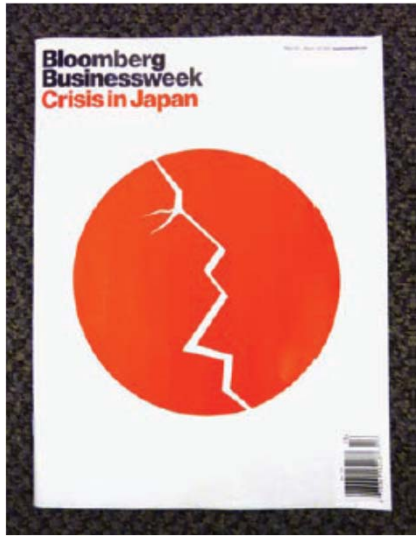
Statistics for Business and Economics

Anderson, Sweeney, Williams, Camm, Cochran

CHAPTER 1
DATA AND STATISTICS

STATISTICS *in* PRACTICE

- Bloomberg *Businessweek* is one of the most widely read business magazines in the world.
- Bloomberg *Businessweek*, provide an in-depth reports contain statistical facts and summaries that help the reader understand the business and economic information.
- Bloomberg *Businessweek* also uses statistics and statistical information in managing its own business.



Chapter 1 Data and Statistics

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

Statistics

- The term statistics can refer to numerical facts such as averages, medians, percents, and index numbers 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.

1.1 Applications in Business and Economics

- Accounting
- Finance
- Marketing
- Production
- Economics
- Information Systems

Applications in Business and Economics -1

- **Accounting**
- Public accounting firms use statistical sampling procedures when conducting audits for their clients.
  - For instance, the audit staff selects a subset of the accounts called a sample. After reviewing the accuracy of the sampled accounts, the auditors draw a conclusion as to whether the accounts receivable amount shown on the client's balance sheet is acceptable.

Applications in Business and Economics -2

- **Finance**
- Financial advisors use price-earnings ratios and dividend yields to guide their investment advice.
  - For instance, In the case of stocks, analysts review financial data such as price/earnings ratios and dividend yields. By comparing the information for an individual stock with information about the stock market averages, an analyst can begin to draw a conclusion as to whether the stock is a good investment.

Applications in Business and Economics - 3

- **Marketing**
- Electronic point-of-sale scanners at retail checkout counters are used to collect data for a variety of marketing research applications.
  - For instance, data suppliers such as ACNielsen purchase point-of-sale scanner data from grocery stores, process the data, and then sell statistical summaries of the data to manufacturers. Manufacturers use statistical summaries on promotional activities. Brand managers review statistics and the promotional activity statistics to gain a better understanding of the relationship between promotional activities and sales.

Applications in Business and Economics - 4

- **Production**
- A variety of statistical quality control charts are used to monitor the output of a production process.
  - For instance, an x-bar chart can be used to monitor the average output. Suppose, for example, that a machine fills containers with 12 ounces of a soft drink. Properly interpreted, an x-bar chart can help determine when adjustments are necessary to correct a production process.

Applications in Business and Economics - 5

- **Economics**
- Economists use statistical information in making forecasts about the future of the economy or some aspect of it.
  - For instance, in forecasting inflation rates, economists use statistical information on such indicators as the Producer Price Index, the unemployment rate, and manufacturing capacity utilization. Often these statistical indicators are entered into computerized forecasting models that predict inflation rates

Applications in Business and Economics - 6

- **Information Systems**
- A variety of statistical information helps administrators assess the performance of computer networks.
  - For instance, Statistics such as the mean number of users on the system, the proportion of time any component of the system is down, and the proportion of bandwidth utilized at various times of the day are examples of statistical information that help the system administrator better understand and manage the computer network.

1.2 Data

- Data and Data Sets
- Elements, Variables, and Observations
- Scales of Measurement
- Categorical and Quantitative Data
- Cross-Sectional and Time series Data

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 o as the data set for the study.

Elements, Variables, and Observations -1

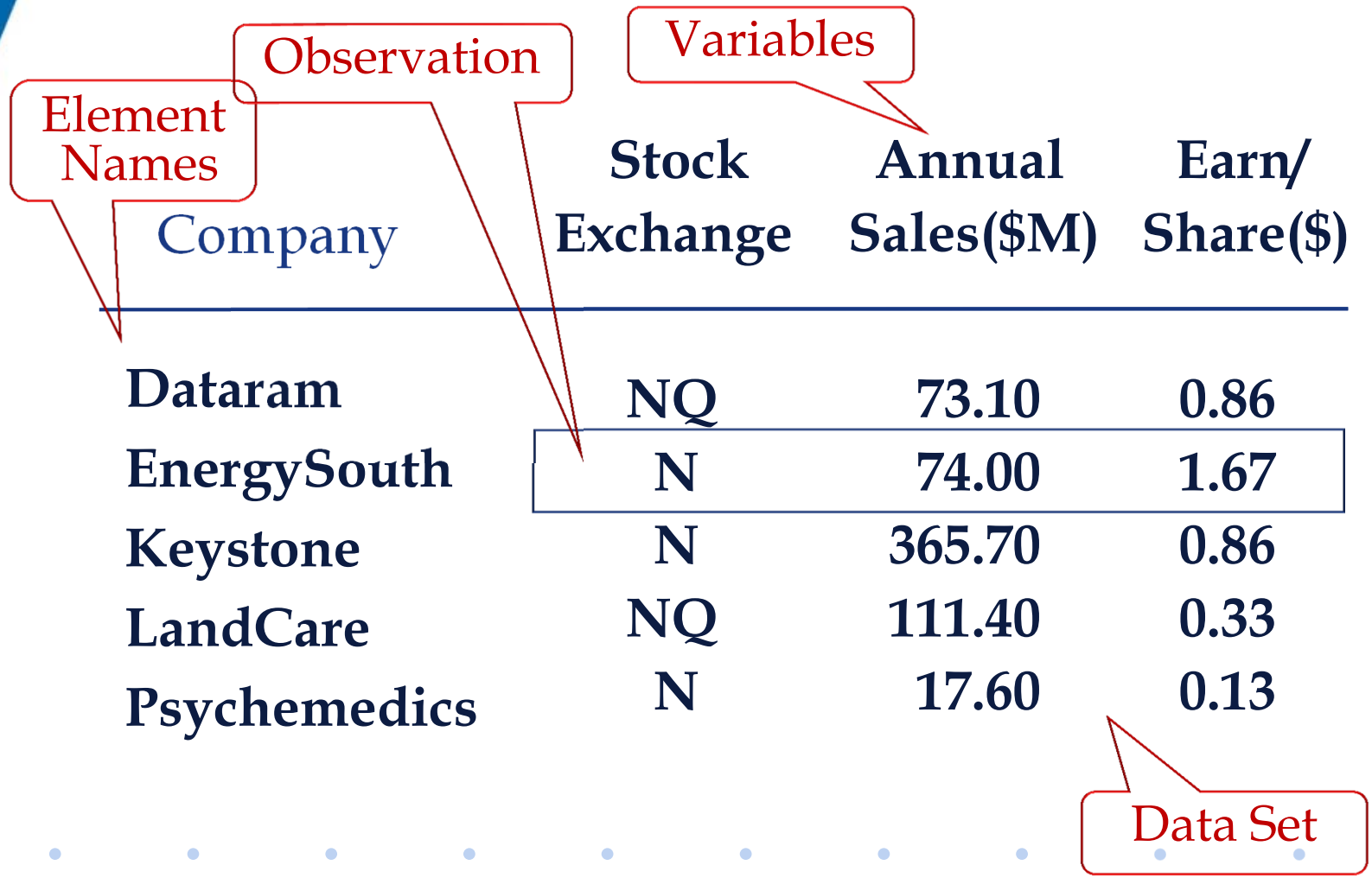
- Elements are the entities on which data are collected.
- A variable is a characteristic of interest for the elements.
- The set of measurements obtained for a particular element is called an observation.
- A data set with n elements contains n observations.
- The total number of data values in a complete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

Elements, Variables, and Observations -2

Nation	WTO Status	Per Capita GDP ($)	Trade Deficit ($1000s)	Fitch Rating	Fitch Outlook
Armenia	Member	5,400	2,673,359	BB−	Stable
Australia	Member	40,800	−33,304,157	AAA	Stable
Austria	Member	41,700	12,796,558	AAA	Stable
Azerbaijan	Observer	5,400	−16,747,320	BBB−	Positive
Bahrain	Member	27,300	3,102,665	BBB	Stable
Belgium	Member	37,600	−14,930,833	AA+	Negative
Brazil	Member	11,600	−29,796,166	BBB	Stable
Bulgaria	Member	13,500	4,049,237	BBB−	Positive

- the data set contains 8 elements.
- five variables: WTO status, Per Capita GDP($), Trade Deficit ($1000s), Fitch Rating, Fitch Outlook.
- observations: the first observation (Armenia) Member, 5400, 2673359, BB-, Stable.

Data, Data Sets, Elements, Variables, and Observations



Scales of Measurement -1

Scales of measurement include:

- Nominal
- Interval
- Ordinal
- 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

- **Nominal**

Data are labels or names used to identify an attribute of the element.

A nonnumeric label or numeric code may be used.

For example, gender, ID number, "WTO Status" in Table 1.1

nominal data can be recorded using a numeric code. we could use "0" for female, and "1" for male.

Scales of Measurement -3

- **Nominal**

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 -4

- **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.

For example, questionnaire: a repair service rating of excellent, good, or poor. Ordinal data can be recorded using a numeric code. We could use 1 for excellent, 2 for good, and 3 for poor.

The scale of measurement for the Fitch Rating is ordinal. The rating labels which range from AAA to F can be rank ordered from best credit rating AAA to poorest credit rating F.

Scales of Measurement -5

- **Ordinal**

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 -6

- **Interval**

The data have the properties of ordinal data, and the interval between observations is expressed in terms of a fixed unit of measure.

Example: SAT scores, temperature

Interval data are always numeric.

Scales of Measurement -7

- **Interval**

Example:

three students with SAT math scores of 620, 550, and 470 can be ranked or ordered in terms of best performance to poorest performance in math.

The differences between the scores are meaningful.

For instance, student 1 scored 620 - 550 = 70 points more than student 2, while
student 2 scored 550 - 470 = 80 points more than student 3.

Scales of Measurement -8

- **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

- **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 are categorical or quantitative.

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

Categorical Data -1

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 analyses are rather limited

Categorical Data -2

- We can summarize categorical data by counting the number of observations in each category or by computing the proportion of the observations in each category.
- However, even when the categorical data are identified by a numerical code, arithmetic operations such as addition, subtraction, multiplication, and division do NOT provide meaningful results.

Quantitative Data

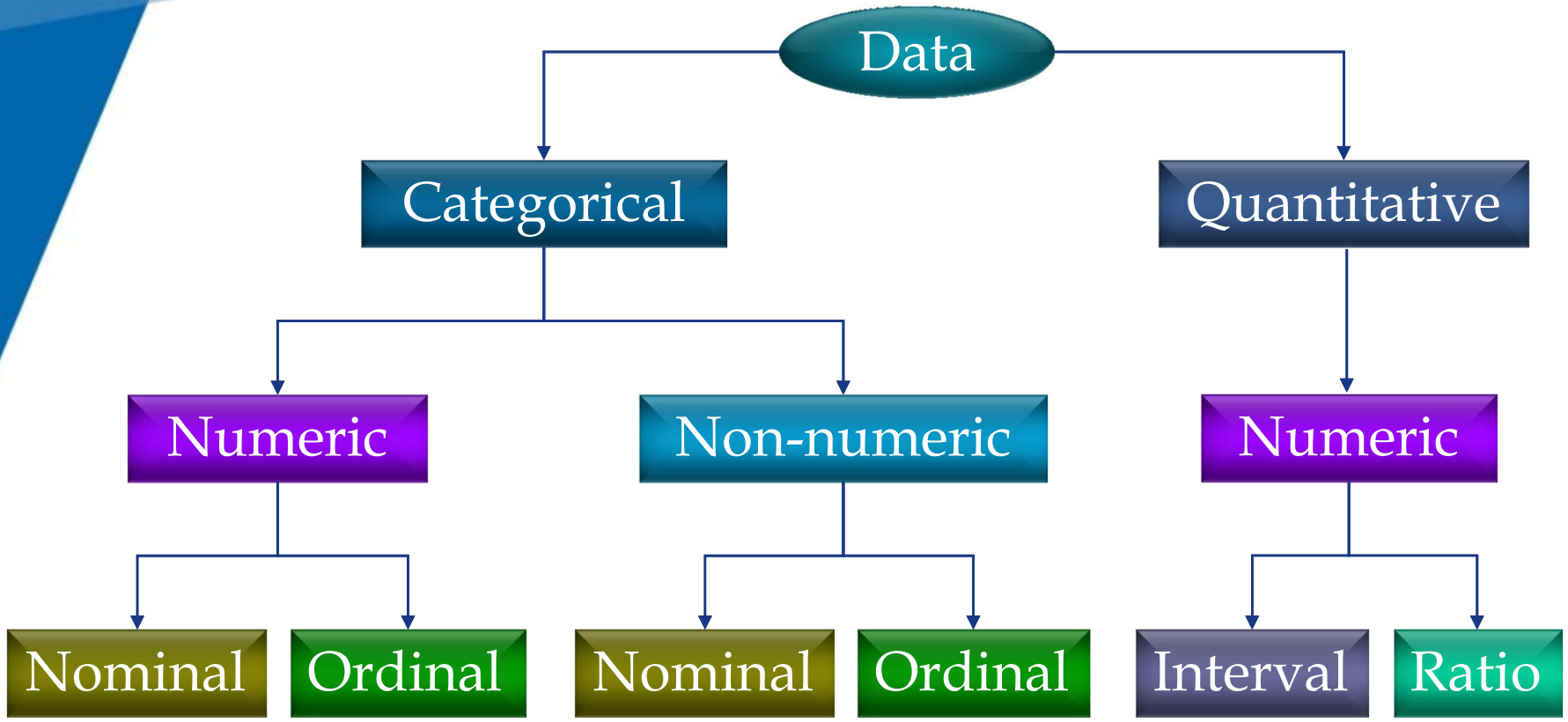
Quantitative data indicate how many or how much:

- 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



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 2012 in each of the counties of Ohio

Time Series Data -1

Time series data are collected over several time periods.

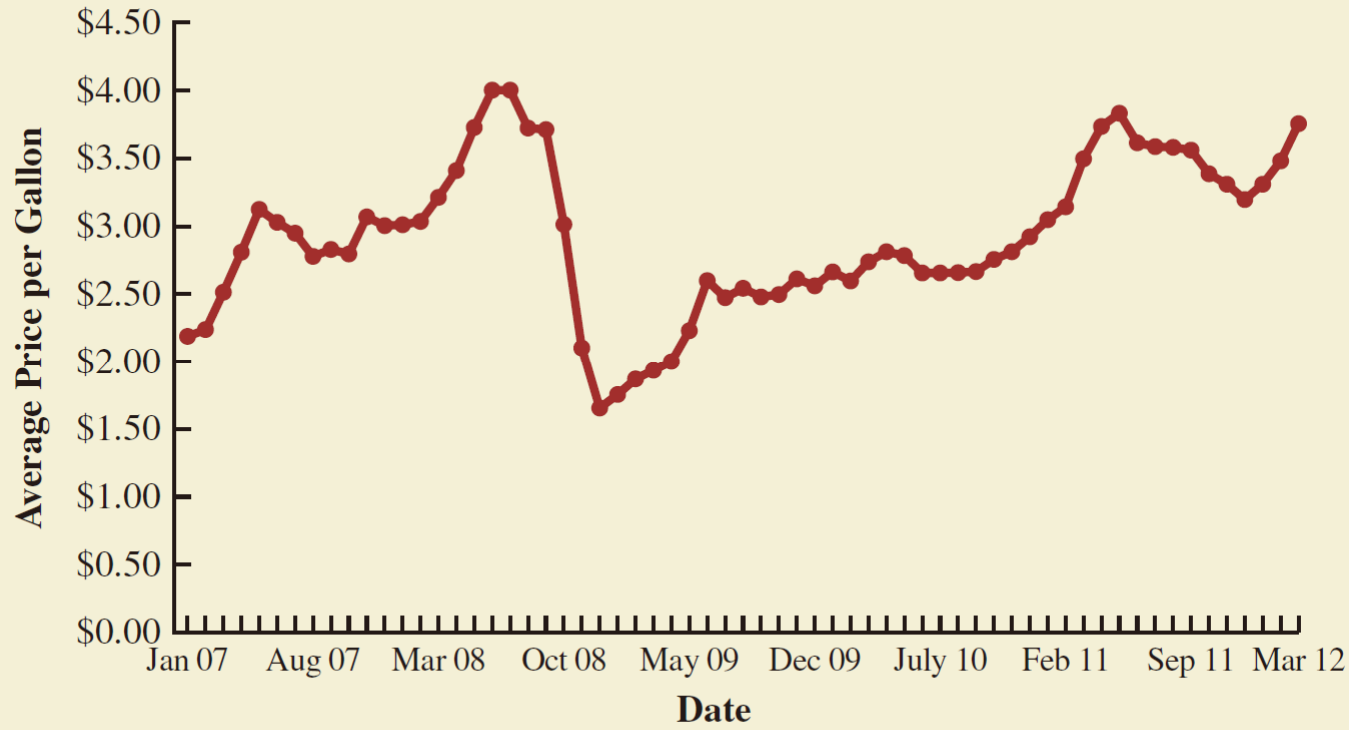
Example: data detailing the number of building permits issued in Lucas County, Ohio in each of the last 36 months

Graphs of time series help analysts understand

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

Time Series Data -2

- **Graph of Time Series Data**



Source: Energy Information Administration, U.S. Department of Energy, March 2012.

1.3 Data Sources

- Existing Sources
- Statistical Studies
- Data Acquisition Errors

Data Sources -1

- **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

Internet – more and more firms

Data Sources -2

- 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

- 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 Sources -4

- **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 Sources -5

- Experimental study, a 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.
- For example, a pharmaceutical firm might be interested in conducting an experiment to learn how a new drug affects blood pressure.
  - Blood pressure is the variable of interest in the study.
  - The dosage level of the new drug is another variable that is hoped to have a causal effect on blood pressure.

Data Sources -6

- **Statistical Studies - Observational**

In observational (nonexperimental) studies no attempt is made to control or influence the variables of interest. a survey is a good example

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 -7

- **Example: A customer opinion questionnaire used by Chops City Grill in Naples, Florida**

Chops CITY GRILL

Date: _____ Server Name: _____

Our customers are our top priority. Please take a moment to fill out our survey card, so we can better serve your needs. You may return this card to the front desk or return by mail. Thank you!

SERVICE SURVEY	Excellent	Good	Average	Fair	Poor
Overall Experience	❑	❑	❑	❑	❑
Greeting by Hostess	❑	❑	❑	❑	❑
Manager (Table Visit)	❑	❑	❑	❑	❑
Overall Service	❑	❑	❑	❑	❑
Professionalism	❑	❑	❑	❑	❑
Menu Knowledge	❑	❑	❑	❑	❑
Friendliness	❑	❑	❑	❑	❑
Wine Selection	❑	❑	❑	❑	❑
Menu Selection	❑	❑	❑	❑	❑
Food Quality	❑	❑	❑	❑	❑
Food Presentation	❑	❑	❑	❑	❑
Value for $ Spent	❑	❑	❑	❑	❑

What comments could you give us to improve our restaurant?

Thank you, we appreciate your comments. —The staff of Chops City Grill.

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.

Data Acquisition Errors

- We should always be aware of the possibility of data errors in statistical studies. Using erroneous data can be worse than not using any data at all.
- An error in data acquisition occurs whenever the data value obtained is not equal to actual value that would be obtained with a correct procedure.
- For example,
  - In writing the age of a 24-year-old person as 42.
  - A respondent shown to be 22 years of age but reporting 20 years of work experience.

1.4 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.

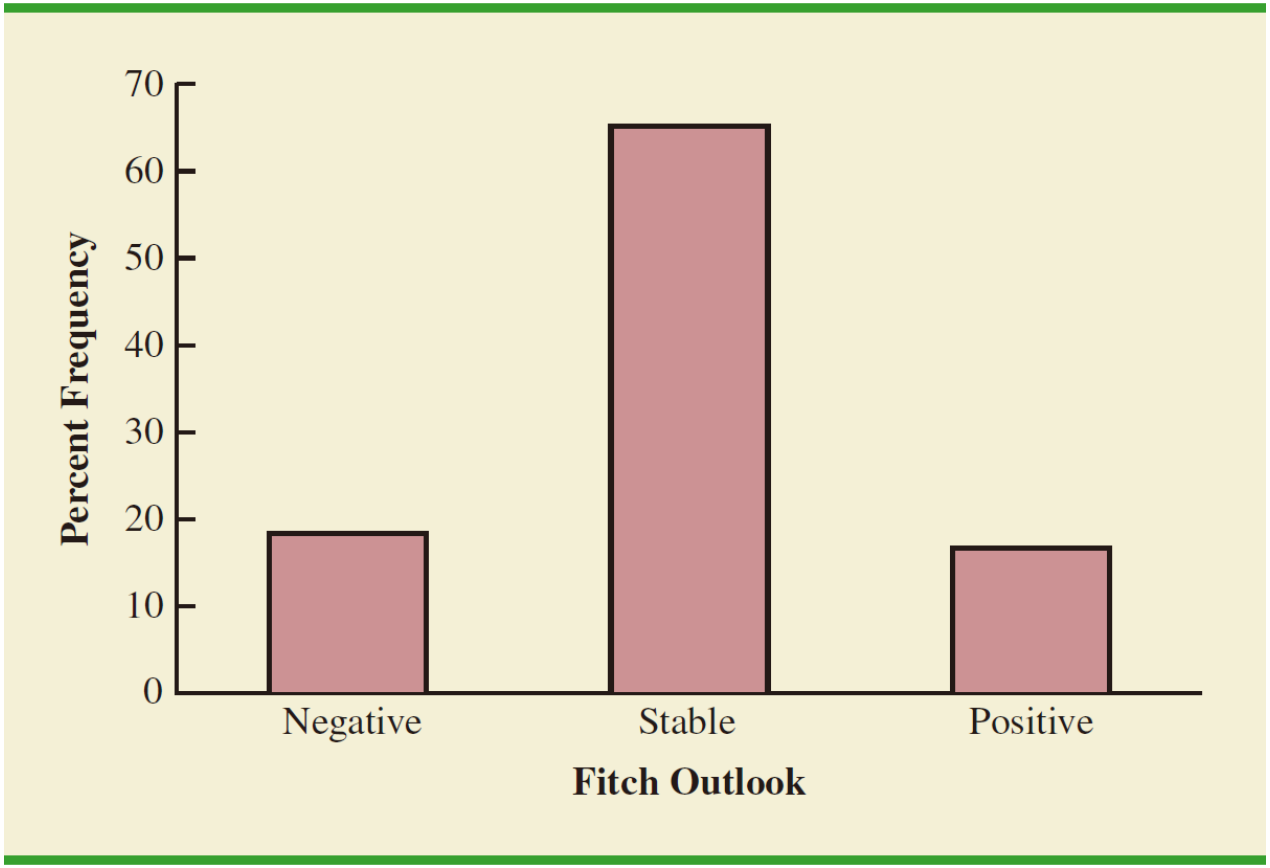
Descriptive Statistics -1

- A tabular summary of the data set in Table 1.1
- Frequencies and Percent Frequencies for The Fitch Credit Rating Outlook of 60 Nations

Fitch Outlook	Frequency	Percent Frequency (%)
Positive	10	16.7
Stable	39	65.0
Negative	11	18.3

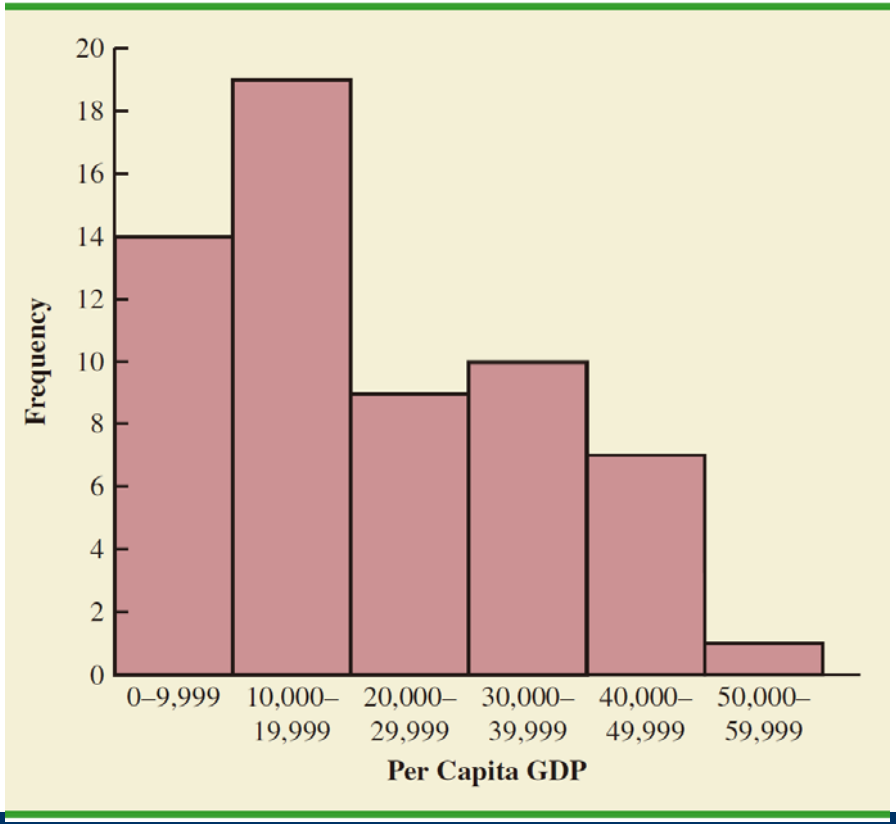
Descriptive Statistics -2

- Bar Chart for The Fitch Credit Rating Outlook of 60 Nations



Descriptive Statistics -3

- A graphical summary of the data for quantitative variable Per Capita GDP in Table 1.1, called a histogram



Example: Hudson Auto Repair -1

- 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

- **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

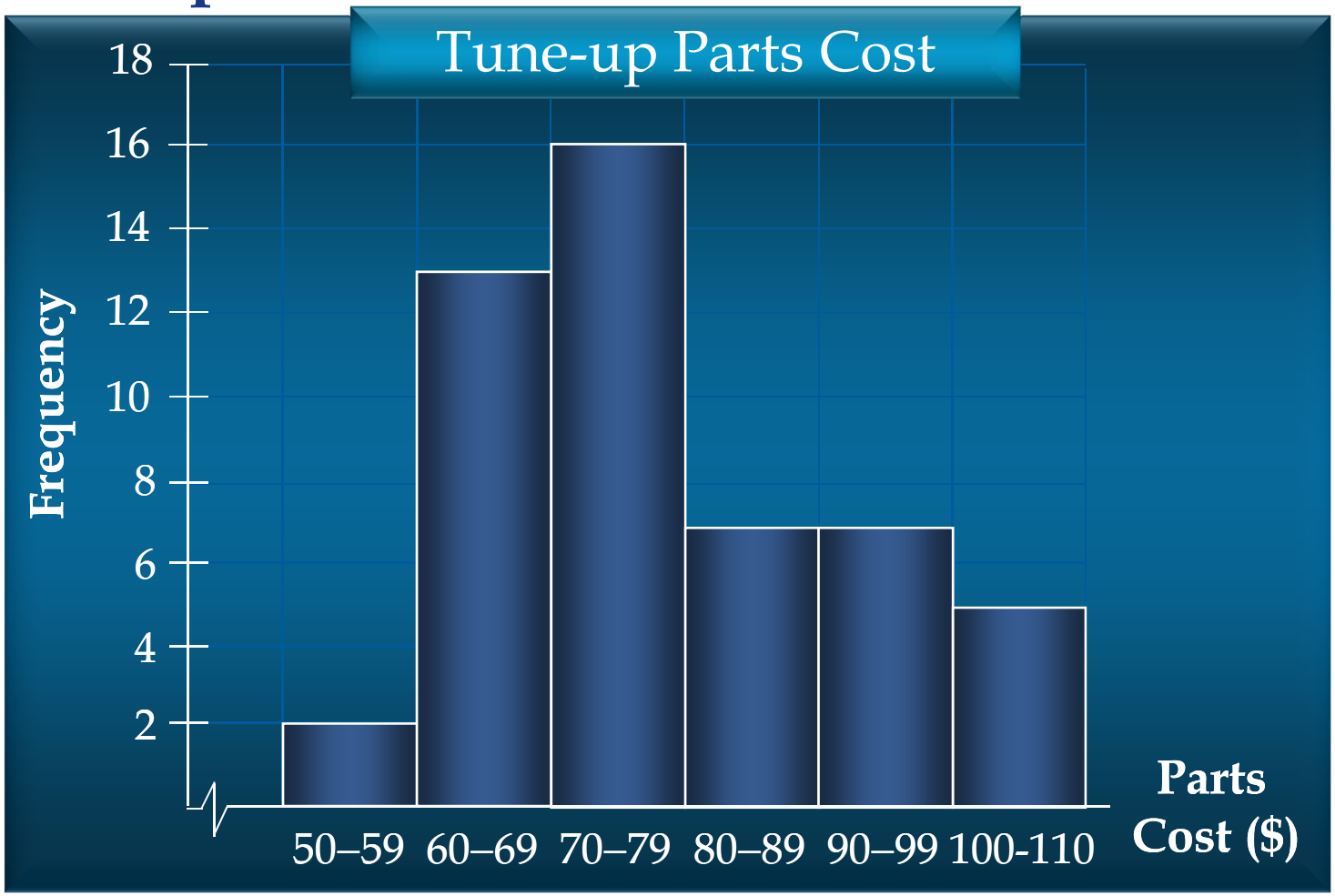
- Example: Hudson Auto

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
	50	100

(2/50)100

Graphical Summary: Histogram

- **Example: Hudson Auto**



Numerical Descriptive Statistics

- The most common numerical descriptive statistic is the average (or mean).
- The average demonstrates a measure of the central tendency, or central location, of the data for a variable.
- Hudson's average cost of parts, based on the 50 tune-ups studied, is $79 (found by summing the 50 cost values and then dividing by 50).
- A numerical summary of the data set in Table 1.1, an average Per Capita GDP of $21,387.

1.5 Statistical Inference -1

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

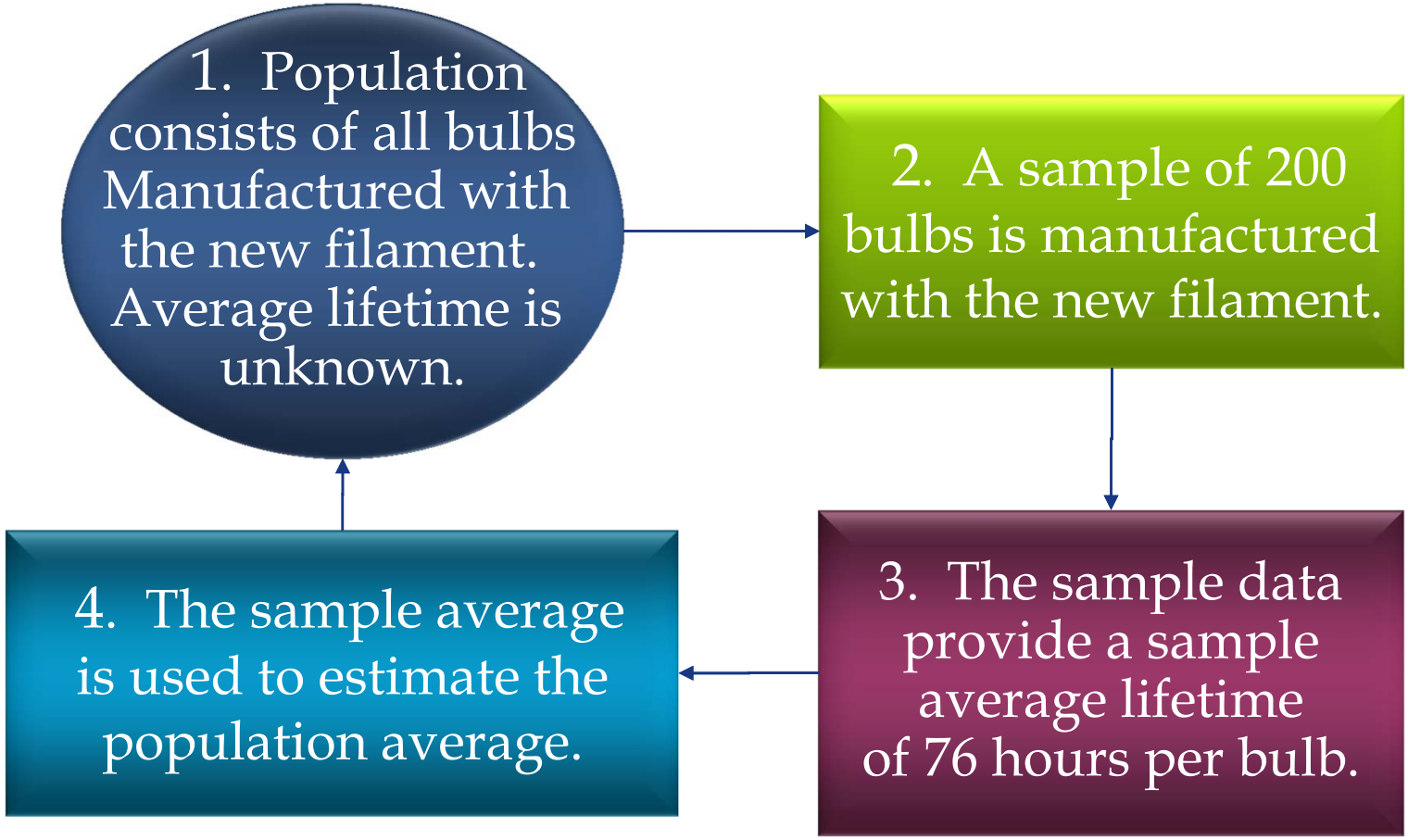
Census – collecting data for the entire population

Sample survey – collecting data for a sample

Statistical Inference -2

- **Example: Norris Electronics.**
  - Norris manufactures a high-intensity lightbulb.
  - To increase the useful life of the lightbulb, the product design group developed a new lightbulb filament. In this case, the population is defined as all lightbulbs that could be produced with the new filament.
  - To evaluate the advantages of the new filament, 200 bulbs with the new filament were manufactured and tested.
  - Data collected from this sample showed the number of hours each lightbulb operated before filament burnout.

Process of Statistical Inference



1.6 Computers and Statistical Analysis

- 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 can downloaded in either Minitab or Excel format.
- Also, the Excel add-in StatTools can be downloaded from the website.

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 touch screen monitors.
- Wal-Mart captures data on 20-30 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.

1.7 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 discover relationships in the data and predict future outcomes, … prompted by only general, 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.

1.8 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
- 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

- The American Statistical Association developed the report "Ethical Guidelines for Statistical Practice".
- The report contains 67 guidelines organized into eight 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