Name ______________________________

Chapter 2 Learning Objectives	**Section**	**Related Example on Page(s)**	**Relevant Chapter Review Exercise(s)**	**Can I do this?**
Find and interpret the percentile of an individual value within a distribution of data.	2.1	86	R2.1	
Estimate percentiles and individual values using a cumulative relative frequency graph.	2.1	87, 88	R2.2	
Find and interpret the standardized score (*z*-score) of an individual value within a distribution of data.	2.1	90, 91	R2.1	
Describe the effect of adding, subtracting, multiplying by, or dividing by a constant on the shape, center, and spread of a distribution of data.	2.1	93, 94, 95	R2.3	
Estimate the relative locations of the median and mean on a density curve.	2.2	Discussion on 106–107	R2.4	
Use the 68–95–99.7 rule to estimate areas (proportions of values) in a Normal distribution.	2.2	111	R2.5	
Use Table A or technology to find (i) the proportion of *z*-values in a specified interval, or (ii) a *z*-score from a percentile in the standard Normal distribution.	2.2	114, 115, Discussion on 116	R2.6	
Use Table A or technology to find (i) the proportion of values in a specified interval, or (ii) the value that corresponds to a given percentile in any Normal distribution.	2.2	118, 119, 120	R2.7, R2.8, R2.9	
Determine whether a distribution of data is approximately Normal from graphical and numerical evidence.	2.2	122, 123, 124	R2.10, R2.11	

2.1: Identifying location in a distribution: percentiles and *z*-scores
Read 84-85

What is a percentile?

> **DEFINITION: Percentile**
>
> The ***p*th percentile** of a distribution is the value with *p* percent of the observations less than it.

Percentile	Percent
A percentile compares one observation to all the observations in the ordered group. A percentile compares one outcome relative to the other outcomes in the group. A percentile is a rank.	A percent compares an earned score to the highest score possible (or the earned number of points to the total number of points possible). A percent shows a part-whole relationship.

A college professor gave her 100 students a test recently and here are their scores (percent correct):

38	40	43	45	47	47	48	49	49	50	51	51	51	52	52	53	54	54	54	54
56	58	59	60	60	61	62	62	62	63	63	64	64	64	65	66	67	67	68	69
70	71	71	71	71	71	72	72	72	72	72	72	73	73	74	75	75	75	75	76
76	78	79	79	79	80	80	81	82	83	83	84	85	85	86	86	88	88	88	89
90	91	91	91	92	92	92	94	94	94	95	95	96	96	97	97	97	98	100	100

Circle the grade for the student who earned a 40% on this professor's test. What would you say about that student's performance?

Circle the grade for the student who has a score at the 40^{th} percentile. What would you say about that student's performance?

On a test, does a student's percentile have to be the same as the student's percent correct?

Wins in Major League Baseball

The stemplot below shows the number of wins for each of the 30 Major League Baseball teams in 2014.

6	467
7	001333677799
8	234578889
9	004668

Key: 7|1 represents a team with 71 wins.

This is called a stemplot.

6	4
6	67
7	001333
7	677799
8	234
8	578889
9	004
9	668

Key: 7|1 represents a team with 71 wins.

This is called a stemplot with split stem;
it just spreads things out.

Calculate and interpret the percentiles for…

the St. Louis Cardinals, who had 90 wins

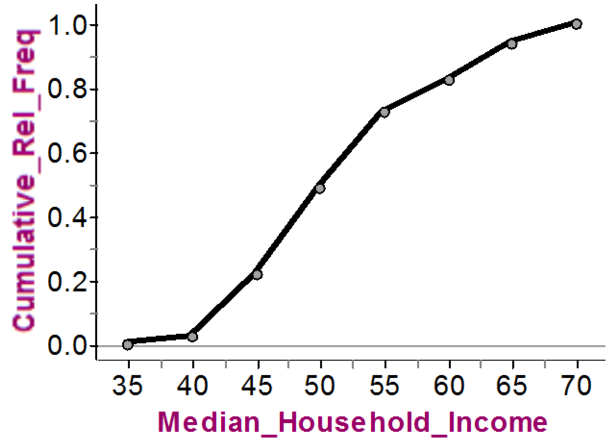
the Milwaukee Brewers, who had 82 wins

the Los Angeles Angels, who had 98 wins

Read 86–88

State Median Household Incomes

Here is a cumulative relative frequency graph showing the distribution of median household incomes for the 50 states and the District of Columbia.



a) California, with a median household income of $57,445, is at what percentile?

Interpret this value.

b) What is the 25th percentile for this distribution?

What is another name for this value?

c) Where is the graph the steepest?

What does this indicate about the distribution?

Bridgette Jordan of Sandoval, Illinos (as of 2014), a town about 86 miles east of St. Charles, is one of the shortest women in the world, standing at 27 inches. Robert Wadlow was born in Alton, Illinois and passed away in 1940, at age 22, with a height of 107.1 inches. Obviously, Bridgette is shorter than most women and Robert was taller than most men —but whose height is more unusual, relatively speaking? That is, relative to other adults, who is taller? We'll say that that women have a mean of 64 in. and a standard deviation of 2.5 in. and that the mean height of men is 69.5 in. with a standard deviation of 2.8 in.

Read 89–91

This is the calculation for a standardized score (z-score):

$$z = \frac{x - \text{mean}}{\text{standard deviation}}$$

The z-score tells us how many standard deviations an observation is above or below the mean & z-scores have no units.

Positive z-scores indicate that an observation is higher than the mean & negative z-scores indicate that an observation is lower than the mean.
The mean for z-scores is 0 and the standard deviation is 1.

Home run kings

Year	Player	HR	Mean	SD
1927	Babe Ruth	60	7.2	9.7
1961	Roger Maris	61	18.8	13.4
1998	Mark McGwire	70	20.7	12.7
2001	Barry Bonds	73	21.4	13.2

The single-season home run record for major league baseball has been set just three times since Babe Ruth hit 60 home runs in 1927. Roger Maris hit 61 in 1961, Mark McGwire hit 70 in 1998 and Barry Bonds hit 73 in 2001. In an absolute sense, Barry Bonds had the best performance of these four players, because he hit the most home runs in a single season. However, in a relative sense this may not be true. Baseball historians suggest that hitting a home run has been easier in some eras than others. This is due to many factors, including quality of batters, quality of pitchers, hardness of the baseball, dimensions of ballparks, and possible use of performance-enhancing drugs. To make a fair comparison, we should see how these performances rate relative to others hitters during the same year. Calculate the standardized score for each player and compare.

In 2001, Arizona Diamondback Mark Grace's home run total had a standardized score of $z = –0.48$. Interpret this value and calculate the number of home runs he hit.

HW #19: page 99 (1, 3, 5, 9, 11, 13, 15)

2.1 Transforming Data and Density Curves

Refer to pages 92–97. Also notice what happens on the Fathom document or applet shown in class.
What are the effects of these “linear transformations”?

	Shape	Measures of Center and Location (mean, median, quartiles, percentiles)	Measures of Spread (range, IQR, standard deviation)
Adding on or subtracting off a constant for each observation			
Multiplying or dividing each observation by a constant			

In July 2015, St. Charles County Cab Company charged an initial fee of \$2.50 plus \$2 per mile. In equation form, *fare* = 2.50 + 2(*miles*). Genevieve walks everywhere or takes the cab around St. Charles. She analyzed the distribution of her fares for July and found it to be skewed to the right with a mean of \$16.45 and a standard deviation of \$2.20.

a) What are the mean and standard deviation of the lengths of Genevieve’s cab rides in miles?

b) Suppose that we standardize Genevieve’s fares for July. Describe the shape, center, and spread of the distribution.

HW #20 page 99 (17–31 odd)

2.2 Density Curves and Normal Distributions

Read 104-107

What is a density curve?

Why do we use density curves?

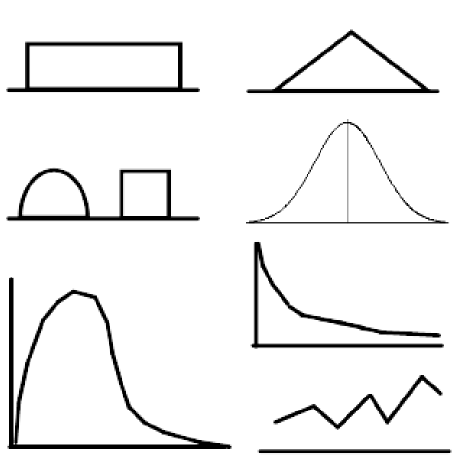
DEFINITION: Density curve

A **density curve** is a curve that

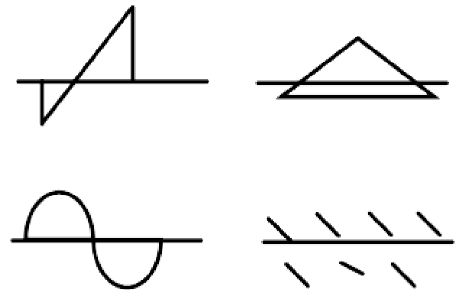
- is always on or above the horizontal axis, and
- has area exactly 1 underneath it.

A density curve describes the overall pattern of a distribution. The area under the curve and above any interval of values on the horizontal axis is the proportion of all observations that fall in that interval.

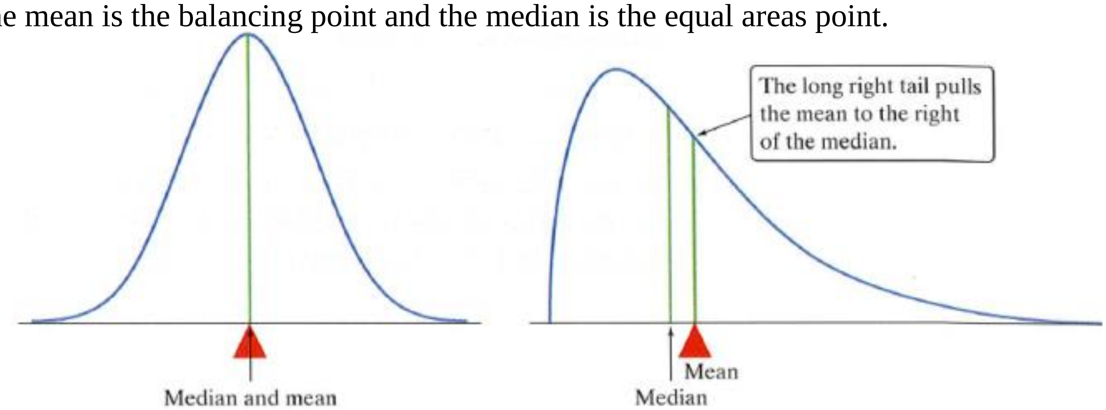
These are density curves:

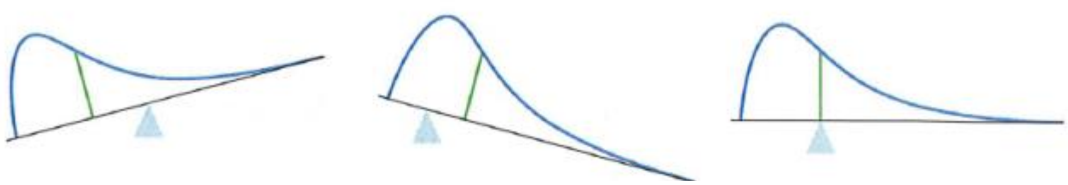


These are not density curves:



To identify the mean and median of a density curve:
The mean is the balancing point and the median is the equal areas point.





DISTINGUISHING THE MEDIAN AND MEAN OF A DENSITY CURVE

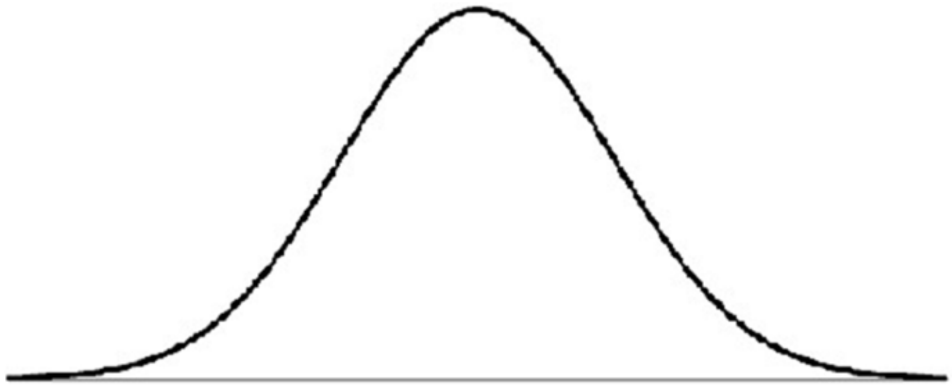
The **median** of a density curve is the equal-areas point, the point that divides the area under the curve in half.

The **mean** of a density curve is the balance point, at which the curve would balance if made of solid material.

The median and mean are the same for a symmetric density curve. They both lie at the center of the curve. The mean of a skewed curve is pulled away from the median in the direction of the long tail.

Read 108-109

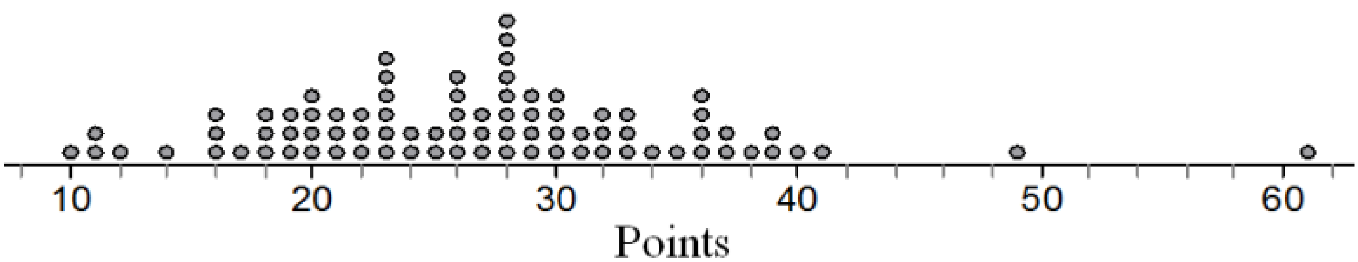
One source says that the heights of 11-year-old females are approximately Normally distributed with a mean of 59 inches and a standard deviation of 3 inches. Label the distribution (below) with the mean and the points one, two, and three standard deviations from the mean.



_____ _____ _____ _____ _____ _____ _____

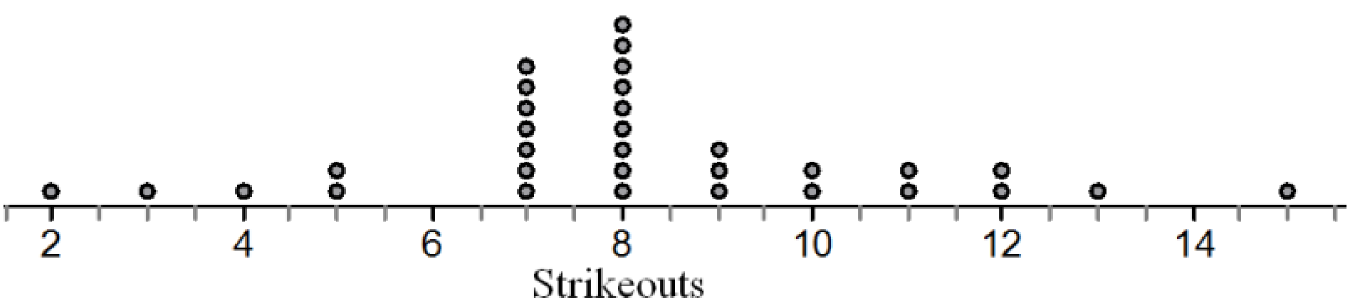
Here is a dotplot of Kobe Bryant's point totals for each of the 82 games in the 2008-2009 regular season. The mean of this distribution is 26.8 with a standard deviation of 8.6 points. In what percentage of games did he score within one standard deviation of his mean?

…within two standard deviations?



Here is a dotplot of Tim Lincecum's strikeout totals for each of the 32 games he pitched in during the 2009 regular season. The mean of this distribution is 8.2 with a standard deviation of 2.8. In what percentage of games were his strikeouts within one standard deviation of his mean?

…within two standard deviations?



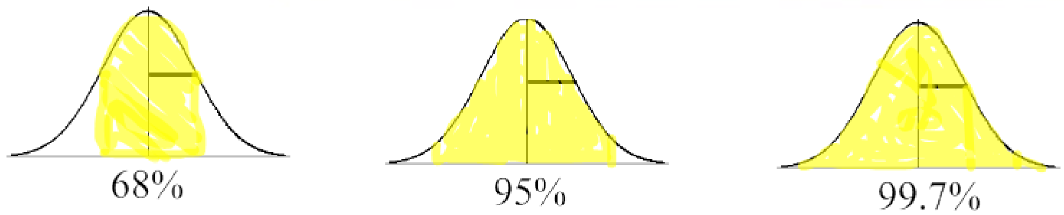
Read 109-112
This is the 68-95-99.7 rule:

Notice that this applies only for <u>Normal distributions</u>.

DEFINITION: The 68–95–99.7 rule

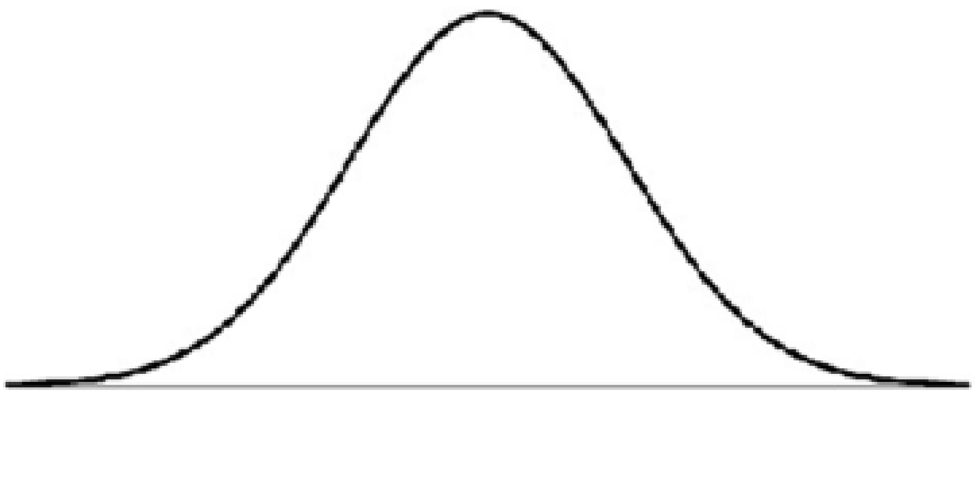
In a Normal distribution with mean μ and standard deviation σ:

- Approximately **68%** of the observations fall within σ of the mean μ.
- Approximately **95%** of the observations fall within 2σ of the mean μ.
- Approximately **99.7%** of the observations fall within 3σ of the mean μ.



Chebyshev's inequality, though interesting, is NOT something you have to know.

Using the earlier example, about what percentage of 11-year-old girls will be over 62 inches tall?



About what percentage of 11-year-old girls will be between 53 and 56 in. tall?

The distribution of blood glucose levels (after 4 hours of fasting and measured in mg/dL) is approximately Normal and the middle 95% of scores are between 70 and 110. What are the mean and standard of this distribution?

____ ____ ____ ____ ____ ____ ____

Can you calculate the percent of scores that are above 80? Explain.

HW #21: page 128 (33–45 odd)
Read 112-114

This is the
standard Normal distribution:

DEFINITION: Standard Normal distribution

The **standard Normal distribution** is the Normal distribution with mean 0 and standard deviation 1 (Figure 2.14).

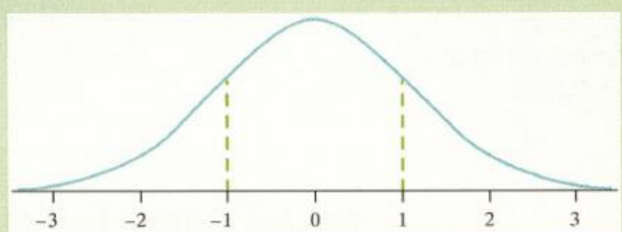


FIGURE 2.14 The standard Normal distribution.

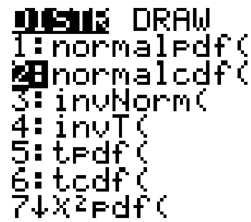
If a variable x has any Normal distribution $N(\mu, \sigma)$ with mean μ and standard deviation σ, then the standardized variable

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

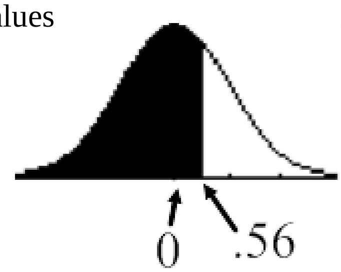
has the standard Normal distribution $N(0,1)$.

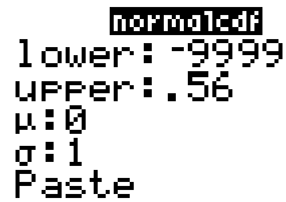
Notice that if we transform
a Normal distribution with
mean, μ, and
standard deviation, σ,
we get the
standard Normal distribution
with mean, 0, and
standard deviation, 1.

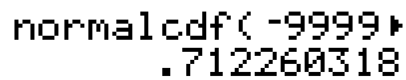
To find the proportion of observations from the standard Normal distribution that are: (a) less than 0.56
Press 2nd VARS to access the distributions (DISTR) menu



Choose 2:normalcdf and enter the values







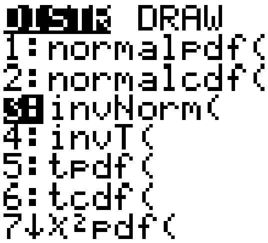
(b) greater than –1.14

(c) greater than 3.79

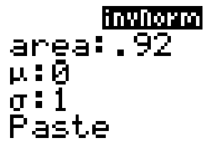
(d) between 0.46 and 1.84

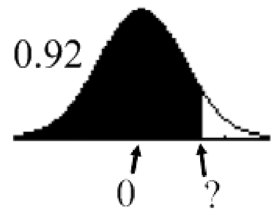
We will **<u>NEVER</u>** use 1:normalpdf
So, what happens if you forget how to use the calculator? Help yourself out-- go do a search online!

Now in reverse... In the standard Normal distribution, 92% of the observations are less than what value?
Press 2nd VARS to access the distributions (DISTR) menu



Choose 3:invNorm and enter the values





Tracy fasted for 4 hours and had her blood glucose checked. Her level was at the 85th percentile. How many standard deviations above the mean is that?

What percent of a Normal distribution is within .6745 standard deviations of the mean?

What do we call the locations at roughly z=-.6745 and at roughly z=.6745?

Serving Speed
In the 2008 Wimbledon tennis tournament, Rafael Nadal averaged 115 miles per hour (mph) on his first serves. Assume that the distribution of his first serve speeds is Normal with a mean of 115 mph and a standard deviation of 6 mph.
(a) About what proportion of his first serves would you expect to exceed 120 mph?

(b) What percent of Rafael Nadal's first serves are between 100 and 110 mph?

(c) The fastest 30% of Nadal's first serves go at least what speed?

(d) What is the *IQR* for the distribution of Nadal's first serve speeds?

(e) A different player has a standard deviation of 8 mph on his first serves and 20% of his serves go less than 100 mph. If the distribution of his serve speeds is approximately Normal, what is his average first serve speed?

HW #22: page 128 (47-59 odd)
2.2: Using the Calculator for Normal Calculations
What do you need to show for credit? (page 118)

HOW TO FIND AREAS IN ANY NORMAL DISTRIBUTION

Step 1: State the distribution and the values of interest. Draw a Normal curve with the area of interest shaded and the mean, standard deviation, and boundary value(s) clearly identified.

Step 2: Perform calculations—show your work! Do one of the following: (i) Compute a *z*-score for each boundary value and use Table A or technology to find the desired area under the standard Normal curve; or (ii) use the normalcdf command and label each of the inputs.

Step 3: Answer the question.

Serve speeds, x, are approx. normal.
$\mu = 115 \quad \sigma = 6$
$x > 120$
$z = \frac{120-115}{6} = .83$
Area for $x > 120$ or $z > .83$ is .2023
About 20.23% of Nadal's serves are over 115mph.

Suppose that along a certain stretch of Interstate Highway in one city, cars travel with a mean velocity of 65 miles per hour (mph) and a standard deviation of 5 mph and that the distribution of speeds can be modeled by a Normal distribution.

(a) About what proportion of the cars will travel over 80 mph?

(b) About what proportion of the cars will travel less than 55 mph?

(c) About what proportion of the cars will travel between 63 and 67 mph?

(d) What is the 30^{th} percentile of this highway's distribution of vehicle speeds?

(e) What car speeds would be considered low outliers for this stretch of highway?

(f) Suppose that in a different city, along a similar stretch of Interstate, vehicles have a mean velocity of 60 mph and 40% of the cars go less than 55 mph. What is his standard deviation of the car speeds, assuming the distribution of speeds can be modeled by a Normal distribution?

HW #23 page 132 (53, 63, 65, 69–74)
2.2 Assessing Normality
Read 121
The measurements listed below describe the useable capacity (in cubic feet) of a sample of 36 side-by-side refrigerators. (Source: *Consumer Reports*, May 2010) The mean for this set is 15.825 cubic feet and the standard deviation is 1.217 cubic feet. Use the 68-95-99.7 Rule to decide whether you think this data has an approximately normal distribution.

12.9	13.7	14.1	14.2	14.5	14.5	14.6	14.7	15.1	15.2	15.3	15.3
15.3	15.3	15.5	15.6	15.6	15.8	16.0	16.0	16.2	16.2	16.3	16.4
16.5	16.6	16.6	16.6	16.8	17.0	17.0	17.2	17.4	17.4	17.9	18.4

Read 122-124

Normal probability plots are optional. Read about them on page 123, if you want another way to assess whether a distribution is approximately Normal.

INTERPRETING NORMAL PROBABILITY PLOTS

If the points on a **Normal probability plot** lie close to a straight line, the data are approximately Normally distributed. Systematic deviations from a straight line indicate a non-Normal distribution. Outliers appear as points that are far away from the overall pattern of the plot.

HW #24: page 136: Chapter review exercises
Chapter 2 Review/FRAPPY
FRAPPY: 2011 #1: Football players
HW #25: page 137: AP Statistics Practice Test (T2.2: choices are 1, 2, 3, 4, 5; skip T2.6)
Chapter 2 Test