



Quantitative Techniques

Statistics

DR. ANCHAL GUPTA



WHAT IS STATISTICS?

- ❑ **Statistics** is a science that helps us make better decisions in business and economics as well as in other fields.
- ❑ **Statistics** deals with aggregation of facts affected to a marked extent by multiplicity of causes, numerically expressed, enumerated or estimated according to reasonable standards of accuracy in systematic manner for a predetermined purpose and placed in relation to each other.

Functions of Statistics



- It presents facts in systematic / definite form.
- It simplifies mass of figures.
- It facilitates comparison.
- It helps in formulating and testing hypothesis.
- It helps in prediction and forecasting.
- Helps in Suitable policy formulation.

Scope of Statistics

❓ Stats & Business

- Production, Sales, Purchase, Forecasting, Quality control etc.

❓ Stats and different departments

- Finance, Transport, Defense, Insurance etc.

❓ Stats & Research

- -Statistical methods are used in conducting experiments for different type of soils, medicine and public health.

Statistics in Business

Accounting — auditing and cost estimation

Economics — local, regional, national, and international economic performance

Finance — investments and portfolio management

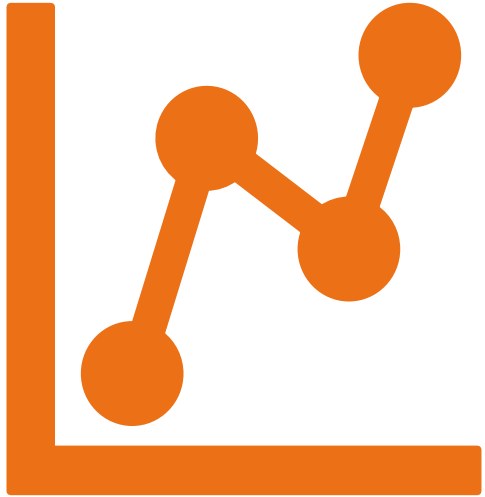
Management — human resources, compensation, and quality management

Management Information Systems — performance of systems which gather, summarize, and disseminate information to various managerial levels

Marketing — market analysis and consumer research

International Business — market and demographic analysis

Using Statistics (Two Categories)



Descriptive Statistics

- Collect
- Organize
- Summarize
- Display
- Analyze



Inferential Statistics

- Predict and forecast values of population parameters
- Test hypotheses about values of population parameters
- Make decisions

Data & 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 & 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

The diagram illustrates the structure of a data set. It features a table with callouts identifying its components:

- Element Names:** A callout pointing to the 'Company' column.
- Observation:** A callout pointing to a single row of data.
- Variables:** A callout pointing to the columns 'Stock Exchange', 'Annual Sales(\$M)', and 'Earn/Share(\$)'.
- Data Set:** A callout pointing to the entire table.

Company	Stock Exchange	Annual Sales(\$M)	Earn/Share(\$)
Dataram	NQ	73.10	0.86
EnergySouth	N	74.00	1.67
Keystone	N	365.70	0.86
LandCare	NQ	111.40	0.33
Psychemedics	N	17.60	0.13

Scales of Measurement

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.

Levels of Data Measurement

Nominal – In nominal measurement the values just "name" the attribute uniquely.

- No ordering of the cases is implied.
- For example, a person gender is nominal. males vs. females (Mutually Exclusive)
- Another example is religion – Catholic, Protestant, Muslim, etc. (No quantitative data)
- Dichotomous (2 categories only)

Nominal Level Data

Numbers are used to classify or categorize

Example: Employment Classification

- 1 for Educator
- 2 for Construction Worker
- 3 for Manufacturing Worker

Levels of Data Measurement

Ordinal - A variable is ordinal measurable if ranking/ordering is possible for values of the variable.

- Non-numeric concepts like satisfaction, happiness, discomfort(Likert Scale)
- Preference scales are typically ordinal – how much do you like this cereal? Like it a lot, somewhat like it, neutral, somewhat dislike it, dislike it a lot.

Ordinal Level Data

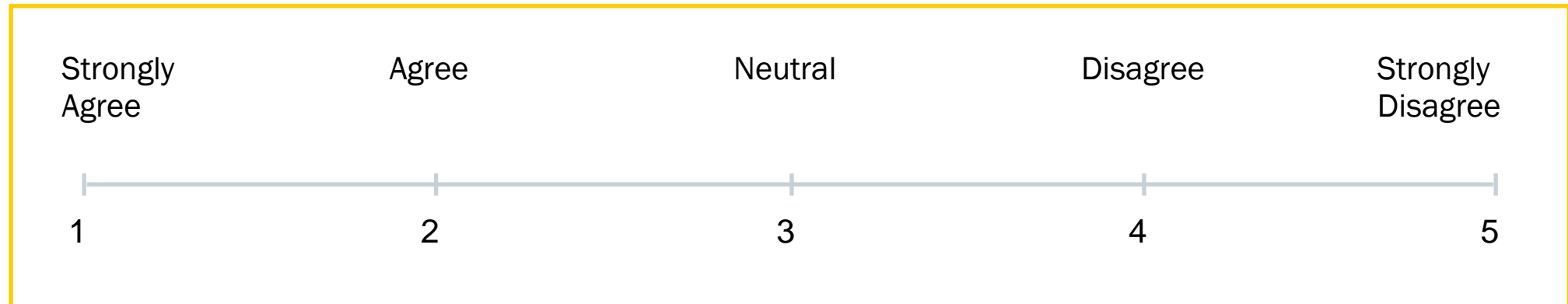
Numbers are used to indicate rank or order

Example: Ranking productivity of employees

Example: Position within an organization

- 1 for President
- 2 for Vice President
- 3 for Plant Manager
- 4 for Department Supervisor
- 5 for Employee

Ordinal Data



Levels of Data Measurement

Interval - In interval measurement the distance between attributes *does* have meaning.

- Numerical data typically fall into this category (Add/subtract)
- For example, when measuring temperature (in Fahrenheit), the distance from 30-40 is same as the distance from 70-80. The interval between values is interpretable.

Interval Level Data

Interval Level data - Distances between consecutive integers are equal

- Relative magnitude of numbers is meaningful
- Differences between numbers are comparable
- Location of origin, zero, is arbitrary
- Vertical intercept of unit of measure transform function is not zero

Example: Celsius Temperature, time

Levels of Data Measurement

Ratio – in ratio measurement there is always a reference point that is meaningful (either 0 for rates or 1 for ratios)

- This means that you can construct a meaningful fraction (or ratio) with a ratio variable. (descriptive statistics)
- In applied social research most "count" variables are ratio, for example, the number of clients in past six months.

Ratio Level Data

Highest level of measurement

- Relative magnitude of numbers is meaningful
- Differences between numbers are comparable
- Location of origin, zero, is absolute (natural)
- Vertical intercept of unit of measure transform function is zero

Examples: Height, Weight, and Volume

Example: Monetary Variables, such as Profit and Loss, Revenues, Expenses, Financial ratios - such as P/E Ratio, Inventory Turnover, and Quick Ratio.

Data measurement with Questionnaire

These questions will result in what level of data measurement?

1. How long ago were you released from the hospital?

2. Which type of unit were you in for most of your stay?

Coronary care

Intensive care

Maternity care

Medical unit

Pediatric/children's unit

Surgical unit

Data measurement with Questionnaire

3. In choosing a hospital, how important was the hospital's location? (circle one)

Very Important Somewhat important Not Very important Not at All Important

4. How serious was your condition when you were first admitted to the hospital?

Critical Serious Moderate Minor

5. Rate the skill of your doctor:

__Excellent __Very Good __Good __Fair __Poor

PRACTICE EXAMPLE

Classify each of the following as nominal, ordinal, interval, or ratio data.

- a. The time required to produce each tire on an assembly line
- b. The number of quarts of milk a family drinks in a month
- c. The ranking of four machines in your plant after they have been designated as excellent, good, satisfactory, and poor
- d. The telephone area code of clients in the United States
- e. The age of each of your employees
- f. The dollar sales at the local pizza shop each month
- g. An employee's identification number
- h. The response time of an emergency unit

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 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:

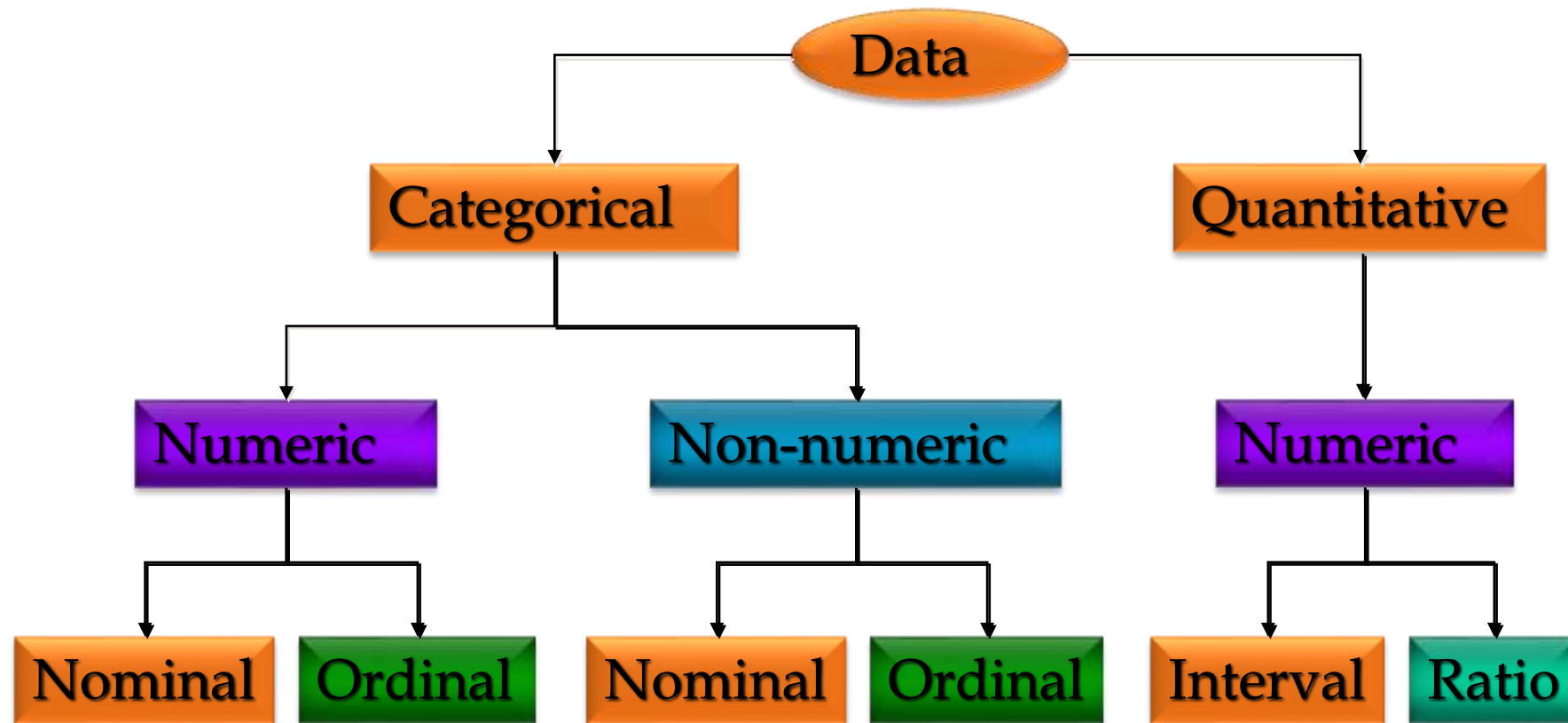
discrete, if measuring how many

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 currency rate of different countries today

Time Series Data

Time series data are collected over several time periods.

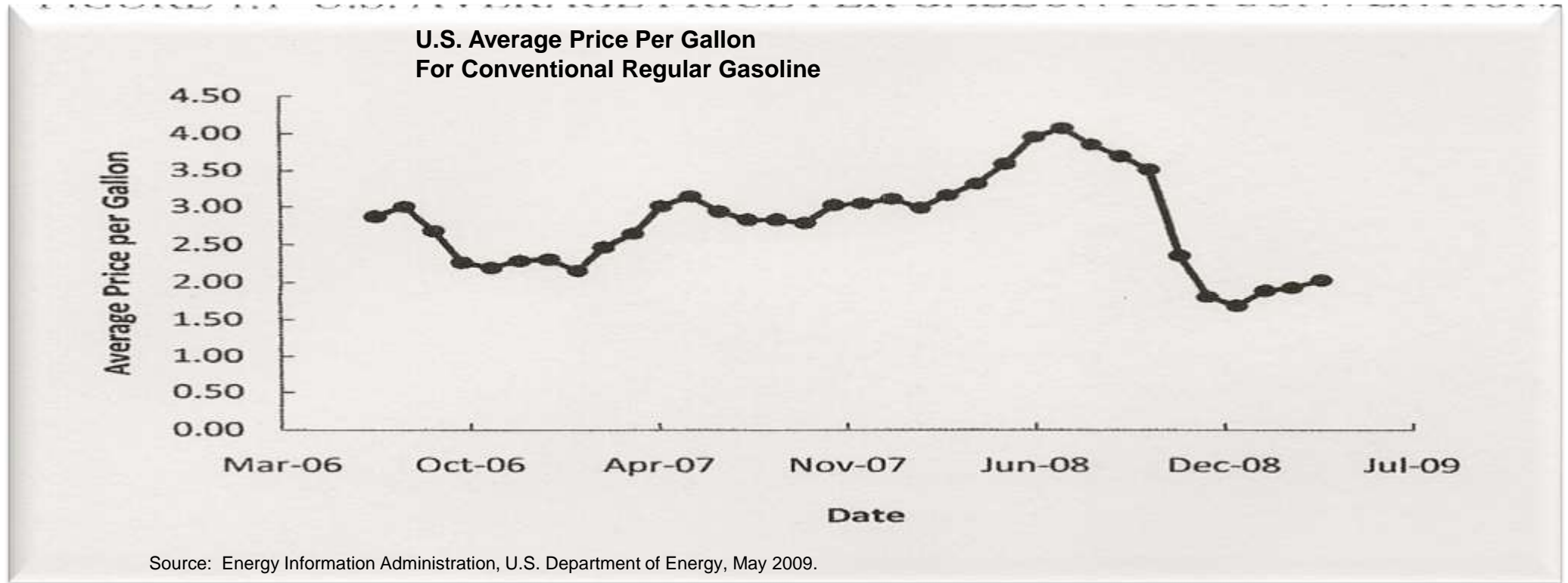
Example: data detailing the U.S. average price per gallon of regular gasoline between 2009 to 2014

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

□ Graph of Time Series Data



Data Sources

Existing Sources (Secondary Sources)

Internal company records - almost any department

Business database services - KPEMG, Clarivate

Government agencies - Indian government

Industry associations - Travel Industry Association
of America

Special-interest organizations - Graduate Management
Admission Council

Internet - more and more firms

Data Sources

□ Data Available From Internal Company Records

Record	Some of the Data Available
<u>Employee records</u>	name, address, social security number
<u>Production records</u>	part number, quantity produced, direct labor cost, material cost
<u>Inventory records</u>	part number, quantity in stock, reorder level, economic order quantity
<u>Sales records</u>	product number, sales volume, sales volume by region
<u>Credit records</u>	customer name, credit limit, accounts receivable balance
<u>Customer profile</u>	age, gender, income, household size

Data Sources

□ Data Available From Selected Government Agencies

Government Agency

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 (Primary 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. (Under controlled conditions)

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

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

Researchers observe group of customers at Walmart such as length of time customer spends on shopping, the gender of the customer, the amount spent and so on

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.

PRACTICE EXAMPLE

The U.S. Department of Energy provides fuel economy information for a variety of motor vehicles. A sample of 10 automobiles is shown in Table 1.6 (Fuel Economy website, February 22, 2008). Data show the size of the automobile (compact, midsize, or large), the number of cylinders in the engine, the city driving miles per gallon, the highway driving miles per gallon, and the recommended fuel (diesel, premium, or regular).

- a. How many elements are in this data set?
- b. How many variables are in this data set?
- c. Which variables are categorical and which variables are quantitative?
- d. What type of measurement scale is used for each of the variables?

TABLE 1.6 FUEL ECONOMY INFORMATION FOR 10 AUTOMOBILES

Car	Size	Cylinders	City MPG	Highway MPG	Fuel
Audi A8	Large	12	13	19	Premium
BMW 328Xi	Compact	6	17	25	Premium
Cadillac CTS	Midsize	6	16	25	Regular
Chrysler 300	Large	8	13	18	Premium
Ford Focus	Compact	4	24	33	Regular
Hyundai Elantra	Midsize	4	25	33	Regular
Jeep Grand Cherokee	Midsize	6	17	26	Diesel
Pontiac G6	Compact	6	15	22	Regular
Toyota Camry	Midsize	4	21	31	Regular
Volkswagen Jetta	Compact	5	21	29	Regular

Solutions

- a. The ten elements are the ten cars
- b. 5 variables: Size, Cylinders, City MPG, Highway MPG, and Fuel
- c. Categorical variables: Size and Fuel
Quantitative variables: Cylinders, City MPG, and Highway MPG

d.

Variable	Measurement Scale
Size	Ordinal
Cylinders	Ratio
City MPG	Ratio
Highway MPG	Ratio
Fuel	Nominal

PRACTICE EXAMPLE

Table 1.7 shows data for seven colleges and universities. The endowment (in billions of dollars) and the percentage of applicants admitted are shown (USA Today, February 3, 2008). The state each school is located in, the campus setting, and the NCAA Division for varsity teams were obtained from the National Center of Education Statistics website, February 22, 2008..

- a. How many elements are in the data set?
- b. How many variables are in the data set?
- c. Which of the variables are categorical and which are quantitative?
- d. What type of measurement scale is used for each of the variables?

TABLE 1.7 DATA FOR SEVEN COLLEGES AND UNIVERSITIES

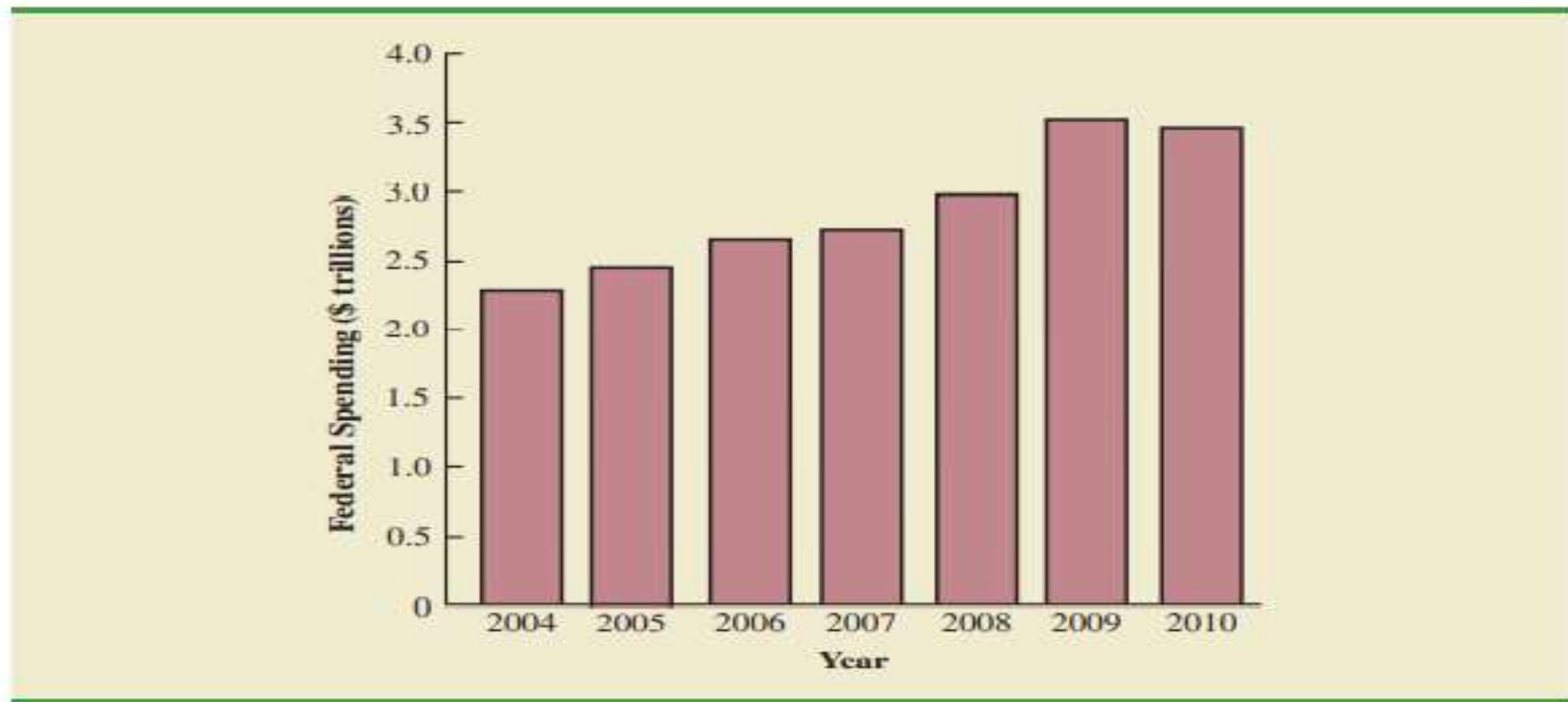
School	State	Campus Setting	Endowment (\$ billions)	% Applicants Admitted	NCAA Division
Amherst College	Massachusetts	Town: Fringe	1.7	18	III
Duke	North Carolina	City: Midsize	5.9	21	I-A
Harvard University	Massachusetts	City: Midsize	34.6	9	I-AA
Swarthmore College	Pennsylvania	Suburb: Large	1.4	18	III
University of Pennsylvania	Pennsylvania	City: Large	6.6	18	I-AA
Williams College	Massachusetts	Town: Fringe	1.9	18	III
Yale University	Connecticut	City: Midsize	22.5	9	I-AA

PRACTICE EXAMPLE

Figure 1.8 provides a bar chart showing the amount of federal spending for the years 2004 to 2010 (Congressional Budget Office website, May 15, 2011).

- a. What is the variable of interest?
- b. Are the data categorical or quantitative?
- c. Are the data time series or cross-sectional?
- d. Comment on the trend in federal spending over time

FIGURE 1.8 FEDERAL SPENDING



Case Study (Ken Black, 6th Ed, p. 15)

Digiornio Pizza: Introducing a frozen pizza to compete with carry out

Home Assignments

Collection of student Demographics Data

From book Statistics for Business and Economics (13th edition)

By Anderson/ Sweeney/ Williams/ Camm/ Cochran

P-24 (Q2, Q4)

P-27 (Q-13,15)