.5 standard deviations of the mean? What are the commute times within 1.5 standard deviations of the mean?

(c) What is the minimum percentage of commuters who have commute times between 3 minutes and 51.6 minutes?

41. Comparing Standard Deviations The standard deviation of batting averages of all teams in the American League is 0.008. The standard deviation of all players in the American League is 0.02154. Why is there less variability in team batting averages?

42. Linear Transformations Benjamin owns a small Internet business. Besides himself, he employs nine other people. The salaries earned by the employees are given next in thousands of dollars (Benjamin's salary is the largest, of course):

$$30, 30, 45, 50, 50, 50, 55, 55, 60, 75$$

(a) Determine the range, population variance, and population standard deviation for the data.

(b) Business has been good! As a result, Benjamin has a total of $25,000 in bonus pay to distribute to his employees. One option for distributing bonuses is to give each employee (including himself) $2,500. Add the bonuses under this plan to the original salaries to create a new data set. Recalculate the range, population variance, and population standard deviation. How do they compare to the originals?

(c) As a second option, Benjamin can give each employee a bonus of 5% of his or her original salary. Add the bonuses under this second plan to the original salaries to create a new data set. Recalculate the range, population variance, and population standard deviation. How do they compare to the originals?

(d) As a third option, Benjamin decides not to give his employees a bonus at all. Instead, he keeps the $25,000 for himself. Use this plan to create a new data set. Recalculate the range, population variance, and population standard deviation. How do they compare to the originals?

43. Resistance and Sample Size Each of the following three data sets represents the IQ scores of a random sample of adults. IQ scores are known to have a mean and median of 100. For each data set, determine the sample standard deviation. Then recompute the sample standard deviation assuming that the individual whose IQ is 106 is accidentally recorded as 160. For each sample size, state what happens to the standard deviation. Comment on the role that the number of observations plays in resistance.

Sample of Size 5				
106	92	98	103	100

Sample of Size 12					
106	92	98	103	100	102
98	124	83	70	108	121

Sample of Size 30					
106	92	98	103	100	102
98	124	83	70	108	121
102	87	121	107	97	114
140	93	130	72	81	90
103	97	89	98	88	103

44. Identical Values Compute the sample standard deviation of the following test scores: 78, 78, 78, 78. What can be said about a data set in which all the values are identical?

45. Blocking and Variability Recall that blocking refers to the idea that we can reduce the variability in a variable by segmenting the data by some other variable. The given data

represent the recumbent length (in centimeters) of a sample of 10 males and 10 females who are 40 months of age.

Males		Females	
104.0	94.4	102.5	100.8
93.7	97.6	100.4	96.3
98.3	100.6	102.7	105.0
86.2	103.0	98.1	106.5
90.7	100.9	95.4	114.5

Source: National Center for Health Statistics

(a) Determine the standard deviation of recumbent length for all 20 observations.
(b) Determine the standard deviation of recumbent length for the males.
(c) Determine the standard deviation of recumbent length for the females.
(d) What effect does blocking by gender have on the standard deviation of recumbent length for each gender?

46. Coefficient of Variation The coefficient of variation is a measure that allows for the comparison of two or more variables measured on a different scale. It measures the relative variability in terms of the mean and does not have a unit of measure. The lower the coefficient of variation is, the less the data vary. The coefficient of variation is defined as

$$CV = \frac{\text{standard deviation}}{\text{mean}} \cdot 100\%$$

For example, suppose the senior class of a school district has a mean ACT score of 23 with a standard deviation of 4 and a mean SAT score of 1,100 with a standard deviation of 150. Which test has more variability? The coefficient of variation for the ACT exam is $\frac{4}{23} \cdot 100\% = 17.4\%$; the coefficient of variation for the SAT exam is $\frac{150}{1100} \cdot 100\% = 13.6\%$. The ACT has more variability.

(a) Suppose the systolic blood pressure of a random sample of 100 students before exercising has a sample mean of 121, with a standard deviation of 14.1. The systolic blood pressure after exercising is 135.9, with a standard deviation of 18.1. Is there more variability in systolic blood pressure before exercise or after exercise? After
(b) An investigation of intracellular calcium and blood pressure was conducted by Erne, Bolli, Buergisser, and Buehler in 1984 and published in the *New England Journal of Medicine*, 310:1084–1088. The researchers measured the free calcium concentration in the blood platelets of 38 people with normal blood pressure and 45 people with high blood pressure. The mean and standard deviation of the normal blood pressure group were 107.9 and 16.1, respectively. The mean and standard deviation of the high blood pressure group were 168.2 and 31.7, respectively. Is there more variability in free calcium concentration in the normal blood pressure or the high blood pressure group?

47. Mean Absolute Deviation Another measure of variation is the mean absolute deviation. It is computed using the formula

$$MAD = \frac{\sum |x_i - \bar{x}|}{n}$$

Compute the mean absolute deviation of the data in Problem 17 and compare the results with the sample standard deviation.

48. Coefficient of Skewness Karl Pearson developed a measure that describes the skewness of a distribution, called the **coefficient of skewness**. The formula is

$$\text{Skewness} = \frac{3(\text{mean} - \text{median})}{\text{standard deviation}}$$

The value of this measure generally lies between −3 and +3. The closer the value lies to −3, the more the distribution is skewed left. The closer the value lies to +3, the more the distribution is skewed right. A value close to 0 indicates a symmetric distribution. Find the coefficient of skewness of the following distributions and comment on the skewness.

(a) Mean = 50, median = 40, standard deviation = 10
(b) Mean = 100, median = 100, standard deviation = 15
(c) Mean = 400, median = 500, standard deviation = 120
(d) Compute the coefficient of skewness for the data in Problem 29.
(e) Compute the coefficient of skewness for the data in Problem 30.

49. Diversification A popular theory in investment states that you should invest a certain amount of money in foreign investments to reduce your risk. The risk of a portfolio is defined as the standard deviation of the rate of return. Refer to the following graph, which depicts the relation between risk (standard deviation of rate of return) and reward (mean rate of return).

How Foreign Stocks Benefit a Domestic Portfolio
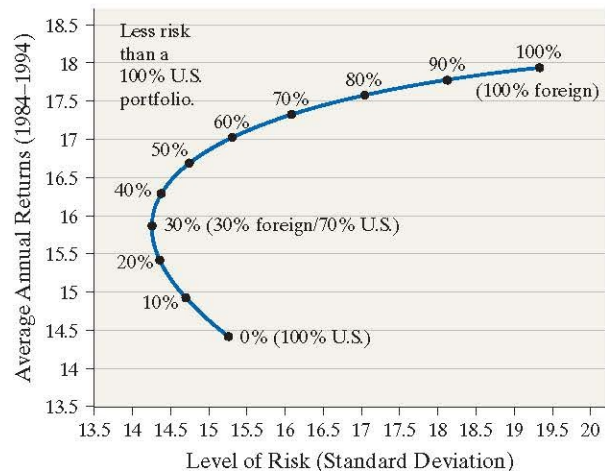
Source: T. Rowe Price

(a) Determine the average annual return and level of risk in a portfolio that is 10% foreign.
(b) Determine the percentage that should be invested in foreign stocks to best minimize risk.
(c) Why do you think risk initially decreases as the percent of foreign investments increases?
(d) A portfolio that is 30% foreign and 70% American has a mean rate of return of about 15.8%, with a standard deviation of 14.3%. According to Chebyshev's inequality, at least 75% of returns will be between what values? According to Chebyshev's inequality, at least 88.9% of returns will be between what two values? Should an investor be surprised if she has a negative rate of return? Why?