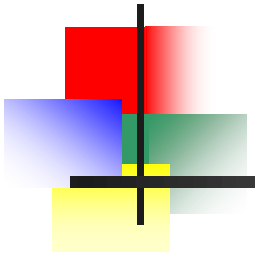
Statistics for Managers Using Microsoft® Excel 4th Edition



Chapter 1

Introduction and Data Collection

Chapter Goals

After completing this chapter, you should be able to:

- Explain key definitions:
  - Population vs. Sample
  - Parameter vs. Statistic
  - Primary vs. Secondary Data
  - Descriptive vs. Inferential Statistics
- Describe key data collection methods
- Describe different sampling methods
  - Probability Samples vs. Nonprobability Samples
- Select a random sample by computer generation
- Identify types of data and levels of measurement
- Describe the different types of survey error

Why a Manager Needs to Know about Statistics

To know how to:

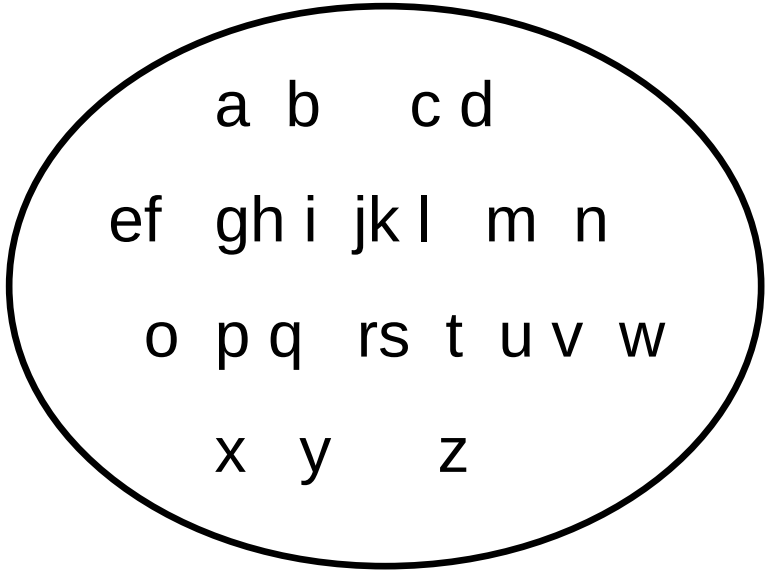
- properly present information (describe things)
- draw conclusions about populations based on sample information (make decisions)
- improve processes
- obtain reliable forecasts

Key Definitions

- A population is the collection of all items or things under consideration –people or objects
- A sample is a portion of the population selected for analysis
- A parameter is a summary measure that describes a characteristic of the population
- A statistic is a summary measure computed from a sample

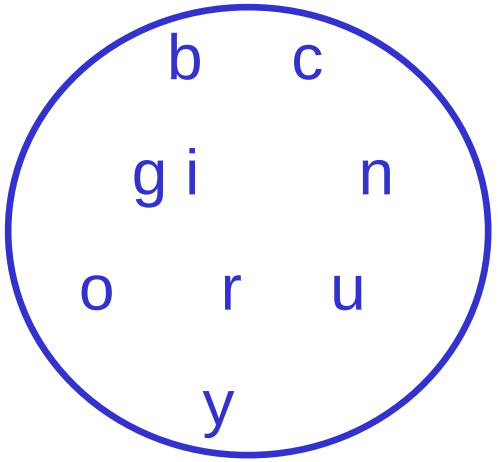
Population vs. Sample

Population



Measures used to describe the population are called parameters

Sample



Measures computed from sample data are called statistics

Key Definitions

- A survey is the gathering of data about a particular group of people or items
- A census is a survey of the entire population
- A sample is a survey of a portion of the population

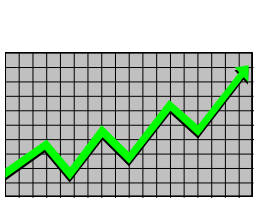
Two Branches of Statistics

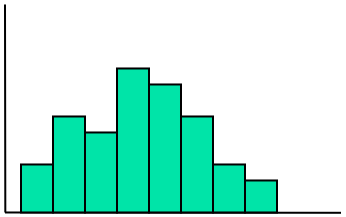
- Descriptive statistics
  - Collecting, summarizing, and describing data
- Inferential statistics
  - Drawing conclusions and/or making decisions concerning a population based only on sample data

Descriptive Statistics

- Collect data
  - e.g. Survey

- Present data
  - e.g. Tables and graphs





- Characterize data
  - e.g. Sample mean = $\frac{\sum X_i}{n}$

Inferential Statistics

- Estimation
  - e.g.: Estimate the population mean weight using the sample mean weight
- Hypothesis testing
  - e.g.: Test the claim that the population mean weight is over 120 pounds

Drawing conclusions and/or making decisions concerning a population based on sample results.

Why We Need Data

- To provide input to study a situation
- To measure performance of service or production processes
- To evaluate conformance to standards
- To assist in formulating alternative courses of action
- To satisfy curiosity

Data Sources

Primary

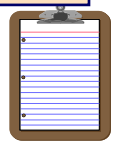
Data Collection

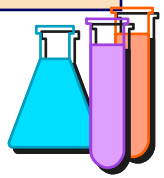
- **Observation**
- **Survey**
- **Experimentation**

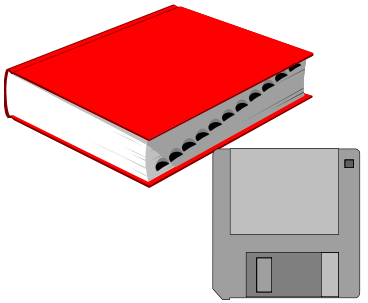
Secondary

Data Compilation

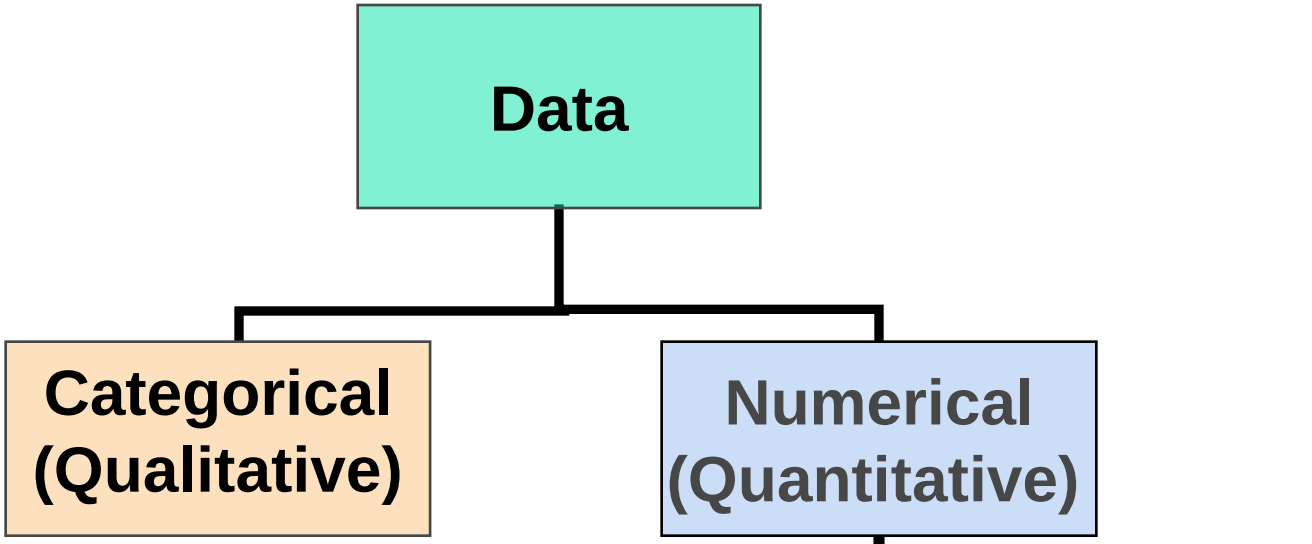
- **Print or Electronic**







Types of Data



Examples:

- Marital Status
- Political Party
- Eye Color

(Defined categories)

Examples:

- Number of Children
- Defects per hour

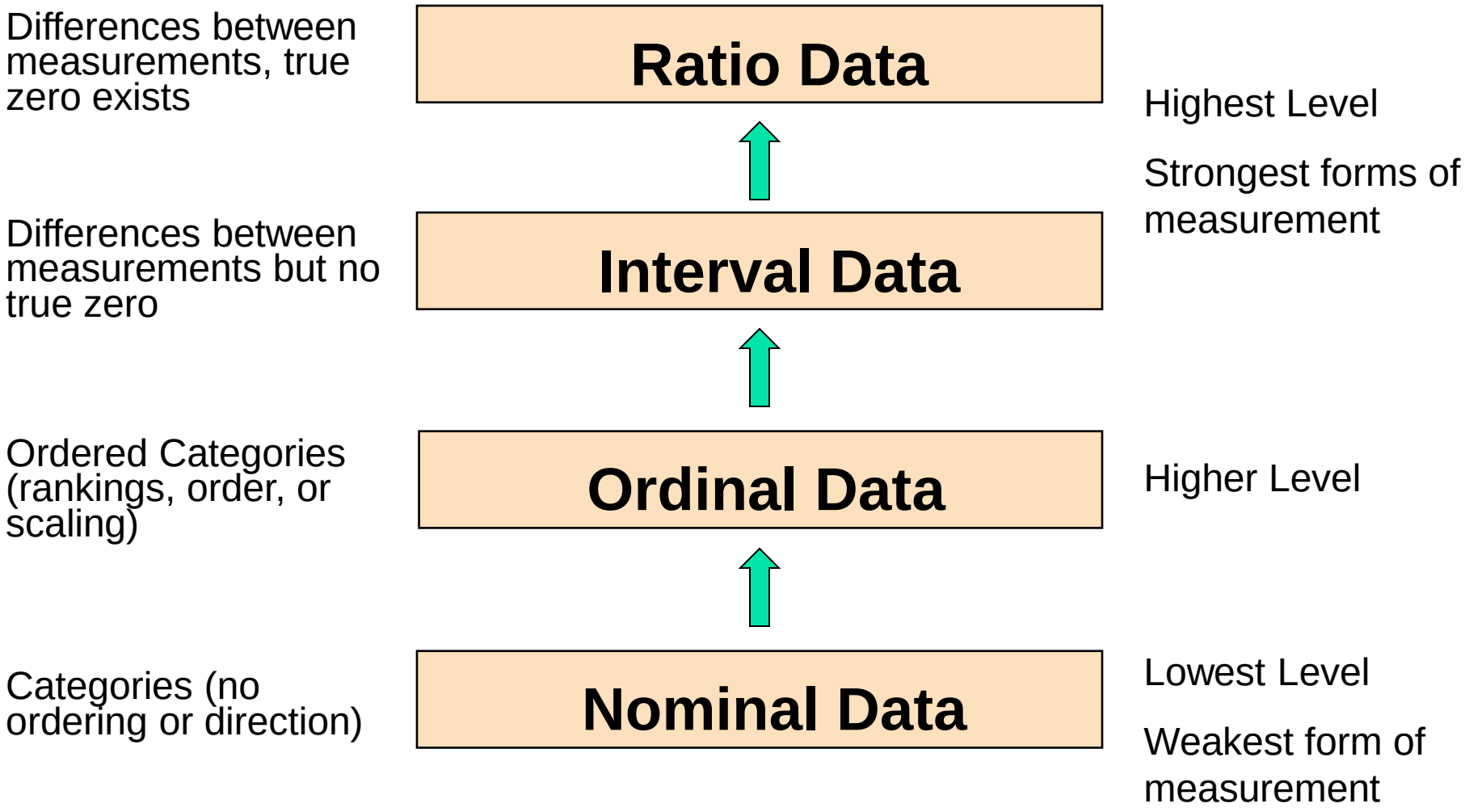
(Counted items)

Examples:

- Weight
- Voltage

(Measured characteristics)

Levels of Measurement and Measurement Scales



Example Data

Subject	Name	Height	Income	Gender	Eye color
1	Mary	62	10,350	Female	Blue
2	John	72	30,500	Male	Brown
3	Jill	64	35,600	Female	Green
4	Donna	59	20,700	Female	Brown
5	Sam	73	15,300	Male	Blue
6	Bill	70	52,800	Male	Black
7	Mario	71	19,400	Male	Blue
8	Carol	73	12,500	Female	Brown
9	Betty	70	30,200	Female	Brown
10	Linda	68	22,700	Female	Brown

Data in Frequency Distributions

Height

Category	Frequency
>54 to ≤ 60	1
>60 to ≤ 66	2
>66 to ≤ 72	5
>72 to ≤ 78	2

Income

Category	Frequency
≤ 20K	4
>20K to ≤ 50K	5
> 50K	1

Gender

Category	Frequency
Female	6
Male	4

Eye Color

Category	Frequency
Black	1
Blue	3
Brown	5
Green	1

Statistical Data

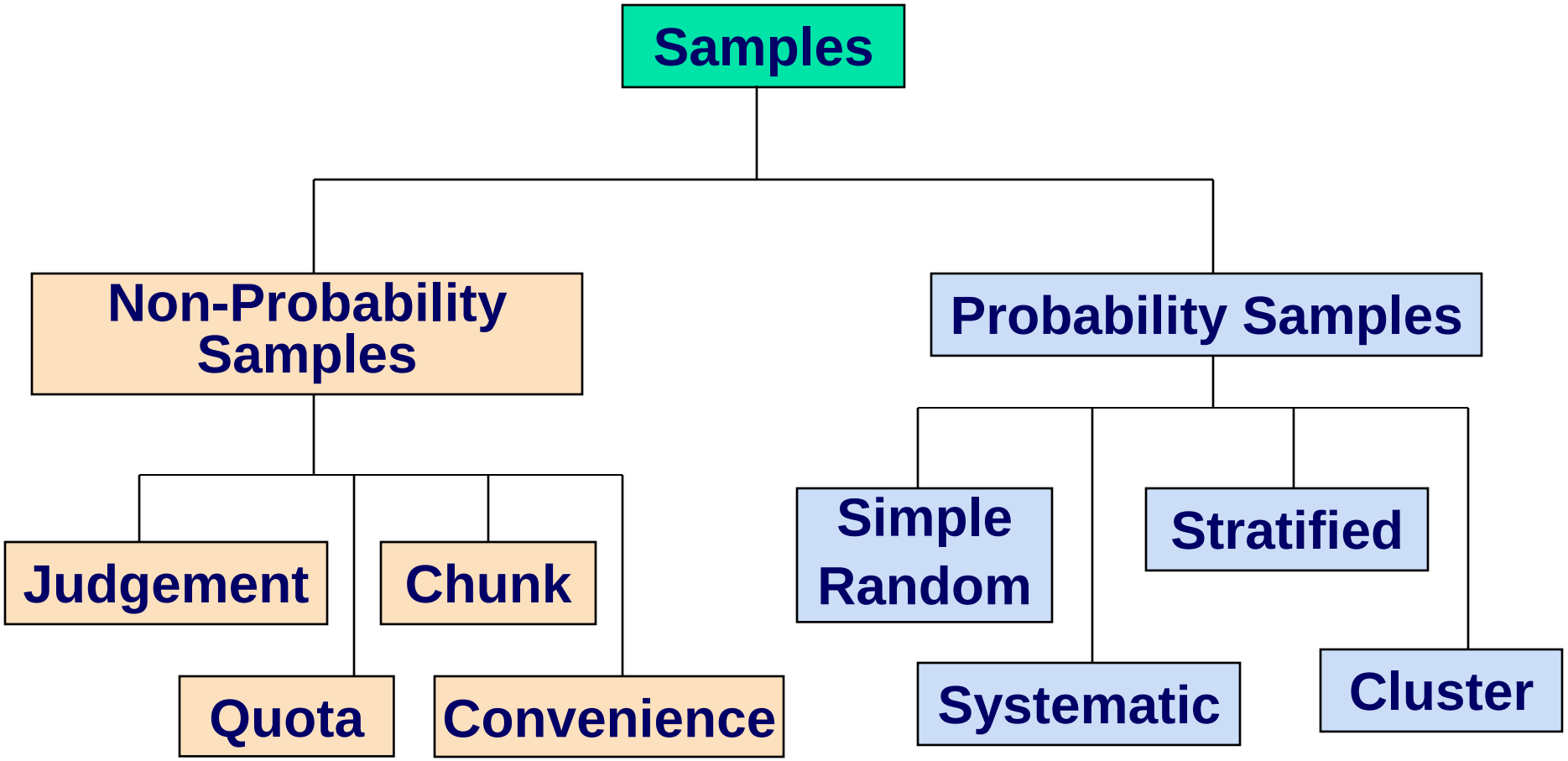
- Numerical Data can be gathered as grouped or converted after gathering.
- Categorical data is by nature always grouped
- Classes for numerical data are usually a range of values
- Classes for categorical data are usually single valued
- Numerical data is usually grouped for graphical presentation

Reasons for Drawing a Sample

- Less time consuming than a census
- Less costly to administer than a census

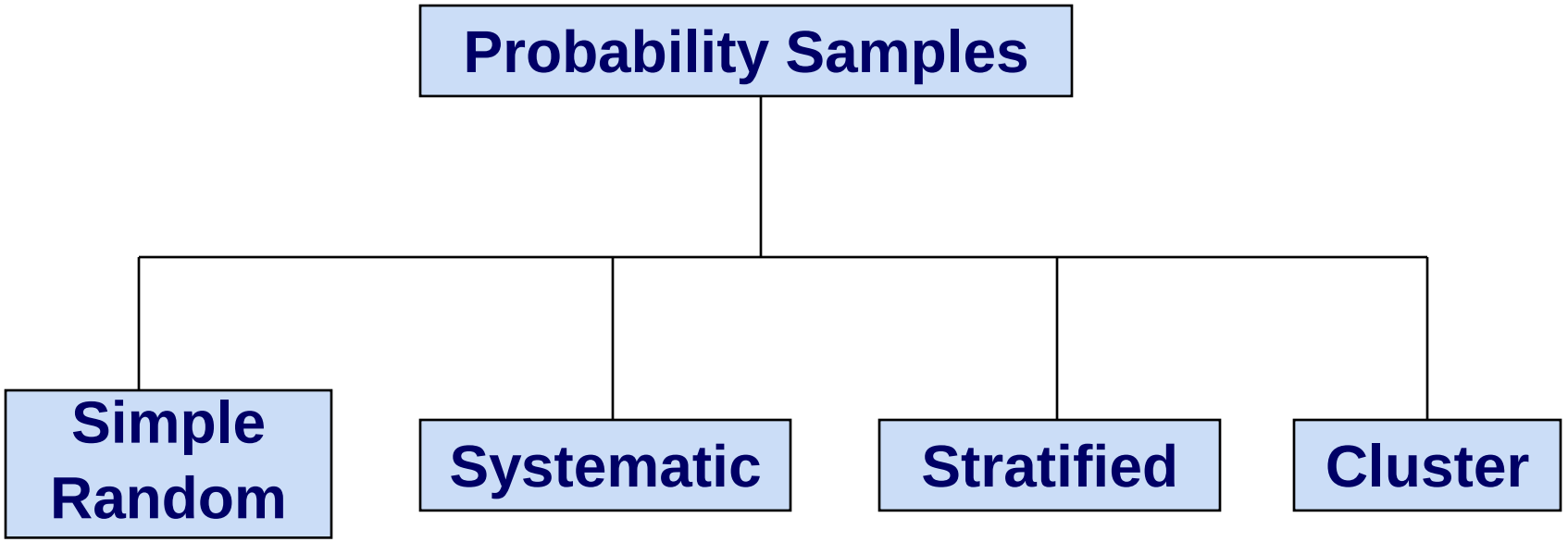
Types of Samples Used

(continued)



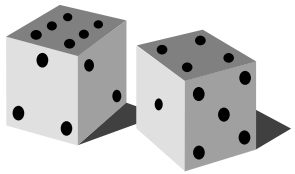
Probability Sampling

- Items in the sample are chosen based on known probabilities



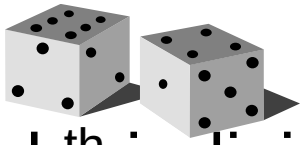
Simple Random Samples

- Every individual or item from the frame has an equal chance of being selected
- Selection may be with replacement or without replacement
- Samples obtained from computer random number generators



Systematic Samples

- 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

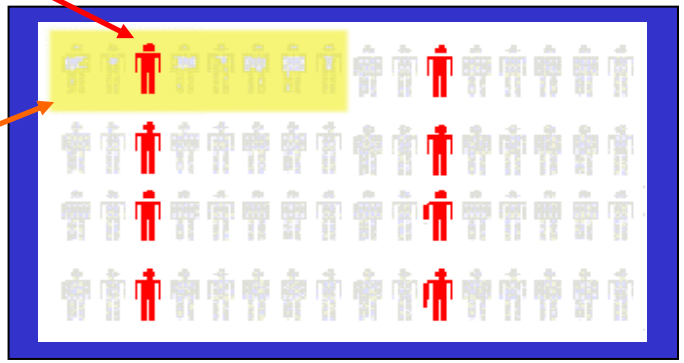


N = 64

n = 8

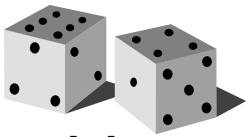
k = 8

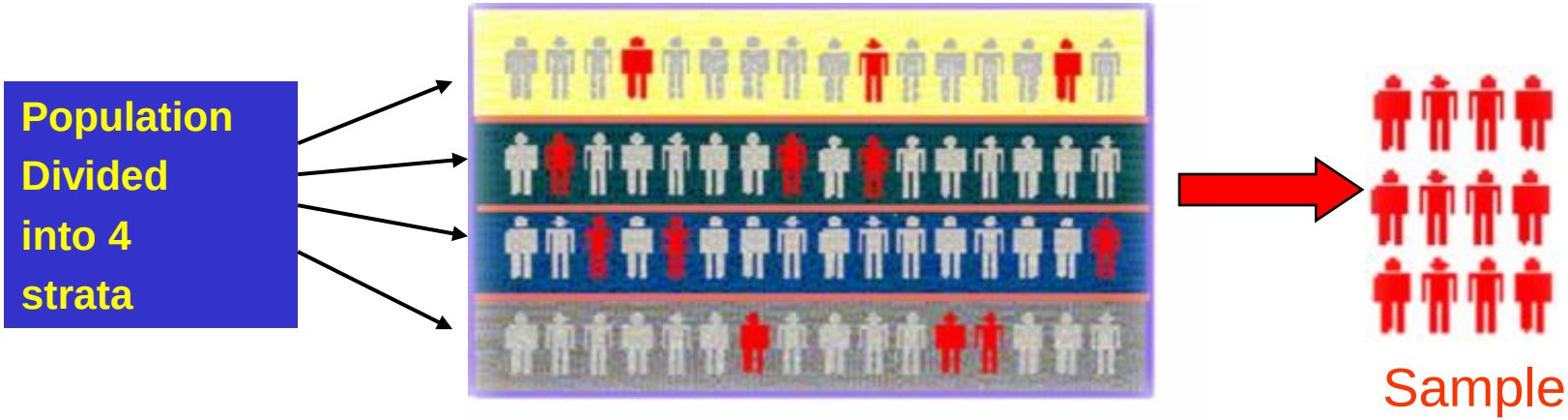
First Group



Stratified Samples

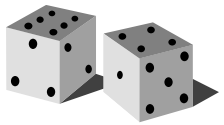
- Population divided into two or more subgroups (called *strata*) according to some common characteristic
- Simple random sample selected from each subgroup
- Samples from subgroups are combined into one

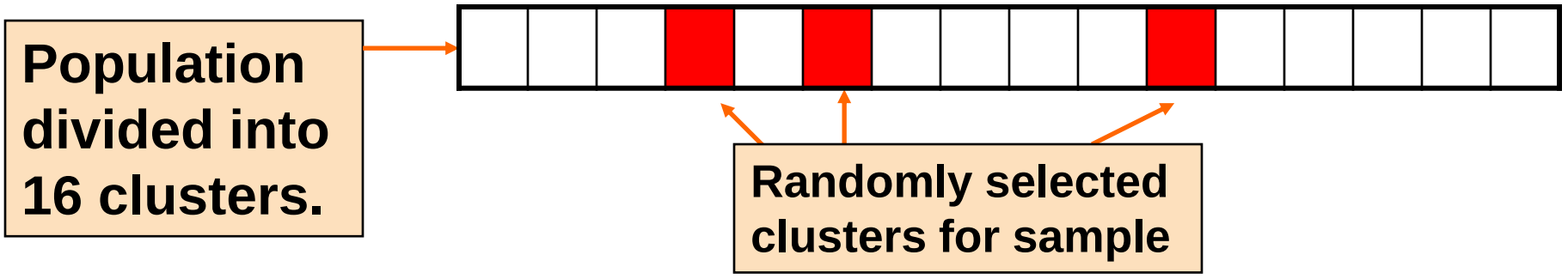




Cluster Samples

- Population is divided into “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





Advantages and Disadvantages

- Simple random sample and systematic sample
  - Simple to use
  - May not be a good representation of the population's underlying characteristics that have small probabilities
- 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

- Coverage error or selection bias
  - Exists if some groups are excluded from the frame and have no chance of being selected
- Non response 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

Evaluating Survey Worthiness

- What is the purpose of the survey?
- Is the survey based on a probability sample?
- Are there coverage errors – (appropriate frame)?
- Is there Non-response error – (follow up)
- Is there Measurement error – (good questions elicit good responses)
- Is the sampling error acceptable – (always exists)

Chapter Summary

- Reviewed why a manager needs to know statistics
- Introduced key definitions
- Examined descriptive vs. inferential statistics
- Described different types of samples
- Reviewed data types and measurement levels
- Examined survey worthiness and types of survey errors