

What is the difference between a population and a sample?

Authored by
stats writer

December 18, 2025

RECOMMENDED CITATION

stats writer (2025). *What is the difference between a population and a sample?*.
PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=107737>

Understanding the fundamental distinction between a population and a sample is essential for anyone working in statistics or data science. Simply put, a **population** is the complete set of all possible observations, measurements, or items under study from a given group. Conversely, a **sample** is defined as a smaller, manageable subset drawn from that larger population. Because collecting data on every member of a massive population is often impossible, researchers utilize samples to gather data and subsequently make scientific inferences about the totality of the group. A sample is always smaller than the population and is the primary tool used in statistical estimation.

In the realm of statistical analysis, data collection is generally driven by the necessity to answer specific research questions. These questions require us to investigate characteristics of an entire group, which we formally define as the population.

Consider the types of investigative queries that drive researchers to define their scope of study:

Determining the median household income for all residents residing in a specific metropolitan area, such as **Miami, Florida**.

Calculating the average (mean) weight across an entire species or specified group of **turtles**.

Estimating the proportion or **percentage** of eligible voters within a certain county who express support for proposed legislation.

In all these scenarios, our ultimate goal is to generate conclusions about the entire group--the population. However, instead of attempting the arduous task of measuring every single individual element, we strategically select and collect data only from a **sample**, which serves as a representative subset or portion of the overall group.

Population: Defined as every single possible individual element, measurement, or observation that the researcher is interested in studying.

Sample: A carefully selected subset or specific portion drawn from the larger population.

Contrasting Population and Sample through Real-World Examples

To solidify this core concept, let us revisit the introductory examples and clearly delineate which entity represents the population and which represents the sample in each context. The size difference between the population (often denoted N) and the sample (often denoted n) highlights the necessity of using sampling techniques for practical research.

Example 1: Median Household Income in Miami, Florida

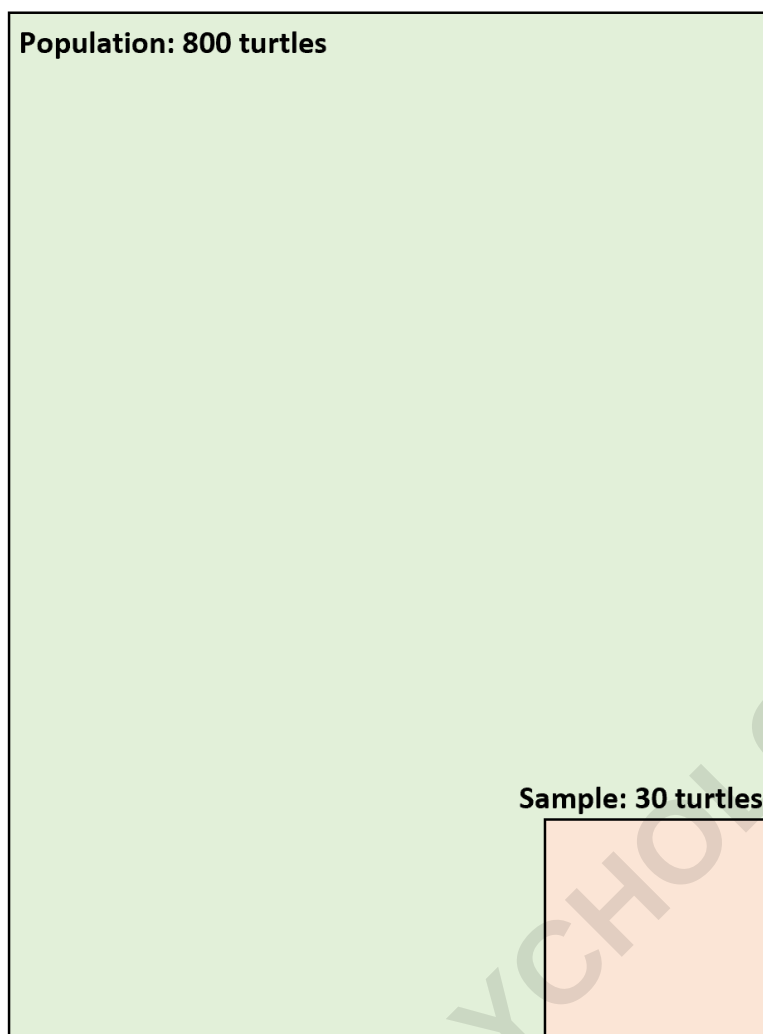
If the focus is the entire city, the **population** would encompass all 500,000 households residing

there. Due to constraints, researchers might only manage to collect income data on a small fraction, perhaps a **sample** consisting of 2,000 total households. This small group is then used to estimate the characteristics of the larger group.



Example 2: Mean Weight of a Turtle Population

For ecological studies, if the defined population includes 800 turtles in a specific habitat, attempting to capture and weigh every single one can be highly challenging. Instead, a researcher may choose a **sample** of 30 turtles, calculating the mean weight of this subset to extrapolate information about all 800 turtles.



Example 3: Resident Support for Local Legislation

If a county has 50,000 eligible residents, this constitutes the entire **population** whose views are of interest. To gauge the overall sentiment, political analysts will often survey a significantly smaller **sample**, perhaps 1,000 residents, and calculate the proportion of support based on those responses to estimate the percentage for the entire county.

The Core Rationale: Why Sampling is Necessary

While analyzing an entire population provides perfect accuracy, doing so is almost always impractical or impossible in real-world statistical research. Statistical analysis overwhelmingly relies on data collected from a sample due to severe limitations concerning resources, time, and feasibility. Understanding these constraints is crucial to appreciating the fundamental role of sampling in modern scientific inquiry and decision-making processes.

The primary reasons researchers opt for sampling methods over complete population census

collection are summarized below:

Time Constraints and Data Validity: It is exceptionally **time-consuming** to collect data from every individual within a large population. For instance, obtaining the median income for every household in a major metropolitan area could take months or even years. By the time the data collection is complete, economic conditions may have changed, rendering the initial findings obsolete or irrelevant to the current state of the population.

Financial Costs: Conducting a complete census of a population is highly **costly**. Employing the necessary personnel, developing the infrastructure, and managing the logistics required to survey every individual element demands significant financial investment. Utilizing a sample drastically reduces these expenses while still providing statistically viable results that meet research objectives.

Logistical and Physical Feasibility: In numerous instances, it is simply **unfeasible** to collect data from every member. If a population consists of migratory animals, such as turtles scattered across a wide area, tracking down and measuring every single individual is practically impossible. Sampling provides a realistic and ethically sound approach to gathering necessary data when complete enumeration is unattainable.

By strategically collecting data from well-chosen samples, researchers are able to acquire information about the target population much more rapidly and cost-effectively. Furthermore, if the sample is carefully selected to be truly representative of the population, the findings drawn from this smaller subset can be generalized to the entire group with a high degree of statistical confidence.

Defining the Representative Sample

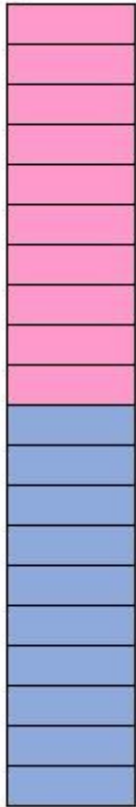
When researchers draw a sample from a population, the goal is for that sample to function as a perfect "mini version" of the entire group. This concept defines a **representative sample**: a subset where the key characteristics of the individuals in the sample closely mirror the corresponding characteristics found in the overall population. Achieving representativeness is paramount because it dictates the validity of generalizing findings from the sample back to the population.

For example, imagine a study aiming to understand the movie preferences of 5,000 students in a large school district. Surveying every student is too time-consuming, so a researcher selects a sample of 100 students. If the overall student population is exactly 50% female and 50% male, a truly representative sample of 100 students should also maintain that 50/50 gender ratio.

If the study yields a sample consisting of 90% male students and only 10% female students, this sample would be severely **non-representative** concerning gender distribution. Such a biased

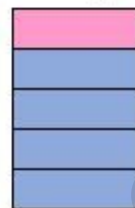
sample could lead to flawed conclusions about the movie preferences of the entire school district population, as the male demographic is significantly overrepresented, distorting the overall results.

Population

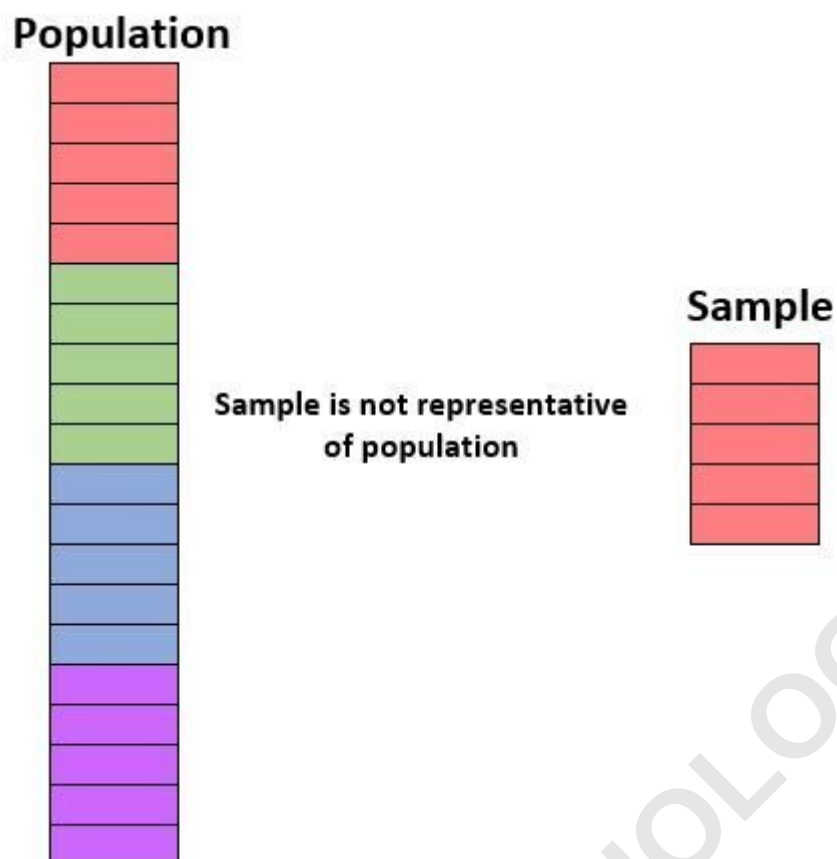


**Sample is not representative
of population**

Sample



Similarly, if the population is structurally balanced across grade levels (freshmen, sophomores, juniors, and seniors), but the chosen sample consists exclusively of freshmen, the results will not accurately reflect the diversity of opinions or characteristics present across the entire school body. A representative sample must accurately reflect all critical strata within the population.



When a sample successfully captures the characteristics of the target population, we can confidently generalize the observed findings--be they averages, proportions, or correlations--from the sample back to the entire group, maximizing the utility and validity of the research effort.

Methodologies for Achieving Representative Sampling

To maximize the likelihood of obtaining a representative sample and minimize sampling bias, researchers employ various rigorous sampling methods. These methodologies, particularly those based on probability theory, ensure that every individual within the target population has a known and, ideally, equal probability of being included in the sample, which is key to statistical validity.

The most effective methods for probability sampling, designed to create that unbiased "mini version" of the population, include:

Simple Random Sampling: This method involves selecting individuals purely through random chance, often utilizing tools like random number generators or truly random selection processes where every unit in the population is assigned a unique identifier. This ensures complete fairness, as every combination of individuals has an equal chance of being selected, thus maximizing the probability of a representative outcome.

Systematic Random Sampling: This technique requires ordering all members of the population sequentially, usually in a list. Researchers then select a random starting point and proceed to choose every n th member (e.g., every 10th person) to be included in the sample. This approach is efficient, straightforward to execute, and provides broad, systematic coverage across the population list.

Stratified Random Sampling: This method is highly valuable when the population naturally divides into distinct groups, or strata (e.g., age groups, geographic regions, or academic classifications). The population is first split into these non-overlapping groups, and then a random sample is drawn proportionally or disproportionately from each subgroup. This ensures that all key segments of the population are appropriately represented in the final sample composition, which is particularly useful when analyzing heterogeneous populations.

By carefully applying these probability sampling techniques, researchers ensure that every unit in the population has a non-zero, quantifiable chance of inclusion. This methodological rigor is what allows the study's results, derived from the small sample, to be confidently extrapolated to draw meaningful inferences about the totality of the population.