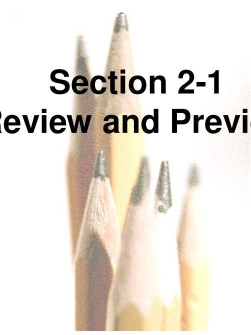
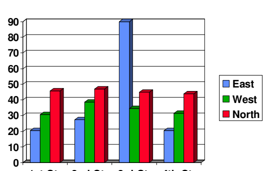
Chapter 2
Summarizing and Graphing Data

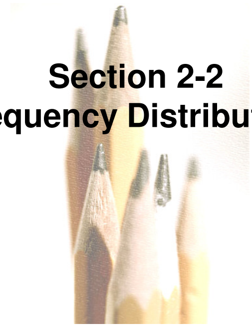


Section 2-1
Review and Preview

Preview
Important Characteristics of Data

1. Center: A representative or average value that indicates where the middle of the data set is located.
2. Variation: A measure of the amount that the data values vary.
3. Distribution: The nature or shape of the spread of data over the range of values (such as bell-shaped, uniform, or skewed).
4. Outliers: Sample values that lie very far away from the vast majority of other sample values.
5. Time: Changing characteristics of the data over time.





Section 2-2
Frequency Distributions

Key Concept

When working with large data sets, it is often helpful to organize and summarize data by constructing a table called a frequency distribution, defined later. Because computer software and calculators can generate frequency distributions, the details of constructing them are not as important as what they tell us about data sets. It helps us understand the nature of the *distribution* of a data set.

Definition

- **Frequency Distribution (or Frequency Table)**

  shows how a data set is partitioned among all of several categories (or classes) by listing all of the categories along with the number of data values in each of the categories.

Pulse Rates of Females and Males

Original Data

Table 2-1 Pulse Rates (beats per minute) of Females and Males

Females																			
76	72	88	60	72	68	80	64	68	68	80	76	68	72	96	72	68	72	64	80
64	80	76	76	76	80	104	88	60	76	72	72	88	80	60	72	88	88	124	64
Males																			
68	64	88	72	64	72	60	88	76	60	96	72	56	64	60	64	84	76	84	88
72	56	68	64	60	68	60	60	56	84	72	84	88	56	64	56	56	60	64	72

Frequency Distribution
Pulse Rates of Females

Table 2-2 Pulse Rates of Females

Pulse Rate	Frequency
60–69	12
70–79	14
80–89	11
90–99	1
100–109	1
110–119	0
120–129	1

The *frequency* for a particular class is the number of original values that fall into that class.

Frequency Distributions

Definitions

Lower Class Limits

are the smallest numbers that can actually belong to different classes

Table 2-2 Pulse Rates of Females

Pulse Rate	Frequency
60-69	12
70-79	14
80-89	11
90-99	1
100-109	1
110-119	0
120-129	1

Lower Class Limits

Upper Class Limits

are the largest numbers that can actually belong to different classes

Table 2-2 Pulse Rates of Females

Pulse Rate	Frequency
60-69	12
70-79	14
80-89	11
90-99	1
100-109	1
110-119	0
120-129	1

Upper Class Limits

Class Boundaries

are the numbers used to separate classes, but without the gaps created by class limits

Table 2-2 Pulse Rates of Females

Pulse Rate	Frequency
60-69	12
70-79	14
80-89	11
90-99	1
100-109	1
110-119	0
120-129	1

Class Boundaries: 59.5, 69.5, 79.5, 89.5, 99.5, 109.5, 119.5, 129.5

Class Midpoints

are the values in the middle of the classes and can be found by adding the lower class limit to the upper class limit and dividing the sum by two

Class Midpoints

Table 2-2 Pulse Rates of Females

Class Midpoint	Pulse Rate	Frequency
64.5	60–69	12
74.5	70–79	14
84.5	80–89	11
94.5	90–99	1
104.5	100–109	1
114.5	110–119	0
124.5	120–129	1

Class Width

is the difference between two consecutive lower class limits or two consecutive lower class boundaries

Class Width

Table 2-2 Pulse Rates of Females

Class Width	Pulse Rate	Frequency
	60–69	12
10	70–79	14
10	80–89	11
10	90–99	1
10	100–109	1
10	110–119	0
10	120–129	1

Reasons for Constructing Frequency Distributions

1. Large data sets can be summarized.
2. We can analyze the nature of data.
3. We have a basis for constructing important graphs.

Constructing A Frequency Distribution

1. Determine the number of classes (should be between 5 and 20).
2. Calculate the class width (round up).

$$\text{class width} \approx \frac{\text{(maximum value)} - \text{(minimum value)}}{\text{number of classes}}$$

3. Starting point: Choose the minimum data value or a convenient value below it as the first lower class limit.
4. Using the first lower class limit and class width, proceed to list the other lower class limits.
5. List the lower class limits in a vertical column and proceed to enter the upper class limits.
6. Take each individual data value and put a tally mark in the appropriate class. Add the tally marks to get the frequency.

Relative Frequency Distribution

includes the same class limits as a frequency distribution, but the frequency of a class is replaced with a relative frequencies (a proportion) or a percentage frequency (a percent)

$$\text{relative frequency} = \frac{\text{class frequency}}{\text{sum of all frequencies}}$$

$$\text{percentage frequency} = \frac{\text{class frequency}}{\text{sum of all frequencies}} \times 100\%$$

Relative Frequency Distribution

Table 2-2 Pulse Rates of Females

Pulse Rate	Frequency
60–69	12
70–79	14
80–89	11
90–99	1
100–109	1
110–119	0
120–129	1

Total Frequency = 40

Table 2-3 Relative Frequency Distribution of Pulse Rates of Females

Pulse Rate	Relative Frequency
60–69	30% *
70–79	35%
80–89	27.5%
90–99	2.5%
100–109	2.5%
110–119	0
120–129	2.5%

* 12/40 × 100 = 30%

Cumulative Frequency Distribution

Table 2-2 Pulse Rates of Females

Pulse Rate	Frequency
60-69	12
70-79	14
80-89	11
90-99	1
100-109	1
110-119	0
120-129	1

Table 2-4 Cumulative Frequency Distribution of Pulse Rates of Females

Pulse Rate	Cumulative Frequency
Less than 70	12
Less than 80	26
Less than 90	37
Less than 100	38
Less than 110	39
Less than 120	39
Less than 130	40

Cumulative Frequencies

Frequency Tables

Table 2-2 Pulse Rates of Females

Pulse Rate	Frequency
60-69	12
70-79	14
80-89	11
90-99	1
100-109	1
110-119	0
120-129	1

Table 2-3 Relative Frequency Distribution of Pulse Rates of Females

Pulse Rate	Relative Frequency
60-69	30%
70-79	35%
80-89	27.5%
90-99	2.5%
100-109	2.5%
110-119	0
120-129	2.5%

Table 2-4 Cumulative Frequency Distribution of Pulse Rates of Females

Pulse Rate	Cumulative Frequency
Less than 70	12
Less than 80	26
Less than 90	37
Less than 100	38
Less than 110	39
Less than 120	39
Less than 130	40

Critical Thinking Interpreting Frequency Distributions

In later chapters, there will be frequent reference to data with a normal distribution. One key characteristic of a normal distribution is that it has a "bell" shape.

- **The frequencies start low, then increase to one or two high frequencies, then decrease to a low frequency.**
- **The distribution is approximately symmetric, with frequencies preceding the maximum being roughly a mirror image of those that follow the maximum.**

Gaps

- **Gaps**

  The presence of gaps can show that we have data from two or more different populations. **However, the converse is not true, because data from different populations do not necessarily result in gaps.**

Recap

In this Section we have discussed

- **Important characteristics of data**
- **Frequency distributions**
- **Procedures for constructing frequency distributions**
- **Relative frequency distributions**
- **Cumulative frequency distributions**