

Creating Distributions from Data

Frequency Distributions for Categorical Data

Relative Frequency and Percent Frequency Distributions Frequency Distributions for Quantitative Data Histograms

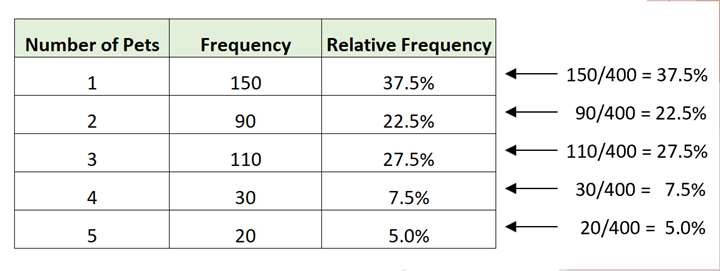
Cumulative Distributions



Creating Distributions from Data

* **Frequency distribution**:

A summary of data that shows the number (frequency) of observations in each of several nonoverlapping classes, typically referred to as **bins.**



Creating Distributions from Data

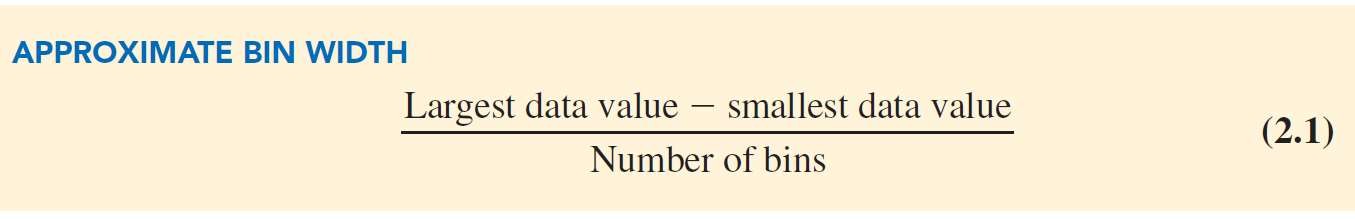
Table 2.5: Relative Frequency and Percent Frequency Distributions of Soft Drink Purchases



Creating Distributions from Data

* **Relative frequency distribution**:
  + A tabular summary of data showing the relative frequency for each bin.
* **Percent frequency distribution**:
  + Summarizes the percent frequency of the data for each bin.
  + Used to provide estimates of the relative likelihoods of different values of a random variable.

|  |  |  |
| --- | --- | --- |
| **Soft Drink** | **Relative**  **Frequency** | **Percent**  **Frequency (%)** |
| Coca-Cola | 0.38 | 38 |
| Diet Coke | 0.16 | 16 |
| Dr. Pepper | 0.10 | 10 |
| Pepsi | 0.26 | 26 |
| Sprite | 0.10 | 10 |
| Total | 1.00 | 100 |



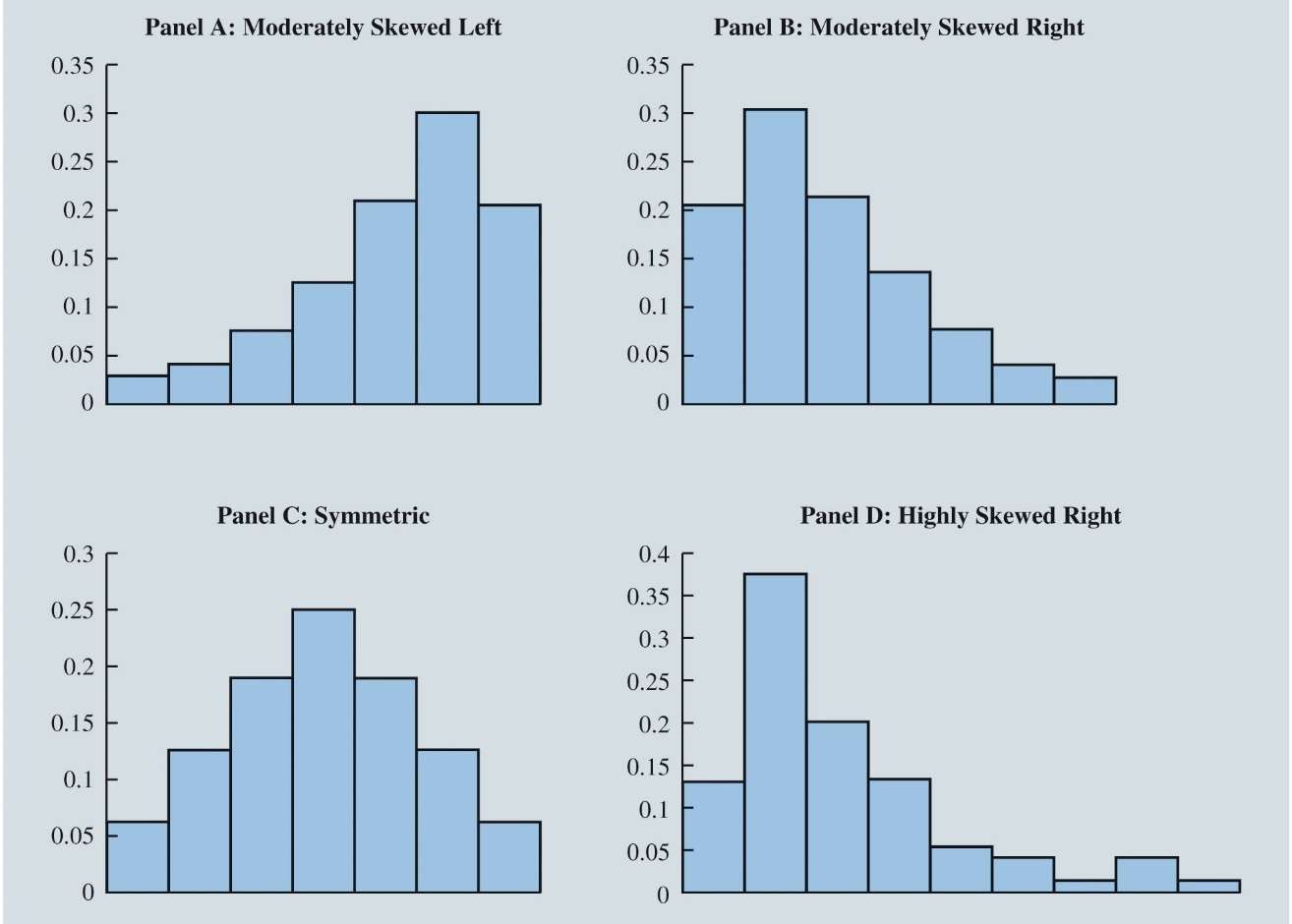
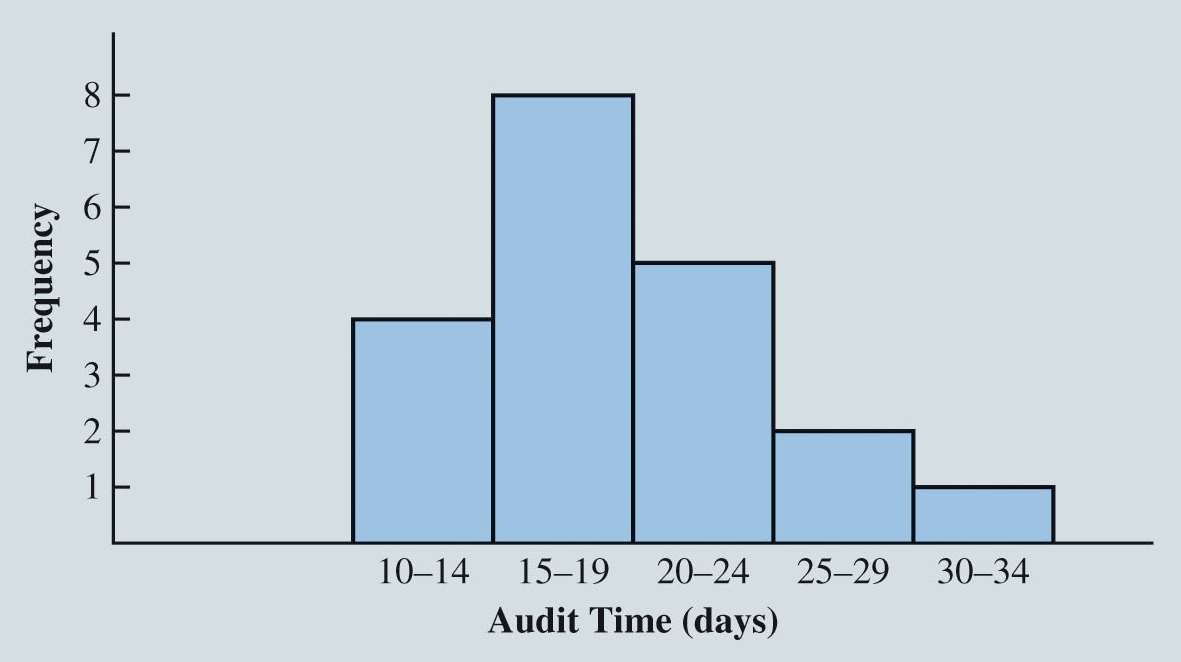
Creating Distributions from Data

* Three steps necessary to define the classes for a frequency distribution with quantitative data:
  1. Determine the number of nonoverlapping bins.
  2. Determine the width of each bin.
  3. Determine the bin limits.



Creating Distributions from Data

* **Histogram:** A common graphical presentation of quantitative data.
* Constructed by placing the variable of interest on the horizontal axis
* Frequency measure (absolute frequency, relative frequency, or percent frequency) on the vertical axis.



Creating Distributions from Data

Figure 2.12: Histogram for the Audit Time Data



Creating Distributions from Data

Histograms (cont.):

* Histograms provide information about the shape, or form, of a distribution.
* **Skewness**: Lack of symmetry.
* Skewness is an important characteristic of the shape of a distribution.



Creating Distributions from Data

* **Cumulative frequency distribution**: A variation of the frequency distribution that provides another tabular summary of quantitative data.
  + Uses the number of classes, class widths, and class limits developed for the frequency distribution.
  + Shows the number of data items with values less than or equal to the upper class limit of each class.



Creating Distributions from Data

Table 2.3: Data from a Sample of 50 Soft Drink Purchases

Coca-Cola Diet Coke Pepsi

Diet Coke Coca-Cola Coca-Cola Dr. Pepper Diet Coke Pepsi Pepsi Coca-Cola Dr. Pepper Sprite Coca-Cola Diet Coke Coca-Cola Coca-Cola

Sprite Coca-Cola Diet Coke Coca-Cola Diet Coke Coca-Cola Sprite Pepsi Coca-Cola Coca-Cola Coca-Cola Pepsi Coca-Cola Sprite

Dr. Pepper Pepsi

Diet Coke

Pepsi Coca-Cola Coca-Cola Coca-Cola Pepsi

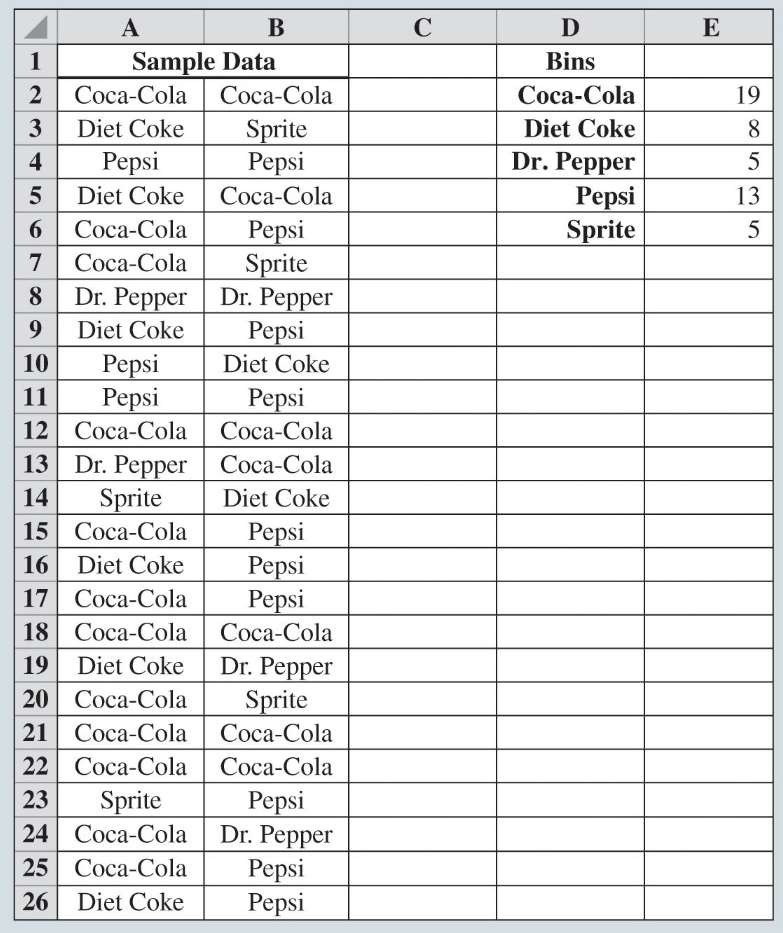
Dr. Pepper Coca-Cola Diet Coke Pepsi Pepsi Pepsi Pepsi Coca-Cola Dr. Pepper Pepsi Sprite



Creating Distributions from Data

Table 2.4: Frequency Distribution of Soft Drink Purchases

* The frequency distribution summarizes information about the popularity of the five soft drinks:
  + Coca-Cola is the leader.
  + Pepsi is second.
  + Diet Coke is third.
  + Sprite and Dr. Pepper are tied for fourth.



|  |  |
| --- | --- |
| **Soft Drink** | **Frequency** |
| Coca-Cola | 19 |
| Diet Coke | 8 |
| Dr. Pepper | 5 |
| Pepsi | 13 |
| Sprite | 5 |
| Total | 50 |



Creating Distributions from Data

Figure 2.10: Creating a Frequency Distribution for Soft Drinks Data in Excel



Creating Distributions from Data

Table 2.6: Year-End Audit Times (Days)

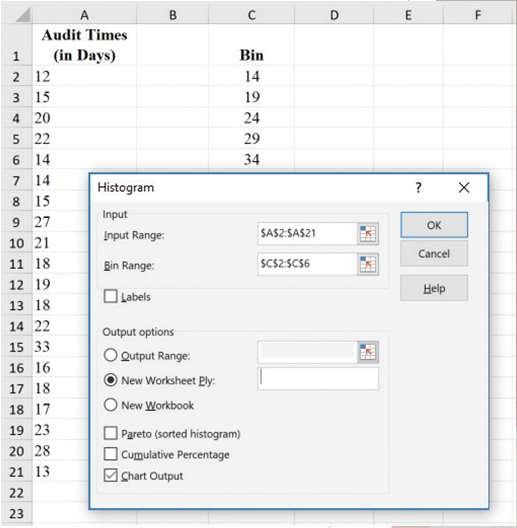
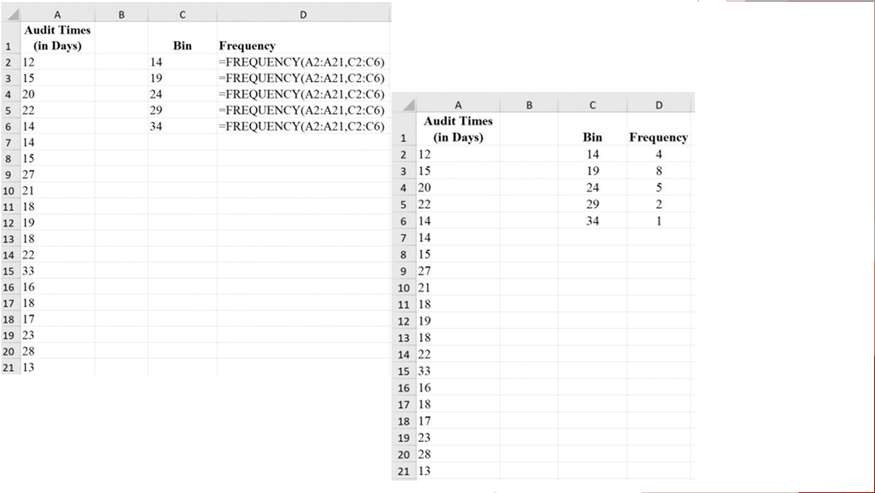


Creating Distributions from Data

Table 2.7: Frequency, Relative Frequency, and Percent Frequency Distributions for the Audit Time Data

|  |  |  |  |
| --- | --- | --- | --- |
| 12 | 14 | 19 | 18 |
| 15 | 15 | 18 | 17 |
| 20 | 27 | 22 | 23 |
| 22 | 21 | 33 | 28 |
| 14 | 18 | 16 | 13 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Audit Times**  **(days)** | **Frequency** | **Relative**  **Frequency** | **Percent**  **Frequency** |
| 10–14 | 4 | 0.20 | 20 |
| 15–19 | 8 | 0.40 | 40 |
| 20–24 | 5 | 0.25 | 25 |
| 25–29 | 2 | 0.10 | 10 |
| 30–34 | 1 | 0.05 | 5 |



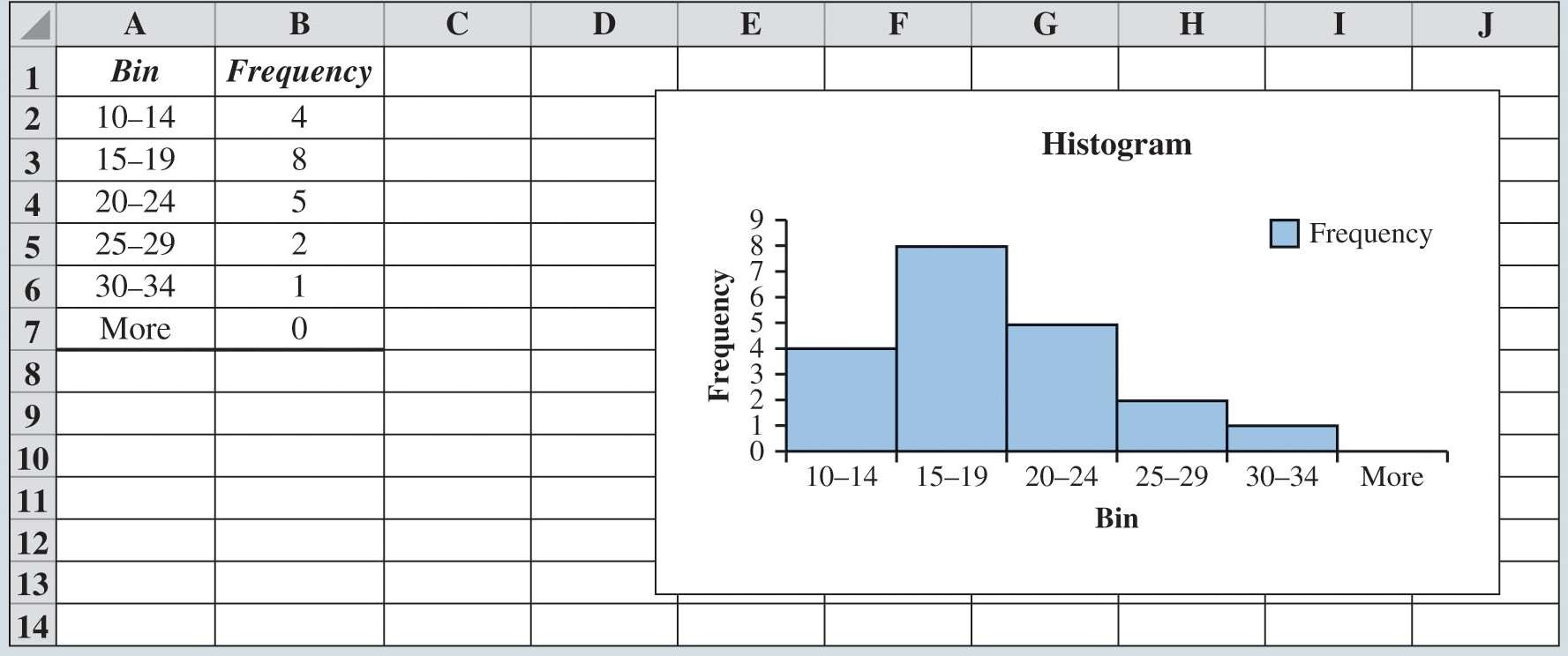
Creating Distributions from Data

Figure 2.11: Using Excel to Generate a Frequency Distribution for Audit Times Data



Creating Distributions from Data

Figure 2.13: Creating a Histogram for the Audit Time Data Using Data Analysis Toolpak in Excel



Creating Distributions from Data

Table 2.8: Cumulative Frequency, Cumulative Relative Frequency, and Cumulative Percent Frequency Distributions for the Audit Time Data



Creating Distributions from Data

Figure 2.14: Completed Histogram for the Audit Time Data Using Data Analysis ToolPak in Excel

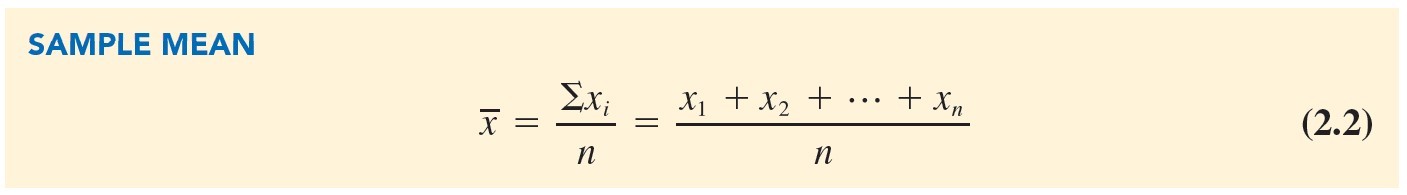
|  |  |  |  |
| --- | --- | --- | --- |
| **Audit Time (days)** | **Cumulative Frequency** | **Cumulative**  **Relative Frequency** | **Cumulative**  **Percent Frequency** |
| Less than or equal to 14 | 4 | 0.20 | 20 |
| Less than or equal to 19 | 12 | 0.60 | 60 |
| Less than or equal to 24 | 17 | 0.85 | 85 |
| Less than or equal to 29 | 19 | 0.95 | 95 |
| Less than or equal to 34 | 20 | 1.00 | 100 |



Measures of Location

Mean (Arithmetic Mean) Median

Mode Geometric Mean



Measures of Location

Mean/Arithmetic Mean

Average value for a variable.

The mean is denoted by

n = sample size.

*x*1 = variable of *x* for the first observation.

*x*2 = variable of *x* for the second observation.

*xi* = variable of *x* for the i*th* observation.

*x*.



Measures of Location

Computation of Sample Mean

* Illustration: Computation of the mean home selling price for the sample of 12 home sales.

*x* = *xi* = *x*1 + *x*2 ++ *x*12

*n*

12

= 138,000 + 254,000 +456,250

12

= 2,639,250 = 219,937.50

12



Measures of Location

* Median:

Value in the middle when the data are arranged in ascending order.

* + Middle value, for an odd number of observations.
  + Average of two middle values, for an even number of observations.



Measures of Location

Computation of Sample Median

* Illustration: When the number of observations are odd:
  + Consider the class size data for a sample of five college classes:

46 54 42 46 32

* + Arrange the class size data in ascending order:

32 42 46 46 54

* + Middlemost value in the data set = 46.
  + Median is 46.



Measures of Location

* Illustration: When the number of observations are even:
  + Consider the data on home sales in Cincinnati, Ohio, Suburb (Table 2.9).
* Arrange the data in ascending order:

108,000 138,000 138,000 142,000 186,000 199,500 208,000 254,000 254,000

257,500 298,000 456,250

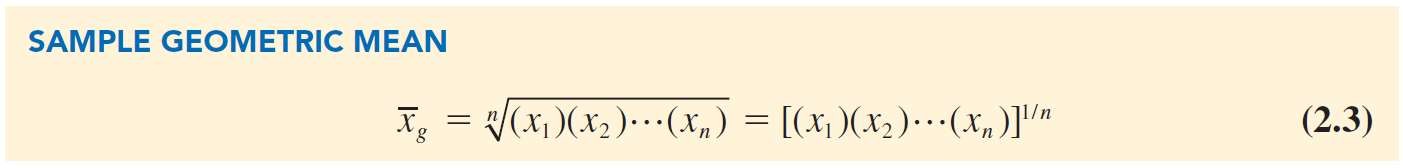
* Median = average of two middle values:

Median =

199,500 + 208,000

2

= 203,750



Measures of Location

* **Mode:**

Value that occurs most frequently in a data set.

* Consider the class size data:

32 42 46 46 54

* Observe: 46 is the only value that occurs more than once.
  + Mode is 46.
* Multimodal data: Data contain at least two modes.
* Bimodal data: Data contain exactly two modes.



Measures of Location

* **Geometric mean:**

A measure of location that is calculated by finding the *n*th root of the product of *n* values

* Used in analyzing growth rates in financial data.

Measures of Location



* We will determine the mean rate of growth for the fund over the 10-year period.

|  |  |  |
| --- | --- | --- |
| **Year** | **Return (%)** | **Growth Factor** |
| 1 | −22.1 | 0.779 |
| 2 | 28.7 | 1.287 |
| 3 | 10.9 | 1.109 |
| 4 | 4.9 | 1.049 |
| 5 | 15.8 | 1.158 |
| 6 | 5.5 | 1.055 |
| 7 | −37.0 | 0.630 |
| 8 | 26.5 | 1.265 |
| 9 | 15.1 | 1.151 |
| 10 | 2.1 | 1.021 |

Percentage Annual Returns and Growth Factors for the Mutual Fund Data:



Measures of Location

Computation of Geometric Mean:

* Product of the growth factors:

$100 (0.779)(1.287)(1.109)(1.049)(1.158)(1.055)(0.630)(1.265)(1.151)(1.021)

= $100(1.335) = $133.45

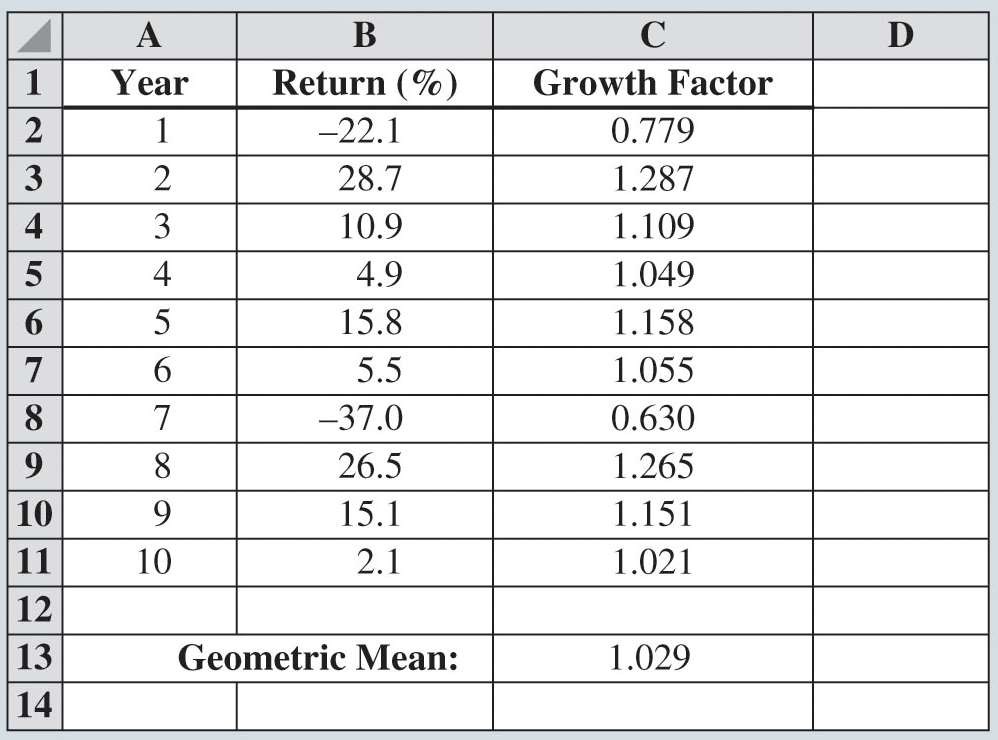
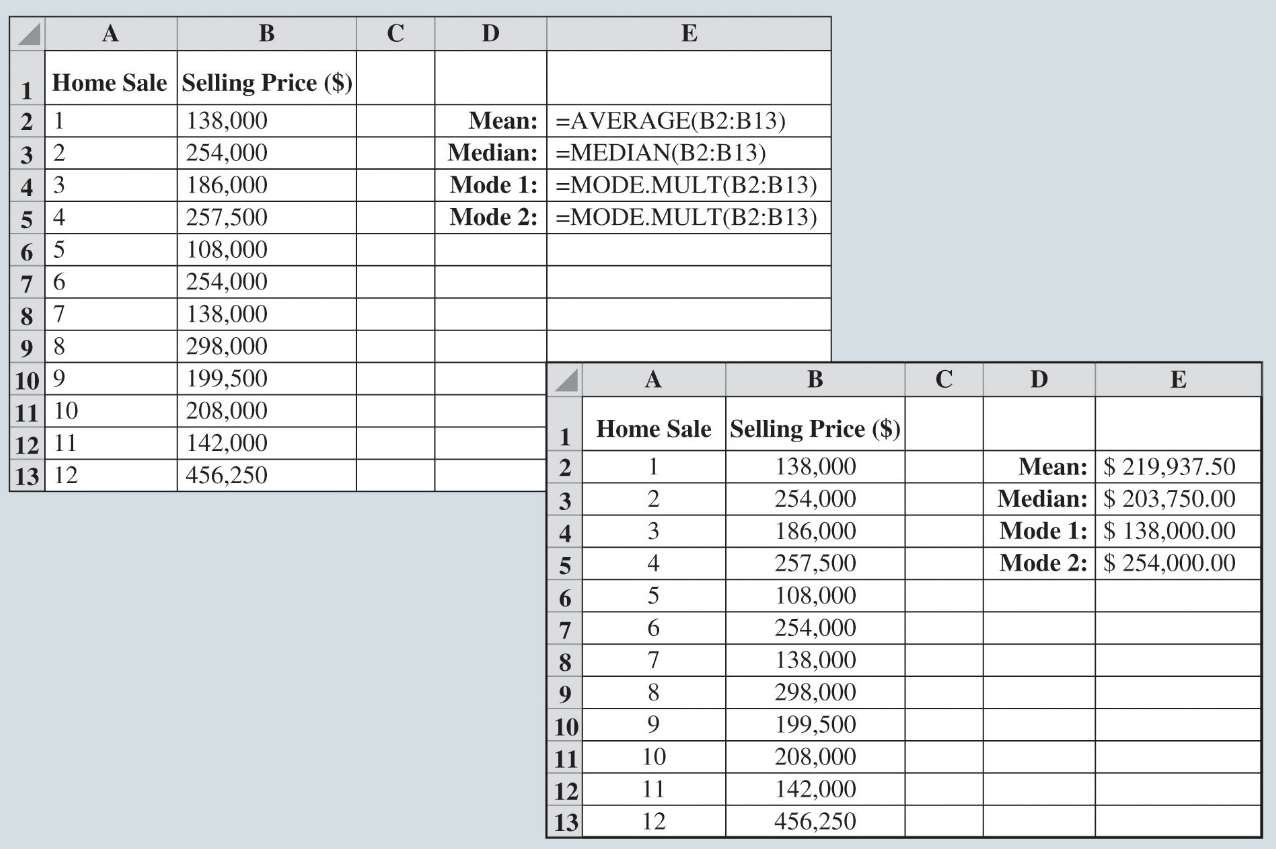
* Geometric mean of the growth factors:

*xg* = 1.335 = 1.029.

10

* Conclude that annual returns grew at an average annual rate of

(1.029  1)100% or 2.9%.



Measures of Location

Figure 2.16: Calculating the Mean, Median, and Modes for the Home Sales Data using Excel



Measures of Location

Figure 2.17: Calculating the Geometric Mean for the Mutual Fund Data Using Excel