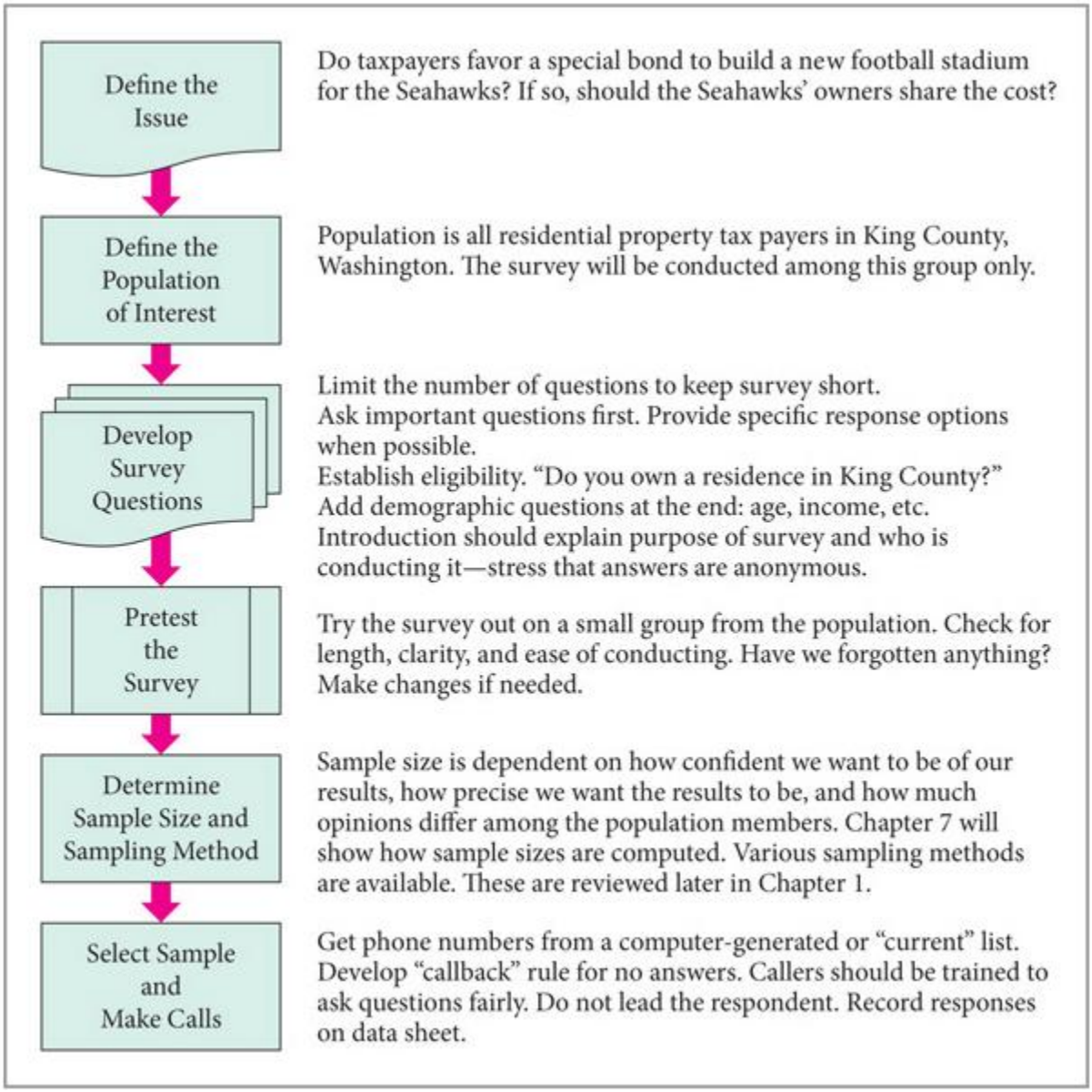
Chapter 1 – Business Statistics

Section 1.2

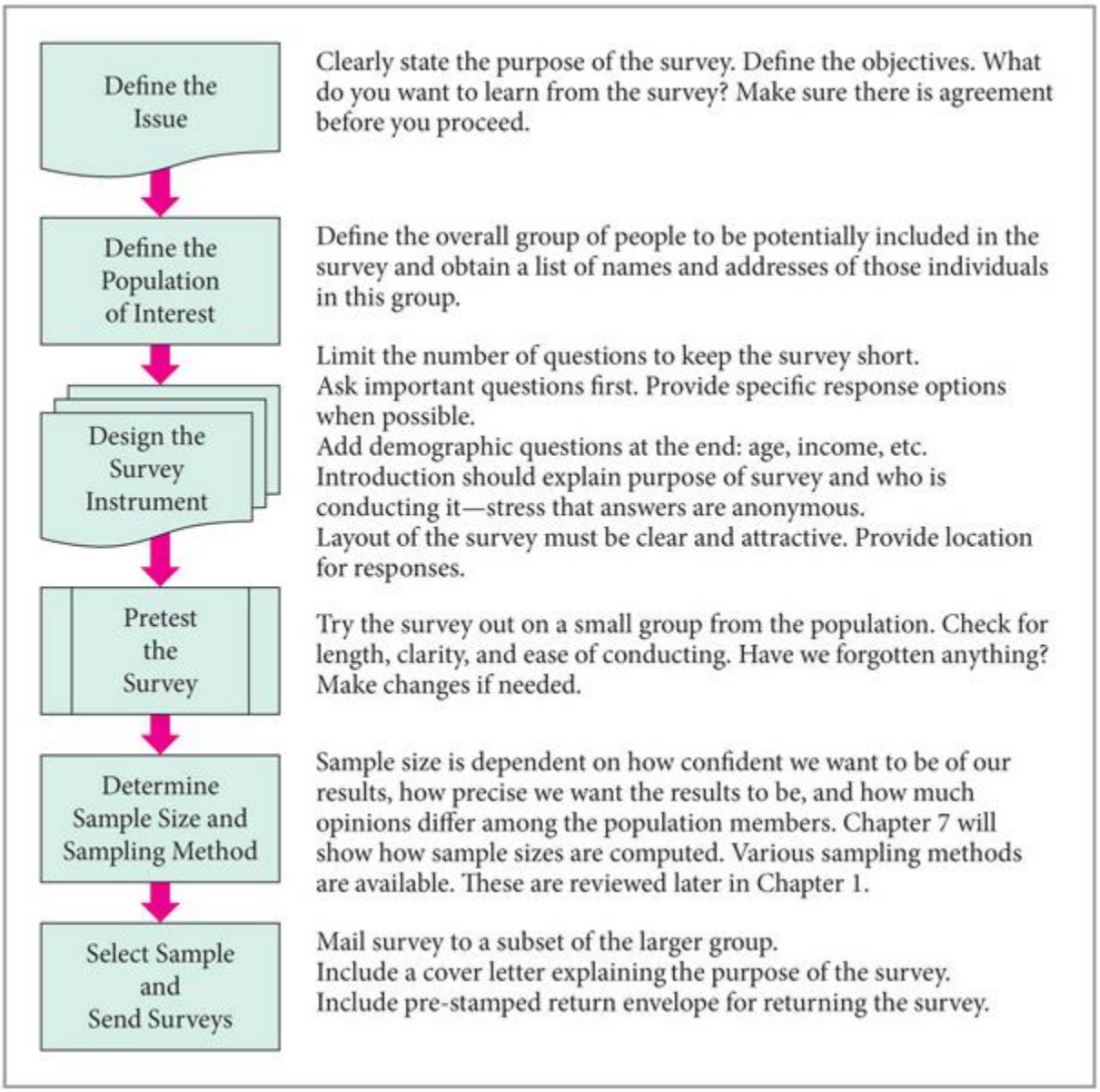
Collection methods

- Experiments
    - Defined the variables of interest. Sets some of the variables and observes another variable. Different levels of smog 02 and observation of the growth of peas.
- Telephone Survey
    - Research company asking you to tell them how you feel about a topic or topics



- Written questionnaire and surveys.

- By Mail (Snail mail)
- By Email
- Links
- Text messages
- Twitter messages
- Can you think of any other methods?



- Direct observation and personal interviews
  - standing in the mall and observing the number of people with their cellphone out.
  - observing the number of car passing through an intersection for 3 pm – 7 pm.

- Can you think of other ways you can observe anything?

Are there other ways to gather data?

- Video from Face Time
- Web crawling
- When pumping Gas
- Scan your finger for identity
- Clicking on a Website to browse or make purchases.

Question types

- Closed ended question - Questions that require the respondent to select from a short list of defined choices.
- Demographic Questions: Questions relating to the respondents' characteristics, backgrounds, and attributes.
- Opened ended - Questions that allow respondents the freedom to respond with any value, words, or statements of their own choosing.
- Structured Interview - Interviews in which the questions are scripted.
- Unstructured Interview - Interviews that begin with one or more broadly stated questions, with further questions being based on the response

Data Collection Issues

- Observer Bias - Data collection through personal observation is also subject to problems.
- Bias - An effect that alters a statistical result by systematically distorting it; different from a random error, which may distort on any one occasion but balances out on the average.
- Interview Bias - in a personal interview, the interviewer can interject bias (either accidentally or on purpose) by the way she asks the questions, by the tone of her voice, or by the way she looks at the subject being interviewed.
- Nonresponse Bias - We stated earlier that mail surveys suffer from a high percentage of unreturned surveys.
- Selection Bias - Bias can be interjected through the way subjects are selected for data collection.
- Measurement Error – Errors in measurements that are recorded.
- Internal Validity - A characteristic of an experiment in which data are collected in such a way as to eliminate the effects of variables within the experimental environment that are not of interest to the researcher. For data to have internal validity, the extraneous factors must be controlled.
- External Validity - A characteristic of an experiment whose results can be generalized beyond the test environment so that the outcomes can be replicated when the experiment is repeated.
- Algorithm Bias

Chapter 1 – Business Statistics

Questions

1-15

For each of the following situations, indicate what type of data collection method you would recommend and discuss why you have made that recommendation:

a. collecting data on the percentage of bike riders who wear helmets.
   Observation would be the most likely method. Observers could be located at various bike routes and observe the number of riders with and without helmets. This would likely be better than asking people if they wear a helmet since the popular response might be to say yes even when they don't always do so.

b. collecting data on the price of regular unleaded gasoline at gas stations in your state
   A telephone survey to gas stations in the state. This could be a cost-effective way of getting data from across the state. The respondent would have the information and be able to provide the correct price.

c. collecting data on customer satisfaction with the service provided by a major U.S. airline

   A written survey of passengers. This could be given out on the plane before the plane lands and passengers could drop the surveys in a box as they de-plane. This method would likely garner higher response rates compared to sending the survey to passengers' mailing address and asking them to return the completed survey by mail.

1-16.

Assume you have received a class assignment to determine the attitude of students in your school toward the school's registration process. What are the validity issues you should be concerned with?

The two types of validity mentioned in the section are internal validity and external validity. For this problem, external validity is easiest to address. It simply means the sampling method chosen will be sufficient to insure the results based on the sample will be able to be generalized to the population of all students. Internal validity would involve making sure the data gathering method, for instance a questionnaire, accurately determines the respondent's attitude toward the registration process.

Chapter 1 – Business Statistics

Section 1.3 Populations, Samples, and Sampling Techniques

Population – Set of all objects or individuals of interest or the measurements obtained from all objects or individuals of interest.

Sample – A subset of the population

Census – an enumeration of the entire set of measurements taken from the whole population.

Sampling Techniques

- Simple Random Sampling – list all population items and assign a number to each item. Then randomly select a number between the beginning and ending number.
- Stratified Random Sampling – The population is divided into (natural) groups called stratus. Then within each stratum a random sample was each population is selected.
- Systematic Random Sample – First determine the k the Population Sample divided by the number in the Sample. If k is not a whole number, round down to the lower integer. Then take a random number between 1 and k, such as r. Then add k to each item until you have the sample size.
  (i, r + k*(i – 1))
- Cluster Random Sample – Random select a group (natural occurring) then taking a measurement or interviewing all the members of the group.

Chapter 1 – Business Statistics

Questions

1-25. Indicate which sampling method would most likely be used in each of the following situations:

a. an interview conducted with mayors of a sample of cities in Florida

   Because the population is spread over a large geographical area, a cluster random sample could be selected to reduce travel costs.

b. a poll of voters regarding a referendum calling for a national value-added tax

   A stratified random sample would probably be used to keep sample size as small as possible.

c. a survey of customers entering a shopping mall in Minneapolis

   Most likely a convenience sample would be used since doing a statistical sample would be too difficult

1-26. A company has 18,000 employees. The file containing the names is ordered by employee number from 1 to 18,000. If a sample of 100 employees is to be selected from the 18,000 using systematic random sampling, within what range of employee numbers will the first employee selected come from?

To determine the range of employee numbers for the first employee selected in a systematic random sample use the following:

$$\text{Part range} = \frac{Population\ Size}{Sample\ Size} = \frac{18000}{100} = 180$$

Thus, the first person selected will come from employees 1-180. Once that person is randomly selected (50), the second person will be 180 higher (230th employee) than the first, and continue to add (180+230 = 310).

1-29. Describe how systematic random sampling could be used to select a random sample of 1,000 customers who have a certificate of deposit at a

commercial bank. Assume that the bank has 25,000 customers who own a certificate of deposit.

From a numbered list of all customers who own a certificate of deposit the bank would need to randomly determine a starting point between 1 and *k*, where *k* would be equal to 25000/1000 = 25. This could be done using a random number table or by having a statistical package or a spreadsheet generate a random number between 1 and 25. Once this value is determined the bank would select that numbered customer as the first sampled customer and then select every 25th customer after that until 100 customers are sampled.

1-36. Give the name of the kind of sampling that was most likely used in each of the following cases:

a. a *Wall Street Journal* poll of 2,000 people to determine the president's approval rating

   Stratified random sampling

b. a poll taken of each of the General Motors (GM) dealerships in Ohio in December to determine an estimate of the average number of Chevrolets not yet sold by GM dealerships in the United States

   Simple random sampling or possibly cluster random sampling

c. a quality-assurance procedure within a Frito-Lay manufacturing plant that tests every 1,000th bag of Fritos Corn Chips produced to make sure the bag is sealed properly

   Systematic random sampling

d. a sampling technique in which a random sample from each of the tax brackets is obtained by the Internal Revenue Service to audit tax returns

   Stratified random sampling

Chapter 1 – Business Statistics

Section 1.4 Data Types and Data Measurement Levels

Qualitative Data - Data whose measurement scale is inherently categorical (labels).

Nominal Data: Labels with no order

- Zip Codes: 93711, 93720, 93710
- Area Codes: 920, 661, 559, 510
- State Names: California, Arizona, Nevada, New York

Ordinal Data: Labels with order

- Address number of a residence: 3615 W. Barstow Ave, 3725 W. Barstow, 3569 Stuart Ave.
- Military Ranks:
  - Seaman Recruit
  - Seaman Apprentice
  - Seaman
  - Petty Officer Third Class
  - Petty Officer Second Class
  - Petty Officer First Class
  - Chief Petty Officer
  - Senior Chief Petty Officer
  - Master Chief Petty Officer
  - Command Master Chief Petty
- Answer to survey questions
  - Strongly Disagree - 0
  - Disagree - 1
  - Neutral - 2
  - Agree - 3
  - Strongly Agree – 4
  - Not Applicable – 5
- Class of school
  - Freshman
  - Sophomore
  - Junior

- Senior

Quantitative: Data whose measurement scale is inherently numeric.

- **Interval Data**: If the distance between two data items can be measured on some scale and the data have ordinal properties (>, <, or =) the data are said to be *interval data*. Thus, interval data allow us to precisely measure the difference between any two values. With ordinal data this is not possible, because all we can say is that one value is larger than another.
  - **Examples**
    - Temperature
      - 0 degrees does not mean absence of temperature
    - Weight Watcher points
      - 0 points does not mean no food
    - IQ Score
- **Ratio Data:** Data that have all the characteristics of interval data but also have a true zero point (at which zero means "none") are called *ratio data*. Ratio measurement is the highest level of measurement.
    - Bank Account
    - Number of Cars that passes in front of an intersection during 3:00 PM – 4:30 PM

Time-Series Data - A set of consecutive data values observed at successive points in time.

- Monthly Unemployment Rate
- Consumer Price Index

Cross-Sectional Data - A set of data values observed at a fixed point in time.

a. Graduation Rate at a community college at the end of Spring 2017.

Chapter 1 – Business Statistics

1-42. For each of the following, indicate whether the data are cross-sectional or time-series:

a. quarterly unemployment rates - **Time-series**

b. unemployment rates by state - **Cross-sectional**

c. monthly sales **– Time-series**

d. employment satisfaction data for a company – **Cross-sectional**

1-44**.** For each of the following variables, indicate the level of data measurement:

a. product rating {1=excellent, 2=good, 3=fair, 4=poor, 5=very poor}

Ordinal – categories with defined order

b. home ownership {own, rent, other
Nominal – categories with no defined order

c. college grade point average

Ratio

d. marital status {single, married, divorced, other}

Nominal – categories with no defined order

1-49**.** You have collected the following information on 15 different real estate investment trusts (REITs). Identify whether the data are cross-sectional or time-series.

a. income distribution by region in 2016

Cross-sectional

b. per share funds from operations (FFO) for the years 2012 to 2016
Time-series

c. number of residential apartment buildings owned on December 31, 2015
Cross-sectional

d. the overall percentage of leased space for the 137 properties in service as of December 31, 2015
Cross-sectional

e. annual dividends paid to shareholders in years 2010–2016

Time-Series