**📘 Sampling Techniques in Statistics & Machine Learning**

**🔍 Why Sampling?**

Sampling allows us to:

* Analyze large populations efficiently
* Reduce costs and time
* Generalize findings from a small group
* Enable statistical inference like confidence intervals and hypothesis tests

**🧪 Types of Sampling**

**I. Probability Sampling (Statistically Sound)**

**1. Simple Random Sampling**

* **Definition**: Every individual has equal chance of selection
* **Use Case**: Survey 100 customers randomly
* ✅ Pros: Unbiased
* ❌ Cons: May miss small subgroups

**2. Stratified Sampling**

* **Definition**: Divide population into groups (strata) and sample proportionally
* **Use Case**: Sample by gender, age, city type
* ✅ Pros: Ensures subgroup representation
* ❌ Cons: Needs population group knowledge

**3. Systematic Sampling**

* **Definition**: Pick every *k-th* item from a list
* **Use Case**: Every 10th user from a CRM database
* ✅ Pros: Easy implementation
* ❌ Cons: Can be biased if list has patterns

**4. Cluster Sampling**

* **Definition**: Divide into clusters, randomly choose a few, sample everyone in them
* **Use Case**: Survey 3 schools out of 50 and question all students
* ✅ Pros: Good for geographically scattered data
* ❌ Cons: Higher sampling error

**II. Non-Probability Sampling (Higher Bias Risk)**

**1. Convenience Sampling**

* **Definition**: Choose easiest-to-reach samples
* **Use Case**: Survey friends or social media followers
* ✅ Pros: Quick and cheap
* ❌ Cons: Highly biased

**2. Judgmental (Purposive) Sampling**

* **Definition**: Researcher selects based on expertise
* **Use Case**: Interview “ideal customers”
* ✅ Pros: Focused insights
* ❌ Cons: Not generalizable

**3. Snowball Sampling**

* **Definition**: Existing participants refer others
* **Use Case**: Reach rare groups like cancer patients
* ✅ Pros: Helps access niche populations
* ❌ Cons: Network bias

| **Method** | **Random?** | **Best For** | **Drawback** |
| --- | --- | --- | --- |
| Simple Random | ✅ | General unbiased sampling | May miss small subgroups |
| Stratified | ✅ | Ensuring subgroup representation | Needs stratification info |
| Systematic | ✅ | Sampling from ordered lists | Risk of pattern bias |
| Cluster | ✅ | Large, dispersed populations | Higher sampling error |
| Convenience | ❌ | Fast insights | Poor generalizability |
| Judgmental | ❌ | Expert-based studies | Subjective bias |
| Snowball | ❌ | Hard-to-reach groups | Non-random referrals |