Section 2-1

Frequency Distributions

TWO TYPES OF STATISTICS

- **Descriptive statistics** summarize or **describe** the important characteristics of data.
- **Inferential statistics** use sample data to make **inferences** (or generalizations) about a population.

IMPORTANT CHARACTERISTICS OF DATA

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

FREQUENCY DISTRIBUTION

A **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 (frequency) of data values in each of them.

QWERTY KEYBOARD WORD RATINGS

An article in *Discover* magazine ("Typecasting" by Scott Kim, *Discover*) suggests that you can measure the ease of typing by using this point rating system:
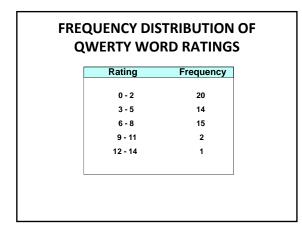
Count each letter on the home row as 0, each letter on the top row as 1, and each letter on the bottom row as 2.
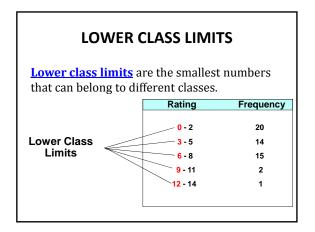
EXAMPLE:
```
S T  A T I S T I C S
0 1  0 1 1 0 1 1 2 0
```

To get the word rating, we sum the numbers. The word rating for "statistics" is 7.

QWERTY KEYBOARD WORD RATINGS

2	2	5	1	2	6	3	3	4	2
4	0	5	7	7	5	6	6	8	10
7	2	2	10	5	8	2	5	4	2
6	2	6	1	7	2	7	2	3	8
1	5	2	5	2	14	2	2	6	3
1	7								

This table shows the word ratings for each of the 52 words in the Preamble to the Constitution.

FREQUENCY DISTRIBUTION OF QWERTY WORD RATINGS

Rating	Frequency
0 - 2	20
3 - 5	14
6 - 8	15
9 - 11	2
12 - 14	1

LOWER CLASS LIMITS

Lower class limits are the smallest numbers that can belong to different classes.

Lower Class Limits

Rating	Frequency
0 - 2	20
3 - 5	14
6 - 8	15
9 - 11	2
12 - 14	1

UPPER CLASS LIMITS

Upper class limits are the largest numbers that can belong to different classes.

Upper Class Limits

Rating	Frequency
0 - 2	20
3 - 5	14
6 - 8	15
9 - 11	2
12 - 14	1

CLASS BOUNDARIES

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

Class Boundaries

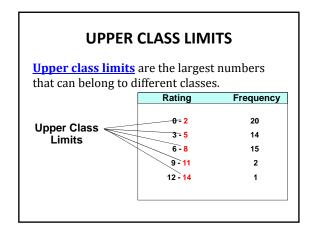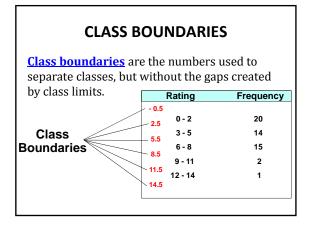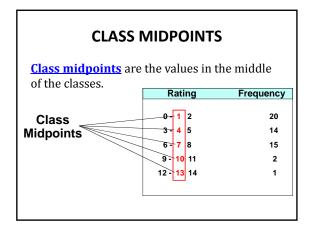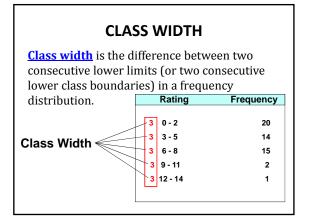| Rating	Frequency
- 0.5	
0 - 2	20
2.5	
3 - 5	14
5.5	
6 - 8	15
8.5	
9 - 11	2
11.5	
12 - 14	1
14.5	

FINDING CLASS BOUNDARIES

To find class boundaries:

1. Find the size of the gap between the upper class limit of one class and the lower class limit of the next class.

2. Add half of that amount to each upper class limit to find the upper class boundaries.

3. Subtract half that amount from each lower class limit to find the lower class boundaries.

CLASS MIDPOINTS

Class midpoints are the values in the middle of the classes.

Class Midpoints

Rating	Frequency
0 - 1 2	20
3 - 4 5	14
6 - 7 8	15
9 - 10 11	2
12 - 13 14	1

CLASS WIDTH

Class width is the difference between two consecutive lower limits (or two consecutive lower class boundaries) in a frequency distribution.

Rating	Frequency
0 - 2	20
3 - 5	14
6 - 8	15
9 - 11	2
12 - 14	1

Class Width

3
3
3
3
3

CONSTRUCTING A FREQUENCY DISTRIBUTION

1. Select the number of classes, usually between 5 and 20.
2. Determine the class width

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

   Round this result to get a convenient number. (It's usually best to round *up*.)
3. Choose the value for the first lower class limit by using either the minimum value or a convenient value slightly below the minimum.
4. Using the first lower class limit and the class width, list the other lower class limits. (Add the class width to the first lower class limit to get the second lower class limit. Add the width to the second lower limit to get the third, and so on.)
5. List the lower class limits in a vertical column and 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 find the total frequency for each class.

RELATIVE FREQUENCY DISTRIBUTION

A **relative frequency distribution** or **percentage frequency distribution** includes the same class limits as a frequency distribution, but relative frequencies are used instead of actual frequencies.

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

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

RELATIVE FREQUENCY DISTRIBUTION

Rating	Frequency
0 - 2	20
3 - 5	14
6 - 8	15
9 - 11	2
12 - 14	1

Rating	Relative Frequency
0 - 2	38.5%
3 - 5	26.9%
6 - 8	28.8%
9 - 11	3.8%
12 - 14	1.9%

20/52 = 38.5%

14/52 = 26.9%

etc.

Total frequency = 52

CUMULATIVE FREQUENCY DISTRIBUTION

A **cumulative frequency** for a class is the sum of the frequencies for that class and *all previous classes*.

Rating	Frequency
0 - 2	20
3 - 5	14
6 - 8	15
9 - 11	2
12 - 14	1

Rating	Cumulative Frequency
Less than 3	20
Less than 6	34
Less than 9	49
Less than 12	51
Less than 15	52

Cumulative Frequencies

CRITICALLY THINKING: USING FREQUENCY DISTRIBUTIONS TO UNDERSTAND DATA

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 should be approximately symmetric, with frequencies preceding the maximum being roughly a mirror image of those that follow the maximum.