Chapter 3 – Graphical and Tabular Displays of Data
Section 4 – Histograms

Objectives
1. Construct and interpret frequency and relative frequency tables.
2. Construct and interpret frequency histograms and relative frequency histograms.
3. Interpret density histograms.
4. Describe the shape of a distribution.
5. Describe the spread of a distribution.
6. Describe the meaning of model.

Vocabulary
1. upper/lower class limit
2. class width
3. frequency distribution of a numerical variable
4. frequency/relative frequency of a class
5. frequency/relative frequency histogram
6. density histogram
7. unimodal/bimodal/multimodal
8. left/right tail
9. skewed left/right, symmetric
10. model

Lesson/Activity
OBJECTIVE 1 − Construct and interpret frequency and relative frequency tables.

1. The endorsements of the 15 athletes with the highest 2015 endorsements are shown in the following table.

Athlete	Sport	Endorsement (millions of dollars)
Tiger Woods	Golf	50
Cristiano Ronaldo	Soccer	27
Usain Bolt	Track and Field	21
Lebron James	Basketball	44
Kobe Bryant	Basketball	26
Kevin Durant	Basketball	35
Rory Mcllroy	Golf	32
Rafael Nadal	Tennis	28
Mahendra Singh Dhoni	Cricket	27
Roger Federer	Tennis	58
Maria Sharapova	Tennis	23
Novak Djokovic	Tennis	31
Lionel Messi	Soccer	22
Phil Mickelson	Golf	44
Neymar da Silva Santos Junior	Soccer	17

Source: Opendorse

Construct a frequency and relative frequency table for the endorsement distribution.

For a class [*a*; *b*], the **lower class limit** is *a* and the **upper class limit** is *b*. To find the **class width** of a class, we subtract the lower class limit of a class from the lower class limit of the next class.

Definition Frequency of a class and relative frequency of a class

The **frequency of a class** is the number of observations in the class. The **relative frequency of a class** is the proportion of the observations in the class.

When using classes, the **frequency distribution of a numerical variable** are the classes together with their frequencies and the **relative frequency distribution of a numerical variable** are the classes together with their relative frequencies.

Sum of Relative Frequencies

For a numerical variable, the sum of the relative frequencies of all the classes is equal to 1.

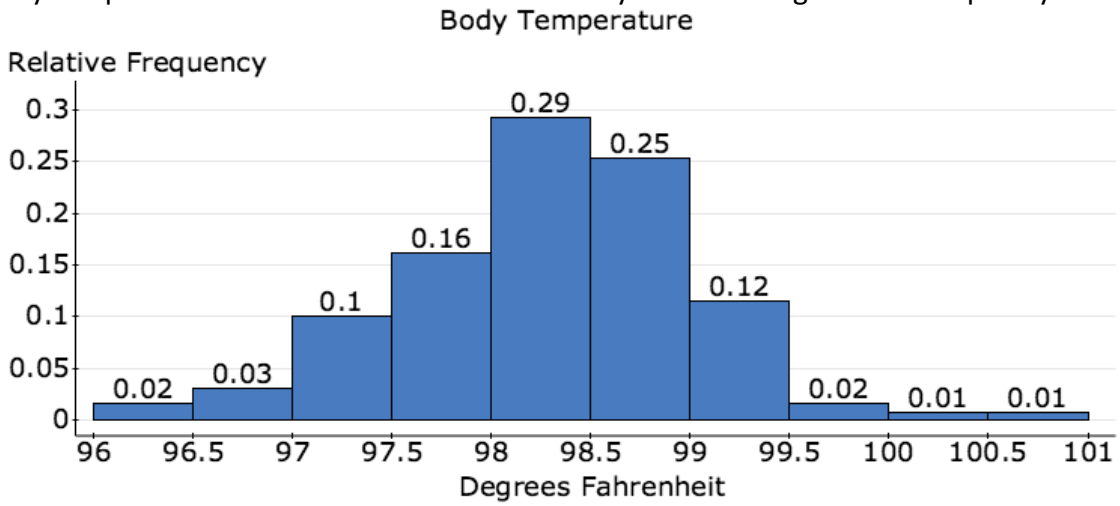
OBJECTIVE 2 – Construct and interpret frequency histograms and relative frequency histograms.

2. Construct a frequency histogram by hand for the endorsement distribution in Problem 1.
3. Construct a relative frequency histogram by hand for the endorsement distribution.
4. Use technology to construct a frequency histogram and a relative frequency histogram for the endorsement distribution.

Histograms can be used with any numerical variable (discrete or continuous) and any number of observations.

Histogram bars can touch, but bar-graph bars never touch.

5. The body temperatures of 130 adults are described by the following relative frequency histogram.



Source: American Statistical Association

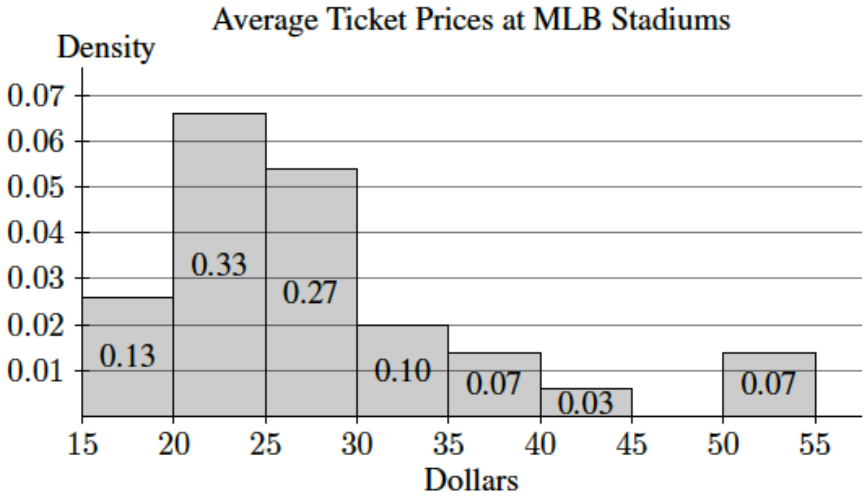
Estimate the proportion of observations that are

a. greater than or equal to 98 degrees Fahrenheit AND less than 99 degrees Fahrenheit.
b. less than 97.5 degrees Fahrenheit.
c. at least 99 degrees Fahrenheit.

OBJECTIVE 3 – Interpret density histograms.

For a **density histogram**, the vertical axis has units called **density** so that the area of each bar is the relative frequency of the bar's class.

6. The average prices of 2014 Major League Baseball (MLB) tickets at the stadiums are described by the following density histogram.



Source: *Team Marketing Report*

a. Find the proportion of stadiums whose average price of 2014 MLB tickets are…
   i. between $30 and $39.99, inclusive.
   ii. less than $20.
   iii. at least $20.

b. Estimate the percentile of a $25 average ticket price. Round to the nearest dollar.

c. The average ticket price at Nationals Park, home of the Washington Nationals, is at the 83rd percentile. Estimate the average ticket price. Round to the nearest dollar.

d. Identify the class which contains the 50th percentile, which is a reasonable measure of the center.

OBJECTIVE 4 – Describe the shape of a distribution.

Definition: Unimodal, bimodal, and multimodal distributions

A distribution is **unimodal** if it has one mound, **bimodal** if it has two mounds, and **multimodal** if it has more than two mounds.

For a unimodal distribution, the **left tail** is the part of the histogram to the left of the 50th percentile and the **right tail** is the part of the histogram to the right of the 50th percentile

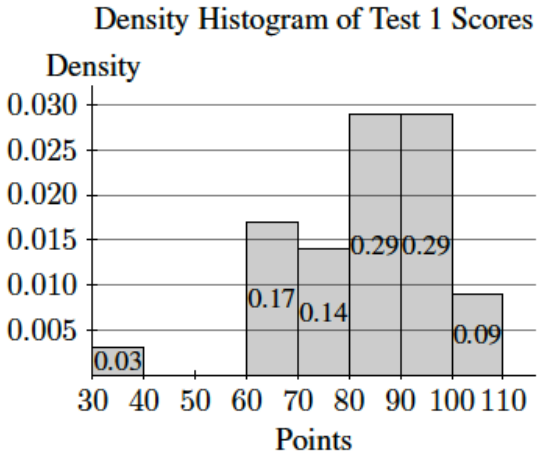
Definition: Skewed-left, skewed-right, and symmetric distributions

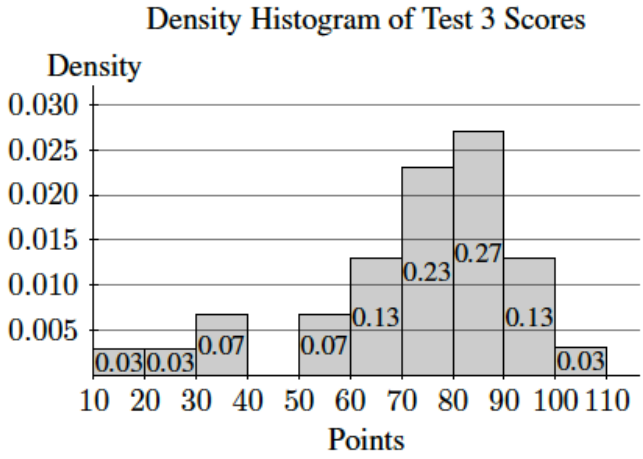
- If the left tail of a unimodal distribution is longer than the right tail, then the distribution is **skewed left**.
- If the right tail of a unimodal distribution is longer than the left tail, then the distribution is **skewed right**.
- If the left tail of a distribution is roughly the mirror image of the right tail, the distribution is **symmetric**.

An observation at the 50th percentile (the approximate center) of a unimodal distribution tends to be a typical observation.

OBJECTIVE 5 – Describe the spread of a distribution.

7. The following two density histograms describe calculus students' test scores (Source: J. Lehmann). The left histogram describes the scores of 35 students on Test 1. The right histogram describes the scores of 30 remaining students on Test 3.





a. For the two distributions identify any outliers.

b. Compare the shapes of the two distributions.

c. Compare the centers of the two distributions. What does that mean in this situation?

d. Compare the spreads of the two distributions. What does that mean in this situation?

e. Give three possible reasons why the distributions are different.

Order of Determining the Four Characteristics of a Distribution with a Numerical Variable

We often determine the four characteristics of a distribution with a numerical variable in the following order:

1. Identify all outliers.
   a. For outliers that stem from errors in measurement or recording, correct the errors, if possible. If the errors cannot be corrected, remove the outliers.
   b. For other outliers, determine whether they should be analyzed in a separate study.

2. Determine the shape. If the distribution is bimodal or multimodal, determine whether subgroups of the data should be analyzed separately.

3. Estimate and interpret the center.

4. Describe the spread.

OBJECTIVE 6 – Describe the meaning of model.

Definition: Model

A **model** is a mathematical description of an authentic situation.
We say the description models the situation.

Diagram	Types of Variables	Benefits
Frequency Bar Graph	One categorical variable	Compare frequencies of categories.
Relative Frequency Bar Graph	One categorical variable	Compare a part to the whole.
Multiple Bar Graph	Two categorical variables	Compare a part to the whole.
Pie Chart	One categorical variable	Compare a part to the whole.
Two-Way Table	Two categorical variables	Compare a part to the whole.
Dotplot	One numerical variable	Describe individual values for a small or medium number of observations.
Stemplot	One numerical variable	Describe individual values for a small number of observations.
Frequency Histogram	One numerical variable	Compare the frequencies of classes.
Relative Frequency Histogram	One numerical variable	Compare a part to the whole.
Density Histogram	One numerical variable	Compare a part to the whole.
Time-Series Plot	Two numerical variables	Find the association between two variables.

Homework/Assessment

1, 3, 5, 9, 10, 11, 13, 17, 19, 23, 25, 27, 33, 43, 45, 49