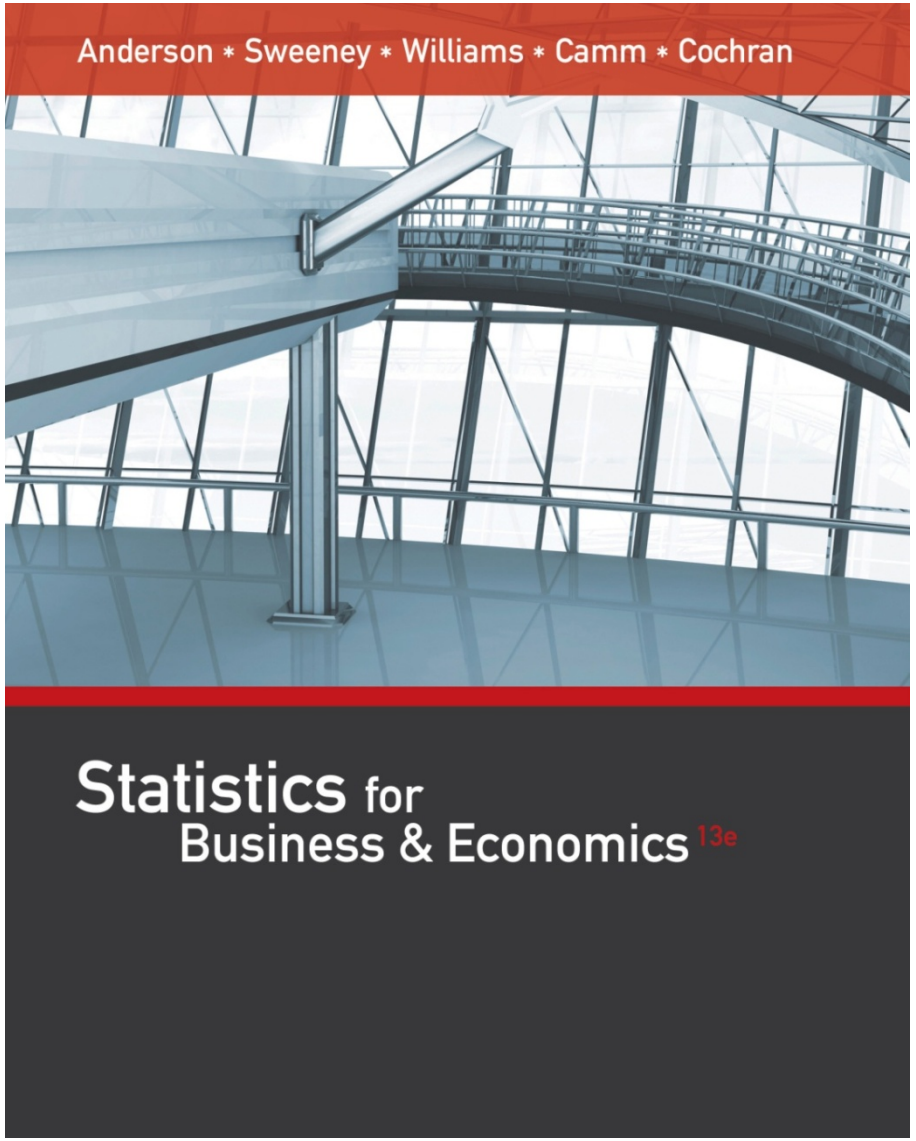
Statistics for Business and Economics (13e)

Anderson, Sweeney, Williams, Camm, Cochran

Slides by John Loucks

St. Edwards University



Chapter 1 - Data and Statistics

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

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

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

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.

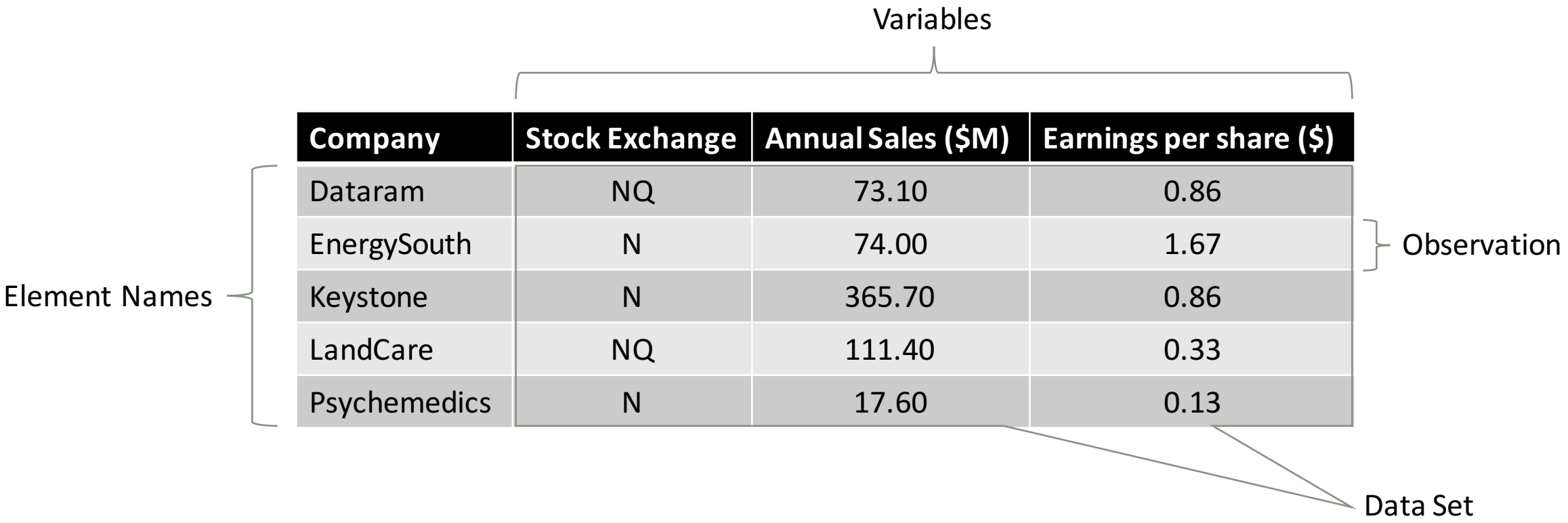
Data and Data Sets

- 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

- 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

Nominal scale

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

Example

Students of a university are classified by the school in which they are enrolled using a nonnumeric label such as Business, Humanities, Education, and so on.

Alternatively, a numeric code could be used for the school variable (e.g. 1 denotes Business, 2 denotes Humanities, 3 denotes Education, and so on).

Scales of Measurement

Ordinal scale

- 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.

Example

Students of a university are classified by their class standing using a nonnumeric label such as Freshman, Sophomore, Junior, or Senior.

Alternatively, a numeric code could be used for the class standing variable (e.g. 1 denotes Freshman, 2 denotes Sophomore, and so on).

Scales of Measurement

Interval scale

- 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.

Example

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

Scales of Measurement

Ratio scale

- Data have all the properties of interval data and the ratio of two values is meaningful.
- Ratio data are always numerical.
- Zero value is included in the scale.

Example:

Price of a book at a retail store is $ 200, while the price of the same book sold online is $100. The ratio property shows that retail stores charge twice the online price.

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 are categorical or quantitative.
- In general, there are more alternatives for statistical analysis when the data are quantitative.

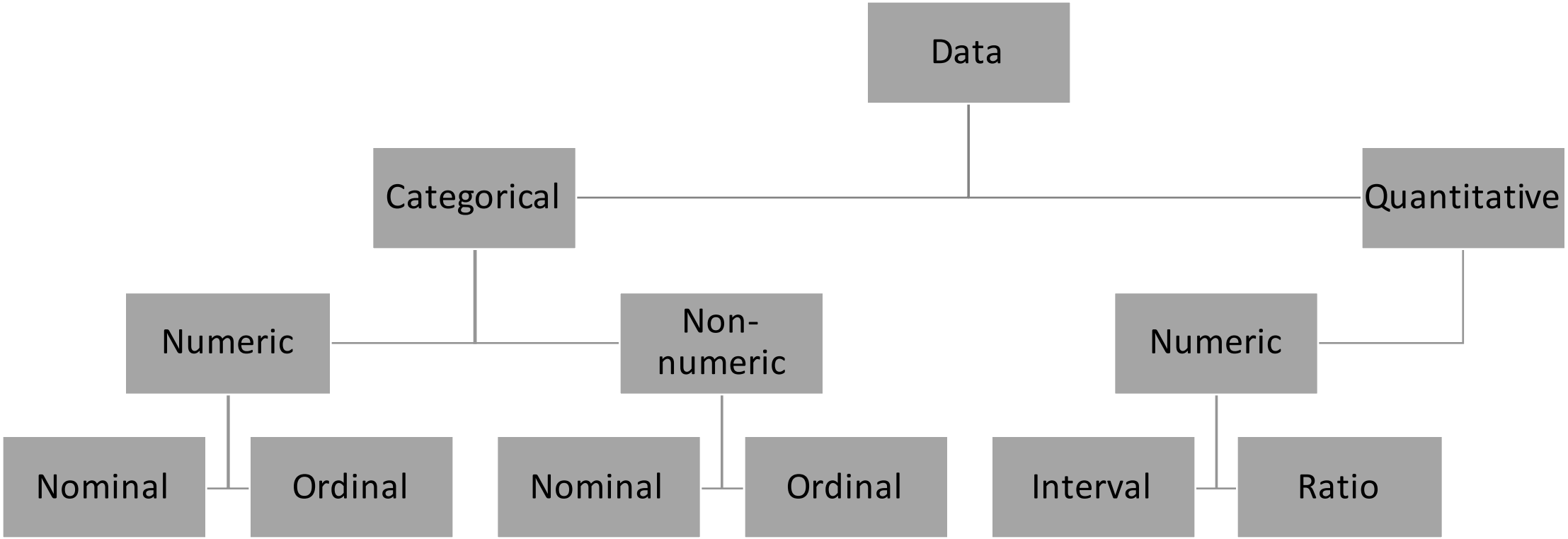
Categorical Data

- Labels or names are 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 analyses are rather limited

Quantitative Data

- Quantitative data indicate how many or how much.
- Quantitative data are always numeric.
- Ordinary arithmetic operations are meaningful for quantitative data.

Scales of Measurement



Cross-Sectional Data

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

Example

Data detailing the number of building permits issued in November 2013 in each of the counties of Ohio.

Time Series Data

Time series data are collected over several time periods.

Example

Data detailing the number of building permits issued in Lucas County, Ohio in each of the last 36 months.

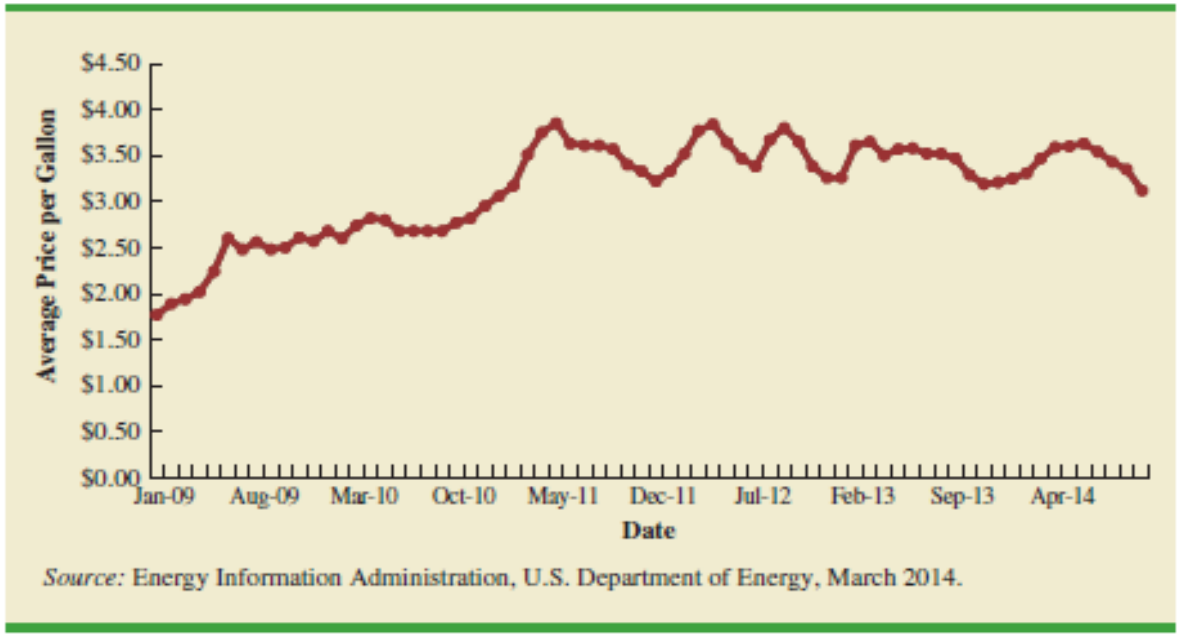
Graphs of time series data help analysts understand

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

Time Series Data

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, March 2014.

Data Sources

Existing Sources

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

Data Sources

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

Data Available From Selected Government Agencies

Government Agency	Web address	Some of the Data Available
Census Bureau	www.census.gov	Population data, number of households, household income
Federal Reserve Board	www.federalreserve.gov	Data on money supply, exchange rates, discount rates
Office of Mgmt. & Budget	www.whitehouse.gov/omb	Data on revenue, expenditures, debt of federal government
Department of Commerce	www.doc.gov	Data on business activity, value of shipments, profit by industry
Bureau of Labor Statistics	www.bls.gov	Customer spending, unemployment rate, hourly earnings, safety record

Data Sources

Statistical Studies – Observational

- In observational (nonexperimental) studies no attempt is made to control or influence the variables of interest.
- Example - Survey
- Studies of smokers and nonsmokers are observational studies because researchers do not determine or control who will smoke and who will not smoke.

Data Sources

Statistical Studies – Experimental

- In experimental studies the variable of interest is first identified. Then one or more other 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- 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 consists 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

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

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

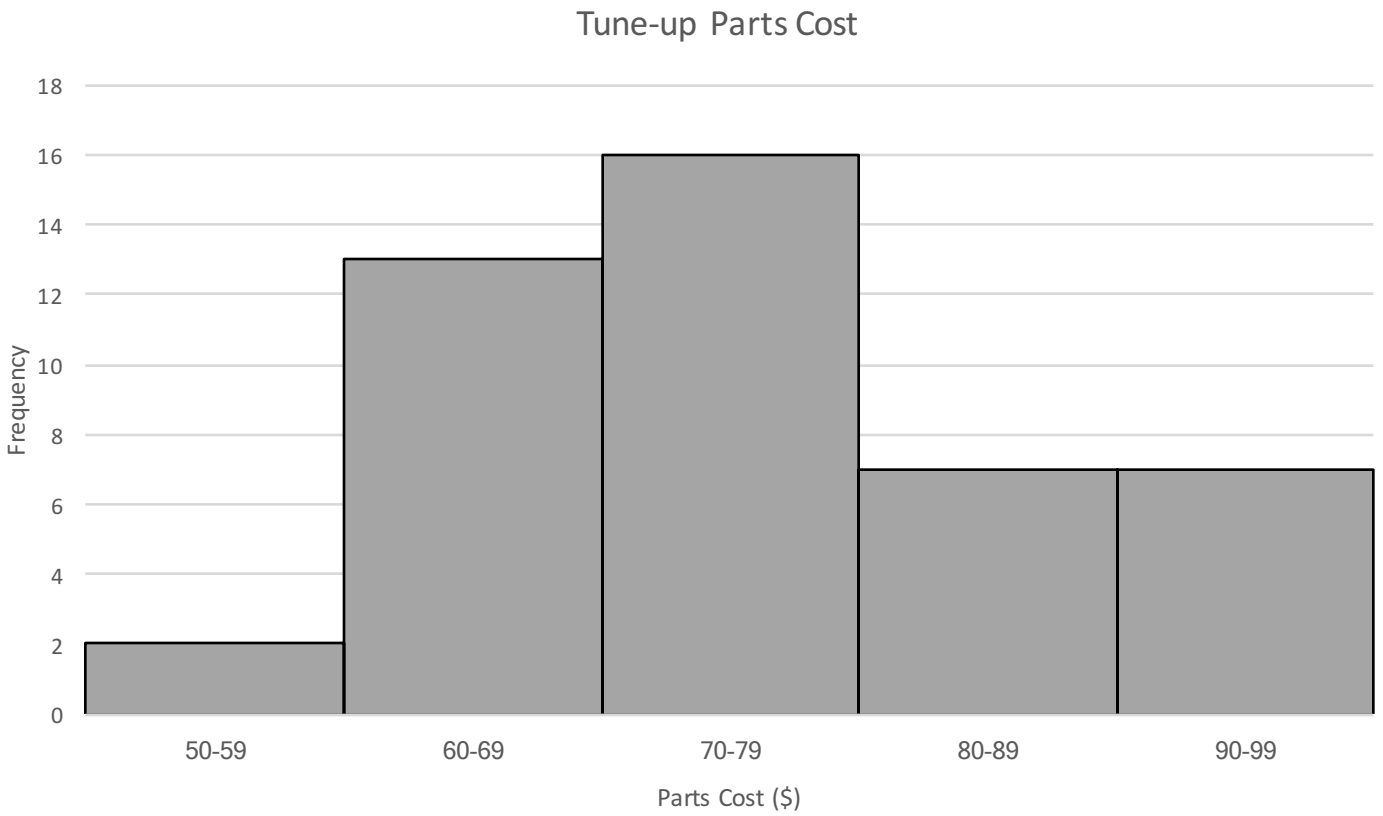
91	78	93	57	75	52	99	80	97	62
71	69	72	89	66	75	79	75	72	76
104	74	62	68	97	105	77	65	80	109
85	97	88	68	83	68	71	69	67	74
62	82	98	101	79	105	79	69	62	73

Tabular Summary: Frequency and Percent Frequency

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

Graphical Summary: Histogram

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

Step 1
- Population consists of all tune ups. Average cost of parts is unknown.

Step 2
- A sample of 50 engine tune-ups is examined.

Step 3
- The sample data provides a sample average parts cost of $79 per tune-up.

Step 4
- The sample average is used to estimate the population average.

Analytics

Analytics is the scientific process of transforming data into insight for making better decisions.

Techniques:

- Descriptive analytics: This describes what has happened in the past.

- Predictive analytics: Use models constructed from past data to predict the future or to assess the impact of one variable on another.

- Prescriptive analytics: The set of analytical techniques that yield a best course of action.

Big data and Data Mining:

Big data: Large and complex data set.

Three V's of Big data:

⇒Volume : Amount of available data

⇒Velocity: Speed at which data is collected and processed

⇒Variety: Different data types

Data warehousing

Data warehousing is the process of capturing, storing, and maintaining the data.

- Organizations obtain large amounts of data on a daily basis by means of magnetic card readers, bar code scanners, point of sale terminals, and touch screen monitors.
- Wal-Mart captures data on 20-30 million transactions per day.
- Visa processes 6,800 payment transactions per second.

Data Mining

- Methods for developing useful decision-making information from large databases.
- 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 discover relationships in the data and predict future outcomes prompted by general and 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).
- Data mining is also 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 overfitting 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

- 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
- One should strive to be fair, thorough, objective, and neutral as you collect, analyze, and present data.
- As a consumer of statistics, one should also be aware of the possibility of unethical behavior by others.

Ethical Guidelines for Statistical Practice

- The American Statistical Association developed the report "Ethical Guidelines for Statistical Practice".
- It contains 67 guidelines organized into 8 topic areas:
  - Professionalism
  - Responsibilities to Funders, Clients, Employers
  - Responsibilities in Publications and Testimony
  - Responsibilities to Research Subjects
  - Responsibilities to Research Team Colleagues
  - Responsibilities to Other Statisticians/Practitioners
  - Responsibilities Regarding Allegations of Misconduct
  - Responsibilities of Employers Including Organizations, Individuals, Attorneys, or Other Clients

End of Chapter 1