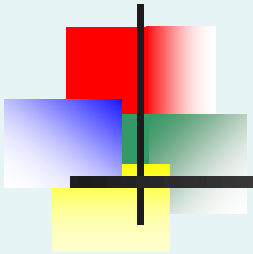
Statistics for Managers Using Microsoft Excel®

7th Edition



Chapter 1

Defining & Collecting Data

Learning Objectives

In this chapter you learn:

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

Types of Variables

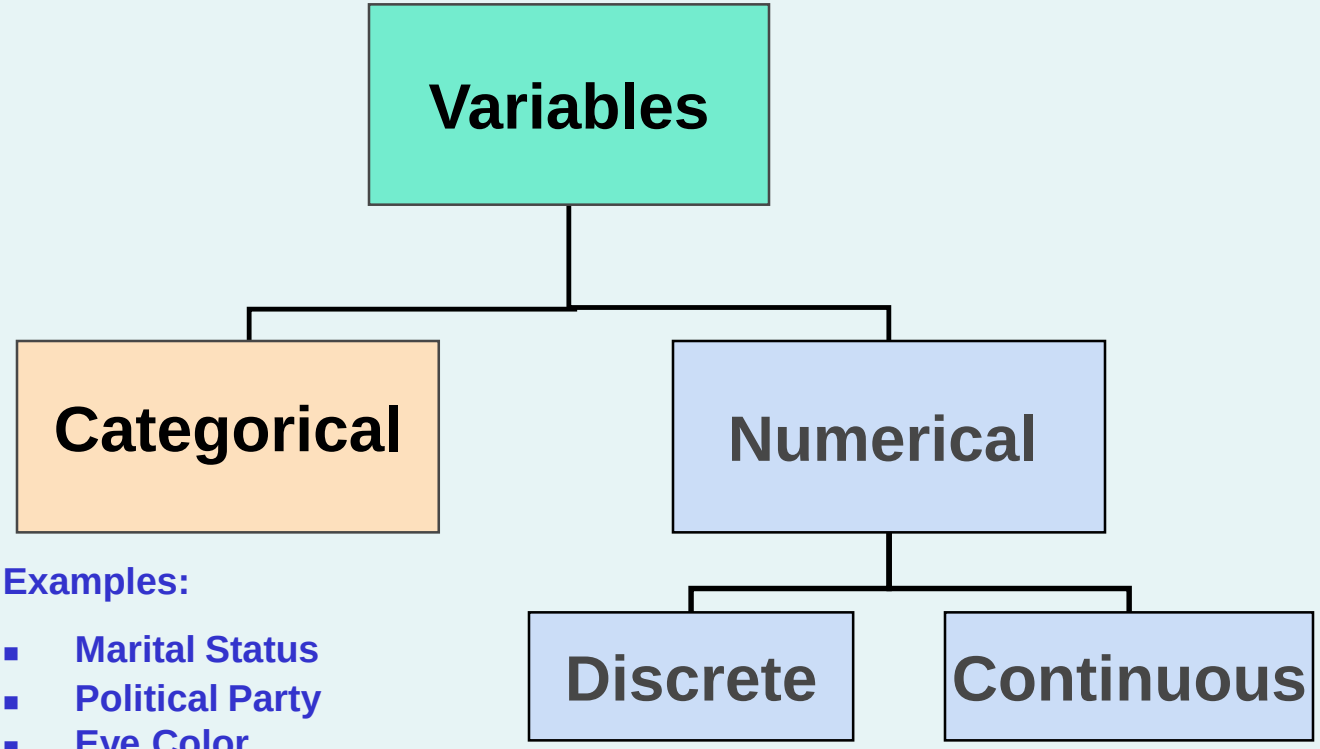
DCOVA

- **Categorical** (***qualitative***) variables have values that can only be placed into categories, such as “yes” and “no.”

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

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)

Levels of Measurement

DCOVA

A **nominal scale** classifies data into distinct categories in which no ranking is implied.

Categorical Variables	*Categories*
Personal Computer Ownership	Yes / No
Type of Stocks Owned	Growth / Value / Other
Internet Provider	AT&T, Verizon, Time Warner Cable

Levels of Measurement (con't.)

DCOVA

An **ordinal scale** classifies data into distinct categories in which ranking is implied

Categorical Variable	*Ordered Categories*
Student class designation	Freshman, Sophomore, Junior, Senior
Product satisfaction	Satisfied, Neutral, Unsatisfied
Faculty rank	Professor, Associate Professor, Assistant Professor, Instructor
Standard & Poor's bond ratings	AAA, AA, A, BBB, BB, B, CCC, CC, C, DDD, DD, D
Student Grades	A, B, C, D, F

Levels of Measurement (con't.)

DCOVA

- An **interval scale** is an ordered scale in which the difference between measurements is a meaningful quantity but the measurements do not have a true zero point.

- A **ratio scale** is an ordered scale in which the difference between the measurements is a meaningful quantity and the measurements have a true zero point.

Interval and Ratio Scales

DCOVA

Numerical Variable	*Level of Measurement*
Temperature (in degrees Celsius or Fahrenheit)	Interval
Standardized exam score (e.g., ACT or SAT)	Interval
Height (in inches or centimeters)	Ratio
Weight (in pounds or kilograms)	Ratio
Age (in years or days)	Ratio
Salary (in American dollars or Japanese yen)	Ratio

Establishing A Business Objective Focuses Data Collection

DCOVA

Examples Of Business Objectives:

- A marketing research analyst needs to assess the effectiveness of a new television advertisement.
- A pharmaceutical manufacturer needs to determine whether a new drug is more effective than those currently in use.
- An operations manager wants to monitor a manufacturing process to find out whether the quality of the product being manufactured is conforming to company standards.
- An auditor wants to review the financial transactions of a company in order to determine whether the company is in compliance with generally accepted accounting principles.

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.

Sources of data fall into five categories

DCOVA

- Data distributed by an organization or an individual
- A designed experiment
- A survey
- An observational study
- Data collected by ongoing business activities

Examples Of Data Distributed By Organizations or Individuals

DCOVA

- Financial data on a company provided by investment services.
- Industry or market data from market research firms and trade associations.
- Stock prices, weather conditions, and sports statistics in daily newspapers.

Examples of Data From A Designed Experiment

DCOVA

- Consumer testing of different versions of a product to help determine which product should be pursued further.
- Material testing to determine which supplier's material should be used in a product.
- Market testing on alternative product promotions to determine which promotion to use more broadly.

Examples of Survey Data

DCOVA

- Political polls of registered voters during political campaigns.
- People being surveyed to determine their satisfaction with a recent product or service experience.

Examples of Data Collected From Observational Studies

DCOVA

- Market researchers utilizing focus groups to elicit unstructured responses to open-ended questions.
- Measuring the time it takes for customers to be served in a fast food establishment.
- Measuring the volume of traffic through an intersection to determine if some form of advertising at the intersection is justified.

Examples of Data Collected From Ongoing Business Activities

DCOVA

- A bank studies years of financial transactions to help them identify patterns of fraud.
- Economists utilize data on searches done via Google to help forecast future economic conditions.
- Marketing companies use tracking data to evaluate the effectiveness of a web site.

Data Is Collected From Either A Population or A Sample

DCOVA

POPULATION

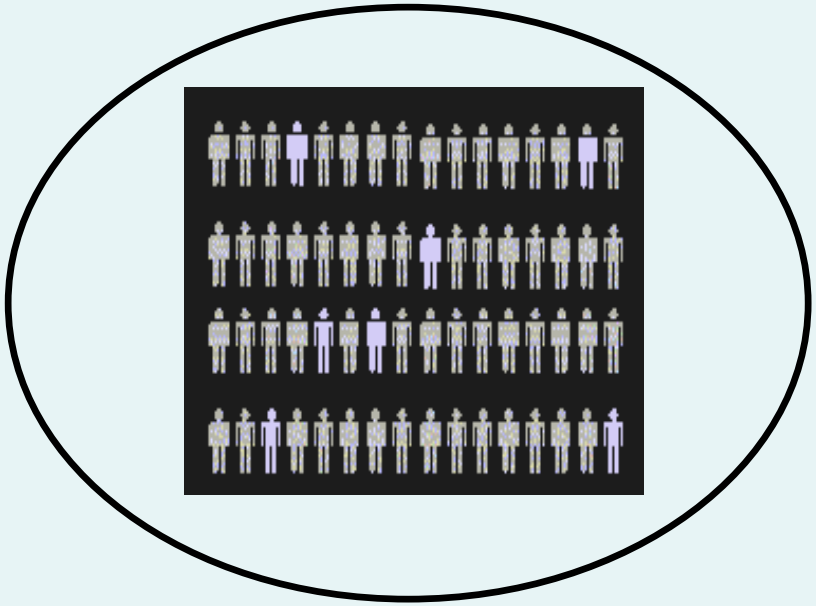
A **population** consists of all the items or individuals about which you want to draw a conclusion. The population is the "large group"

SAMPLE

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

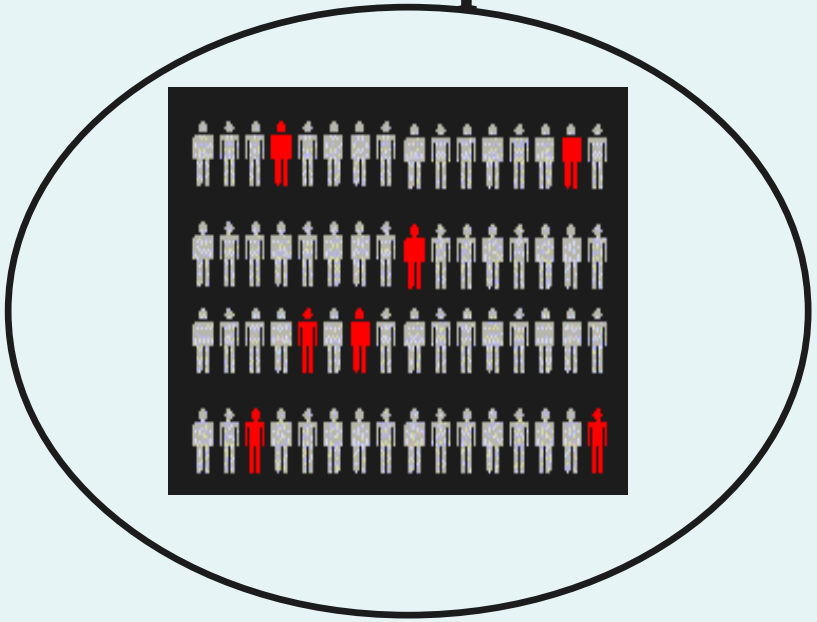
Population vs. Sample

Population



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

Sample



A portion of the population of items or individuals

Data Cleaning Is Often A Necessary Activity When Collecting Data

- Often find “irregularities” in the data
  - Typographical or data entry errors
  - Values that are impossible or undefined
  - Missing values
  - Outliers
- When found these irregularities should be reviewed
- Many statistical software packages will handle irregularities in an automated fashion (Excel does not)

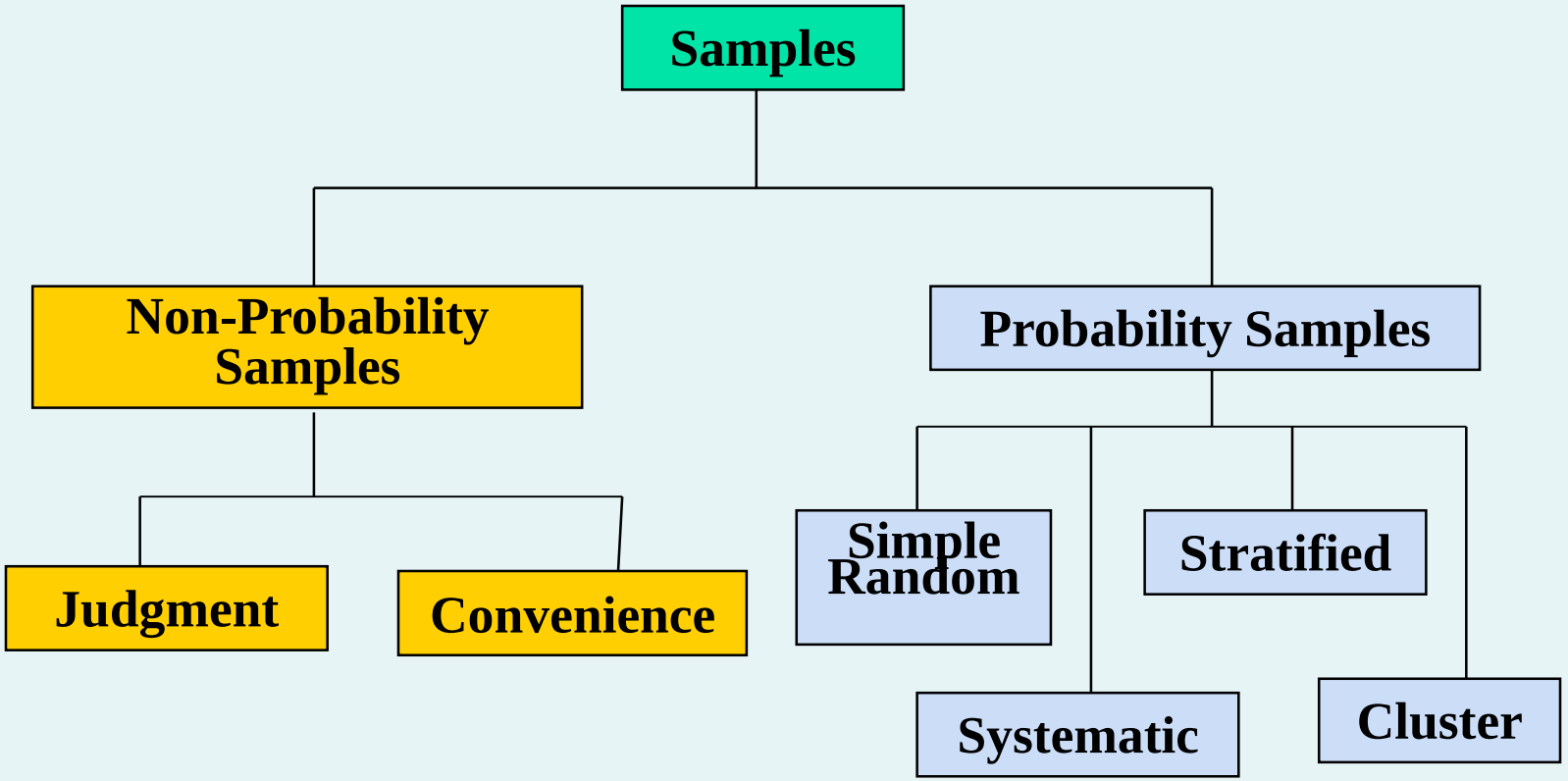
A Sampling Process Begins With A Sampling Frame

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

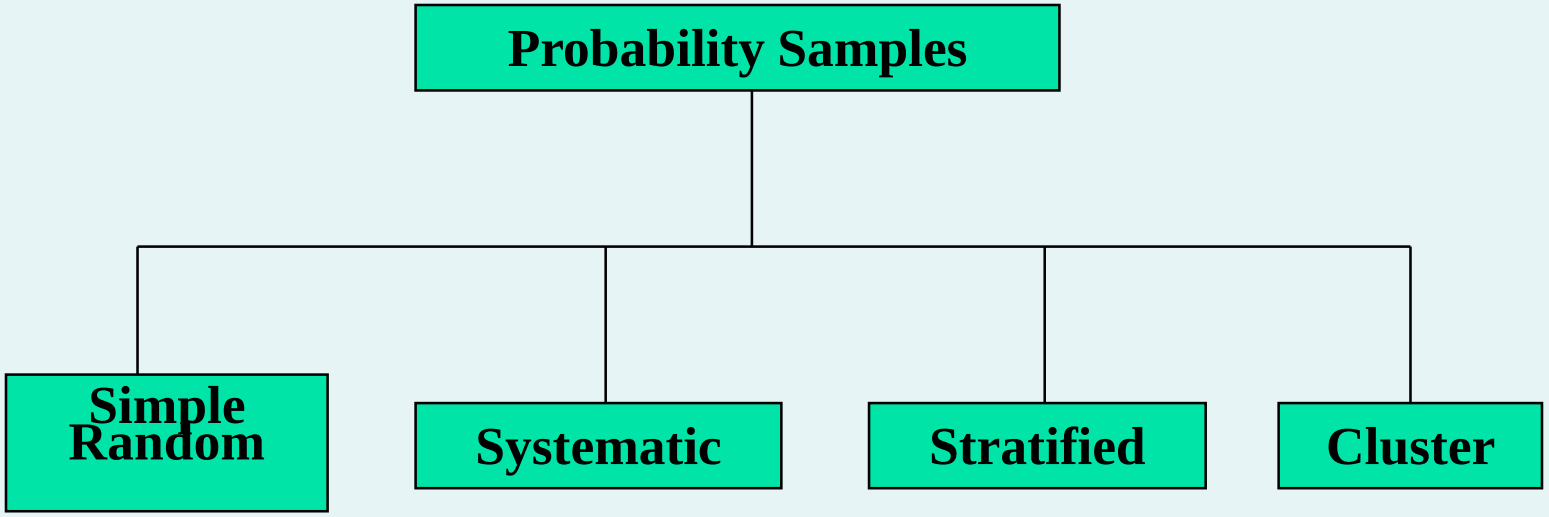
DCOVA

- 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

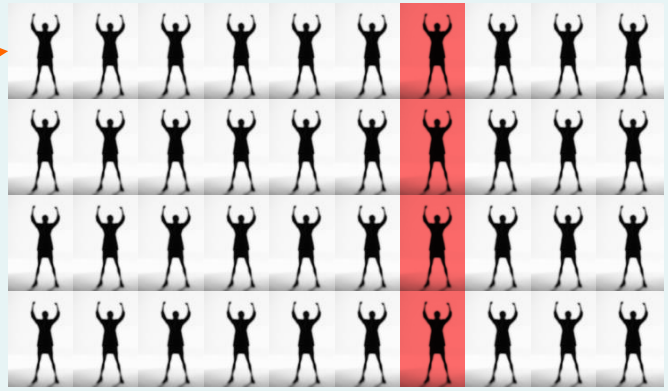
Probability Sample: Systematic Sample

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

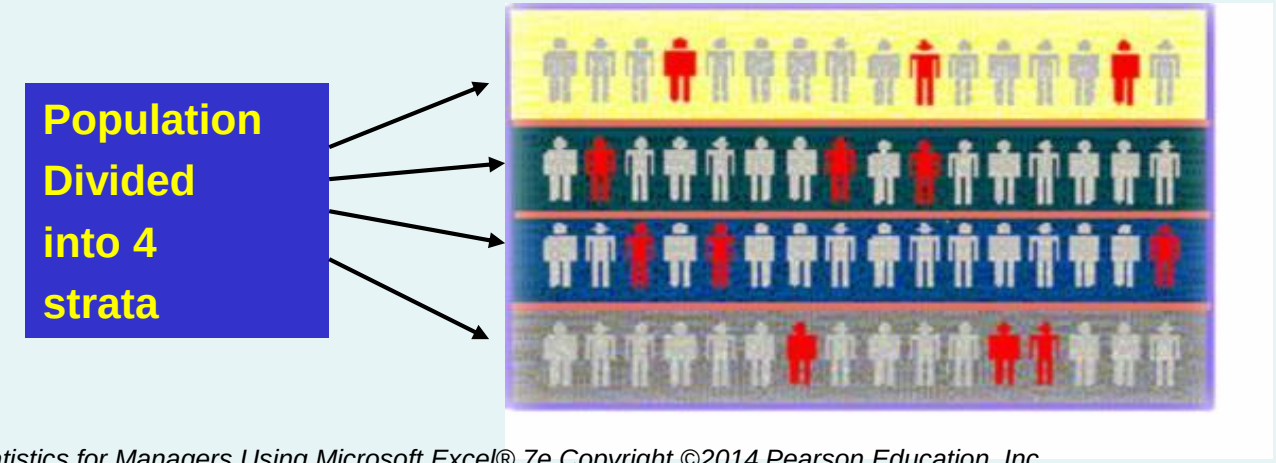
First Group



Probability Sample: Stratified Sample

DCOVA

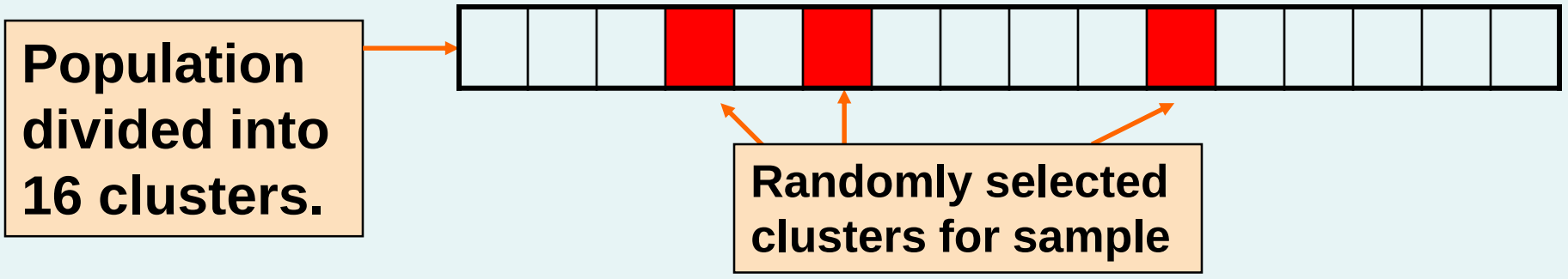
- Divide population into two or more subgroups (called *strata*) according to some common characteristic
- A simple random sample is selected from each subgroup, with sample sizes proportional to strata sizes
- Samples from subgroups are combined into one
- This is a common technique when sampling population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

Evaluating Survey Worthiness

DCOVA

- What is the purpose of the survey?
- Is the survey based on a probability sample?
- Coverage error – appropriate frame?
- Nonresponse error – follow up
- Measurement error – good questions elicit good responses
- Sampling error – always exists

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, respondent error, and interviewer's effects on the respondent ("Hawthorne effect")

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
- The measurement scales of variables
- How to collect data
- The different ways to collect a sample
- The types of survey errors