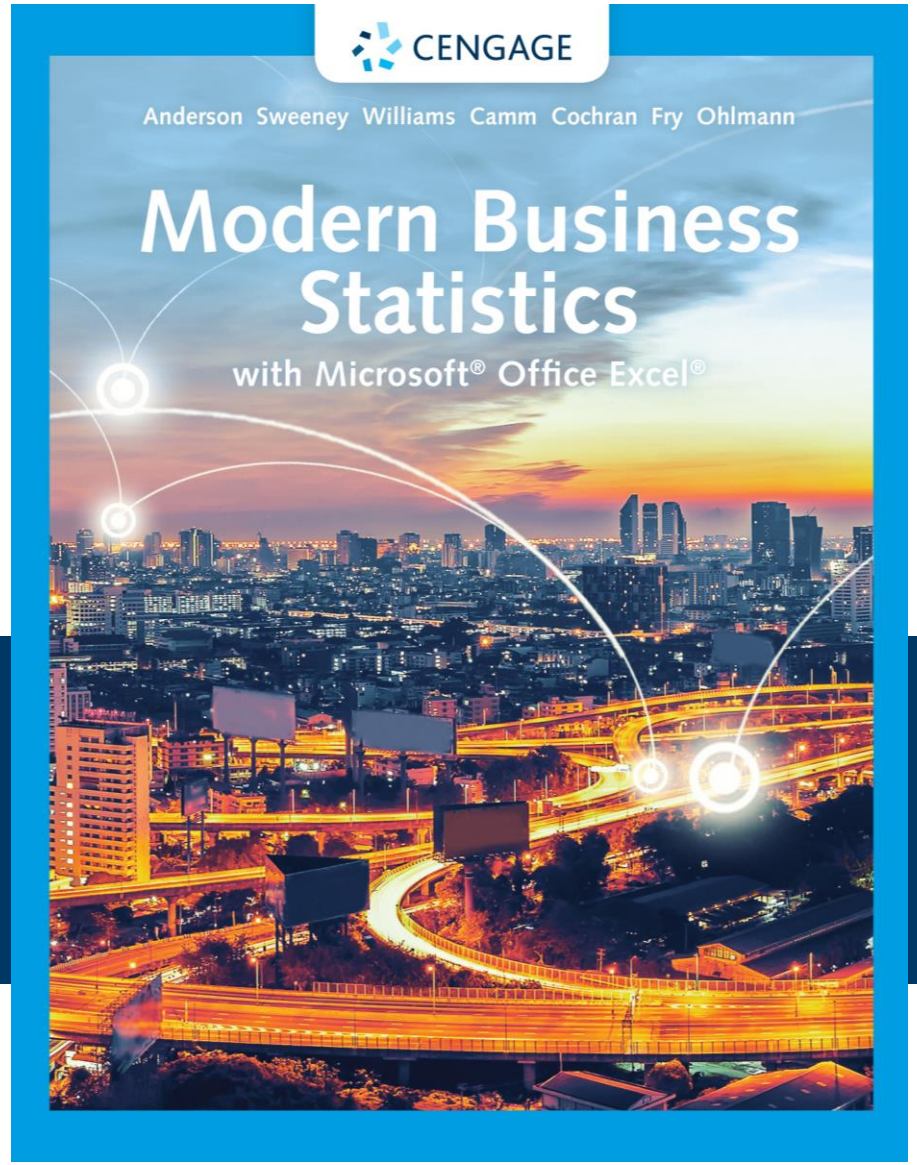


Essentials of Modern Business Statistics, 8e

Anderson, Sweeney, Williams,
Camm, Cochran, Fry, Ohlmann

Chapter 1: Data and Statistics

- Statistics
- Applications in Business and Economics
- Data Sources
- Descriptive Statistics
- Statistical Inference
- Statistical Analysis Using Microsoft Excel
- Analytics
- Big Data and Data Mining
- Ethical Guidelines for Statistical Practice

CENGAGE

What Is Statistics?

- The term “statistics” can refer to numerical facts such as averages, medians, percentages, and maximums that help us understand a variety of business and economic situations.
- Statistics can also refer to the art and science of collecting, analyzing, presenting, and interpreting data.

Applications in Business and Economics (1 of 2)

- Accounting
  - Public accounting firms use statistical sampling procedures when conducting audits for their clients.
- Economics
  - Economists use statistical information in making forecasts about the future of the economy or some aspect of it.
- Finance
  - Financial advisors use price-earnings ratios and dividend yields to guide their investment advice.

Applications in Business and Economics (2 of 2)

- Marketing
  - Electronic point-of-sale scanners at retail checkout counters are used to collect data for a variety of marketing research applications.
- Production
  - A variety of statistical quality control charts are used to monitor the output of a production process.
- Information Systems
  - A variety of statistical information helps administrators assess the performance of computer networks.

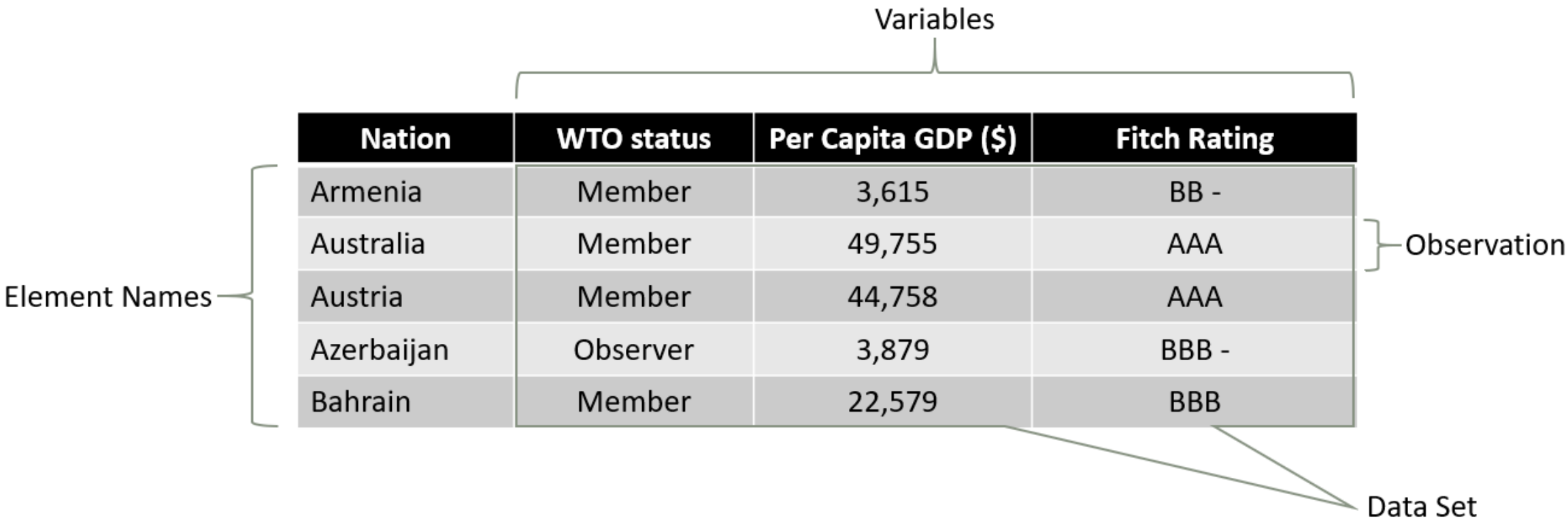
Data and Data Sets

- Data are the facts and figures collected, analyzed, and summarized for presentation and interpretation.
- All the data collected in a particular study are referred to as the data set for the study.

Elements, Variables, and Observations

- Elements are the entities on which data are collected.
- A variable is a characteristic of interest for the elements.
- The set of measurements obtained for a particular element is called an observation.
- A data set with *n* elements contains *n* observations.
- The total number of data values in a complete data set is the number of elements multiplied by the number of variables.

Data, Data Sets, Elements, Variables, and Observations



Scales of Measurement (1 of 10)

- Scales of measurement include
  - Nominal
  - Ordinal
  - Interval
  - Ratio
- The scale determines the amount of information contained in the data.
- The scale indicates the data summarization and statistical analyses that are most appropriate.

Scales of Measurement (2 of 10)

- **Nominal**
  - Data are labels or names used to identify an attribute of the element.
  - A nonnumeric label or numeric code may be used.

Scales of Measurement (3 of 10)

- **Nominal**

  Example:

  The WTO status category for the nations in the previous example is classified using nonnumerical labels—“member” and “observer.”

  Alternatively, a numeric code could be used for the WTO status variable by letting 1 denote a member nation and 2 denote an observer nation.

Scales of Measurement (4 of 10)

- **Ordinal**
  - The data have the properties of nominal data and the order or rank of the data is meaningful.
  - A nonnumeric label or numeric code may be used.

Scales of Measurement (5 of 10)

- **Ordinal**

  Example:

  The nonnumeric rating labels from AAA to F used for Fitch rating. These can be rank ordered from best credit rating AAA to poorest credit rating F.

  Numerical code can also be used—Class rank of a student in school.

Scales of Measurement (6 of 10)

- **Interval**
  - The data have the properties of ordinal data, and the interval between observations is expressed in terms of a fixed unit of measure.
  - Interval data are always numeric.

Scales of Measurement (7 of 10)

- **Interval**

  Example:

  Melissa has a SAT score of 1985, while Kevin has a SAT score of 1880. Melissa scored 105 points more than Kevin.

Scales of Measurement (8 of 10)

- **Ratio**
  - The data have all the properties of interval data, and the ratio of two values is meaningful.
  - Variables such as distance, height, weight, and time use the ratio scale.
  - This scale must contain a zero value that indicates that nothing exists for the variable at the zero point.

Scales of Measurement (9 of 10)

- **Ratio**

  Example:

  Melissa's college record shows 36 credit hours earned, while Kevin's record shows 72 credit hours earned. Kevin has twice as many credit hours earned as Melissa.

Categorical and Quantitative Data

- Data can be further classified as being categorical or quantitative.
- The statistical analysis that is appropriate depends on whether the data for the variable is categorical or quantitative.
- In general, there are more alternatives for statistical analysis when the data are quantitative.

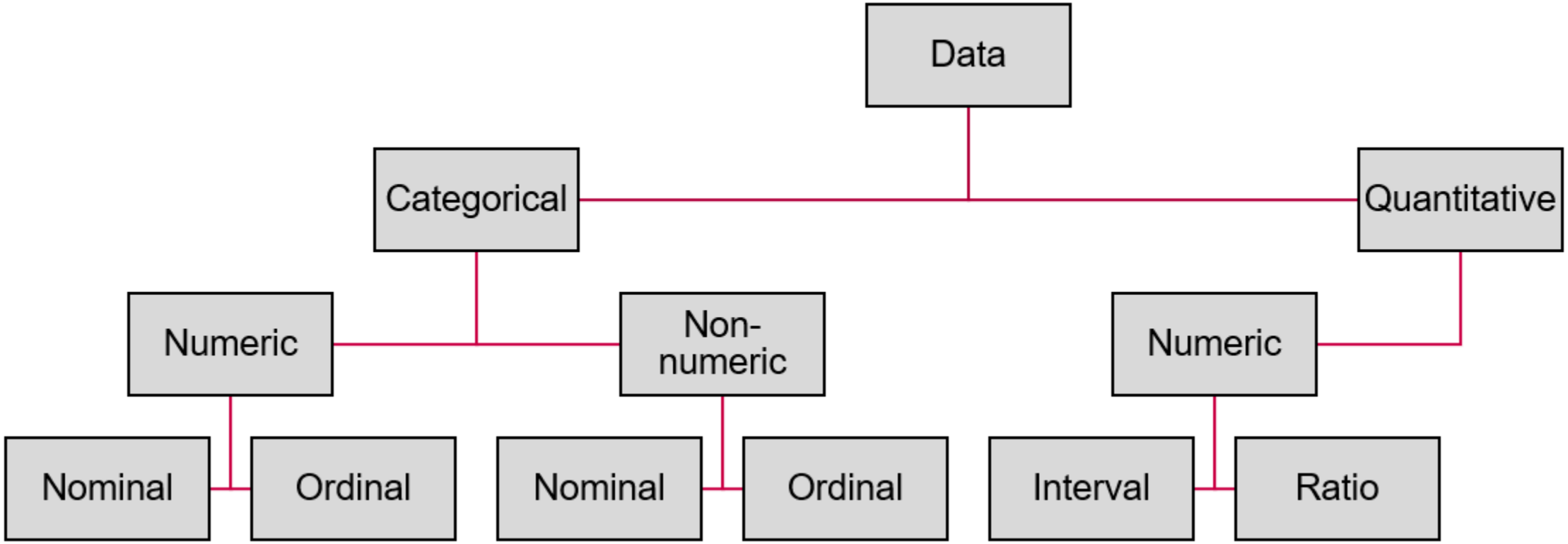
Categorical Data

- Labels or names used to identify an attribute of each element
- Often referred to as qualitative data
- Use either the nominal or ordinal scale of measurement
- Can be either numeric or nonnumeric
- Appropriate statistical analysis is rather limited

Quantitative Data

- Quantitative data indicate how many or how much:
  - discrete, if measuring how many
  - continuous, if measuring how much
- Quantitative data are always numeric.
- Ordinary arithmetic operations are meaningful for quantitative data.

Scales of Measurement (10 of 10)



Cross-Sectional Data

- Cross-sectional data are collected at the same or approximately the same point in time.

  Example:

  Data detailing different variables like status, Per capita GDP, Fitch rating for 60 different WTO nations at the same point in time.

Time Series Data (1 of 2)

- Time series data are collected over several time periods.

  Example:

  U.S average price per gallon of conventional regular gasoline between 2012 and 2018

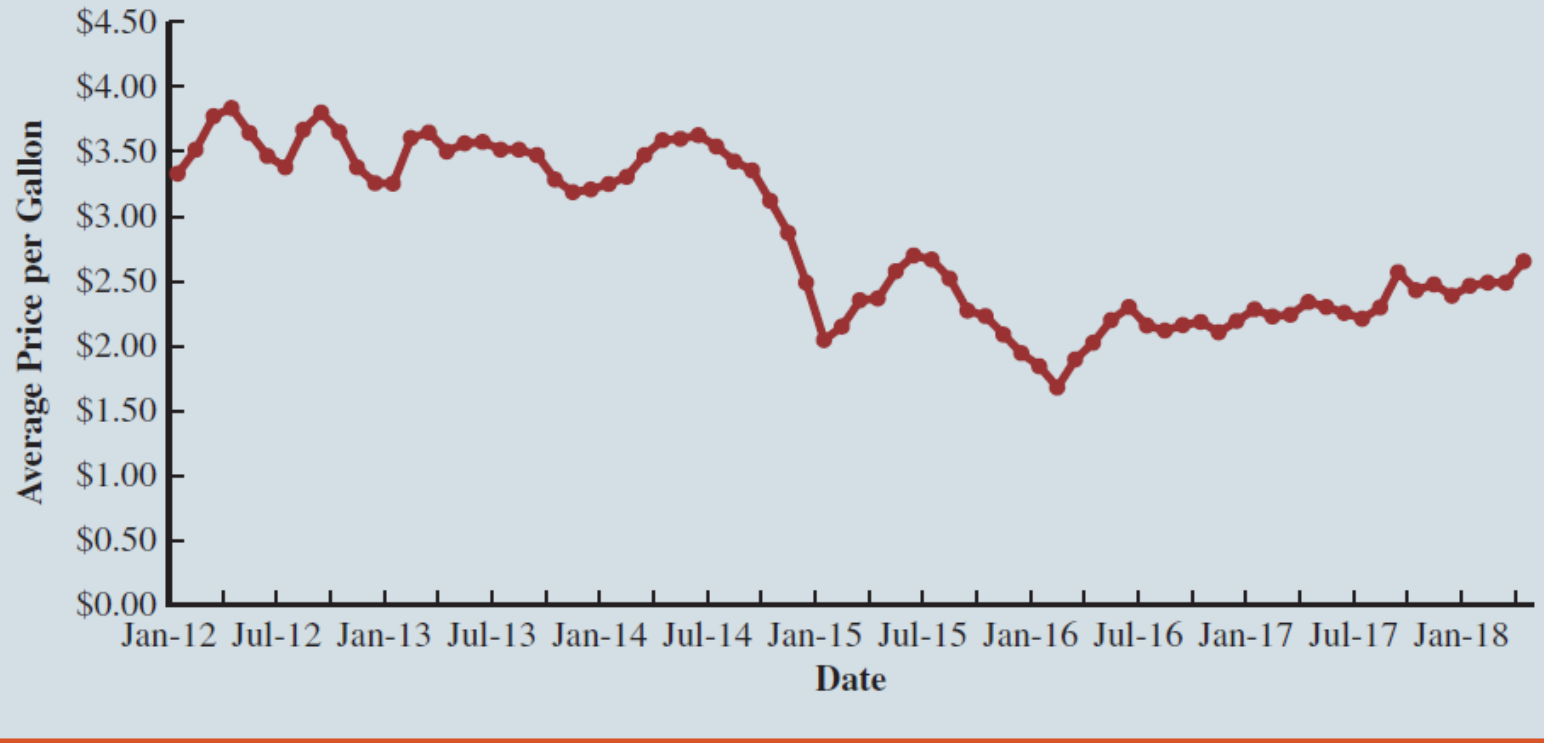
  Graphs of time series help analysts understand:

  - what happened in the past,
  - identify any trends over time, and
  - project future values for the time series.

Time Series Data (2 of 2)

Graph of Time Series Data

FIGURE 1.1 U.S. Average Price per Gallon for Conventional Regular Gasoline



Source: *Energy Information Administration, U.S. Department of Energy.*

Data Sources (1 of 5)

- Existing Sources
  - Internal company records—almost any department
  - Business database services—Dow Jones & Co.
  - Government agencies—U.S. Department of Labor
  - Industry associations—U.S. Travel Association
  - Special-interest organizations—Graduate Management Admission Council (GMAT)
  - Internet—more and more firms

Data Sources (2 of 5)

- Data Available from Internal Company Records

Record	Some of the Data Available
Employee records	Name, address, social security number
Production records	Part number, quantity produced, direct labor cost, material cost
Inventory records	Part number, quantity in stock, reorder level, economic order quantity
Sales records	Product number, sales volume, sales volume by region
Credit records	Customer name, credit limit, accounts receivable balance
Customer profile	Age, gender, income, household size

Data Sources (3 of 5)

- Data Available from Selected Government Agencies

Government Agency	Some of the Data Available
Census Bureau	Population data, number of households, household income
Federal Reserve Board	Data on money supply, exchange rates, discount rates
Office of Mgmt. & Budget	Data on revenue, expenditures, debt of federal government
Department of Commerce	Data on business activity, value of shipments, profit by industry
Bureau of Labor Statistics	Customer spending, unemployment rate, hourly earnings, safety record
DATA.GOV	More than 150,000 data sets including agriculture, consumer education, health, and manufacturing data

Data Sources (4 of 5)

- Statistical Studies—Observational
  - In observational studies, no attempt is made to control or influence the variables of interest.
    - A survey is a good example.
  - An example of an observational study is researchers observing a randomly selected group of customers that enter a Walmart Supercenter to collect data on variables such as time spent in the store, gender of the customer, and the amount spent.

Data Sources (5 of 5)

- Statistical Studies—Experimental
  - In experimental studies, the variable of interest is first identified. Then one or more variables are identified and controlled so that data can be obtained about how they influence the variable of interest.
  - The largest experimental study ever conducted is believed to be the 1954 Public Health Service experiment for the Salk polio vaccine. Nearly two million U.S. children (grades 1 through 3) were selected.

Data Acquisition Considerations

- Time Requirement
  - Searching for information can be time consuming.
  - Information may no longer be useful by the time it is available.
- Cost of Acquisition
  - Organizations often charge for information even when it is not their primary business activity.
- Data Errors
  - Using any data that happen to be available or were acquired with little care can lead to misleading information.

Descriptive Statistics

- Most of the statistical information in newspapers, magazines, company reports, and other publications consist of data that are summarized and presented in a form that is easy to understand.
- Such summaries of data, which may be tabular, graphical, or numerical, are referred to as descriptive statistics.

Example: Hudson Auto Repair (1 of 2)

- The manager of Hudson Auto would like to have a better understanding of the cost of parts used in the engine tune-ups performed in her shop. She examines 50 customer invoices for tune-ups. The costs of parts, rounded to the nearest dollar, are listed on the next slide.

Example: Hudson Auto Repair (2 of 2)

- Sample of Parts Cost ($) for 50 Tune-ups

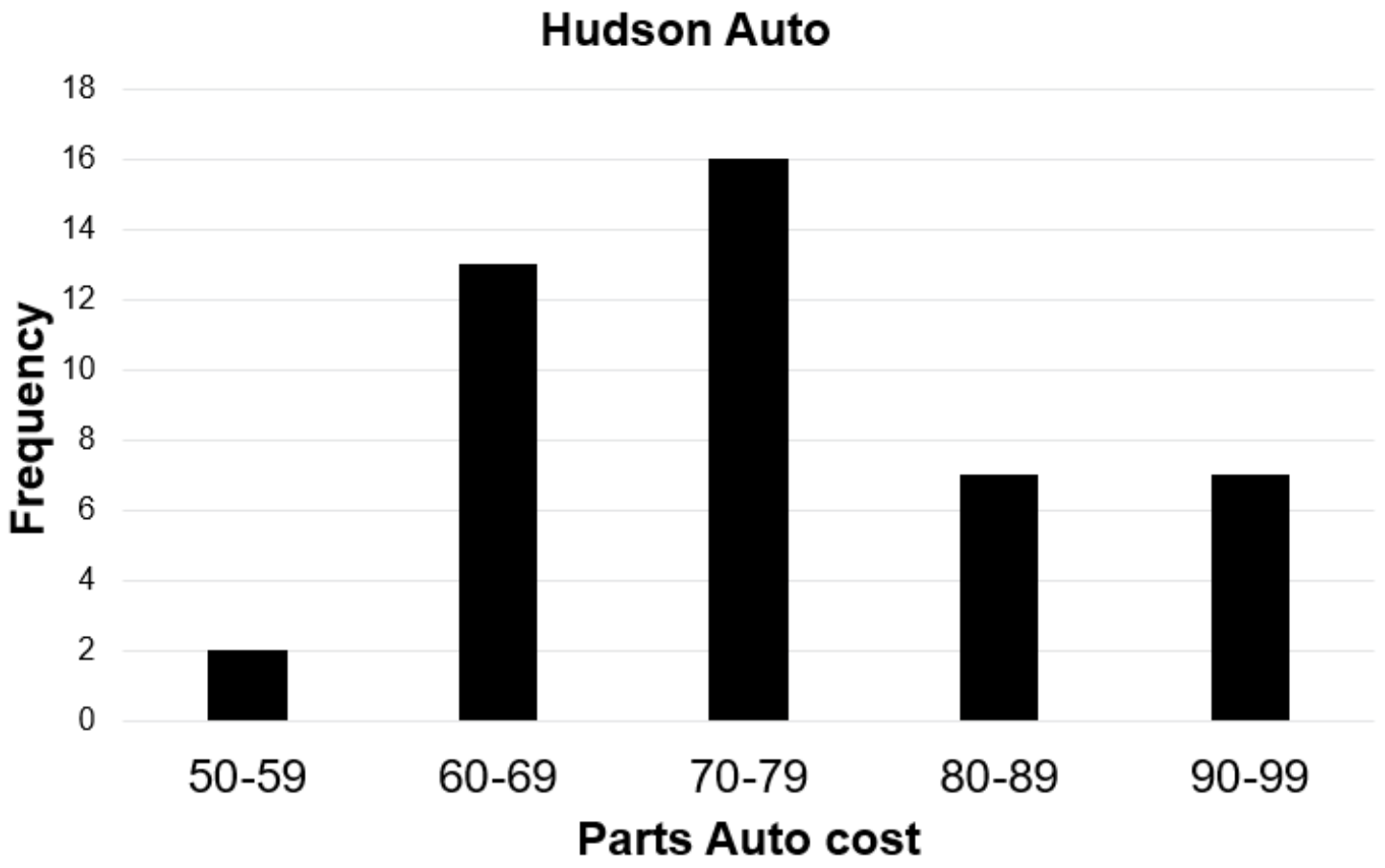
91	78	93	57	75	52	99	80	97	62
71	69	72	89	66	75	79	75	72	76
104	74	62	68	97	105	77	65	80	109
85	97	88	68	83	68	71	69	67	74
62	82	98	101	79	105	79	69	62	73

Tabular Summary: Frequency and Percent Frequency

Parts Cost ($)	Frequency	Percent Frequency
50 to 59	2	4%
60 to 69	13	26%
70 to 79	16	32%
80 to 89	7	14%
90 to 99	7	14%
100 to 109	5	10%
TOTAL	**50**	**100%**

Graphical Summary: Bar Chart

Example:
Hudson Auto



Numerical Descriptive Statistics

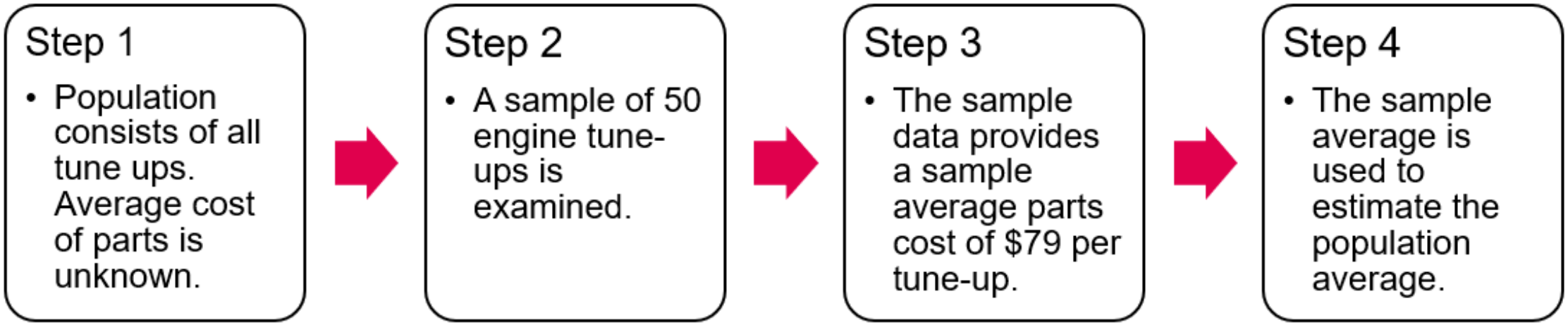
- The most common numerical descriptive statistic is the mean (or average).
- The mean demonstrates a measure of the central tendency, or central location, of the data for a variable.
- Hudson’s mean cost of parts, based on the 50 tune-ups studied, is $79 (found by summing up the 50 cost values and then dividing by 50).

Statistical Inference

- Population: The set of all elements of interest in a particular study.
- Sample: A subset of the population.
- Statistical inference: The process of using data obtained from a sample to make estimates and test hypotheses about the characteristics of a population.
- Census: Collecting data for the entire population.
- Sample survey: Collecting data for a sample.

Process of Statistical Inference

Example: Hudson Auto



Statistical Analysis Using Microsoft Excel (1 of 5)

- Statisticians often use computer software to perform the statistical computations required with large amounts of data.
- Many of the data sets in this book are available on the website that accompanies the book.
- The data sets are in Microsoft Excel format.
- The Excel add-in, StatTools, can be downloaded from the website.

Statistical Analysis Using Microsoft Excel (2 of 5)

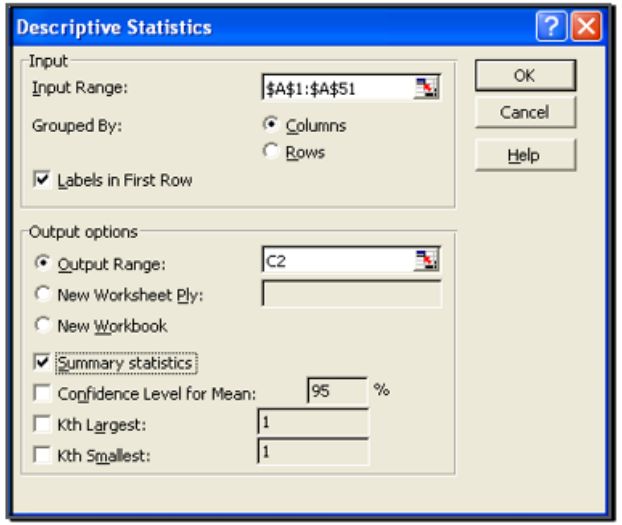
4 tasks might be needed:

Enter/Access Data

Enter Functions and Formulas

Apply Tools

Editing Options



	A
1	Parts Cost
2	91
3	71
4	104
5	85
6	62
7	78
8	69

D	E
Mean	=AVERAGE(A2:A71)
Median	=MEDIAN(A2:A71)
Mode	=MODE.SNGL(A2:A71)
Range	=MAX(A2:A71)-MIN(A2:A71)

Statistical Analysis Using Microsoft Excel (3 of 5)

- Excel Worksheet (showing data)

	A	B	C	D
1	Customer	Invoice #	Parts Cost ($)	Labor cost ($)
2	Sam Abrams	20994	91	185
3	Mary Gagnon	21003	71	205
4	Ted Dunn	21010	104	192
5	ABC appliances	21094	85	178
6	Harry Morgan	21116	62	242
7	Sara Morehead	21155	78	148
8	Vista travel, Inc.	21172	69	165
9	John Williams	21198	74	190

- Note: Rows 10 to 51 are not shown.

Statistical Analysis Using Microsoft Excel (4 of 5)

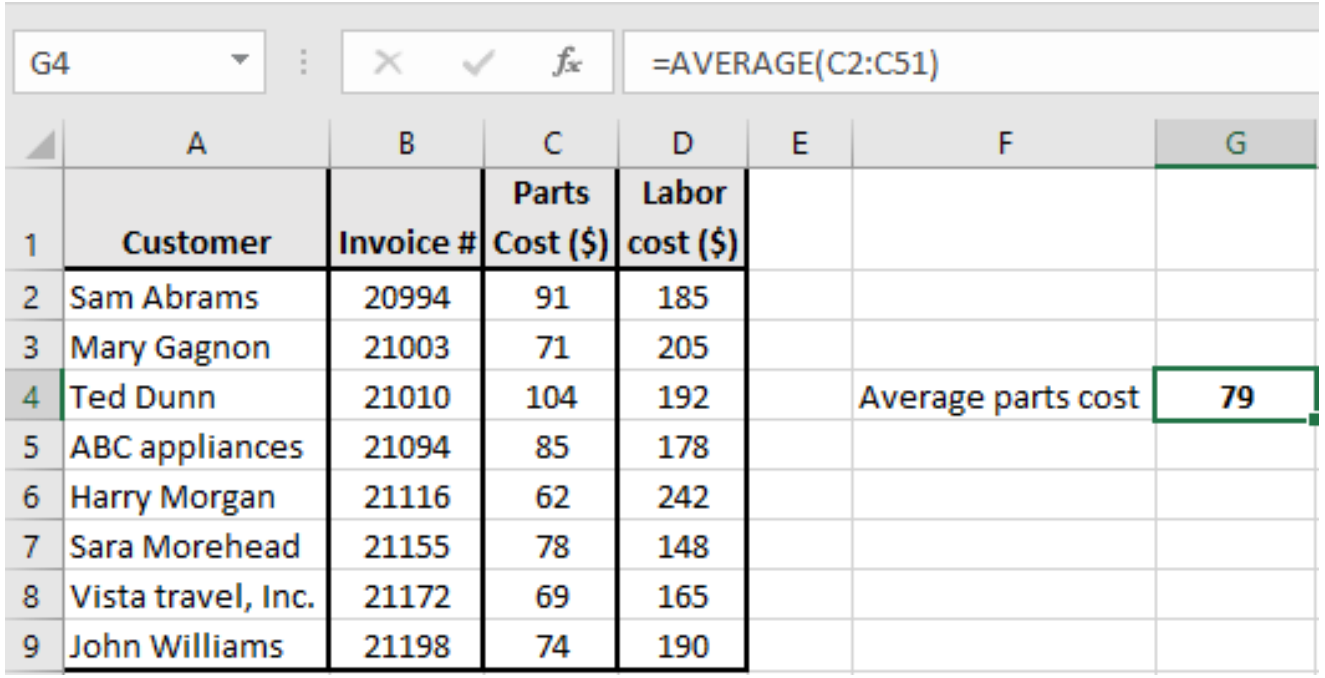
- Excel Formula Worksheet

	A	B	C	D	E	F	G
1	**Customer**	**Invoice #**	**Parts Cost ($)**	**Labor cost ($)**			
2	Sam Abrams	20994	91	185			
3	Mary Gagnon	21003	71	205			
4	Ted Dunn	21010	104	192		Average parts cost	=AVERAGE(C2:C51)
5	ABC appliances	21094	85	178			
6	Harry Morgan	21116	62	242			
7	Sara Morehead	21155	78	148			
8	Vista travel, Inc.	21172	69	165			
9	John Williams	21198	74	190			

- Note: Rows 10 to 51 are not shown.

Statistical Analysis Using Microsoft Excel (5 of 5)

- Excel Value Worksheet



G4 | =AVERAGE(C2:C51)

	A	B	C	D	E	F	G
1	Customer	Invoice #	Parts Cost ($)	Labor cost ($)			
2	Sam Abrams	20994	91	185			
3	Mary Gagnon	21003	71	205			
4	Ted Dunn	21010	104	192		Average parts cost	79
5	ABC appliances	21094	85	178			
6	Harry Morgan	21116	62	242			
7	Sara Morehead	21155	78	148			
8	Vista travel, Inc.	21172	69	165			
9	John Williams	21198	74	190			

- Note: Rows 10 to 51 are not shown.

Analytics

- Scientific process of transforming data into insight for making better decisions.
  - Descriptive analysis—Analytical techniques that describe what happened in the past.
  - Predictive analysis
    - Analytical techniques that use models constructed from past data to predict future.
    - Helps assess the impact the impact of one variable on another
  - Prescriptive analysis—Analytical techniques that yield a best course of action to take.

Big Data and Data Warehousing

- Organizations obtain large amounts of data on a daily basis by means of magnetic card readers, bar code scanners, point-of-sale terminals, and touchscreen monitors. Large and complex data sets are known as big data.
- Walmart captures data on 20 to 30 million transactions per day.
- Visa processes 6,800 payment transactions per second.
- Capturing, storing, and maintaining the data, referred to as data warehousing, is a significant undertaking.

Data Mining

- Analysis of the data in the warehouse might aid in decisions that will lead to new strategies and higher profits for the organization.
- Using a combination of procedures from statistics, mathematics, and computer science, analysts “mine the data” to convert it into useful information.
- The most effective data mining systems use automated procedures to extract information from the data prompted by only general or even vague queries by the user.

Data Mining Applications

- The major applications of data mining have been made by companies with a strong consumer focus such as retail, financial, and communication firms.
- Data mining is used to identify related products that customers who have already purchased a specific product are also likely to purchase (and then pop-ups are used to draw attention to those related products).
- As another example, data mining is used to identify customers who should receive special discount offers based on their past purchasing volumes.

Data Mining Requirements

- Statistical methodology such as multiple regression, logistic regression, and correlation are heavily used.
- Also needed are computer science technologies involving artificial intelligence and machine learning.
- A significant investment in time and money is required as well.

Data Mining Model Reliability

- Finding a statistical model that works well for a particular sample of data does not necessarily mean that it can be reliably applied to other data.
- With the enormous amount of data available, the data set can be partitioned into a training set (for model development) and a test set (for validating the model).
- There is, however, a danger of over fitting the model to the point that misleading associations and conclusions appear to exist.
- Careful interpretation of results and extensive testing is important.

Ethical Guidelines for Statistical Practice (1 of 2)

- In a statistical study, unethical behavior can take a variety of forms including:
  - Improper sampling
  - Inappropriate analysis of the data
  - Development of misleading graphs
  - Use of inappropriate summary statistics
  - Biased interpretation of the statistical results
- You should strive to be fair, thorough, objective, and neutral as you collect, analyze, and present data.
- As a consumer of statistics, you should also be aware of the possibility of unethical behavior by others.

Ethical Guidelines for Statistical Practice (2 of 2)

- The American Statistical Association developed the report “Ethical Guidelines for Statistical Practice.”
- It contains 52 guidelines organized into 8 topic areas:
  - Professional integrity and Accountability
  - Integrity of Data and Methods
  - Responsibilities to Science/Public/Funder/Client
  - Responsibilities to Research Subjects
  - Responsibilities to Research Team Colleagues
  - Responsibilities to Other Statisticians or Statistics Practitioners
  - Responsibilities Regarding Allegations of Misconduct
  - Responsibilities of Employers Including Organizations, Individuals, Attorneys, or Other Clients Employing Statistical Practitioners