Applied Statistics in Business and Economics 6th edition Doane Solutions Manual

SKU: 9781259957598-SOLUTIONS

Chapter 2

Data Collection

2.1 a. Categorical
b. Categorical
c. Discrete numerical
Learning Objective: 02-2

2.2 a. Continuous numerical
b. Discrete numerical
c. Categorical
d. Continuous numerical
Learning Objective: 02-2

2.3 a. Continuous numerical
b. Continuous numerical (often reported as an integer)
c. Categorical
d. Categorical
Learning Objective: 02-2

2.4 Answers will vary.
Learning Objective: 02-2

2.5 a. Cross-sectional
b. Time series
c. Time series
d. Cross-sectional.
Learning Objective: 02-3

2.6 a. Time series
b. Cross-sectional
c. Time series
d. Cross-sectional
Learning Objective: 02-3

2.7 a. Time series
b. Cross-sectional.
c. Time series.
d. Cross-sectional.
Learning Objective: 02-3

2.8 Answers will vary.
Learning Objective: 02-3

2.9 a. Ratio. The number of hits is an integer with zero a possibility.
b. Ordinal. Ranking but difference in ranks is not meaningful.
c. Nominal. Positions on the field have no ranking implied.
d. Interval. Celsius is an interval measure because the zero is not meaningful.
e. Ratio. Salary has a meaningful zero.
f. Ordinal. Ranking but differences are not meaningful.
Learning Objective: 02-4

2.10 a. Ratio. The number of employees is a count and you can have zero employees.
b. Ratio. The number of returns is a count and you can have zero returns.
c. Interval. The temperature difference from 70 degrees to 80 degrees is the same increase as 80 degrees to 90 degrees. However, zero temperature does not mean no temperature exists, therefore it is interval.
d. Nominal. It is not a number and you could not rank order this cashier with others.
e. Ordinal. Ratings of employees generally fall into categories such as "exceeds standards", etc. Therefore, we know it is either nominal or ordinal and since we can rank order this employee with others given their rankings, we can say it is ordinal.
f. Nominal. There is no meaningful zero and distance between social security numbers has no meaning. We also would not rank order based on social security number so this is nominal even though it is a number.
Learning Objective: 02-4

2.11 a. Ratio. The number of passengers is a count and the zero point means absence of passengers.
b. Ratio. The waiting time is a continuous variable and you can have a waiting time equal to zero.
c. Nominal. Brand names are categories with no ranking.
d. Ordinal. Ticket class is a nonnumerical category and there is a ranking implied.
e. Interval. The temperature difference from 70 degrees to 80 degrees is the same increase as 80 degrees to 90 degrees. However, zero temperature does not mean no temperature exists, therefore it is interval.
f. Interval most common answer. Likert scales are typically assumed to be interval. Ordinal is a possible answer if the assumption is that the differences between ratings are not equal.
Learning Objective: 02-4

2.12 a. Ordinal (possibly interval). There is no meaningful zero so we can eliminate ratio. There is a rank order to the "categories" so we can eliminate nominal. With only three responses on the scale most statisticians would call this ordinal meaning the intervals between responses are not equal.
b. Ordinal. There is no meaningful zero so we can eliminate ratio. There is a rank order to the "categories" so we can eliminate nominal. But we cannot assume the difference between Rarely and Often is the same as the difference between Often and Very Often.
c. Nominal. There is no meaningful zero or distance or ranking.
d. Ratio. This is a number not a category and zero has meaning.

Learning Objective: 02-4
Learning Objective: 02-5

2.13 a. Interval, assuming intervals are equal, otherwise ordinal.
b. Yes (assuming interval data)
c. 10 point scale might give too many points and make it hard for guests to choose between.
Learning Objective: 02-4
Learning Objective: 02-5

2.14 a. Interval because it is a ranking with meaningful intervals between scale points.
b. No, we can only say that the difference between 3 and 4 is the same as the difference between 4 and 5.
c. Yes, a 5 point Likert scale would work just as well. In fact, a 5 point scale might be preferred. It might be difficult for customers to differentiate between a 1 and a 2 on 10 point scale whereas a 5 point scale would make it easier for the customer to answer.
Learning Objective: 02-4
Learning Objective: 02-5

2.15 a. Census. You can easily ask each of your friends this question.
b. Census or Sample. If your class is large you might take a sample.
c. Sample. The number of students at a university is too large to take a census.
d. Census. You most likely have fewer than 7 classes so fewer than 7 professors.
Learning Objective: 02-6

2.16 a. Sample. Over the lifetime of your computer you will recharge your battery a very high number of times. A sample makes sense in this case.
b. Census or sample. If your class is large you might take a sample.
c. Sample. The number of students at a university is too large to take a census.
d. Census. You can easily ask each of your friends this question.
Learning Objective: 02-6

2.17 a. Parameter. The S&P is the population.
b. Parameter. Same as above. The S&P is the population.
c. Statistic. We clearly stated a random *sample*.
d. Statistic. This isn't random but it could be considered a sample.
Learning Objective: 02-6

2.18 Use the formula: $N = 20 \times n$
a. $N = 20 \times 10 = 200$
b. $N = 20 \times 50 = 1000$
c. $N = 20 \times 100 = 2000$
Learning Objective: 02-7

2.19 a. Convenience.
b. Systematic.
c. Judgment or biased.
Learning Objective: 02-7

2.20 a. Random
b. Convenience
c. Systematic
Learning Objective: 02-7

2.21 Answers will vary.
Learning Objective: 02-7

2.22 a. There were 24 ages under 30. The proportion is 24/48 = 0.50.
b. Answers will vary.
c. Answers will vary.
Learning Objective: 02-7

2.23 Answers will vary.
Learning Objective: 02-7

2.24 a. Response bias. The students might exaggerate the number of dates they've had.
b. Self-selection bias, coverage error. By only asking folks outside of a church you might get a number that is higher than the number from the general public.
c. Coverage error, self-selection bias. Same reasons as in part b.
Learning Objective: 02-9

2.25 a. Telephone or web. A web-based survey might overestimate the numbers who prefer a web-based course.
b. Direct observation of students on campus.
c. Interview, web, or mail. Response rates would most likely differ with the three methods. Mail surveys tend to have lower response rates.
d. Interview or web.
Learning Objective: 02-9

2.26 a. Mail or interview. You would most likely have a list of customer addresses but you could also just ask customers that come in. A mail survey might have a lower response rate.
b. Direct observation, through customer invoices/receipts.
c. If you track zip codes as well as invoices/receipts this could be done via direct observation of your records. However, mail would be another option if that data is unavailable.
d. Interview since you only have to ask seven employees.
Learning Objective: 02-9

2.27 Version 1: Most would say yes. Version 2: More varied responses.
Learning Objective: 02-9

2.28 Does not include all possible responses or allow for the responder to pick something other than those presented.
Learning Objective: 02-9

2.29 a. Continuous numerical. Age can be measured with fractions.
b. Categorical. Nationality is not a numerical measure.
c. Discrete numerical. We can count the double-faults using integers.
Learning Objective: 02-2

2.30 a. Discrete numerical. We can count the number of spectators using integers.
b. Continuous numerical. The amount of water can be fractions of liters.
c. Categorical. Gender is a category, not a number.
Learning Objective: 02-2

2.31 a. Ordinal. We do have a ranking but the differences between rankings would not be equal.
b. Interval measure if using a noise meter to measures decibels (20dB is not twice as much as 10dB; 0dB does not mean no sound). But if the noise level is based on a word description such as "noisy" or "quiet" then the measurement scale would be ordinal.
c. Ratio because this is a count.
Learning Objective: 02-4

2.32 a. Ratio because this is a count.
b. Ratio if there were a way to measure the actual amount. Most likely this would be an ordinal measure because one would characterize the consumption as high, medium or low.
c. Categorical. Type of vehicle is not a numerical measure and no ranking is implied.
Learning Objective: 02-4

2.33 Q1 Categorical, nominal. Not numerical and no ranking.
Q2 Continuous, ratio. Can take on decimal values and has a clearly defined zero value.
Q3 Continuous, ratio. Can take on decimal values and has a clearly defined zero value.
Q4 Discrete, ratio. Integer, clearly defined zero.
Q5 Categorical, ordinal or interval. Interval if differences are equal.
Learning Objective: 02-4
Learning Objective: 02-5

2.34 Q1 Categorical, ordinal or interval. Interval if differences are equal.
Q2 Discrete, ratio. Integer, clearly defined zero.
Q3 Continuous, ratio. Can take on decimal values and has a clearly defined zero value.
Q4 Discrete, ratio. Integer, clearly defined zero.
Q5 Categorical, ordinal. Ranking but not numerical so no calculations possible.
Learning Objective: 02-4
Learning Objective: 02-5

2.35 Q1 Continuous, ratio. Can take on decimal values and has a clearly defined zero value.
Q2 Discrete, ratio. Integer, clearly defined zero.
Q3 Categorical, ordinal or interval. Interval if differences are equal.
Q4 Categorical, nominal. Binary, no ranking.
Q5 Categorical, ordinal or interval. Interval if differences are equal.
Learning Objective: 02-4
Learning Objective: 02-5

2.36 a. Cross-sectional. A single point in time: end of 2017.
b. Time series. Data is collected over an 8 year time period.
c. Time series. Data collected over 52 weeks.
d. Cross-sectional. Single point in time: Christmas Day 2017.
Learning Objective: 02-3

2.37 a. Time series. Data collected over 31 days in January.
b. Cross-sectional. Single point in time: start of a particular semester.
c. Cross-sectional. Single point in time: summary for a particular week.
d. Time series. Data collected for the past 10 years.
Learning Objective: 02-3

2.38 a. Census. It would be easy enough to count all of them.
b. Sample. It would be too costly to track each can.
c. Census. You can count them all quickly and cheaply.
Learning Objective: 02-6

2.39 a. Census. This is assuming the company can easily generate the value from its human resource center.
b. Sample. Impossible to observe prices of all cans in grocery stores.
c. Census. This should be in Campbell Soup's data base.
Learning Objective: 02-6

2.40 a. Statistic. The data collected at your local supermarket would be a sample for the population of all soup sold by the company.
b. Parameter. The population is all soup sold last year.
c. Statistic. The sample consists of 10 students.
Learning Objective: 02-6

2.41 a. Statistic. The week of visits is the sample.
b. Parameter. The population is all books sold to date.
c. Parameter. The population is all books sold.
Learning Objective: 02-6

2.42 No, a census would be too difficult since this is an infinite population (people can continue to send e-mails).
Learning Objective: 02-6

2.43 a. The patient's complaint.
b. The number of patient visits is discrete numerical. The waiting time is continuous.
Learning Objective: 02-3

2.44 a. Simple Random Sample. It is easy enough to use a computerized random number generator to choose 15 ports of entry.
Learning Objective: 02-7

2.45 No a census could not be used. It would be impossible to ask each taxpayer how much time they spent in preparation. A sample is more appropriate.
Learning Objective: 02-6

2.46 b. Cluster sampling. Easier to define geographic areas within a state where gasoline is sold. Gasoline stations are not everywhere, thus simple random sample or stratified sampling doesn't make sense.
Learning Objective: 02-7

2.47 a. Cluster sampling. It makes sense to take samples from geographic regions.
b. No, population is effectively infinite.
Learning Objective: 02-7

2.48 a. Sample – this information is most likely collected by a survey of a sample of customers.
b. Census – this information can be collected from the point-of-sale system.
c. Sample – this is most likely monitored by sampling coffee served.
d. Census – this can be tracked on the point-of-sale system and will be population data.
Learning Objective: 02-6

2.49 a. Census – this information is collected for all restaurants.
b. Sample – this cannot be tracked for all customers, must be taken from a sample.
c. Sample – this cannot be tracked for all customers, must be taken from a sample.
d. Census – this can be tracked on the point-of-sale system and will be population data.
Learning Objective: 02-6

2.50 Simple random sample or systematic sampling. A simple random sample is always best because it reduces bias. If it is truly random, every major stock fund was equally likely to be chosen. One way to do that is to:

- Create an excel spreadsheet with the funds listed and numbered
- Click on a separate cell and use the excel function =RANDBETWEEN(1,1699).
- This will give you one random number between 1 and 1,699. Whatever that number is can represent the first randomly chosen fund. For example, if the random number is 42, you would select the fund that you had listed under the #42.
- To get the other 20 randomly chosen funds, you would simply drag the bottom right corner of the cell that has the first number, to the next 19 cells below it.

Another way to get a random sample is to use systematic sampling. For example, I might decide to take every 5th fund until I have 20. Pick a random starting point and then take every 20^{th} fund from the starting point.
Learning Objective: 02-7

2.51 a. Cluster sample. Most likely choose businesses within a geographic region then take a random sample within the region.
b. Cluster sample. Most likely choose practices within a geographic region then take a random sample within the region.
c. Simple Random Sample (SRS), fairly accurate.
d. The statistic is most likely based on sales data reported by cigarette companies. While the data does not come from a random sample, this information is available for almost all companies and therefore fairly accurate.
Learning Objective: 02-7

2.52 a. This is a simple random sample. This population is effectively infinite because $n = 780$ and $780 \times 20 = 15{,}600$. This value is much less than $N = 999{,}645$.
Learning Objective: 02-7

2.53 a. Cluster sampling, neighborhoods are natural clusters.
b. Picking a day near a holiday with light trash.
Learning Objective: 02-7
Learning Objective: 02-9

2.54 Yes, the population is effectively infinite. The population size, 30000, is at least 20 times greater than the sample size: $20 \times 600 = 12000 < 30000$.
Learning Objective: 02-6

2.55 a. Yes, the population is effectively infinite because $18 \times 20 < 11{,}000$. b. 1/39 is the value from the sample therefore it is the statistic.
Learning Objective: 02-6

2.56 Because $1200 \times 20 = 24{,}000$ and this value is less than the population we can consider the population effectively infinite.
Learning Objective: 02-6

2.57 Education and income could affect who uses the no-call list.
a. They won't reach those who purchase such services. Same response for b and c.
Learning Objective: 02-9

2.58 For each question, the difficulty is deciding what the possible responses should be and giving a realistic range of responses.
Learning Objective: 02-9

2.59 a. Rate the effectiveness of this professor. 1 – Excellent to 5 – Poor.
b. Rate your satisfaction with the President's economic policy. 1 – Very Satisfied to 5 – Very dissatisfied.
c. How long did you wait to see your doctor? Less than 15 minutes, between 15 and 30 minutes, between 30 minutes and 1 hour, more than 1 hour.
Learning Objective: 02-4
Learning Objective: 02-9

2.60 Ordinal measure. There is no numerical scale and the intervals are not considered equal.
Learning Objective: 02-4
Learning Objective: 02-9

2.61 a. Ordinal.
b. That the intervals are equal.
Learning Objective: 02-4
Learning Objective: 02-9

2.62 a. A binary response scale.
b. A Likert scale would be better.
c. Self-selection bias. People with very bad experiences might respond more often than people with acceptable experiences.
Learning Objective: 02-5
Learning Objective: 02-9

2.63 Answers will vary.
Learning Objective: 02-2

2.64 Answers will vary.
Learning Objective: 02-3

2.65 Answers will vary.
Learning Objective: 02-7

2.66 Answers will vary.
Learning Objective: 02-7

2.67 Answers vary for a-c; most appropriate method is simple random sampling (or stratified based on department).
Learning Objective: 02-7

2.68 We can use the =RANDBETWEEN(1,52) function in excel to get random numbers which will allow us to choose 5 random cards. A stratified sample would not work since the cards are listed in order and spades and hearts are listed first. Therefore, if I chose every 5th card and stopped after 5 cards, I would not have any clubs or diamonds represented. Stratified sampling doesn't really make sense since there is already an equal number of each suit and the same numbers within each suit. Cluster sampling

doesn't make sense since we are not concerned with geographic region. Judgment is not necessary when choosing playing cards, and convenience would be possible but not necessary and we want to avoid it if possible.

Learning Objective: 02-7

2.69 Answers will vary.
Learning Objective: 02-7

2.70 Answers will vary.
Learning Objective: 02-7

2.71 Answers will vary.
Learning Objective: 02-7

A PowerPoint Presentation Package to Accompany

Applied Statistics in Business & Economics, 6th edition

David P. Doane and Lori E. Seward

Prepared by Lloyd R. Jaisingh

Data Collection

Chapter Contents

Chapter Learning Objectives

LO2-1: Use basic terminology for describing data and samples.

LO2-2: Explain the difference between numerical and categorical data.

LO2-3: Explain the difference between time series and cross-sectional data.

LO2-4: Recognize levels of measurement in data and ways of coding data.

LO2-5: Recognize a Likert scale and know how to use it.

Chapter Learning Objectives (continued)

LO2-6: Use the correct terminology for samples and populations.

LO2-7: Explain the common sampling methods and how to implement them.

LO2-8: Find everyday print or electronic data sources.

LO2-9: Describe basic elements of survey types, survey designs, and response scales.

2.1 Variables and Data

LO2-1: Use basic terminology for describing data and samples.

Data Terminology: Observations, Variables, Data Sets

- *Observation*: a single member of a collection of items that we want to study, such as a person, firm, or region.
- *Variable*: a characteristic of the subject or individual, such as an employee's income or an invoice amount.
- *Data set*: consists of all the values of all of the variables for all of the observations we have chosen to observe.

LO2-1: Use basic terminology for describing data and samples (continued).

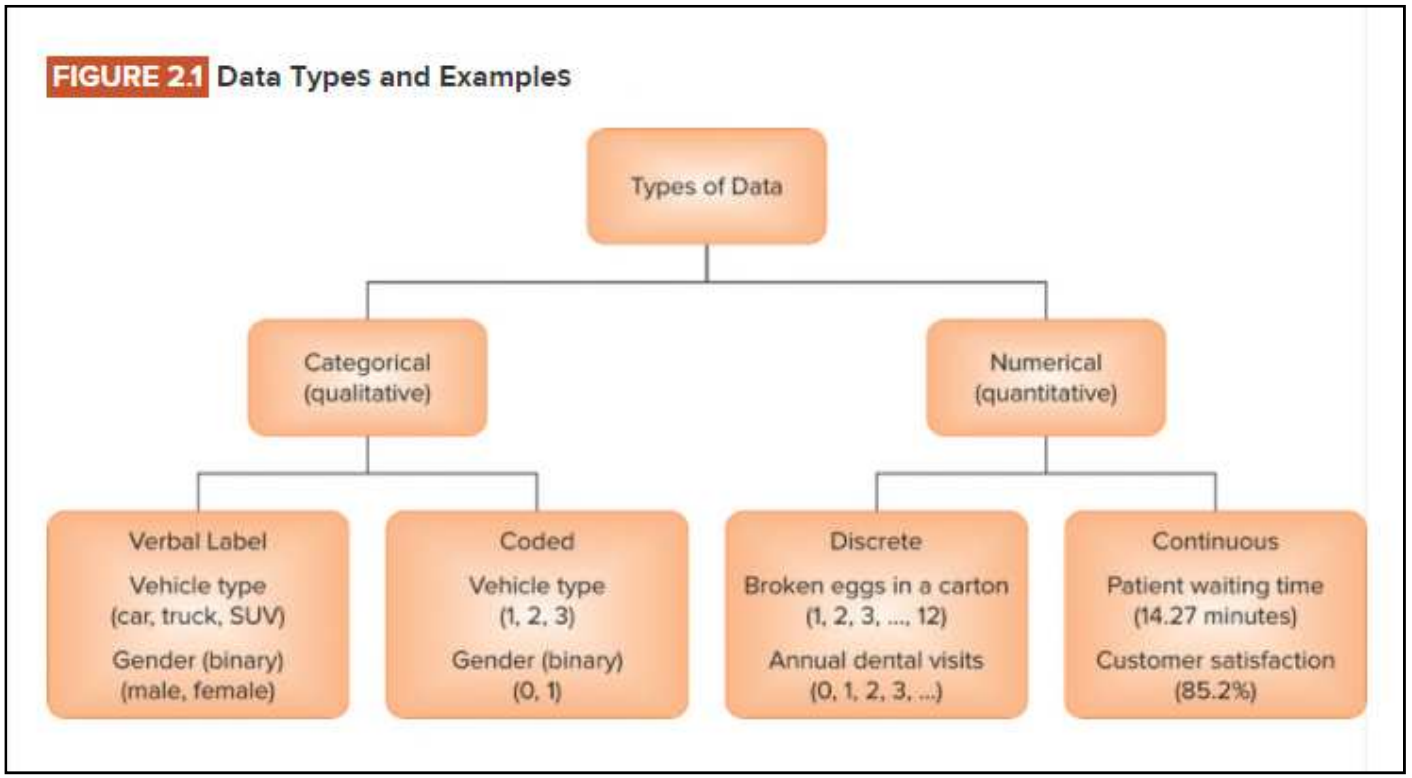
- A data set may consist of many variables.
- The questions that can be explored and the analytical techniques that can be used will depend upon the data type and the number of variables.
- This textbook starts with **univariate data sets** (one variable), then moves to **bivariate data sets** (two variables) and **multivariate data sets** (more than two variables), as illustrated in Table 2.2.

LO2-1: Use basic terminology for describing data and samples (continued, 2).

Table 2.2: Number of Variables and Typical Tasks

Data Set	*Variables*	*Example*	*Typical Tasks*
Univariate	One	Income	Histograms, basic statistics
Bivariate	Two	Income, Age	Scatter plots, correlation
Multivariate	More than two	Income, Age, Gender	Regression modeling

LO2-2: Explain the difference between numerical and categorical data.



Note: Ambiguity is introduced when continuous data are rounded to whole numbers. Be cautious.

LO2-3: Explain the difference between time series and cross-sectional data.

Time Series Data

- Each observation in the sample represents a different equally spaced point in time (e.g., years, months, days).
- The periodicity is the time between observations.
- Periodicity may be annual, quarterly, monthly, weekly, daily, hourly, etc.
- Examples of macroeconomics time series data would include national income, economic indicators, and monetary data.
- Examples of microeconomics time series data would include sales, market share, inventory turnover, etc.
- For time series, we are interested in trends and patterns over time (e.g., personal bankruptcies from 1980 to 2008 as shown in Figure 2.2).

LO2-3: Explain the difference between time series and cross-sectional data (continued).

Cross-Sectional Data

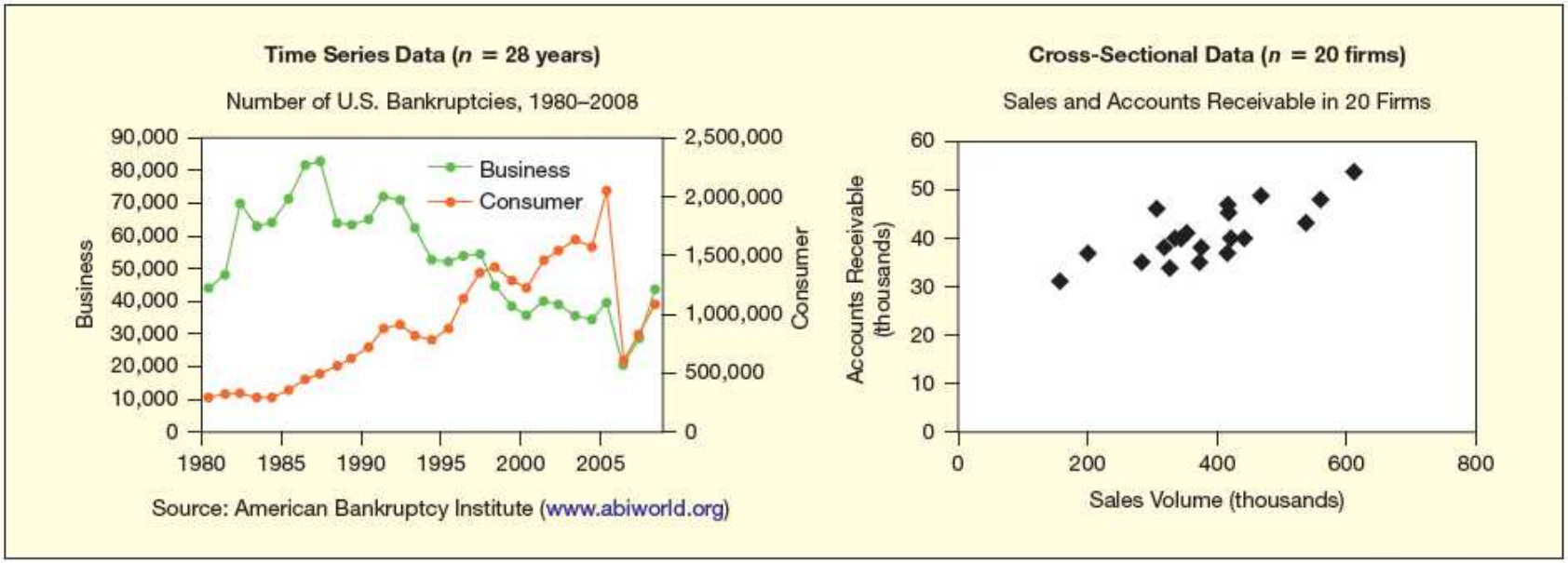
- Each observation represents a different individual unit (e.g., person) at the same point in time (e.g., monthly VISA balances).
- For cross-sectional data, we are interested in:
  - *variation among observations* (e.g., accounts receivable in 20 Subway franchises).
  - *relationships* (e.g., whether accounts receivable are related to sales volume in 20 Subway franchises as shown in Figure 2.2).
- We can combine the two data types to get pooled cross-sectional and time series data.

LO2-3: Explain the difference between time series and cross-sectional data (continued, 2).

Cross-Sectional Data (continued)

FIGURE 2.2

Examples of Time Series versus Cross-Sectional Data



2.2 Level of Measurement

LO2-4: Recognize levels of measurement in data and ways of coding data.

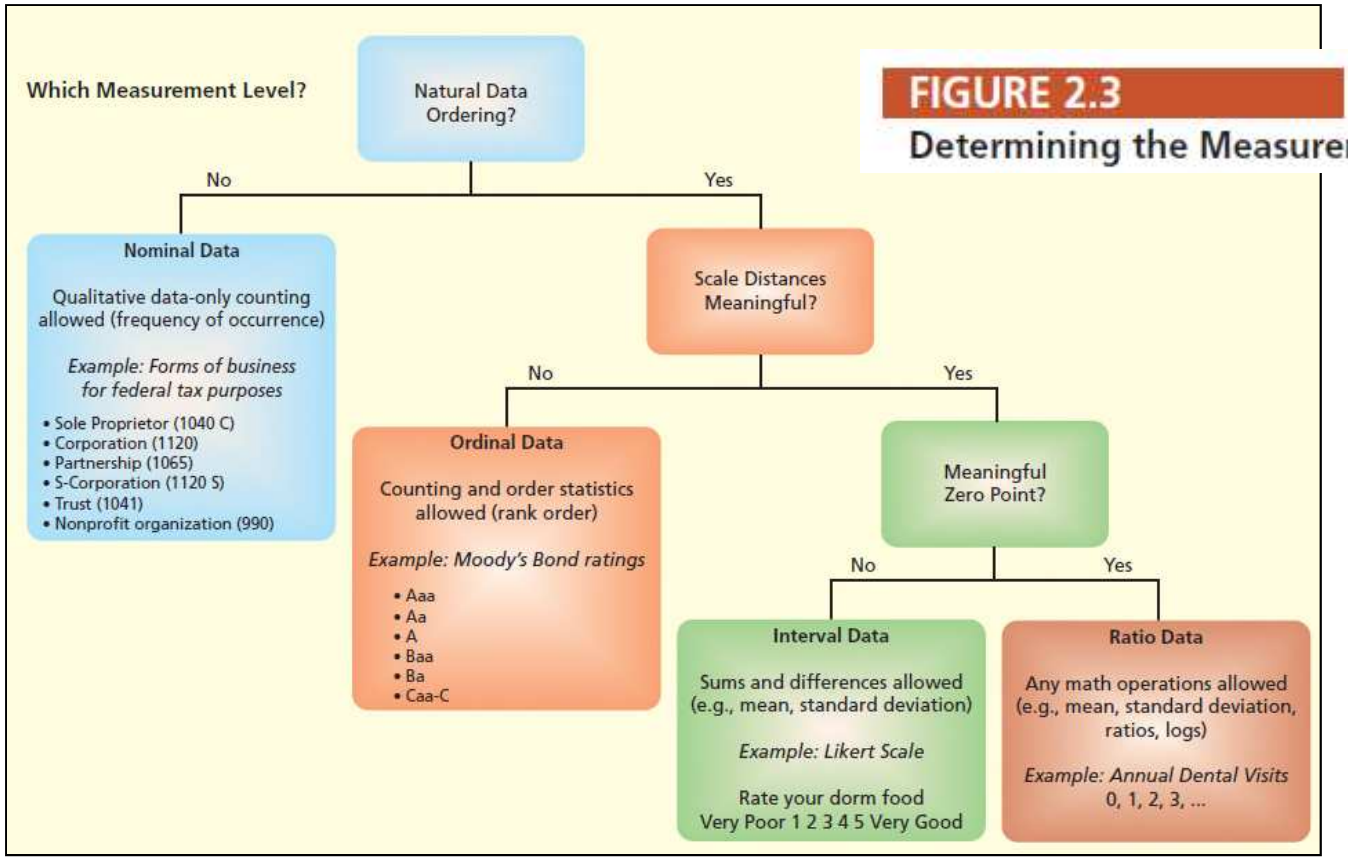


FIGURE 2.3
Determining the Measurement Level

LO2-4: Recognize levels of measurement in data and ways of coding data (continued).

Levels of Measurement

Level of Measurement	Characteristics	Example
Nominal	Categories only	Eye color (blue, brown, green, etc.)
Ordinal	Rank has meaning. No clear meaning to distance	Rarely, never
Interval	Distance has meaning	Temperature (57° Celsius)
Ratio	Meaningful zero exists	Accounts payable ($21.7 million)

LO2-4: Recognize levels of measurement in data and ways of coding data (continued, 2).

Nominal Measurement

- Nominal data merely identify a category.
- Nominal data are qualitative, attribute, categorical or classification data, and can be coded numerically (e.g., 1 = Apple, 2 = Samsung, 3 = Dell, 4 = HP).
- The only mathematical operations are counting (e.g., frequencies) and simple statistics.

Ordinal Measurement

- Ordinal data codes can be ranked (e.g., 1 = Frequently, 2 = Sometimes, 3 = Rarely, 4 = Never).

LO2-4: Recognize levels of measurement in data and ways of coding data (continued, 3).

Ordinal Measurement

- *Distance* between data codes is not meaningful (e.g., distance between 1 and 2, or between 2 and 3, or between 3 and 4 lacks meaning).
- Many useful statistical tests exist for ordinal data, which are especially useful in social science, marketing, and human resource research.

Interval Measurement

- Not only can data be ranked, but it can also have meaningful intervals between scale points (e.g., difference between 60°F and 70°F is same as difference between 20°F and 30°F).

LO2-4: Recognize levels of measurement in data and ways of coding data (continued, 4).

Interval Measurement

- Since intervals between numbers represent *distances*, mathematical operations can be performed (e.g., average).
- Zero point of interval scales is arbitrary, so ratios are not meaningful (e.g., 60°F *is not* twice as warm as 30°F).

Ratio Measurement

- *Ratio data* have all properties of nominal, ordinal, and interval data types and also possess a meaningful zero (absence of quantity being measured).

LO2-4: Recognize levels of measurement in data and ways of coding data (continued, 5).

Ratio Measurement

- Because of this zero point, ratios of data values are meaningful (e.g., $20 million profit is twice as much as $10 million).
- Zero does not have to be observable in the data; it is an absolute reference point.

LO2-5: Recognize a Likert scale and know how to use it.

Likert Scales

- A special case of interval data frequently used in survey research.
- The coarseness of a Likert scale refers to the number of scale points (typically 5 or 7).

LO2-5: Recognize a Likert scale and know how to use it (continued).

Likert Scales (examples)

"College-bound high school students should be required to study a foreign language." (check one)

❑	❑	❑	❑	❑
Strongly Agree	Somewhat Agree	Neither Agree Nor Disagree	Somewhat Disagree	Strongly Disagree

"How would you rate your Internet service provider?" (check one)

❑ Terrible ❑ Poor ❑ Adequate ❑ Good ❑ Excellent

LO2-5: Recognize a Likert scale and know how to use it (continued, 2).

Use the following procedure to recognize data types:

Question	If "Yes"
Q1. Is there a meaningful zero point?	Ratio data (statistical operations are allowed)
Q2. Are intervals between scale points meaningful?	Interval data (common statistics allowed, e.g., means and standard deviations)
Q3. Do scale points represent rankings?	Ordinal data (restricted to certain types of nonparametric statistical tests)
Q4. Are there discrete categories?	Nominal data (only counting allowed, e.g., finding the mode)

LO2-5: Recognize a Likert scale and know how to use it (continued, 3).

Changing Data By Recoding

- In order to simplify data or when exact data magnitude is of little interest, ratio data can be recoded downward into ordinal or nominal measurements (but not conversely).
- For example, recode systolic blood pressure as “normal” (under 130), “elevated” (130 to 140), or “high” (over 140).
- The above recoded data are ordinal (ranking is preserved), but intervals are unequal and some information is lost.

2.3 Sampling Concepts

LO2-6: Use the correct terminology for samples and populations.

Population or Sample?

- A *population* involves all of the items one is interested in. It may be finite (e.g., all of the passengers on a plane) or effectively infinite (e.g., all of the Cokes produced in an ongoing bottling process).
- A *sample* is a subset of the population and involves looking only at some of the items selected from the population.

LO2-6: Use the correct terminology for samples and populations (continued).

Sample or Census?

- A *sample* involves looking only at some items selected from the population.
- A *census* is an examination of all items in a defined population.
- Why can't the United States Census survey every person in the population? – mobility, undocumented workers, budget constraints, incomplete responses, etc.
- The next slide shows a summary of Table 2.4.

LO2-6: Use the correct terminology for samples and populations (continued, 2).

TABLE 2.4 Sample or Census?

Situations Where a Sample May Be Preferred	*Situations Where a Census May Be Preferred*
Infinite Population No census is possible if the population is of indefinite size (an assembly line can keep producing bolts, a doctor can keep seeing more patients).	**Small Population** If the population is small, there is little reason to sample, for the effort of data collection may be only a small part of the total cost.
Destructive Testing The act of measurement may destroy or devalue the item (battery life, vehicle crash tests).	**Large Sample Size** If the required sample size approaches the population size, we might as well go ahead and take a census.
Timely Results Sampling may yield more timely results (checking wheat samples for moisture content, checking peanut butter for salmonella contamination).	**Database Exists** If the data are on disk, we can examine 100% of the cases. But auditing or validating data against physical records may raise the cost.
Accuracy Instead of spreading resources thinly to attempt a census, budget might be better spent to improve training of field interviewers and improve data safeguards.	**Legal Requirements** Banks must count *all* the cash in bank teller drawers at the end of each business day. The U.S. Congress forbade sampling in the 2000 decennial population census.
Cost Even if a census is feasible, the cost, in either time or money, may exceed our budget.	
Sensitive Information A trained interviewer might learn more about sexual harassment in an organization through confidential interviews of a small sample of employees.	

LO2-6: Use the correct terminology for samples and populations (continued, 3).

Parameter or Statistic?

- A *parameter* is a measurement or characteristic of the population.
- A parameter is usually unknown because we can rarely observe the entire population.

LO2-6: Use the correct terminology for samples and populations (continued, 4).

Parameter or Statistic?

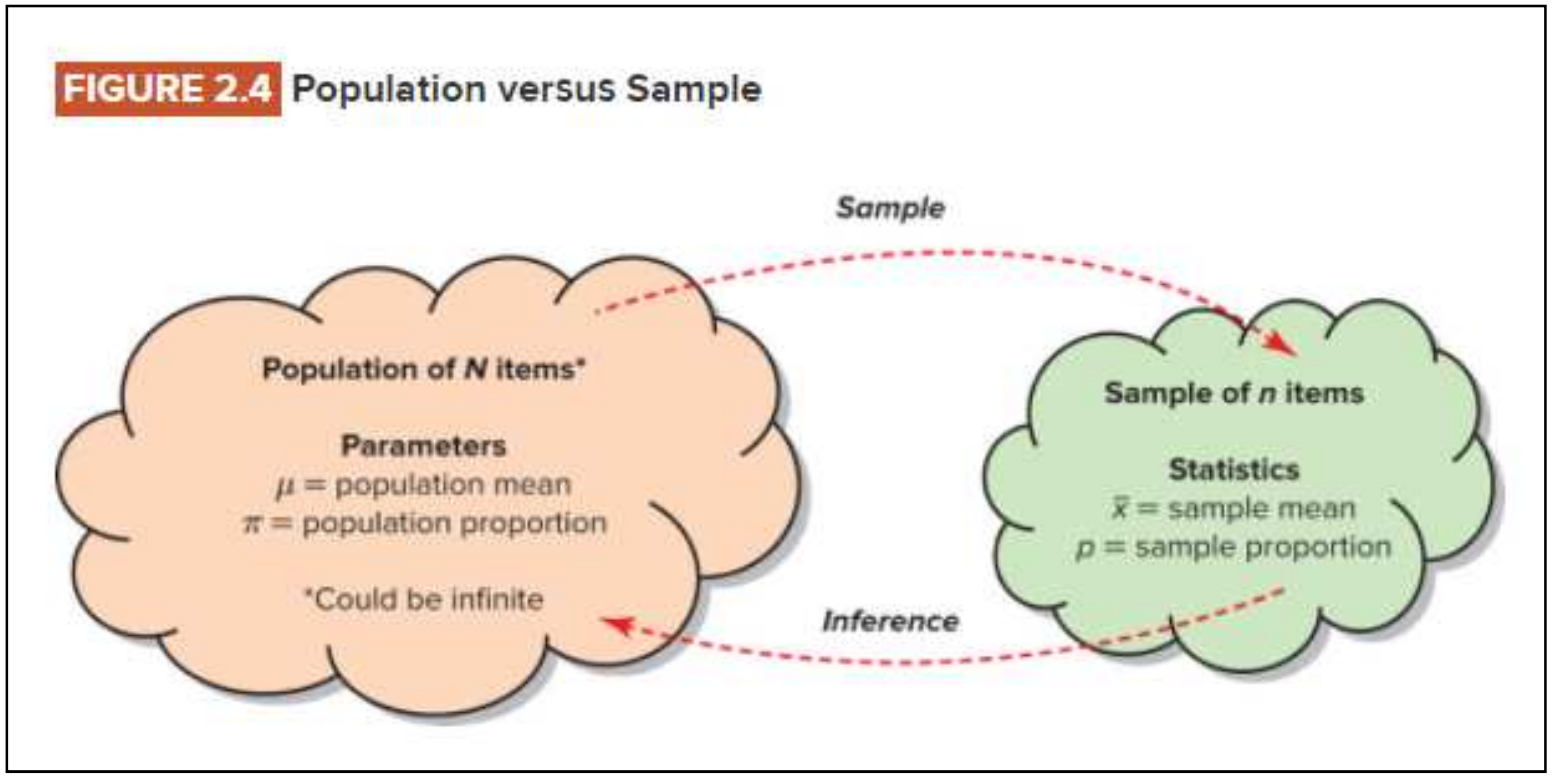
- *Statistic* is a numerical value computed from a sample.
- Statistics can be used as estimates of *parameters* found in the population.
- Symbols are used to represent population parameters and sample statistics.

LO2-6: Use the correct terminology for samples and populations (continued, 5).

- From a sample of *n* items, chosen from a population, we compute **statistics** that can be used as estimates of **parameters** found in the population.
- To avoid confusion, we use different symbols for each parameter and its corresponding statistic.
- Thus, the population mean is denoted by μ (the lowercase Greek letter mu) while the sample mean is $\bar{x}$.
- The population proportion is denoted π (the lowercase Greek letter pi), while the sample proportion is p.

LO2-6: Use the correct terminology for samples and populations (continued, 6).

Figure 2.4 illustrates this idea.



LO2-6: Use the correct terminology for samples and populations (continued, 7).

Target Population

- The population must be carefully specified and the sample must be drawn scientifically so that the sample is representative.
- The *target population* is the population we are interested in (e.g., U.S. gasoline prices).
- The *sampling frame* is the group from which we take the sample (e.g., 115,000 stations).
- The frame should not differ from the target population.

2.4 Sampling Methods

LO2-7: Explain the common sampling methods and how to implement them.

Random Sampling Methods

Simple Random Sample	Use random numbers to select items from a list (e.g., Visa cardholders).
Systematic Sample	Select every *k*th item from a list or sequence (e.g., restaurant customers).
Stratified Sample	Select randomly within defined strata (e.g., by age, occupation, gender).
Cluster Sample	Select random geographical regions (e.g., zip codes) that represent the population.

LO2-7: Explain the common sampling methods and how to implement them (continued).

Simple Random Sample

- We usually denote the population size by *N* and the sample size by *n*.
- In *a simple random sample,* every item in the population of *N* items has the same chance of being chosen in the sample of *n* items.
- A physical experiment to accomplish this would be to write each of the *N* data values on a poker chip, and then to draw *n* chips from a bowl after stirring it thoroughly.

LO2-7: Explain the common sampling methods and how to implement them (continued, 2).

Sampling With or Without Replacement

- With the advent of technology, random samples can easily be generated.
- If we do not allow duplicates when sampling, then we are sampling *without replacement*.
- If we allow duplicates when sampling, then we are sampling *with replacement*.

LO2-7: Explain the common sampling methods and how to implement them (continued, 3).

Sampling With or Without Replacement

- Instinctively most people believe that sampling without replacement is preferred over sampling with replacement because allowing duplicates in our sample seems odd.
- In reality, sampling without replacement can be a problem when our sample size *n* is close to our population size *N*.
- At some point in the sampling process, the remaining items in the population will no longer have the same probability of being selected as the items we chose at the beginning of the sampling process. This could lead to a bias.

LO2-7: Explain the common sampling methods and how to implement them (continued, 4).

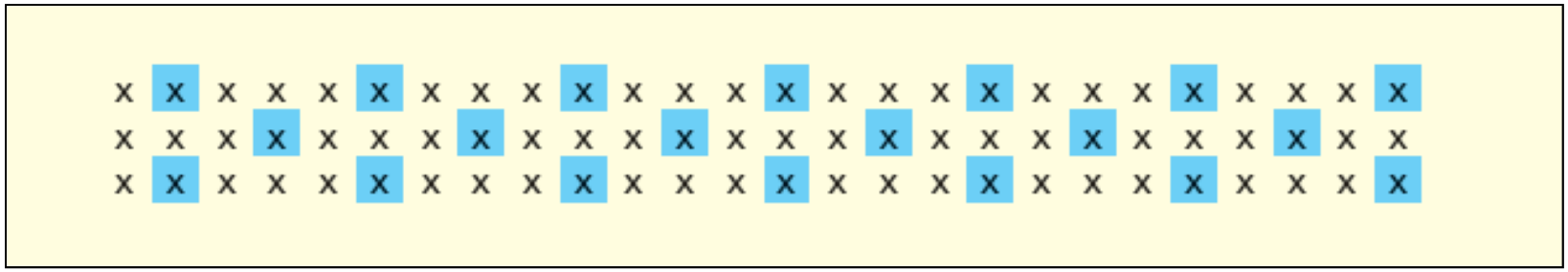
Sampling With or Without Replacement

- When should we worry about sampling without replacement?
- Only when the population is finite and the sample size is close to the population size.
- Note: A common criterion is that a finite population is *effectively infinite* if the sample is less than 5 percent of the population (i.e., if $n/N \leq .05$).
- An equivalent statement is that a population is effectively infinite when it is at least 20 times as large as the sample (i.e., when $N/n \geq 20$).

LO2-7: Explain the common sampling methods and how to implement them (continued, 5).

Systematic Sampling

- Sample by choosing every *kth* item from a list, starting from a randomly chosen entry on the list.
- For example, starting at item 2 (see below), we sample every 4 items to obtain a sample of $n = 20$ items from a list of $N = 78$ items. Note that $N/n = 78/20 \approx 4$ (periodicity).



LO2-7: Explain the common sampling methods and how to implement them (continued, 6).

Stratified Sampling

- Utilizes prior information about the population.
- Applicable when the population can be divided into relatively homogeneous subgroups of known size (*strata*).
- A simple random sample of the desired size is taken within each *stratum*.
- Alternatively, a random sample of the whole population could be taken, and then individual strata estimates could be combined using appropriate weights.

LO2-7: Explain the common sampling methods and how to implement them (continued, 7).

Cluster Sample

- Strata consist of geographical regions.
- *One-stage* cluster sampling – sample consists of all elements in each of *k* randomly chosen subregions (clusters).
- *Two-stage* cluster sampling, first choose *k* subregions (clusters), then choose a random sample of elements within each cluster.

LO2-7: Explain the common sampling methods and how to implement them (continued, 8).

Cluster Sample

Here is an example of 4 elements sampled from each of 3 randomly chosen clusters (two-stage cluster sampling).

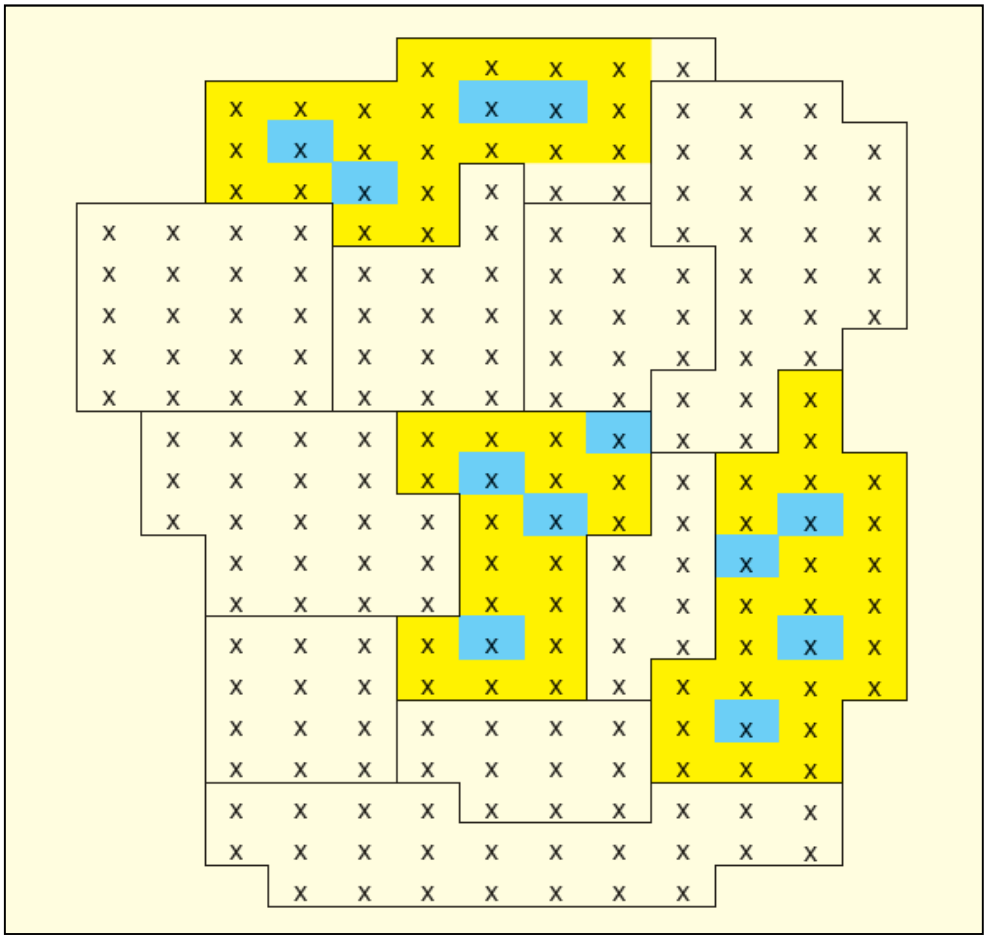


FIGURE 2.7

Two-Stage Cluster Sampling: Randomly choose three clusters, then randomly choose four items in each cluster

LO2-7: Explain the common sampling methods and how to implement them (continued, 9).

Non-random Sampling Methods

Judgment Sample	Use expert knowledge to choose "typical" items (e.g., which employees to interview).
Convenience Sample	Use a sample that happens to be available (e.g., ask coworkers' opinions at lunch).
Focus Groups	In-depth dialog with a representative panel of individuals (e.g., iPod users).

LO2-7: Explain the common sampling methods and how to implement them (continued, 10).

Judgment Sample

- A non-probability sampling method that relies on the expertise of the sampler to choose items that are representative of the population.
- Can be affected by subconscious bias (i.e., *non-randomness* in the choice).
- *Quota sampling* is a special kind of judgment sampling, in which the interviewer chooses a certain number of people in each category.

LO2-7: Explain the common sampling methods and how to implement them (continued, 11).

Convenience Sample

- Take advantage of whatever sample is available at that moment. A quick way to sample.

Focus Groups

- A panel of individuals chosen to be representative of a wider population, formed for open-ended discussion and idea gathering.

LO2-7: Explain the common sampling methods and how to implement them (continued, 12).

Other Data Collection Methods

- Businesses now have many other methods for collecting data with the use of technology.
- Point-of-sale (POS) systems can collect real-time data on purchases at retail or convenience stores, restaurants, and gas stations.
- Many companies use loyalty cards that are swiped during the purchase. These loyalty cards have the customer's information, which can be matched to the purchase just made.

LO2-7: Explain the common sampling methods and how to implement them (continued, 13).

Other Data Collection Methods

- Businesses also send out e-mail surveys to loyal customers on a regular basis to get feedback on their products and services.
- Facebook can track your Internet searches using its software algorithms.
- Google also tracks Internet searches and provides these data through its Google Analytics services.

LO2-7: Explain the common sampling methods and how to implement them (continued, 14).

Sample Size

- The necessary sample size depends on the inherent variability of the quantity being measured and the desired precision of the estimate.
- For example, the caffeine content of Mountain Dew is fairly consistent because each can or bottle is filled at the factory, so a small sample size would suffice to estimate the mean.
- In contrast, the amount of caffeine in an individually brewed cup of Bigelow Raspberry Royale tea varies widely because people let it steep for varying lengths of time, so a larger sample would be needed to estimate the mean.
- The purposes of the investigation, the costs of sampling, the budget, and time constraints also are taken into account in deciding on sample size.

LO2-7: Explain the common sampling methods and how to implement them (continued, 15).

Sources of Error or Bias

- In sampling, the word *bias* refers to a systematic tendency to over- or underestimate a population parameter of interest.
- However, the words *bias* and *error* are often used interchangeably.
- The word *error* generally refers to problems in sample methodology that lead to inaccurate estimates of a population parameter.
- No matter how careful you are when conducting a survey, you will encounter potential sources of error.
- The next slide shows some sources of errors.

LO2-7: Explain the common sampling methods and how to implement them (continued, 16).

Sources of Error or Bias

Source of Error	***Characteristics***
Nonresponse bias	Respondents differ from non-respondents
Selection bias	Self-selected respondents are atypical
Response error	Respondents give false information
Coverage error	Incorrect specification of frame or population
Measurement error	Unclear survey instrument wording
Interviewer error	Responses influenced by interviewer
Sampling error	Random and unavoidable

LO2-7: Explain the common sampling methods and how to implement them (continued, 17).

Sources of Error or Bias

- *Nonresponse bias* occurs when those who respond have characteristics different from those who don't respond.
- For example, people with caller ID, answering machines, blocked or unlisted numbers, or cell phones are likely to be missed in telephone surveys. Because these are generally more affluent individuals, their socioeconomic class may be underrepresented in the poll.

LO2-7: Explain the common sampling methods and how to implement them (continued, 18).

Sources of Error or Bias

- A special case is *selection bias*, a self-selected sample.
- For example, a talk show host who invites viewers to take a web survey about their sex lives will attract plenty of respondents. But those who are willing to reveal details of their personal lives (and who have time to complete the survey) are likely to differ substantially from those who dislike such surveys or are too busy (and probably weren't watching the show anyway).

LO2-7: Explain the common sampling methods and how to implement them (continued, 19).

Sources of Error or Bias

- *Response error* occurs when respondents deliberately give false information to mimic socially acceptable answers, to avoid embarrassment, or to protect personal information.
- *Coverage error* occurs when some important segment of the target population is systematically missed.
- *Measurement error* results when the survey questions do not accurately reveal the construct being assessed.

LO2-7: Explain the common sampling methods and how to implement them (continued, 20).

Sources of Error or Bias

- Interviewer error occurs when the interviewer's facial expressions, tone of voice, or appearance influences the responses.
- *Sampling error* is uncontrollable random error that is inherent in any random sample.
- Even when using a random sampling method, it is possible that the sample will contain unusual responses. This cannot be prevented and is generally undetectable. It is *not* an error on your part.

2.5 Data Sources

LO2-8: Find everyday print or electronic data sources.

One goal of a statistics course is to help you learn where to find data that might be needed. Fortunately, many excellent sources are widely available. Some sources are given in the following table.

Type of Data	***Examples***
U.S. job-related data	U.S. Bureau of Labor Statistics
U.S. economic data	*Economic Report of the President*
Almanacs	*World Almanac, Time Almanac*
Periodicals	*Economist, Bloomberg Businessweek, Fortune, Forbes*
Indexes	*The New York Times, The Wall Street Journal*
Databases	Compustat, Citibase, U.S. Census
World data	*CIA World Factbook*
Web	Google, Yahoo!, MSN

LO2-8: Find everyday print or electronic data sources (continued).

Web Data Sources

A few are given below:

- Bureau of Economic Analysis
- Bureau of Justice Statistics
- Bureau of Labor Statistics
- Central Intelligence Agency
- Economic Report of the President
- Environmental Protection Agency
- Federal Reserve System
- Food and Drug Administration

2.6 Surveys

LO2-9: Describe basic elements of survey types, survey designs, and response scales.

Basic Steps of Survey Research

Step 1: State the goals of the research.

Step 2: Develop the budget (time, money, staff).

Step 3: Create a research design (target population, frame, sample size).

Step 4: Choose a survey type and method of administration.

LO2-9: Describe basic elements of survey types, survey designs, and response scales (continued).

Basic Steps of Survey Research, continued

Step 5: Design a data collection instrument (questionnaire).

Step 6: Pretest the survey instrument and revise as needed.

Step 7: Administer the survey (follow up if needed).

Step 8: Code the data and analyze it.

LO2-9: Describe basic elements of survey types, survey designs, and response scales (continued, 2).

Surveys fall into five general categories: mail, telephone, interview, web, and direct observation. They differ in cost, response rate, data quality, time required, and survey staff training requirements.

Survey Type	***Characteristics***
Mail	Mail requires a well-targeted and current mailing list (people move a lot). Expect low response rates and nonresponse bias (non-respondents differ from those who respond). Zip code lists (often costly) are an attractive option to define strata of similar income, education, and attitudes. To encourage participation, a cover letter should explain the uses of the survey data. Plan for follow-up mailings.

LO2-9: Describe basic elements of survey types, survey designs, and response scales (continued, 3).

Surveys fall into five general categories: mail, telephone, interview, web, and direct observation. They differ in cost, response rate, data quality, time required, and survey staff training requirements.

Survey Type	***Characteristics***
Telephone	Random dialing yields low response and is poorly targeted. Purchased phone lists help reach the target population, though a low response rate still is typical (disconnected phones, caller screening, answering machines, work hours, no-call lists). Other sources of nonresponse bias include the growing number of cell phones, non-English speakers, and distrust caused by scams.

LO2-9: Describe basic elements of survey types, survey designs, and response scales (continued, 4).

Surveys fall into five general categories: mail, telephone, interview, web, and direct observation. They differ in cost, response rate, data quality, time required, and survey staff training requirements.

Survey Type	***Characteristics***
Interviews	Interviewing is expensive and time-consuming, yet a trade-off between sample size for high-quality results may be worth it. Interviewers must be well-trained—an added cost. Interviewers can obtain information on complex or sensitive topics (e.g., gender discrimination in companies, birth control practices, diet and exercise).

LO2-9: Describe basic elements of survey types, survey designs, and response scales (continued, 5).

Surveys fall into five general categories: mail, telephone, interview, web, and direct observation. They differ in cost, response rate, data quality, time required, and survey staff training requirements.

Survey Type	***Characteristics***
Web	Web surveys are growing in popularity but are subject to nonresponse bias because they miss those who feel too busy, don't own computers, or distrust your motives (scams and spam). This type of survey works best when targeted to a well-defined interest group on a question of self-interest (e.g., views of CPAs on Sarbanes-Oxley accounting rules, frequent flyer views on airline security).

LO2-9: Describe basic elements of survey types, survey designs, and response scales (continued, 6).

Surveys fall into five general categories: mail, telephone, interview, web, and direct observation. They differ in cost, response rate, data quality, time required, and survey staff training requirements.

Survey Type	***Characteristics***
Direct Observation	Observation can be done in a controlled setting (e.g., psychology lab) but requires informed consent, which can change behavior. Unobtrusive observation is possible in some non-lab settings (e.g., what percentage of airline passengers carry on more than two bags, what percentage of SUVs carry no passengers, what percentage of drivers wear seat belts).

LO2-9: Describe basic elements of survey types, survey designs, and response scales (continued, 7).

Questionnaire Design

- Use a lot of white space in layout.
- Begin with short, clear instructions.
- State the survey purpose
- Assure anonymity.
- Instruct on how to submit the completed survey.
- Break survey into naturally occurring sections.
- Let respondents bypass sections that are not applicable (e.g., “if you answered no to question 7, skip directly to Question 15”).

LO2-9: Describe basic elements of survey types, survey designs, and response scales (continued, 8).

Questionnaire Design

- Pretest and revise as needed.
- Keep as short as possible.

Types of Questions

- Open-ended
- Fill-in-the-blank
- Check boxes
- Ranked choices
- Pictograms
- Likert scale

LO2-9: Describe basic elements of survey types, survey designs, and response scales (continued, 9).

Question Wording

The way a question is asked has a profound influence on the response. For example:

1. Shall state taxes be cut?
2. Shall state taxes be cut, if it means reducing highway maintenance?
3. Shall state taxes be cut, if it means firing teachers and police?

LO2-9: Describe basic elements of survey types, survey designs, and response scales (continued, 10).

Question Wording

Make sure you have covered all the possibilities.

- For example: Are you married? ❑ Yes ❑ No

Overlapping classes or unclear categories are a problem. What if your father is deceased or is 45 years old?

How old is your father?

❑ 35 – 45
❑ 45 – 55
❑ 55 – 65
❑ 65 or older

LO2-9: Describe basic elements of survey types, survey designs, and response scales (continued, 11).

Data Quality

- Responses are usually coded numerically (e.g., 1 = male, 2 = female).
- Missing values are typically denoted by special characters (e.g., blank, “.” or “*”).
- Discard questionnaires that are flawed or missing many responses.
- Watch for multiple responses, outrageous or inconsistent replies, or out-of-range answers.
- Follow up if necessary and always document your data-coding decisions.

LO2-9: Describe basic elements of survey types, survey designs, and response scales (continued, 12).

Survey Software

- Designing and creating a survey is much easier than it used to be.
- Software is available that automates much of the process, allowing you to use different question formats, skip questions and move to a new section, easily visualize the layout, and other features.
- Because most surveys are now administered online, survey software also includes features that allow the respondent to remain anonymous if warranted and prevent respondents from taking the survey twice.
- One of the most commonly used free applications is SurveyMonkey.

LO2-9: Describe basic elements of survey types, survey designs, and response scales (continued, 13).

Survey Software

- Qualtrics offers a 14-day free trial but is used in larger enterprises such as universities.
- Many other software applications exist that offer similar features to SurveyMonkey or Qualtrics.
- It is important to remember that survey design, creation, and administration require thoughtful preparation and planning in order to capture the right information and to ensure a high response rate.