

Sampling & Data Collection

Prepared by Dr. Hesham Mousli

MHA613
Statistics in Healthcare

Lecture 2



INTENDED OBJECTIVES

1. Grasp the different sampling techniques used in healthcare management.
2. Explore various data collection methods and their impact on decision making.
3. Learn the importance of validity and reliability in data measurement.
4. Apply sampling techniques and data quality concepts to real-world healthcare management scenarios.



Chapter

02

SAMPLING TYPES

➤ Random vs. Non-random Samples:

1. Non-random Samples:

- **Characteristics:**

- ✓ May not represent the entire population.
- ✓ Often biased and not suitable for statistical inference.

SAMPLING TYPES

➤ Random vs. Non-random Samples:

1. Non-random Samples:

- **Examples:**

- ✓ Patients in an ER on a specific afternoon.
- ✓ Hospitals in select cities (Buffalo, Syracuse, Albany) for a state-wide study.
- ✓ Patients of a specific physician on cholesterol drugs.

SAMPLING TYPES

➤ Random vs. Non-random Samples:

2. Random Samples:

- **Characteristics:**

- ✓ A sample where every member of the target population has a known and equal probability of being selected.

Chapter

02

SAMPLING TYPES

➤ Importance of Random Sampling:

- **Why Random Sampling Matters?**

- A. **Accuracy:** Provides a representative snapshot of the target population.
- B. **Bias Reduction:** Minimizes selection bias compared to non-random sampling methods.
- C. **Statistical Validity:** Ensures that findings can be generalized to the broader population.

A. SIMPLE RANDOM SAMPLING

- **Definition:** Simple random sampling primarily ensures that every possible sample of a given size has an equal chance of being selected.
- **Process:**
 - ✓ **Assign Numbers:** Number each element in the population.
 - ✓ **Random Selection:** Use random number tables or software (*e.g., Excel*) to select samples.

Chapter
02

A. SIMPLE RANDOM SAMPLING

- **Example:** For a population of 30,000 families and a desired sample of 100, every combination of 100 families has the same probability of being chosen.



B. SYSTEMATIC RANDOM SAMPLING

- **Definition:** Systematic random sampling involves dividing the population into equal-sized subsets and then selecting one observation from each subset.
- **Process:**
 - ✓ **Divide:** Create subsets of equal size.
 - ✓ **Random Selection:** Choose one family from the first subset.
 - ✓ **Systematic Selection:** Pick corresponding families from other subsets.

B. SYSTEMATIC RANDOM SAMPLING

- **Example:** For a population of 30,000 families and a desired sample of 100, ***divide the population into 100 subsets of 300 families each***. Select ***one family randomly*** from the ***first subset*** and then select ***corresponding families*** from the remaining subsets.



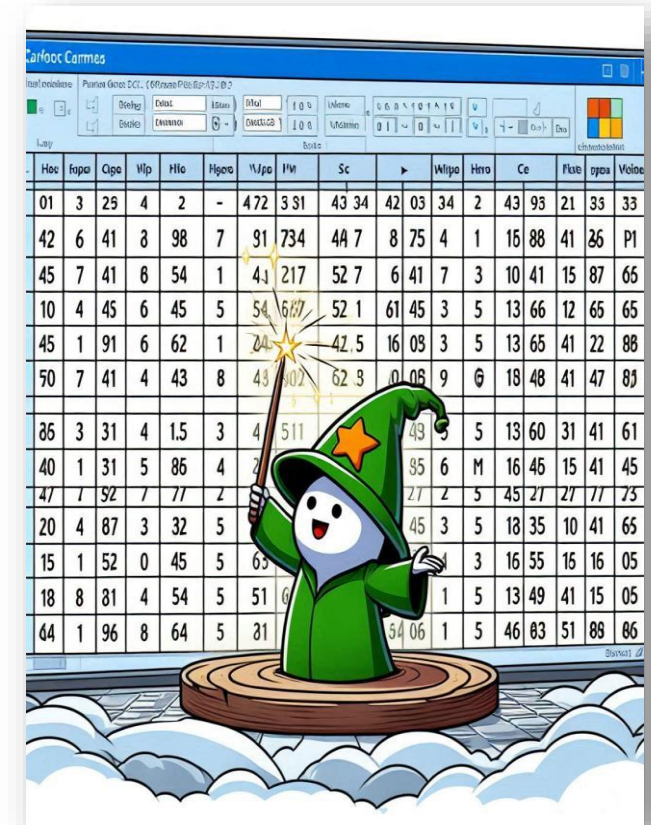
KEY POINTS IN SAMPLING

- **Key Point 1:** In Simple Random Sampling, every sample of the desired size has an equal chance of selection, ensuring unbiased representation.
- **Key Point 2:** Systematic Random Sampling results in fixed samples once the first selection is made. There are 300 possible samples based on the choice of the first family.

Chapter
02

KEY POINTS IN SAMPLING

- **Random number generators**, such as those in software like Excel, are used to produce random sequences that help ensure each item in a population has an equal chance of being selected for a sample.



SAMPLING TYPES

3. Stratified Sampling:

- **Definition:** Dividing the population into distinct groups (***strata***) and sampling a specified proportion from each group.
- **Sampling Process:**
 - ✓ **Strata:** Groups can be based on characteristics like age or income.
 - ✓ **Sampling Method:** Simple random or systematic sampling within each stratum.

SAMPLING TYPES

3. Stratified Sampling:

- **Example:** In a study on educational attainment, the population is divided into strata based on education levels (e.g., high school, bachelor's, master's). A random sample is drawn from each educational level.

SAMPLING TYPES

4. Cluster Sampling:

- **Definition:** Dividing the population into clusters and sampling in two or more stages.
- **Sampling Process:**
 - ✓ **Stage 1:** Randomly select clusters.
 - ✓ **Stage 2:** Sample all or a subset of members from the selected clusters.

SAMPLING TYPES

4. Cluster Sampling:

- **Example:** In a survey of 100 families from 30,000, ZIP code areas are selected first. Families within those ZIP codes are then randomly sampled.

Stratified Sampling Vs. Cluster Sampling

➤ **Stratified Sampling:**

- **Scope:** All strata represented. Best for detailed and precise estimates across all groups.
- **Precision:** Higher precision with proportional representation.
- **Example:** Sampling students from different academic programs to ensure all programs are represented.

Stratified Sampling Vs. Cluster Sampling

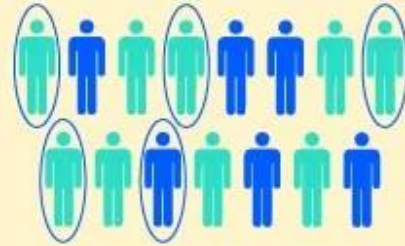
➤ **Cluster Sampling**

- **Scope:** Only selected clusters are represented. Best for logistical ease and cost efficiency when dealing with large populations.
- **Efficiency:** Cost-effective and less time-consuming but may increase variability.
- **Example:** Selecting schools (clusters) in various districts and surveying students within those schools.

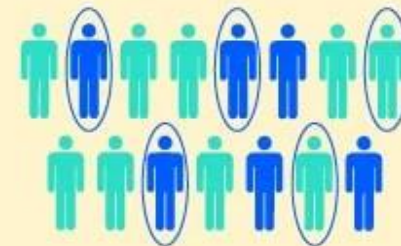
Chapter
02

SAMPLING TYPES SUMMARY

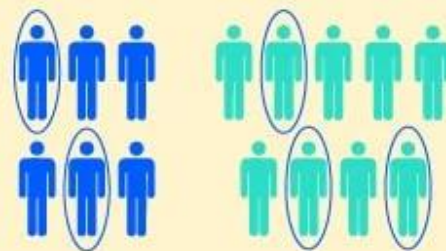
SIMPLE RANDOM SAMPLE



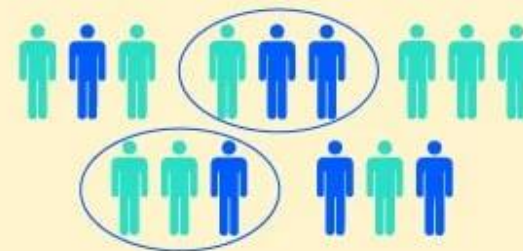
SYSTEMATIC SAMPLE



STRATIFIED SAMPLE



CLUSTER SAMPLE



Validity vs. Reliability

1. Definition of Validity:

- **Meaningfulness of Data:** Validity refers to how well the data truly represents the phenomenon under investigation. It ensures that the variable being measured accurately reflects the underlying issue.



Validity vs. Reliability

➤ Examples of Invalid Data Measurement:

➤ Teen Pregnancy Statistics:

- **Common Metric:** Reported number of babies born to unmarried mothers under 18 years old.

- **Validity Issues:**

Excludes: Terminated pregnancies, unwanted births to teen mothers who married to avoid stigma, and out-of-state births.

Result: Provides an incomplete and invalid picture of the teen pregnancy problem.

Validity vs. Reliability

➤ Examples of Invalid Data Measurement:

➤ Quality of Care:

- **Common Metric:** Malpractice complaints.
- **Validity Issues:**

Underreported Issues: Many negligent practices do not result in court cases, and most cases favor the doctor.

Result: Incomplete measure of care quality.

.

Validity vs. Reliability

2. Definition of Reliability: Accuracy of measurement and repeatable precision in the quantification process.

- **Errors in Recording Data:**
- **Issue:** Primary data is rarely displayed; errors in data transformation and recording are common.
- **Example:** Observations may be misread or recorded incorrectly, particularly in analysis with extensive data transformations.

Validity vs. Reliability

➤ Examples of Data Unreliability:

- **Inaccurate Patient Records:**
- **Issue:** Errors in manual entry of patient information, such as mistyping names, dates, or conditions.
- **Example:** A hospital records a patient's diagnosis incorrectly, leading to inappropriate treatment or misinterpretation of health statistics.
- **Flawed Data Impact:** Research using hospital data may show inaccurate prevalence rates of diseases or conditions, which can misguide resource allocation.

Chapter

02

DATA ACCESS & COLLECTION

➤ Data Access:

- **Data can be acquired in two primary ways:**
 - I. Direct Collection by Investigator (primary data).
 - II. Secondary Data

I. Direct Data Collection Methods

- **Questionnaires:**

Filled out by respondents (e.g., patients, hospital CEOs).

- **Interviews:**

Structured or semi-structured schedules with respondent answers.

- **Observation:**

Researcher observes and records details on a predetermined schedule.

II. Secondary Data

- **Definition:** Data collected for purposes other than the current study or analysis, but available for research use.
- **Examples:** Treatment effectiveness, resource utilization, vaccination coverage, impact of nurse-to-patient ratios on patient outcomes.

II. Secondary Data

➤ Sources:

- **Patient Records:** With appropriate clearances and informed consent.
- **Operational Data:** From hospitals or health care organizations.
- **Public Domain Data:** Statistics like county median income, poverty rates, or low birth weight rates.

DATA ACCESS & COLLECTION

➤ Advantages of Direct Data Collection:

- Tailored to specific needs of the investigator.
- **Example:** Directly asking a patient's age vs. converting date of birth from secondary data.

Chapter

02

DATA ACCESS & COLLECTION

➤ Challenges with Secondary Data:

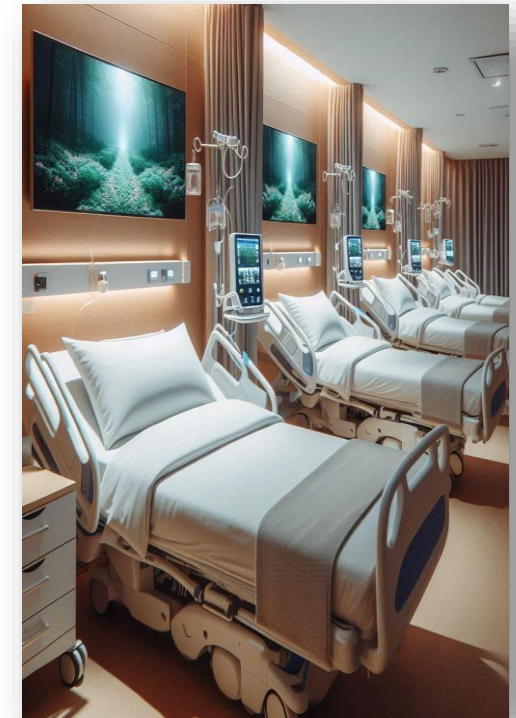
- Data may not align perfectly with current research needs.
- **Example:** The need for converting *date of birth* into *age* for statistical analysis purposes.

CASE SCENARIO (1)

- **Optimizing Hospital Bed Utilization:**
 - A large urban hospital is experiencing frequent bed shortages, leading to delays in patient admissions and reduced operational efficiency. The hospital's management team wants to analyze bed utilization rates to identify patterns and optimize the allocation of beds across departments.

CASE SCENARIO (1)

- **Optimizing Hospital Bed Utilization:**
- **Sampling Technique:**
 - **Systematic Random Sampling:** The hospital management team decides to systematically select every **5th** patient discharged over the past year to analyze their **length of stay**, **bed usage**, and **department**.

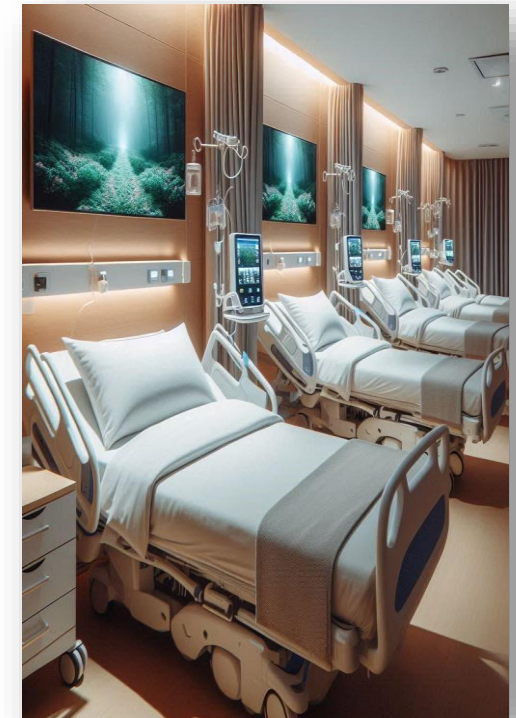


CASE SCENARIO (1)

➤ Optimizing Hospital Bed Utilization:

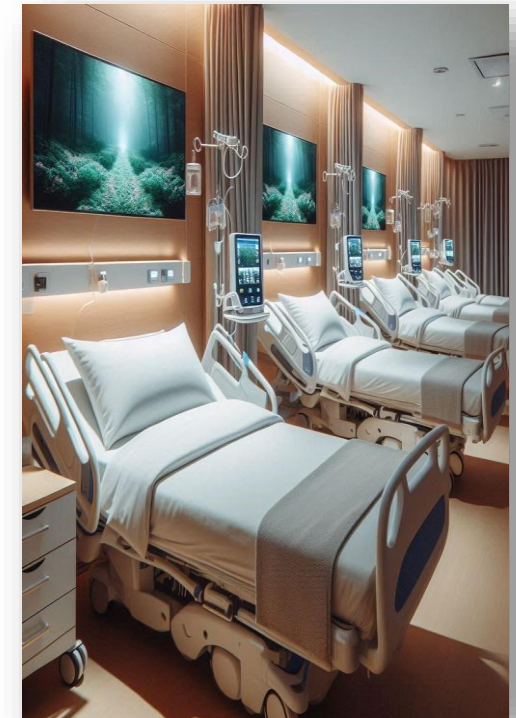
- **Validity and Reliability Considerations:**

- **Validity:** The management ensures that the study design controls for factors such as patient acuity and admission type (**emergency vs. elective**) to accurately identify patterns related to bed utilization.
- **Reliability:** Data on patient admissions, length of stay, and bed allocation are consistently recorded using the hospital's electronic health records (EHR) system to ensure repeatable and reliable findings.



CASE SCENARIO (1)

- **Optimizing Hospital Bed Utilization:**
 - **Outcome:** Systematic sampling improves bed utilization insights, enabling targeted strategies for discharge planning and resource allocation, enhancing efficiency.



CASE SCENARIO (2)

➤ **Enhancing Pharmacy Inventory Management:**

- A large healthcare system is experiencing frequent stock outs of critical medications in its pharmacies, leading to disruptions in patient care and increased operational costs due to last-minute procurement. The healthcare system's management wants to assess the effectiveness of its current inventory management practices.

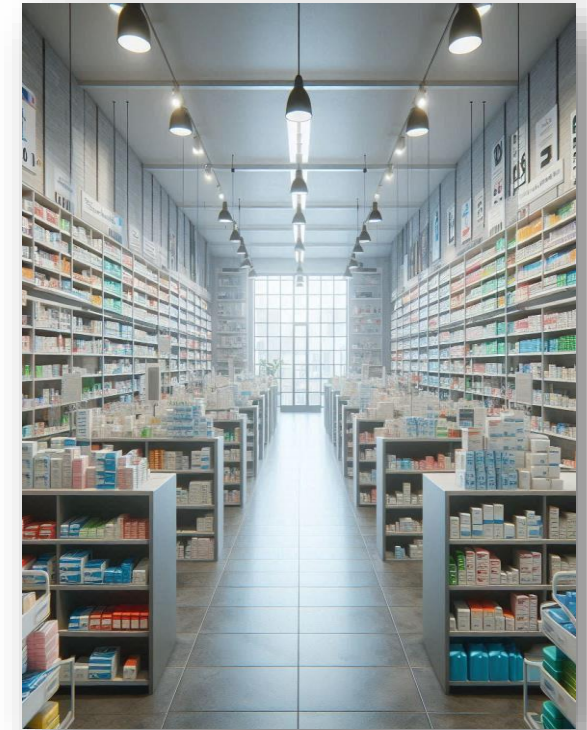
CASE SCENARIO (2)

➤ Enhancing Pharmacy Inventory Management:

- **Sampling Technique:**
- **Stratified Random Sampling:** The management team stratifies the pharmacy inventory by medication type (e.g., essential, non-essential, controlled substances) and selects a random sample from each stratum to evaluate inventory levels, reorder processes, and stock out frequency.

Chapter
02**CASE SCENARIO (2)**

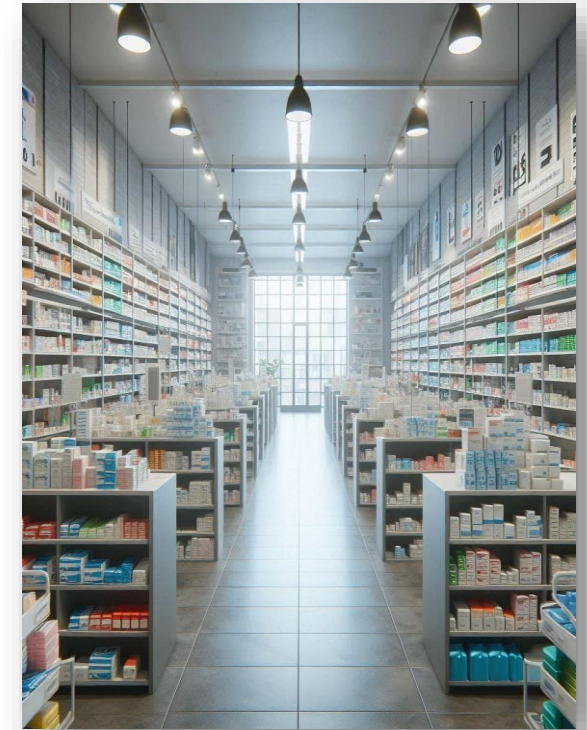
- **Enhancing Pharmacy Inventory Management:**
- **Validity and Reliability Considerations:**
 - **Validity:** The sampling process ensures that the study accurately reflects the different categories of medications, focusing on those most critical to patient care.
 - **Reliability:** Inventory data, including reorder points and stock levels, are consistently monitored and recorded using the pharmacy's inventory management system, ensuring reliable and repeatable analysis.



Chapter
02

CASE SCENARIO (2)

- **Enhancing Pharmacy Inventory Management:**
 - **Outcome:** Stratified sampling improves medication inventory management by identifying gaps, reducing stock outs, and enhancing operational efficiency in pharmacy management.



TO WRAP UP

- Random sampling methods help produce results that are unbiased and can be generalized to the broader population.
- Non-random sampling may introduce bias, limiting the generalizability and applicability of research findings.
- Validity ensures that the data collected accurately reflects the phenomenon or variable being studied.

TO WRAP UP

- Reliability ensures that the measurements and data collection processes are consistent and can be repeated with similar results.
- High-quality data is essential for producing reliable research results that are trustworthy and robust.
- Reliable and high-quality research data is key to guiding and shaping effective healthcare policies and practices.



Thank You