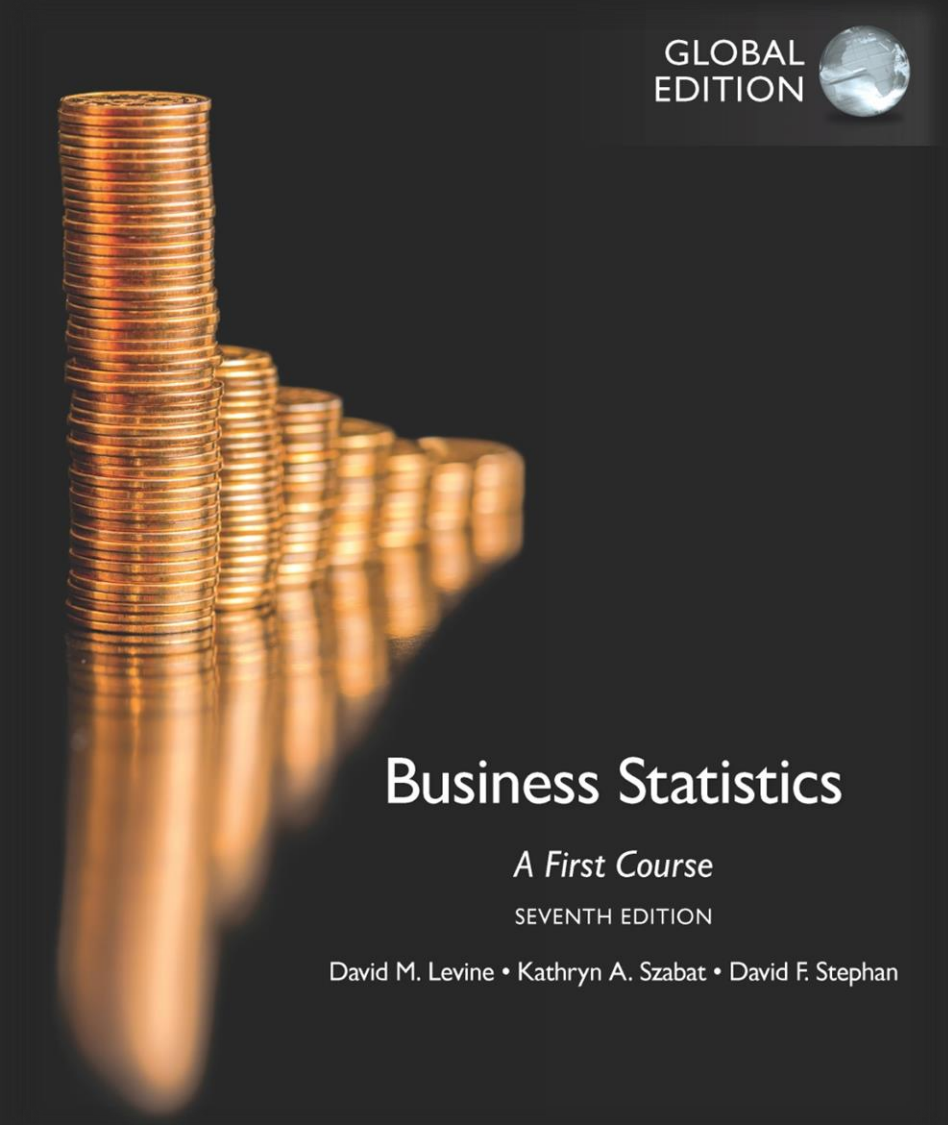


Chapter 1

Defining and Collecting Data

Objectives

In this chapter you learn:

- To understand issues that arise when defining variables.
- How to define variables
- How to collect data
- To identify different ways to collect a sample
- Understand the types of survey errors

Classifying Variables By Type

DCOVA

- **Categorical** (***qualitative***) variables take categories as their values such as “yes”, “no”, or “blue”, “brown”, “green”.

- **Numerical** (***quantitative***) variables have values that represent a counted or measured quantity.
  - **Discrete** variables arise from a ***counting process***
  - **Continuous** variables arise from a ***measuring process***

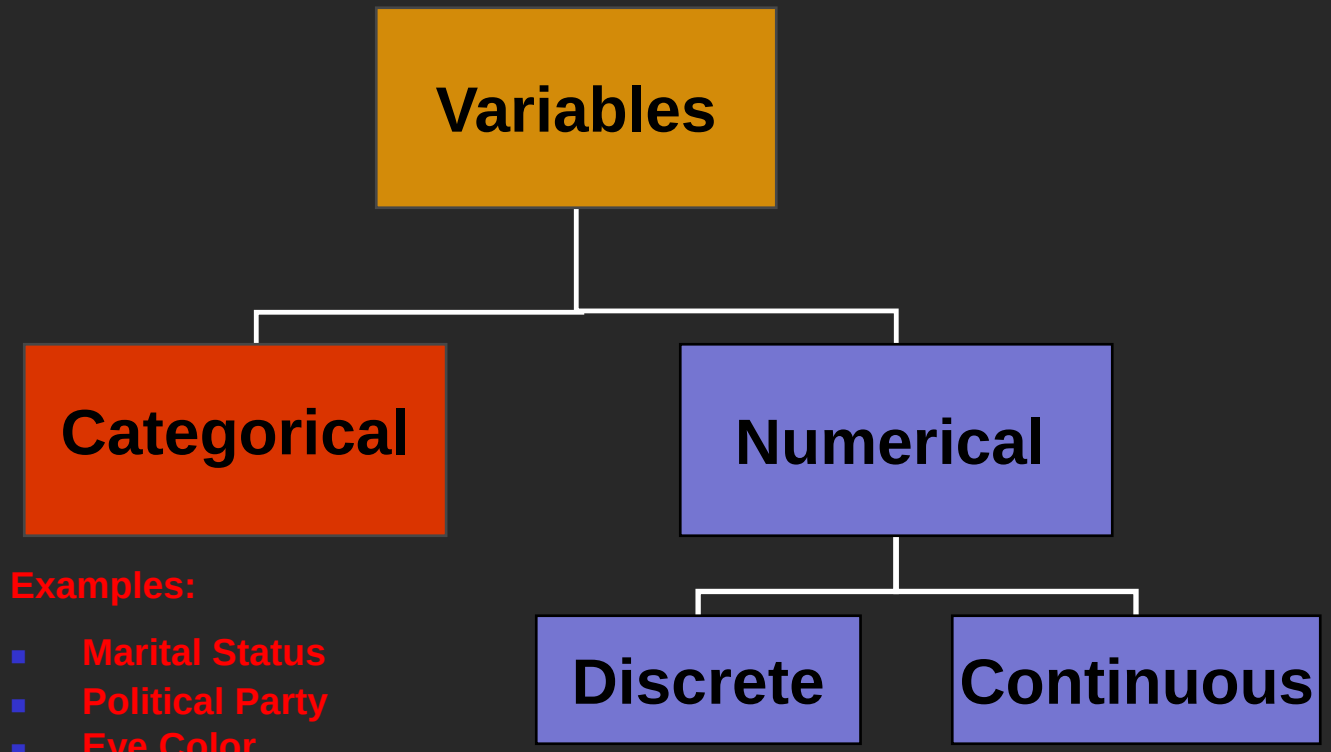
Examples of Types of Variables

DCOVA

Question	Responses	Variable Type
Do you have a Facebook profile?	Yes or No	Categorical (Qualitative)
How many text messages have you sent in the past three days?	----------	Numerical (discrete)
How long did the mobile app update take to download?	----------	Numerical (continuous)

Types of Variables

DCOVA



Examples:

- Marital Status
- Political Party
- Eye Color

(Defined categories)

Examples:

- Number of Children
- Defects per hour

(Counted items)

Examples:

- Weight
- Voltage

(Measured characteristics)

Collecting Data Correctly Is A Critical Task

DCOVA

- Need to avoid data flawed by biases, ambiguities, or other types of errors.
- Results from flawed data will be suspect or in error.
- Even the most sophisticated statistical methods are not very useful when the data is flawed.

Sources of Data

DCOVA

- **Primary Sources**: The data collector is the one using the data for analysis
  - Data from a political survey
  - Data collected from an experiment
  - Observed data
- **Secondary Sources**: The person performing data analysis is not the data collector
  - Analyzing census data
  - Examining data from print journals or data published on the internet.

Data Is Collected From Either A Population or A Sample

DCOVA

POPULATION

A **population** consists of all the items or individuals about which you want to draw a conclusion. The population is the "large group"

SAMPLE

A **sample** is the portion of a population selected for analysis. The sample is the "small group"

Population vs. Sample

DCOVA

Population

All the items or individuals about which you want to draw conclusion(s)

Sample

A portion of the population of items or individuals

Collecting Data Via Sampling Is Used When Selecting A Sample Is

DCOVA

- Less time consuming than selecting every item in the population.
- Less costly than selecting every item in the population.
- Less cumbersome and more practical than analyzing the entire population.

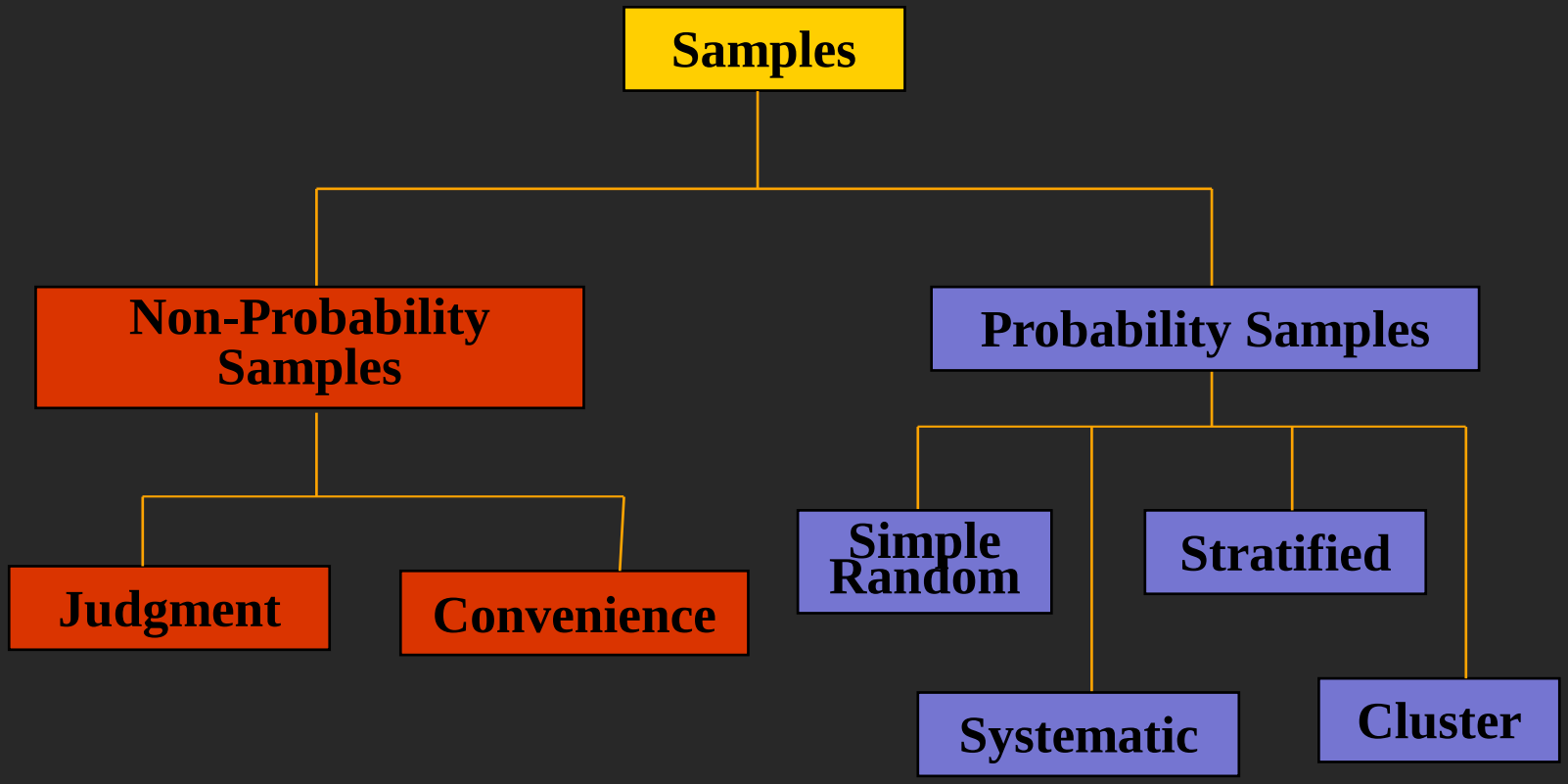
A Sampling Process Begins With A Sampling Frame

DCOVA

- The sampling frame is a listing of items that make up the population
- Frames are data sources such as population lists, directories, or maps
- Inaccurate or biased results can result if a frame excludes certain portions of the population
- Using different frames to generate data can lead to dissimilar conclusions

Types of Samples

DCOVA



Types of Samples: Nonprobability Sample

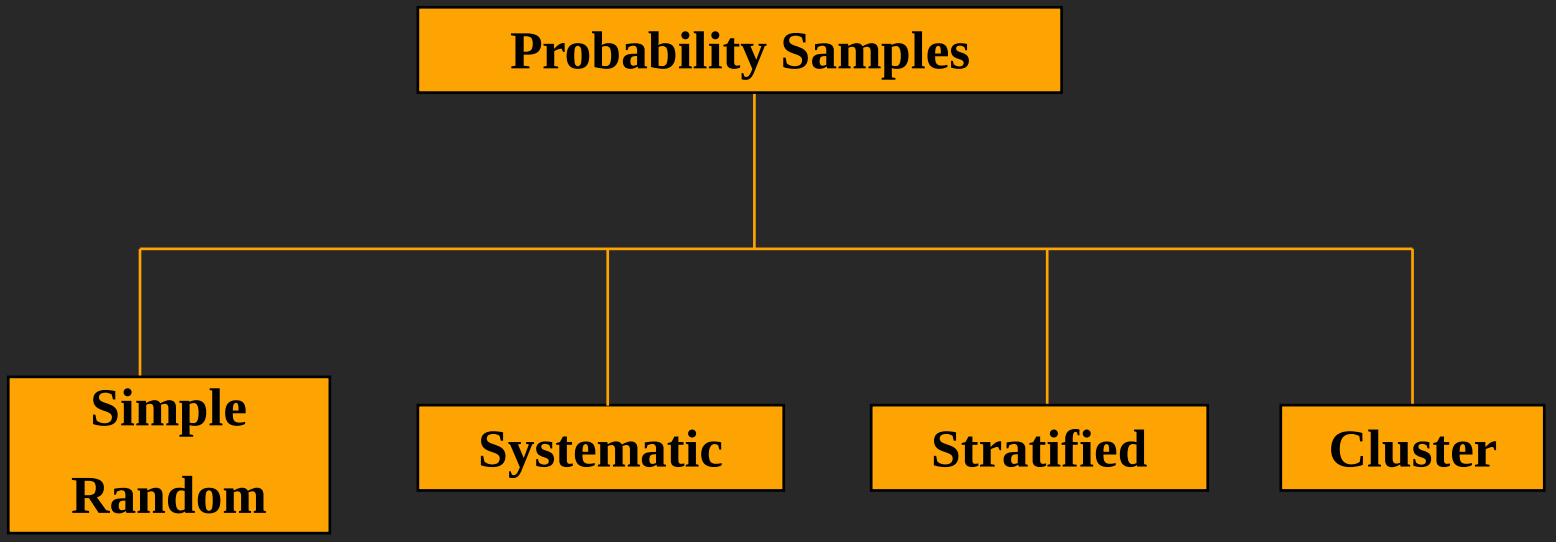
DCOVA

- In a nonprobability sample, items included are chosen without regard to their probability of occurrence.
  - In **convenience sampling**, items are selected based only on the fact that they are easy, inexpensive, or convenient to sample.
  - In a **judgment sample,** you get the opinions of pre-selected experts in the subject matter.

Types of Samples: Probability Sample

DCOVA

- In a **probability sample**, items in the sample are chosen on the basis of known probabilities.



Probability Sample: Simple Random Sample

DCOVA

- Every individual or item from the frame has an equal chance of being selected
- Selection may be with replacement (selected individual is returned to frame for possible reselection) or without replacement (selected individual isn't returned to the frame).
- Samples obtained from table of random numbers or computer random number generators.

Selecting a Simple Random Sample Using A Random Number Table

DCOVA

Sampling Frame For Population With 850 Items	
Item Name	**Item #**
Bev R.	001
Ulan X.	002
.	.
.	.
.	.
.	.
Joann P.	849
Paul F.	850

Portion Of A Random Number Table

49280 88924 35779 00283 81163 07275
11100 02340 12860 74697 96644 89439
09893 23997 20048 49420 88872 08401

The First 5 Items in a simple random sample

Item # 492
Item # 808
Item # 892 -- does not exist so ignore
Item # 435
Item # 779
Item # 002

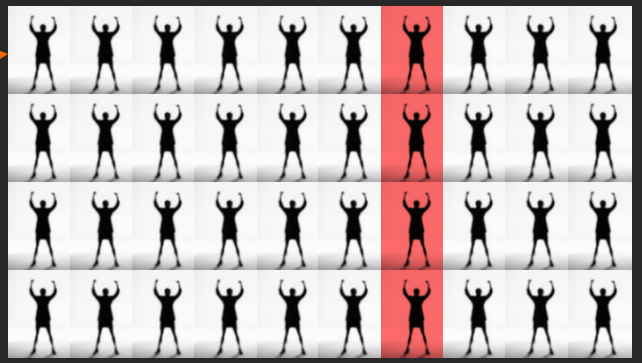
Probability Sample: Systematic Sample

DCOVA

- Decide on sample size: n
- Divide frame of N individuals into groups of k individuals: k=N/n
- Randomly select one individual from the 1st group
- Select every kth individual thereafter

N = 40
n = 4
k = 10

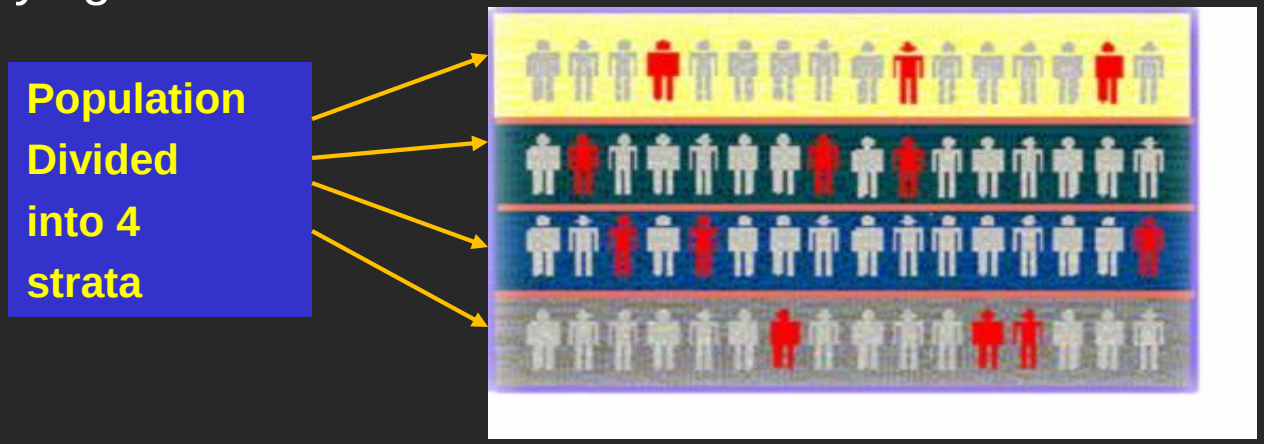
First Group



Probability Sample: Stratified Sample

DCOVA

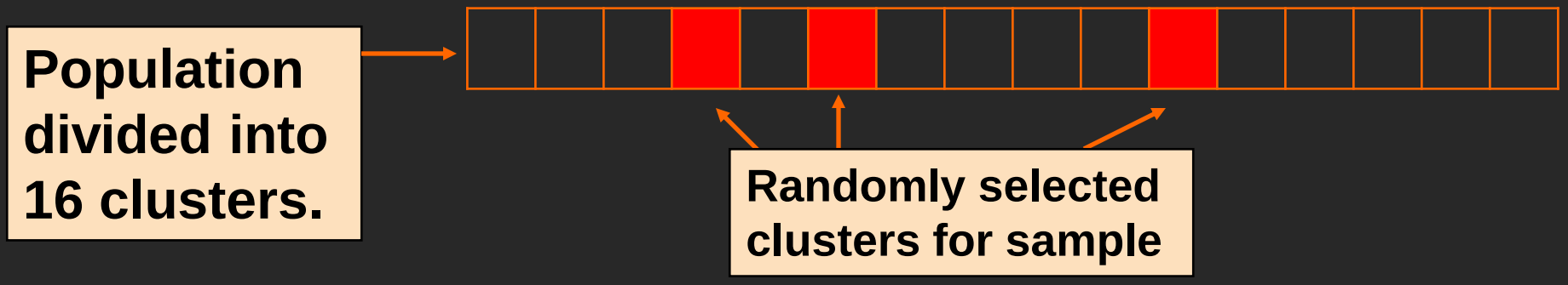
- Divide population into two or more subgroups (called *strata*) according to some common characteristic
- A simple random sample is selected from each subgroup, with sample sizes proportional to strata sizes
- Samples from subgroups are combined into one
- This is a common technique when sampling population of voters, stratifying across racial or socio-economic lines.



Probability Sample
Cluster Sample

DCOVA

- Population is divided into several “clusters,” each representative of the population
- A simple random sample of clusters is selected
- All items in the selected clusters can be used, or items can be chosen from a cluster using another probability sampling technique
- A common application of cluster sampling involves election exit polls, where certain election districts are selected and sampled.



Probability Sample: Comparing Sampling Methods

DCOVA

- Simple random sample and Systematic sample
  - Simple to use
  - May not be a good representation of the population's underlying characteristics
- Stratified sample
  - Ensures representation of individuals across the entire population
- Cluster sample
  - More cost effective
  - Less efficient (need larger sample to acquire the same level of precision)

Types of Survey Errors

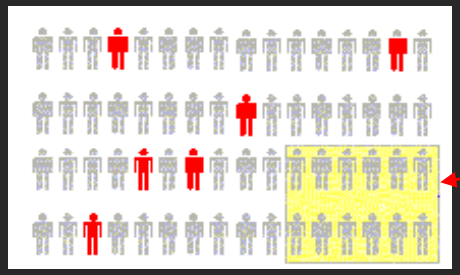
DCOVA

- Coverage error or selection bias
  - Exists if some groups are excluded from the frame and have no chance of being selected
- Nonresponse error or bias
  - People who do not respond may be different from those who do respond
- Sampling error
  - Variation from sample to sample will always exist
- Measurement error
  - Due to weaknesses in question design and / or respondent error

Types of Survey Errors

DCOVA
(continued)

- Coverage error



Excluded from frame

- Nonresponse error

Follow up on nonresponses

- Sampling error

Random differences from sample to sample

- Measurement error

Bad or leading question

Chapter Summary

In this chapter we have discussed:

- The types of variables used in statistics
- How to collect data
- The different ways to collect a sample
- The types of survey errors