Chapter 2 Exploring Data with Tables and Graphs
Frequency Distributions for Organizing and Summarizing Data (2.1)

Objectives

In this lesson, we will learn how to

- organize large data sets using frequency distributions;
- identify class limits, class boundaries, class midpoints, and class width;
- construct a frequency distribution from raw data;
- create and interpret relative frequency and cumulative frequency distributions;
- use frequency distributions to recognize important features of data, such as shape and gaps.

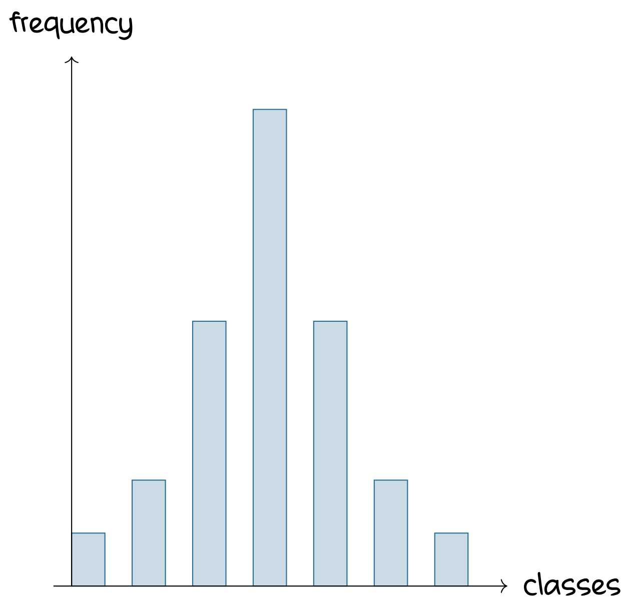
1. Key Idea

When working with a large data set, it can be difficult to see the overall pattern by looking only at the raw values.

A frequency distribution helps organize and summarize data so that we can better understand the nature of the distribution.

Frequency Distribution

A frequency distribution, also called a frequency table, shows how data are divided among several categories or classes.
It lists each class together with the number of data values in that class. That number is called the frequency.



A frequency distribution helps us see the overall shape.

2. Important Vocabulary

Parts of a Frequency Distribution

Lower class limits
The smallest numbers that can belong to each class.

Upper class limits
The largest numbers that can belong to each class.

Class boundaries
The numbers used to separate the classes without gaps between the classes.

Class midpoints
The values in the middle of the classes.

Class width
The difference between two consecutive lower class limits, or two consecutive lower class boundaries.

For a class such as

$$15 - 29,$$

the lower class limit is 15 and the upper class limit is 29.

The class midpoint is

$$\frac{15+29}{2} = 22.$$

3. Constructing a Frequency Distribution

Procedure

To construct a frequency distribution:

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

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

3. Round the class width to a convenient number, usually rounding

up.

4. Choose the first lower class limit. This can be the minimum value or a convenient value below the minimum.

5. Use the class width to list the remaining lower class limits.

6. Determine the upper class limits.

7. Tally the data values and find the frequency for each class.

4. Example: Commute Times in Los Angeles

Example 1: The following are daily commute times, in minutes, for Los Angeles. Construct a frequency distribution using 7 classes.

18	25	45	75	60	40	25	8	50	10	10	30	15	25	50	20	30
20	45	30	60	30	20	15	30	60	30	15	35	40	5	30	40	20
10	45	30	15	25	25	5	90	30	15	60	20	60	30	25	25	

Solution

Step 1: Use 7 classes.
Step 2: The minimum value is 5 and the maximum value is 90. Therefore,

$$\frac{90-5}{7} = \frac{85}{7} \approx 12.14.$$

We round up to a convenient class width of

$$15.$$

Step 3: Choose 0 as the first lower class limit because it is convenient and below the minimum value of 5.
Step 4: Add 15 repeatedly to get the lower class limits:

$$0, 15, 30, 45, 60, 75, 90.$$

Step 5: The frequency distribution is:

Daily Commute Time (minutes)	Frequency
$0 - 14$	6
$15 - 29$	18
$30 - 44$	14
$45 - 59$	5
$60 - 74$	5
$75 - 89$	1
$90 - 104$	1

5. Interpreting the Frequency Distribution

The frequency distribution from Example 1 tells us how commute times are spread across the classes.

What We Can See

- The most common commute times are between 15 and 29 minutes.
- Most commute times are less than 45 minutes.
- Only a few commute times are very long, such as 75 minutes or more.

6. Relative Frequency Distribution

Sometimes we want to compare data sets with different total sizes. In that case, frequencies alone may not be enough.

A relative frequency distribution replaces each frequency by a relative frequency or a percentage.

Relative Frequency

$$\text{relative frequency} = \frac{\text{class frequency}}{\text{total frequency}}.$$

A percentage frequency is found by multiplying the relative frequency by 100%.

Example 2: Using the commute-time frequency distribution from Example 1, complete the relative frequency distribution.

Daily Commute Time (minutes)	Frequency	Relative Frequency
$0-14$	6	
$15-29$	18	
$30-44$	14	
$45-59$	5	
$60-74$	5	
$75-89$	1	
$90-104$	1	

Solution

There are

$$6+18+14+5+5+1+1=50$$

total commute times.
So each relative frequency is found by dividing by 50.

Daily Commute Time (minutes)	Frequency	Relative Frequency
$0-14$	6	12%
$15-29$	18	36%
$30-44$	14	28%
$45-59$	5	10%
$60-74$	5	10%
$75-89$	1	2%
$90-104$	1	2%

The percentages add to 100%.

7. Comparing Relative Frequency Distributions

Relative frequencies make it easier to compare data sets that have different total numbers of observations.

Example 3: The table compares commute times in New York and Boise.

Daily Commute Time (minutes)	New York	Boise
$0 - 14$	8.6%	30.3%
$15 - 29$	20.3%	45.5%
$30 - 44$	24.8%	17.0%
$45 - 59$	17.2%	3.5%
$60 - 74$	18.5%	2.2%
$75 - 89$	3.3%	0.3%
$90 - 104$	4.7%	0.3%
$105 - 119$	0.1%	0.0%
$120 - 134$	0.0%	0.0%
$135 - 149$	2.5%	0.9%

What does the table suggest about commute times in the two cities?

Solution

Boise commute times appear to be lower than New York commute times.
For Boise, the first two classes have

$$30.3\% + 45.5\% = 75.8\%$$

of all commute times.
For New York, the first two classes have only

$$8.6\% + 20.3\% = 28.9\%.$$

This suggests that commute times are much shorter in Boise than in New York.

8. Cumulative Frequency Distribution

A cumulative frequency distribution shows the running total of frequencies.

Cumulative Frequency

The cumulative frequency for a class is the sum of the frequency for that class and all previous classes.

Example 4: Complete the cumulative frequency distribution for the Los Angeles commute times.

Daily Commute Time (minutes)	Cumulative Frequency
Less than 15	
Less than 30	
Less than 45	
Less than 60	
Less than 75	
Less than 90	
Less than 105	

Solution

Using the frequencies

$$6, 18, 14, 5, 5, 1, 1,$$

we compute the running totals.

Daily Commute Time (minutes)	Cumulative Frequency
Less than 15	6
Less than 30	24
Less than 45	38
Less than 60	43
Less than 75	48
Less than 90	49
Less than 105	50

9. Critical Thinking: Understanding the Shape

Frequency distributions help us think about the shape of the data.

Approximately Normal Shape

A distribution is roughly bell-shaped when:

- frequencies start low;
- frequencies increase to one or two high frequencies;
- frequencies then decrease to low frequencies;
- the distribution is approximately symmetric.

Time	Frequency
0-14	2
15-29	4
30-44	10
45-59	18
60-74	10
75-89	4
90-104	2

This pattern starts low, increases to a maximum, and then decreases again.

10. Gaps in Frequency Distributions

Sometimes a frequency distribution contains a gap, where one or more classes have frequency 0.

Gaps

The presence of gaps can suggest that the data may come from two or more different populations.
However, the converse is not always true: data from different populations do not always create visible gaps.

Example 5: The table shows a frequency distribution of weights, in grams, of randomly selected pennies.

Weight of Penny (grams)	Frequency
$2.40 - 2.49$	18
$2.50 - 2.59$	19
$2.60 - 2.69$	0
$2.70 - 2.79$	0
$2.80 - 2.89$	0
$2.90 - 2.99$	2
$3.00 - 3.09$	25
$3.10 - 3.19$	8

What does the large gap suggest?

Solution

The large gap between the lighter pennies and heavier pennies suggests that the data may come from two different populations.
One explanation is that pennies made before 1983 and pennies made after 1983 have different compositions.

- Pennies made before 1983 were mostly copper.
- Pennies made after 1983 were mostly zinc.

This change in composition helps explain the gap in the weights.

Summary

Key Ideas

- A frequency distribution organizes data into classes and frequencies.
- Class limits describe the smallest and largest values allowed in each class.
- Class midpoints are found by averaging the lower and upper class limits.
- Class width is the difference between consecutive lower class limits.
- Relative frequency distributions use proportions or percentages.
- Cumulative frequency distributions show running totals.
- Frequency distributions help us recognize shape, unusual patterns, and gaps.