

Dr. Saeed A. Dobbah Alghamdi

---

Department of Statistics

Faculty of Sciences

Building 90, Office 26F41

King Abdulaziz University

<http://saalghamdy.kau.edu.sa>



Dr. Saeed's Website

# Statistics for Business & Economics

---

David R. Anderson  
Dennis J. Sweeney  
Thomas A. Williams

*Copyright © South-Western Cengage Learning All rights reserved.*

# Descriptive Statistics: Tabular and Graphical Presentations

## Chapter 2





# Learning Objectives

1. For single qualitative variables: frequency, relative frequency and percentage frequency distributions; bar charts and pie charts.
2. For single quantitative variables: frequency, relative frequency and percentage frequency distributions; cumulative frequency, relative cumulative frequency and percentage cumulative frequency distributions; stem-and-leaf plots, histograms and cumulative distribution plots (ogives).
3. For pairs of qualitative and quantitative data: cross-tabulations, with row and column percentages.
4. For pairs of quantitative variables: scatter diagrams.

# Frequency Table

**FREQUENCY TABLE** is a listing of qualitative data values (classes) along with their corresponding frequencies or counts.

**EXAMPLE:** The following data is the blood type of 28 patients admitted to a hospital in a given period of time.

A	B	B	AB	O	O	A	O	O	B	A	B	O	AB
O	AB	B	B	A	A	O	B	B	O	O	O	A	O

Classes	Frequency	Percent
A	6	$\frac{6}{28} \times 100 = 21\%$
B	8	29%
O	11	39%
AB	3	11%
<b>Total</b>	<b>28</b>	<b>100%</b>

$$\frac{\text{Frequency}}{\text{Total}} \times 100 = \frac{f}{\sum f} \times 100 = \frac{f}{n} \times 100$$

Sample Size



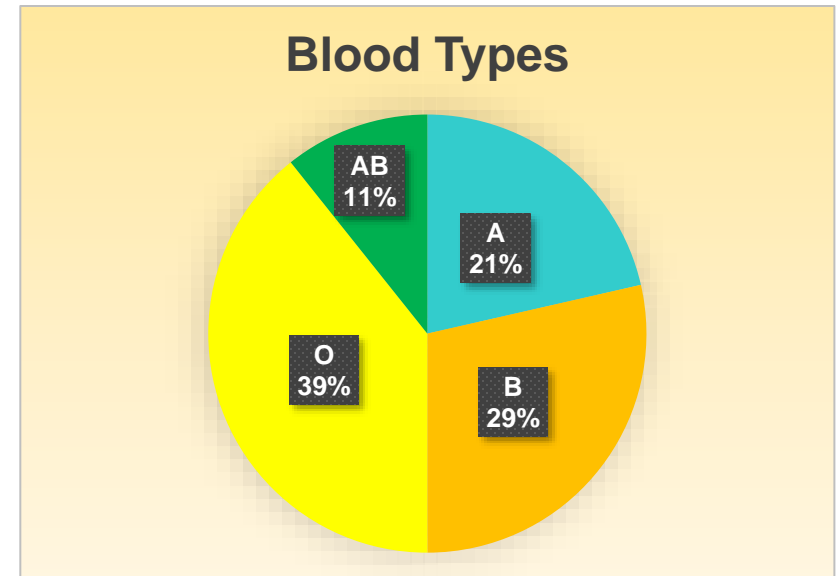
# Pie Charts

**PIE CHART** is a circle that is divided into sections according to the frequency of each class in the frequency table.

Each section in the pie chart is determined by its degree,  $Degree = \frac{f}{n} \times 360$ .

It shows the proportion or percent that each class represents of the total number of frequencies. It is the best graph for displaying the nominal level of qualitative data.

Class	Frequency	Degree
A	6	$\frac{6}{28} \times 360 = 77$
B	8	103
O	11	141
AB	3	39
<b>Total</b>	<b>28</b>	<b>360</b>

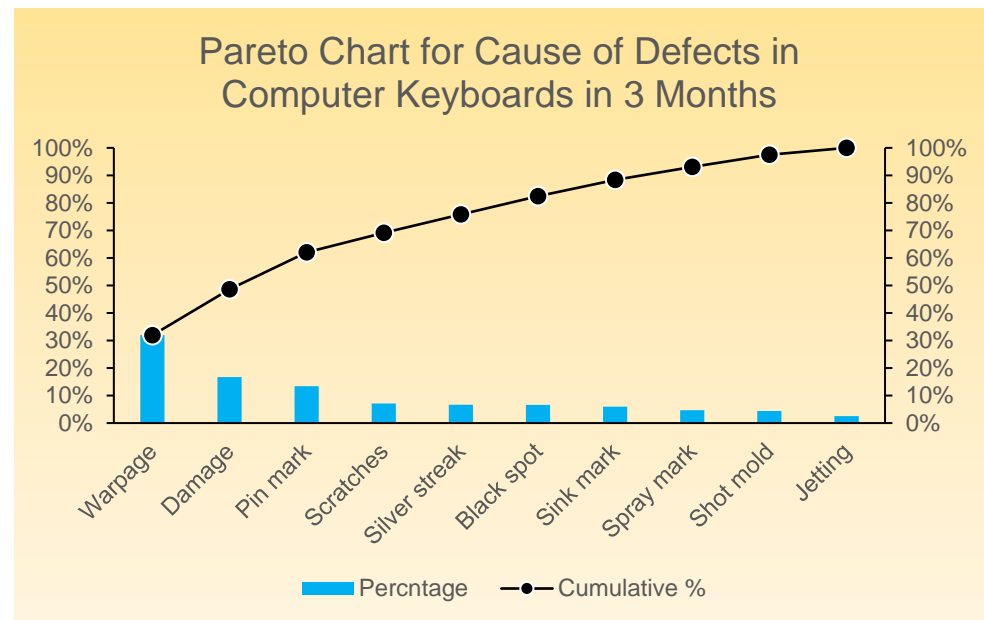


# Pareto Charts

**PARETO CHART** is vertical bars of various heights in descending order that reflect the frequencies or percentages of the classes along with a cumulative percentage line on the same chart.

It is the best graph for displaying qualitative data when we want to separate the “vital few” from the “trivial many”. It enables us to focus on the important categories.

Cause	Frequency	Percentage	Cumulative %
Warpage	1988	32%	32%
Damage	1039	17%	49%
Pin mark	834	13%	62%
Scratches	443	7%	69%
Silver streak	414	7%	76%
Black spot	412	7%	82%
Sink mark	371	6%	88%
Spray mark	292	5%	93%
Shot mold	274	4%	97%
Jetting	156	3%	100%
Total	6223	100%	

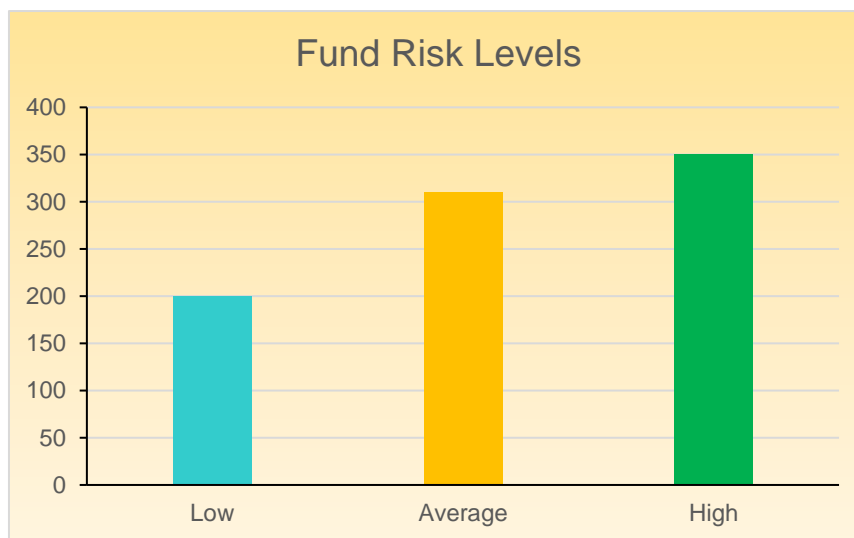


# Bar Charts

**BAR CHART** is vertical bars of various heights that reflect the frequencies of the classes. The classes are reported on the horizontal axis and the class frequencies on the vertical axis.

It shows the amount, frequency or percentage, of values falling into a class and allows us to compare the amounts in different classes. It is the best graph for displaying the ordinal level of qualitative data or discrete type of data.

<b>Fund Risk Level</b>	<b>Number of Funds</b>
Low	200
Average	310
High	350
<b>Total</b>	<b>860</b>





# Frequency Distribution

**FREQUENCY DISTRIBUTION** is a listing of quantitative data values either individually or by intervals along with their corresponding frequencies or counts.

If the range of the data set is small, classes consisting individual value should be used.

**EXAMPLE:** The following data represent the number of miles per gallon (mpg) that 30 selected four-wheel-drive sports utility vehicles obtained in city driving.

12 17 12 14 16 18 16 18 12 16  
 17 15 15 16 12 15 16 16 12 14  
 15 12 15 15 19 13 16 18 16 14

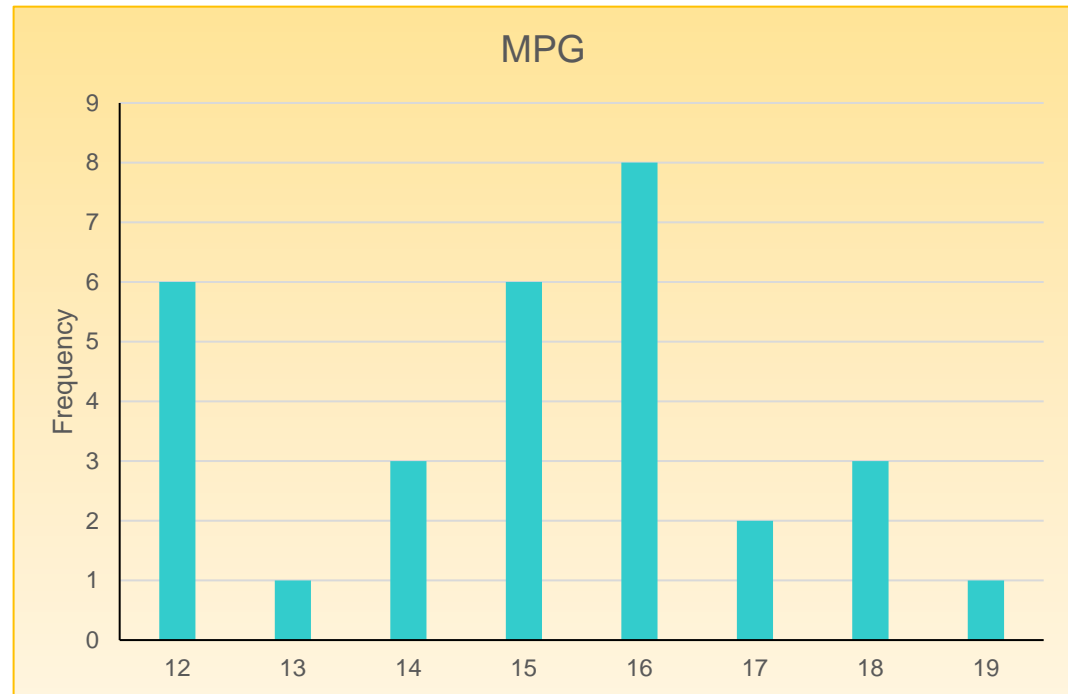
MPG	12	13	14	15	16	17	18	19
Frequency	6	1	3	6	8	2	3	1

# Bar Charts

**BAR CHART** is vertical bars of various heights that reflect the frequencies of the classes. The classes are reported on the horizontal axis and the class frequencies on the vertical axis.

It shows the amount, frequency or percentage, of values falling into a class and allows us to compare the amounts in different classes. It is the best graph for displaying the ordinal level of qualitative data or discrete type of data.

MPG	Frequency
12	6
13	1
14	3
15	6
16	8
17	2
18	3
19	1



# Frequency Distribution

**FREQUENCY DISTRIBUTION** is a listing of quantitative data values either individually or by intervals along with their corresponding frequencies or counts.

If the range of the data set is large, the data must be grouped into classes that have intervals with more than one unit in width.

Systolic Blood Pressure	Frequency	Class Midpoint
90 up to 100	1	95
100 up to 110	4	105
110 up to 120	17	115
120 up to 130	12	125
130 up to 140	5	135
140 up to 150	0	145
150 up to 160	1	155

**Class interval:** The class interval is obtained by subtracting the lower limit of a class from the lower limit of the next class.

**Class frequency:** The number of observations in each class.

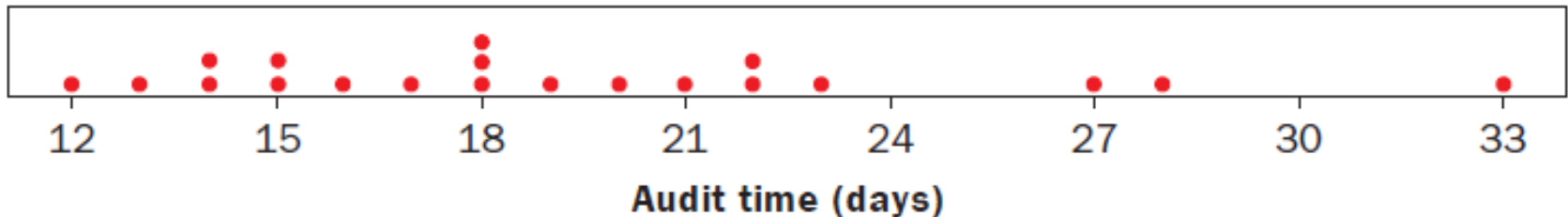
**Class midpoint:** A point that divides a class into two equal parts. This is the average of the upper and lower class limits.

# Dot plot

A **dot plot** is a horizontal axis shows the range of values for the observations where each data value is represented by a dot placed above the axis.

**TABLE 2.4** Year-end audit times (in days)

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



# Stem-and-Leaf

- **Stem-and-leaf display** is a statistical technique to present a set of data. Each numerical value is divided into two parts. The leading digit(s) becomes the stem and the trailing digit the leaf. The stems are located along the vertical axis, and the leaf values are stacked against each other along the horizontal axis.
- Two disadvantages to organizing the data into a frequency distribution:
  - (1) The exact identity of each value is lost
  - (2) Difficult to tell how the values within each class are distributed.

## EXAMPLE

Listed in Table 4–1 is the number of 30-second radio advertising spots purchased by each of the 45 members of an Automobile Dealers Association last year. Organize the data into a stem-and-leaf display. Around what values do the number of advertising spots tend to cluster? What is the fewest number of spots purchased by a dealer? The largest number purchased?

**TABLE 4–1** Number of Advertising Spots Purchased by Members of the Greater Buffalo Automobile Dealers Association

96	93	88	117	127	95	113	96	108	94	148	156
139	142	94	107	125	155	155	103	112	127	117	120
112	135	132	111	125	104	106	139	134	119	97	89
118	136	125	143	120	103	113	124	138			

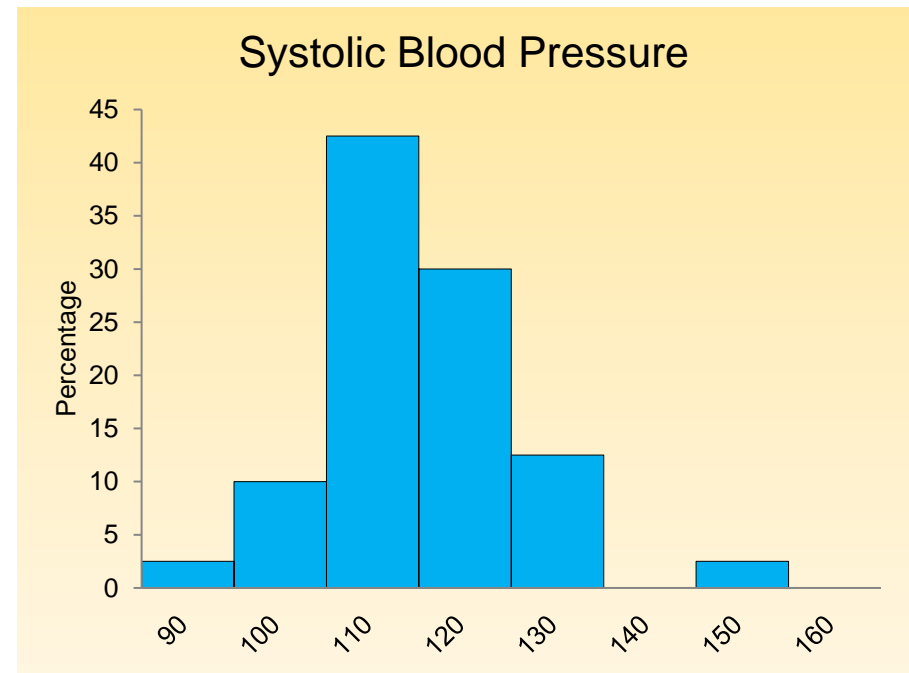
Frequency	Stem	Leaf	
2	8	8 9	
7	9	3 4 4 5 6 6 7	
6	10	3 3 4 6 7 8	
9	11	1 2 2 3 3 7 7 8 9	
8	12	0 0 4 5 5 5 7 7	
7	13	2 4 5 6 8 9 9	
3	14	2 3 8	
3	15	5 5 6	
45			

# Histogram

**HISTOGRAM** is a graph of vertical bars with various heights that reflect the frequencies of the classes. The classes are marked on the horizontal axis and the class frequencies on the vertical axis. There are no gaps between adjacent bars as in a bar chart.

It shows the amount, frequency or percentage, of values falling into a class interval and allows us to compare the amounts in different classes. It is the best graph for displaying the continuous type of data.

Systolic Blood Pressure	Frequency	%
90 up to 100	1	2.5
100 up to 110	4	10
110 up to 120	17	42.5
120 up to 130	12	30
130 up to 140	5	12.5
140 up to 150	0	0
150 up to 160	1	2.5

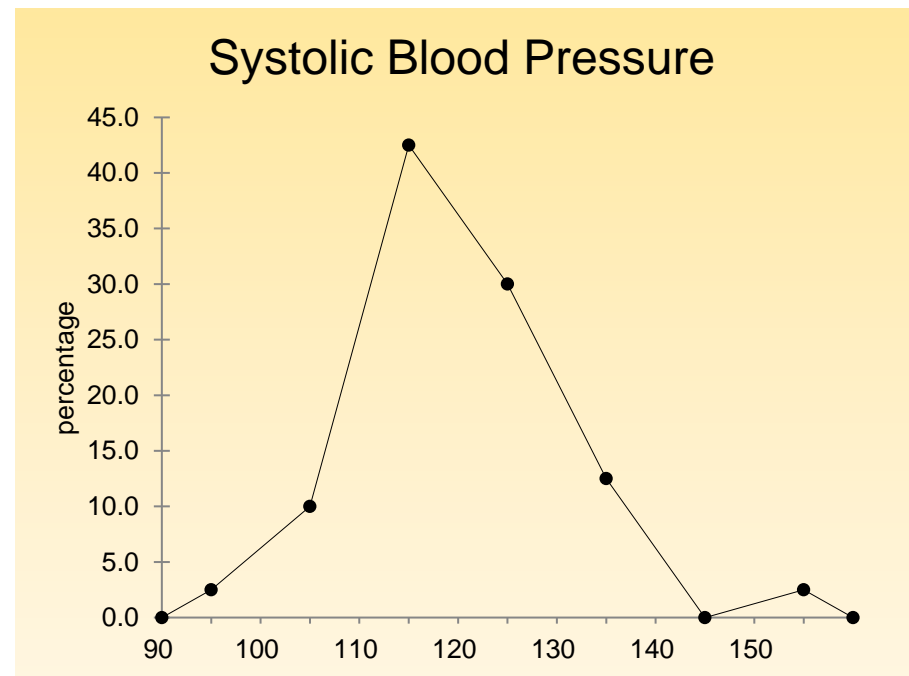


# Frequency Polygon

**FREQUENCY POLYGON** is a graph that displays the data by using lines that connect points plotted for the frequencies at the midpoints of the classes. The heights of the points reflect the frequencies of the classes.

It shows the amount, frequency or percentage, of values falling into a class interval and allows us to show the shape of a distribution. It is the best graph for displaying the continuous type of data.

Systolic Blood Pressure	Frequency	%
90 up to 100	1	2.5
100 up to 110	4	10
110 up to 120	17	42.5
120 up to 130	12	30
130 up to 140	5	12.5
140 up to 150	0	0
150 up to 160	1	2.5



# Histogram Versus Frequency Polygon

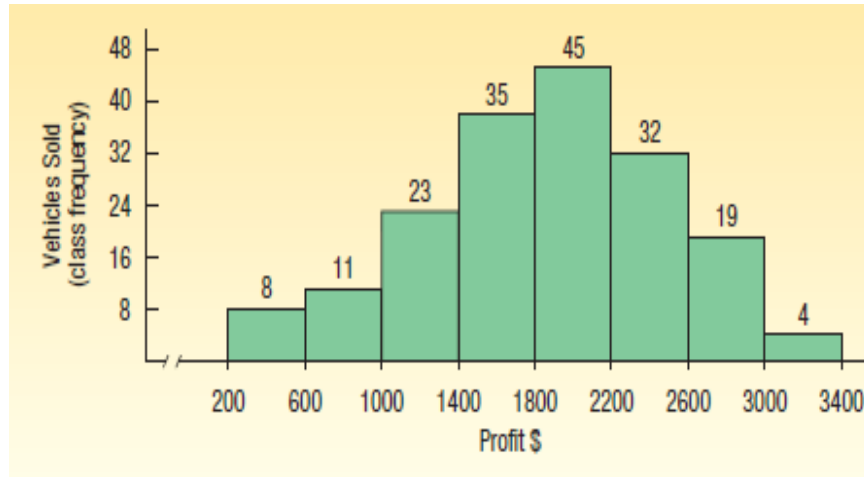


CHART 2-6 Distribution of Profit at Applewood Auto Group and Fowler Motors

- Both provide a quick picture of the main characteristics of the data (highs, lows, points of concentration, etc.)
- The histogram has the advantage of depicting each class as a rectangle, with the height of the rectangular bar representing the number in each class.
- The frequency polygon has an advantage over the histogram. It allows us to compare directly two or more frequency distributions.



# Relative Class Frequencies

- Class frequencies can be converted to **relative class frequencies** to show the fraction of the total number of observations in each class.
- A relative frequency captures the relationship between a class total and the total number of observations.

**TABLE 2–2** Relative Frequency Table of Vehicles Sold By Type At Whitner Autoplex Last Month

Vehicle Type	Number Sold	Relative Frequency
Domestic	50	0.625
Foreign	30	0.375
Total	<u>80</u>	<u>1.000</u>

Relative frequency for the number of miles per gallon (mpg) that 30 selected four-wheel-drive sports utility vehicles

MPG	12	13	14	15	16	17	18	19	Total
Frequency	6	1	3	6	8	2	3	1	30
Relative Frequency	0.2	0.03	0.1	0.2	0.27	0.07	0.1	0.03	1

# Relative Frequency Distribution

To convert a frequency distribution to a *relative* frequency distribution, each of the class frequencies is divided by the total number of observations.

**TABLE 2–8** Relative Frequency Distribution of Profit for Vehicles Sold Last Month at Applewood Auto Group

Profit	Frequency	Relative Frequency	Found by
\$ 200 up to \$ 600	8	.044	8/180
600 up to 1,000	11	.061	11/180
1,000 up to 1,400	23	.128	23/180
1,400 up to 1,800	38	.211	38/180
1,800 up to 2,200	45	.250	45/180
2,200 up to 2,600	32	.178	32/180
2,600 up to 3,000	19	.106	19/180
3,000 up to 3,400	4	.022	4/180
Total	180	1.000	

**NOTE:** Graphs of relative frequencies are used instead of frequencies when the proportion of data values that fall into a given class is more important than the actual number of data values that fall into that class.

# Cumulative Frequency Distribution

**CUMULATIVE FREQUENCY DISTRIBUTION** is a distribution that shows the number of data values less than or equal to a specific value (usually an upper boundary).

The **cumulative frequency** for a class is the sum of the frequencies for that class and all the previous classes which gives an ascending cumulative frequency.

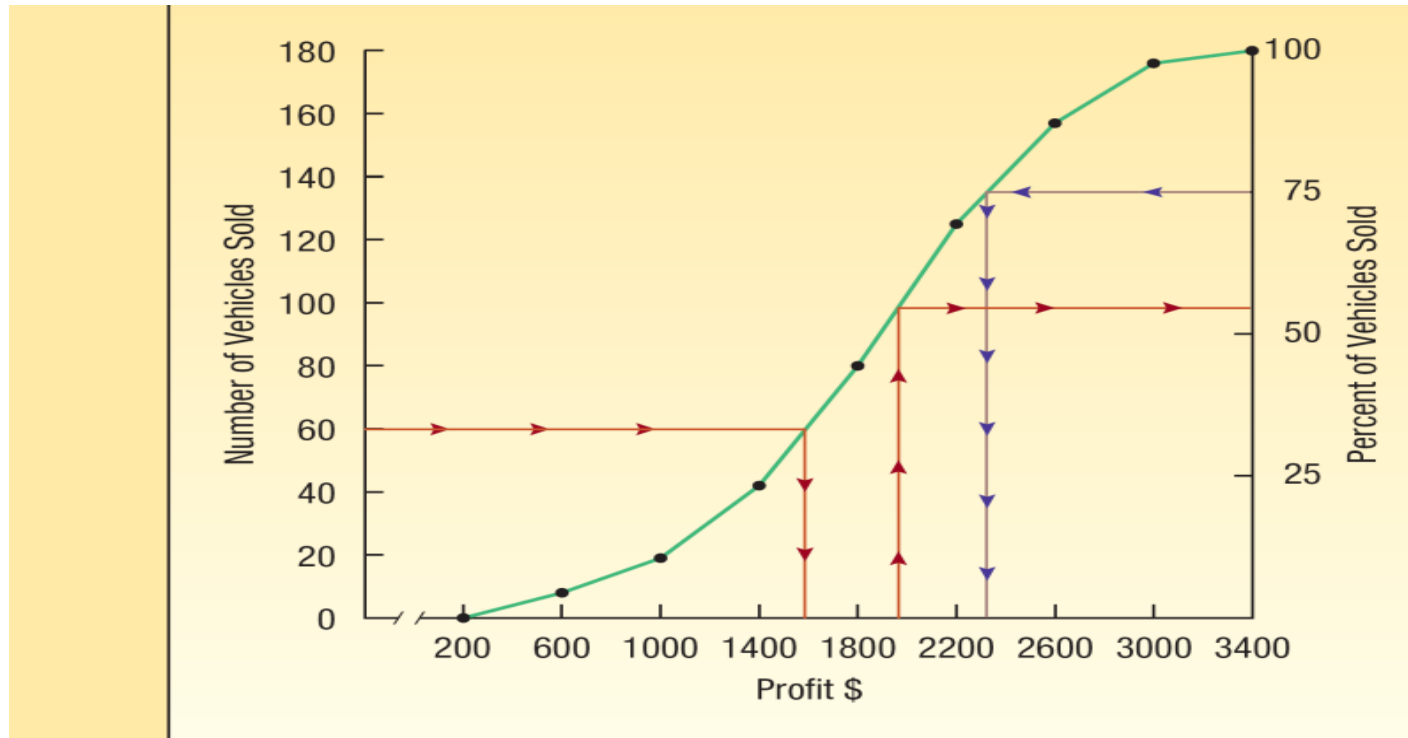
**TABLE 2–9** Cumulative Frequency Distribution for Profit on Vehicles Sold Last Month at Applewood Auto Group

Profit	Frequency	Cumulative Frequency	Found by
\$ 200 up to \$ 600	8	8	8
600 up to 1,000	11	19	$8 + 11$
1,000 up to 1,400	23	42	$8 + 11 + 23$
1,400 up to 1,800	38	80	$8 + 11 + 23 + 30$
1,800 up to 2,200	45	125	$8 + 11 + 23 + 30 + 45$
2,200 up to 2,600	32	157	$8 + 11 + 23 + 30 + 45 + 32$
2,600 up to 3,000	19	176	$8 + 11 + 23 + 30 + 45 + 32 + 19$
3,000 up to 3,400	4	180	$8 + 11 + 23 + 30 + 45 + 32 + 19 + 4$
Total	180		

# Cumulative Frequency Polygon (ogive)

**CUMULATIVE FREQUENCY GRAPH** or **ogive** represents the cumulative frequencies for the classes in a **frequency distribution** of continuous data.

Ogives are useful in determining the number of values below or above a specific value or between two specific values within the range of the data set.



**CHART 2-7** Cumulative Frequency Distribution for Vehicle Profit

# Cross-tabulation

A **cross-tabulation** is a tabular summary of data for two variables.

Consider the following data from a consumer restaurant review, based on a sample of 300 restaurants in a large European city.

**TABLE 2.10** Cross-tabulation of quality rating and meal price for 300 restaurants

Quality rating	Meal price				Total
	€10–19	€20–29	€30–39	€40–49	
Disappointing	42	40	2	0	84
Good	34	64	46	6	150
Excellent	2	14	28	22	66
<b>Total</b>	78	118	76	28	300

# Scatter Diagram

A **scatter diagram** is a graphical presentation of the relationship between two quantitative variables, and a **trend line** is a line that provides an approximation of the relationship.

Consider the advertising sales relationship for a hi-fi equipment store. On ten occasions during the past three months, the store used weekend television commercials to promote sales at its stores. The managers want to investigate whether a relationship exists between the number of commercials shown and sales at the store during the following week.

