Chapter 2 Exploring Data with Tables and Graphs
Histograms (2.2)

Objectives

In this lesson, we will learn how to

- describe what a histogram is and how it differs from a frequency table;
- interpret the shape, center, spread, and outliers of a data set using a histogram;
- recognize common distribution shapes, including normal, skewed right, and skewed left;
- use normal quantile plots to help assess normality.

1. Why Histograms Are Useful

A frequency distribution is useful for summarizing data in a table.

A histogram is even easier to interpret because it displays the data visually.

Key Idea

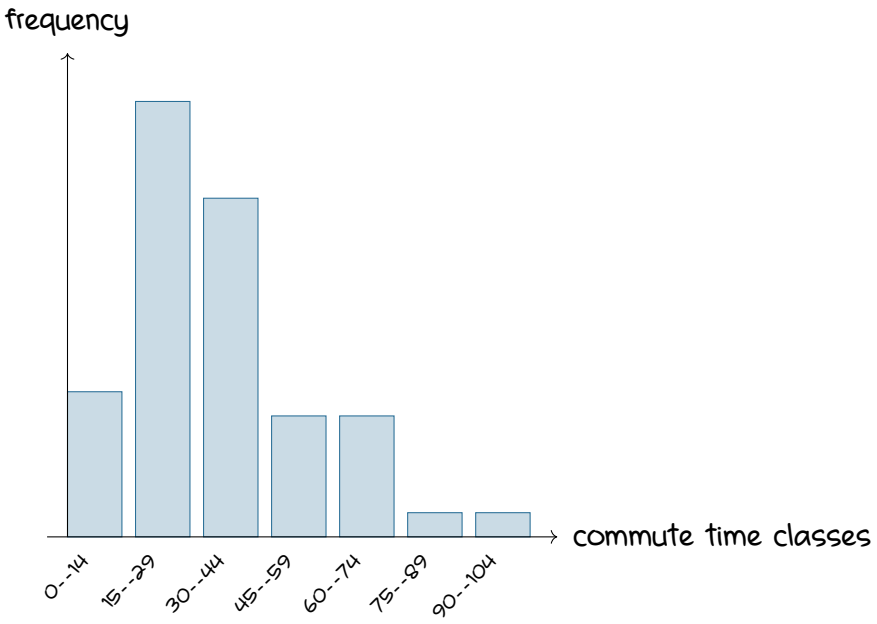
A histogram helps us see the overall pattern of a data set more quickly than a table of numbers.

2. Histogram

Histogram

A histogram is a graph consisting of bars of equal width drawn adjacent to each other, unless there are gaps in the data.

- The horizontal scale represents classes of quantitative data values.
- The vertical scale represents frequencies.
- The height of each bar corresponds to the frequency for that class.



Important Uses of a Histogram

A histogram can help us identify:

- the shape of the distribution;

- the approximate center of the data;
- the spread or variation of the data;
- possible outliers.

3. Relative Frequency Histogram

Relative Frequency Histogram

A relative frequency histogram has the same shape and horizontal scale as a regular histogram, but the vertical scale uses relative frequencies instead of actual frequencies.

For example, instead of showing that a class has frequency 18, we may show that the class contains

$$\frac{18}{50} = 0.36 = 36\%$$

of the data values.

Example 1: Suppose a data set has 50 commute times. One class has frequency 14. Find the relative frequency for that class.

Solution

The relative frequency is

$$\frac{14}{50} = 0.28.$$

As a percentage,

$$0.28 = 28\%.$$

So this class contains 28% of all commute times.

4. Critical Thinking: Interpreting Histograms

When interpreting a histogram, it is helpful to think about CVDOT.

CVDOT

- C: Center -- Where is the middle of the data?
- V: Variation -- How spread out are the values?
- D: Distribution -- What is the overall shape?
- O: Outliers -- Are there any values far away from the others?
- T: Time -- Are there changes over time or patterns related to time?

Example 2: A histogram has one high peak near the middle, and the bars decrease on both sides in a roughly symmetric way. What shape does this suggest?

Solution

This suggests that the data are approximately bell-shaped. In many statistics settings, this may suggest an approximately normal distribution.

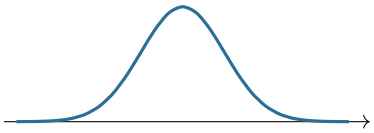
5. Common Distribution Shapes

Normal Distribution

Normal Distribution

If a histogram is roughly bell-shaped and approximately symmetric, we say that the data may have an approximately normal distribution.

roughly bell-shaped and symmetric

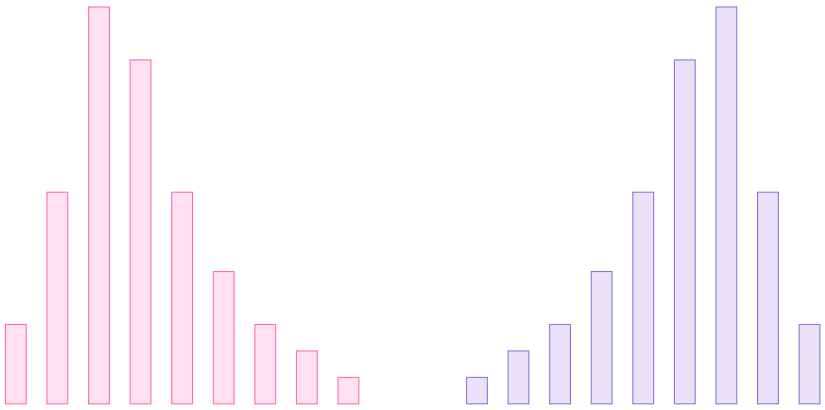


Skewness

Skewness

A distribution is skewed if it is not symmetric and extends more to one side than to the other.

- Data skewed to the right have a longer right tail.
- Data skewed to the left have a longer left tail.



Skewed right Skewed left

Example 3: A histogram has most of its bars on the left side, but a few values stretch far to the right. Is the distribution skewed left or skewed right?

Solution

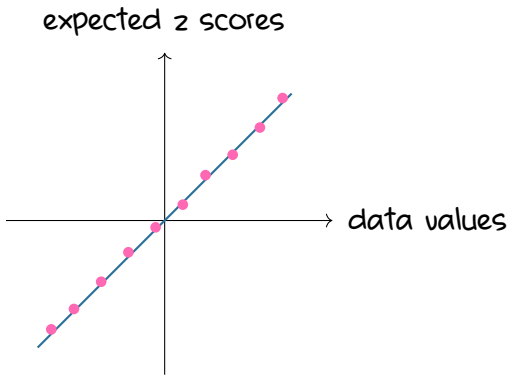
The distribution is skewed right because the long tail extends to the right.

6. Assessing Normality with Normal Quantile Plots

A histogram is useful, but sometimes we also use a normal quantile plot to help decide whether data appear to come from a normal distribution.

Normal Quantile Plot

A normal quantile plot is used to compare sample data with what we would expect from a normal distribution.



Normal quantile plot: points close to a straight-line pattern

Criteria for Assessing Normality

Normal distribution: The points in the normal quantile plot are reasonably close to a straight line, and the points do not show another systematic pattern.

Not normal: The population distribution is likely not normal if either of the following occurs:

- the points do not lie reasonably close to a straight line;
- the points show a systematic pattern that is not a straight-line pattern.

Example 4: A normal quantile plot shows points that curve upward rather than following a straight-line pattern. What should we conclude?

Solution

Because the points show a systematic curved pattern instead of a straight-line pattern, the data do not appear to come from a normal distribution.

7. Summary

Key Ideas

- A histogram is a graph of quantitative data using adjacent bars.
- The horizontal axis shows classes; the vertical axis shows frequencies or relative frequencies.
- Histograms help us understand center, variation, shape, and outliers.
- A normal distribution is roughly bell-shaped and symmetric.
- A right-skewed distribution has a long right tail; a left-skewed distribution has a long left tail.
- A normal quantile plot supports normality when the points are reasonably close to a straight line.