Sample: a subset of a population that is examined in order to determine information about the entire population.

Types of Samples: Note that all statistical sampling approaches have the common goal that chance, rather than human choice, is used to select the sample.

- **Cluster Sample:** a sampling approach in which entire groups (i.e., clusters) are chosen at random; a census is taken of each cluster. *Each cluster should be representative of the entire population.* All clusters should be heterogeneous and similar to each other. The problem with cluster samples is that the clusters are often not homogeneous and representative.

- **Convenience Sample:** a sample of individuals who are conveniently available. Convenience samples often fail to be representative.

- **Multistage Sample:** a sampling approach that combines several sampling methods. Example: stratify the country by geographic region; randomly select cities from each region; interview a cluster of residents from each city. Care should be taken at each step not to introduce bias.

- **Simple Random Sample (SRS):** a sample of size $n$ in which each set of $n$ elements has an equal chance of being selected. This is the standard against which other sampling methods are measured.

- **Stratified Random Sample:** the population is divided into subgroups (i.e., strata), and random samples are taken from each subgroup. This is better than a simple random sample if the strata are relatively homogeneous and different from each other. It results in reduced sampling variability, and can point out differences in responses among groups.

- **Systematic Sample:** individuals are selected systematically from a sampling frame (e.g., every $10^{th}$ person). Can be representative if there is no relationship between the order of the sampling frame and the variables of interest.

**Randomization:** each member of a population is given a fair, random chance of selection in the sample. This reduces bias in a sample.

**Biased Sample:** one that over- or under-emphasizes some characteristics of the population. It is caused by poor design and is not reduced as sample size increases.

**Types of Bias**

- **Voluntary Response Bias:** occurs when sample participants are self-selected volunteers (i.e., those willing to participate).

- **Undercoverage Bias:** occurs when some members of the population are inadequately covered in a sample.

- **Nonresponse Bias:** occurs when respondents to a survey differ in meaningful ways from non-respondents.

- **Response Bias:** occurs when the question is asked in such a way that it influences the response.
**Sample Size:** the number of individuals in a sample.

**Required sample size** does **NOT** depend on the size of the population (as long as the population is large enough and our sample is less than 10% of the population).

**Representative Sample:** A sample whose statistics accurately reflect the corresponding population parameters.

**Sampling Frame:** a list of individuals from which the sample is drawn.

**Sampling Variability:** the natural tendency of randomly drawn samples to differ from one another. Note: sampling variability is not a problem.

**Pilot:** A small trial run of a survey used to determine if the questions are clear.

**Population:** the entire group of individuals that we hope to learn about.

**Census:** examination of information about every member of a population. This is the best approach when the population is small and accessible.

Why not do a census all the time?
- Difficult or expensive to complete.
- Populations rarely stand still. A census takes time and the population changes during it.
- A census is more complex than a sample.

**Parameter:** a descriptive measure (using a numerical value) of the population, e.g., $\mu$, $\sigma$. Also called a **population parameter**.

**Statistic:** a descriptive measure (using a numerical value) of a sample, e.g., $\bar{x}$, $s$. Also called a **sample statistic**.

<table>
<thead>
<tr>
<th>Key Statistics and Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Name</strong></td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>Standard Deviation</td>
</tr>
<tr>
<td>Correlation</td>
</tr>
<tr>
<td>Regression Coefficient</td>
</tr>
<tr>
<td>Proportion</td>
</tr>
</tbody>
</table>

**The Valid Survey**

- What do I want to know?
- Am I asking the right respondents (i.e., do I have the right sampling frame)?
- Am I asking the right questions? Ask only questions that help you learn what you want to know. Be specific. In each question, either give a set of alternative answers (i.e., multiple choice) or ask for a numerical response, if possible. Ask questions in a neutral way (i.e., avoid bias).
- What will I do with the answers: will they address what I want to know?