**UNIT 11**
**SAMPLING AND DATA ANALYSIS**

## 1. WHAT IS SAMPLING IN MARKET RESEARCH

Sampling is the systematic process of selecting a subset of elements from a larger group (the population) with the expectation that the information gathered from the subset will allow for accurate conclusions about the entire group. Since we can't talk to every potential customer, we select a manageable, representative group (the sample). If this sample is chosen scientifically, their opinions act as a reliable proxy for the opinions of the entire market.

## 2. SAMPLING TERMINOLOGY

- **Population (or Universe):** The entire group of people or objects of interest that the researcher wishes to study (e.g., All potential buyers of electric cars in a city).
- **Element:** The unit about which information is sought (e.g., A single electric car buyer).
- **Sampling Unit:** An element, or a group of elements, available for selection in the sampling process (e.g., A specific household or census block).
- **Sampling Frame:** A physical list or database of all the sampling units in the population from which the sample is drawn (e.g., A list of all motor vehicle registrations).
- **Sample:** The actual subset of elements drawn from the sampling frame that participates in the study.
- **Census:** A data collection process that includes **every single element** in the population.

## 3. NEED FOR SAMPLING

Sampling is essential in market research for several practical and analytical reasons:

1. **Cost Savings:** Surveying a sample is significantly **cheaper** than conducting a census.
2. **Time Savings:** Data collection and processing are much **faster** for a small sample.
3. **Feasibility:** The population may be infinite, geographically dispersed, or unknown, making a census impossible.
4. **Accuracy:** A well-executed sample can sometimes provide *more accurate* results than a poorly executed census because resources are concentrated on reducing **non-sampling error**.

## 4. CHARACTERISTICS OF A GOOD SAMPLE

A good sample ensures the results are reliable and projectable.

- **Representativeness:** The sample must accurately **reflect the characteristics** of the population (e.g., proportional balance of age, gender, or income).
- **Accuracy:** The results obtained from the sample closely approximate the results that would have been obtained from a census.

- **Precision:** The results exhibit minimal **random sampling error**. Precision is increased by increasing the sample size.

## 5. SAMPLING DESIGNS

**Sampling Design** refers to the master plan or methodology for selecting the sample elements. The main choice is between two categories:

| Type | Key Feature | Explanation |
|---|---|---|
| **Probability Sampling** | Every element in the population has a known, non-zero chance of being selected. | Statistical Generalization. This is the scientific standard. It allows us to calculate sampling error and project the results to the entire population. |
| **Non-Probability Sampling** | The selection of elements is based on the researcher's subjective judgment or convenience. The probability of selection is unknown. | Quick Insights. Useful for exploratory research or tight budgets. We cannot generalize results or measure the sampling error. |

## 6. SAMPLING CONCEPTS

### A. Parameter vs. Statistic

- **Parameter:** A numerical characteristic of the **population** (e.g., the average age of *all* customers). It is usually unknown.
- **Statistic:** A numerical characteristic of the **sample** (e.g., the average age of the *500 sampled* customers). It is used to estimate the parameter.

### B. Sampling Error

The difference between the sample statistic and the population parameter, caused by the sample not being a perfect representation of the population. It's the natural, unavoidable mismatch that happens because you're only looking at a subset. It is the error inherent in all probability sampling and is reduced by increasing the sample size.

### C. Non-Sampling Error (Systematic Error)

Errors that arise from sources other than the sampling procedure (e.g., interviewer bias, measurement error, non-response error, or flaws in questionnaire design). Human or design mistakes (e.g., a poorly phrased question). It cannot be reduced by simply increasing the sample size.

## 7. CALCULATION OF SAMPLE SIZE

The calculation of sample size () for estimating population characteristics relies on balancing three factors: Precision, Confidence, and Variability.

- **Level of Precision (Tolerance,** E**):** The desired margin of error (e.g., ±5%).
- **Level of Confidence** (Z)**:** The probability that the confidence interval will contain the true population parameter (95% confidence uses Z=1.96).

- **Degree of Variability ($\sigma$ or $\pi$):** The heterogeneity or variance of the population (how varied the responses are). Higher variability requires a larger sample.

**Formula for Sample Size (Estimating Mean):**

$$n=Z^2\sigma^2/E^2$$

- n: Required Sample Size
- Z: Z-score (Standard error constant for the confidence level)
- $\sigma$: Population standard deviation (must be estimated)
- E: Maximum allowable error (Precision)

## 8. TYPES OF SAMPLE DESIGNS

### A. Non-Probability Sampling Methods (Bias Risk is High)

1. **Convenience Sampling:**
    - Selection of sampling units that are easily accessible to the researcher.
    - **Example:** Surveying people who walk past a certain point in a mall.
    - **Use:** Exploratory research, pre-testing.
2. **Judgmental Sampling:**
    - The researcher uses their own expert judgment to select sample members who are considered knowledgeable or representative.
    - **Example:** Interviewing only experts or key decision-makers in a specific industry.
    - **Use:** Qualitative research, small pilot studies.
3. **Quota Sampling:**
    - Ensures the sample matches the population based on pre-specified control characteristics (e.g., gender, age). Interviewers fill quotas until the target number for each subgroup is reached, but the selection *within* the quota is non-random.
    - **Use:** Most common non-probability method; often used when time and budget are limited.
4. **Snowball Sampling:**
    - Initial respondents are selected, and then they refer other potential respondents with similar, often rare characteristics.
    - **Use:** Finding low-incidence or highly specialized populations (e.g., customers of a very niche product).

## B. Probability Sampling Methods (Allows Statistical Inference)

### 1. Simple Random Sampling (SRS):

- Each element in the population has an equal and known probability of being selected.
- Selection is typically done using a random number generator on a complete sampling frame.
- High external validity, but costly and requires a perfect list of the population.

### 2. Systematic Sampling:

- The sample is chosen by selecting a random starting point and then picking every kth element in succession (where k is the skip interval).
- Select a random start, then select elements at fixed intervals (e.g., 5th, 10th, 15th, etc.).
- Simpler and faster than SRS, but risky if there is a hidden periodicity (pattern) in the sampling frame.

### 3. Stratified Sampling:

- The population is divided into mutually exclusive, homogeneous subgroups (strata), and a random sample is then drawn from *each* stratum.
- The goal is to ensure adequate representation of all subgroups.
- **Types:** Proportionate (sample size in strata matches population size) or Disproportionate (sample size is weighted based on variability or importance).

### 4. Cluster Sampling:

- The population is divided into mutually exclusive and exhaustive subgroups (clusters), and a random sample of *clusters* is selected. All elements *within* the selected clusters are then surveyed.
- The goal is to increase efficiency and reduce cost, typically geographically.
- Cost-effective due to reduced travel, but generally less precise than SRS because elements within a cluster are often similar to each other.