

What is the 10% Condition in Statistics: Definition & Example

Authored by
stats writer

December 21, 2025

RECOMMENDED CITATION

stats writer (2025). *What is the 10% Condition in Statistics: Definition & Example*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=108291>

The 10% Condition is a critical guideline used extensively in inferential statistics, particularly when applying the Normal Approximation to the binomial distribution. Fundamentally, this rule establishes a threshold for the relationship between the analyzed sample size and the total population from which the sample is drawn. Specifically, it mandates that the sample size must not exceed 10% of the overall population size. This prerequisite is essential for maintaining the integrity of statistical inferences and guaranteeing that the sampling process does not significantly alter the probability distribution of the population, thereby ensuring that the sample remains statistically representative of the larger group.

This condition addresses a fundamental statistical challenge related to sampling methodologies. When researchers draw samples from a finite population without replacing the selected elements--a common practice known as sampling without replacement--the trials cease to be truly independent. If the sample constitutes a substantial portion of the population, the composition of the remaining population changes drastically with each selection, distorting subsequent probabilities. For instance, if a population totals 1000 individuals, the 10% Condition restricts the sample size to a maximum of 100 individuals, ensuring that the removal of 100 observations does not fundamentally skew the parameters of the remaining 900.

The adherence to the 10% Condition allows statisticians to proceed with calculations and modeling that rely heavily on the assumption of independence. Although the trials are technically dependent when sampling without replacement from a finite population, the condition provides a robust margin of safety. When the sample size remains small relative to the population (less than 10%), the resulting dependency introduced by the sampling method is negligible, meaning the computed probabilities closely approximate those derived under perfect independence. This practical relaxation simplifies complex calculations, allowing researchers to leverage powerful statistical tools, such as the Normal Approximation, which are contingent upon the independence assumption.

The Foundation: Understanding Bernoulli Trials

At the heart of the statistical methods governed by the 10% Condition lies the concept of a **Bernoulli trial**. Defined rigorously, a Bernoulli trial is a random experiment characterized by two crucial properties: first, it must possess only two mutually exclusive outcomes, conventionally labeled "success" or "failure"; and second, the probability of "success" must remain constant across every repetition of the experiment. This foundational model forms the basis for the binomial distribution, which allows statisticians to calculate the probability of obtaining a specific number of successes within a fixed number of independent trials. Understanding these trials is paramount because many real-world scenarios, from quality control inspections to public opinion polling, can be modeled effectively using this dichotomous framework.

A classic and highly intuitive example of a Bernoulli trial is the simple act of flipping a fair coin. In this scenario, the outcomes are strictly limited to two possibilities--Heads or Tails. If we designate Heads as the "success" and Tails as the "failure," the probability of success, $P(\text{Heads})$, is consistently 0.5 on every single flip, provided the coin is unbiased. However, Bernoulli processes extend far beyond coin flips; they encompass any situation where the outcome is binary and the underlying probability of success, denoted as p , remains unchanged. For instance, determining whether a randomly selected component is defective or non-defective, or whether a survey respondent approves or disapproves of a policy, all fit the structure of a sequence of Bernoulli trials.

When dealing with a large volume of these individual Bernoulli trials, calculating exact binomial probabilities can become computationally cumbersome. Consequently, statisticians frequently rely on approximation methods to simplify the analysis. The most common and powerful approximation tool for large numbers of Bernoulli trials is the application of the Normal Distribution, often referred to as the Normal Approximation to the binomial. This approximation is exceptionally useful because it converts the discrete binomial distribution into a continuous normal distribution, simplifying probability calculations considerably. Nevertheless, the theoretical validity of this substitution is rigidly dependent upon one core statistical assumption: the absolute independence of all trials.

The Critical Requirement of Independence

The concept of statistical independence implies that the outcome of any single trial in a sequence has absolutely no influence on the outcome of any other subsequent trial. This condition is inherently met in situations involving sampling with replacement or when the population is theoretically infinite, such as the coin flip example. If trials are dependent, meaning the outcome of one trial alters the probability for the next, the mathematical formulas derived for binomial and normal distributions break down, leading to inaccurate probability estimates and flawed statistical inferences. Therefore, when attempting to use the Normal Distribution to model a series of Bernoulli trials, the prerequisite of independence is non-negotiable from a theoretical standpoint.

The reliance on independence is particularly important in inferential statistics, where the goal is to generalize findings from a small sample back to the entire population. If the sampling process itself introduces systematic bias or alters the fundamental probabilities (i.e., dependency), the resulting confidence intervals and hypothesis tests will be unreliable. Statisticians must rigorously assess whether the data collection process adheres to this assumption. When sampling from practical, real-world populations--which are almost always finite--perfect independence is rarely achievable unless the sample size is extremely small or the sampling method involves replacement, a method often impractical in sociological or industrial studies.

Addressing Dependency: Sampling Without Replacement

In most practical research contexts, data collection involves sampling without replacement. This means that once an element (a person, item, or observation) is selected from the population and included in the sample, it cannot be selected again. While this ensures that no element is counted multiple times, it fundamentally violates the requirement for perfect independence. The act of removing an element from a finite population changes the composition of the remaining pool, thus altering the probability of success for subsequent selections. This shift is minimal when the population is enormous relative to the sample, but it becomes pronounced as the sample size increases relative to the population size.

Consider a small example: selecting two red marbles from a jar containing five red and five blue marbles (Population $N=10$). The probability of the first selection being red is $5/10$ (0.5). If we do not replace the first marble, the population is now 9 marbles. If the first was red, the probability of the second selection also being red is now $4/9$ (approximately 0.444). The two selections are clearly dependent, as the result of the first trial directly influenced the probability distribution of the second. This dependency creates a dilemma: how can we use powerful independent models, like the Normal Approximation, when our real-world sampling introduces dependency?

The 10% Condition Defined

The 10% Condition: As long as the sample size (n) is less than or equal to 10% of the total population size (N), researchers can safely proceed by making the assumption that the Bernoulli trials are independent, even though sampling occurred without replacement.

The 10% Condition serves as a pragmatic bridge between theoretical requirements and practical sampling methods. This "rule of thumb" dictates that if the sample size, n , does not exceed one-tenth of the population size, N , the finite population correction factor--a statistical adjustment designed to account for sampling without replacement--will be close enough to 1 that its omission introduces negligible error. This mathematical tolerance allows statisticians to utilize simpler, independence-based models, such as the Normal Approximation or standard error formulas derived from independent assumptions, without materially sacrificing the accuracy of their results. This is known as **The 10% Condition**.

The condition is therefore not a statement about perfect independence, but rather an empirical threshold for acceptable dependency. By keeping the sample size small relative to the population, the depletion effect caused by sampling without replacement is minimized. If only a small fraction of the population is removed, the probability distribution of the remaining population is only slightly perturbed, allowing us to treat the trials as if they were independent for all practical purposes of inference. This powerful allowance is why the 10% Condition is foundational to introductory and advanced statistics alike, especially in hypothesis testing and the construction of confidence

intervals for proportions.

Intuition Behind The 10% Condition

To truly develop an intuition regarding the efficacy of the 10% Condition, we must visualize how the probability shift diminishes as the population size grows relative to the sample size. The core issue is the change in conditional probability. If a sample represents a substantial chunk of the population, the act of selecting an item drastically alters the proportions remaining. Conversely, if the population is vast, removing a small number of items makes the probability of selecting subsequent items nearly identical to the probability before the selection occurred. This principle underpins the allowance provided by the condition.

Let us consider the previous scenario, but formalized: Suppose a classroom represents our finite population, and we know the true proportion of students who prefer football over basketball is 50%. Let the random variable X denote the number of students randomly selected who prefer football. We are interested in the probability that all four randomly selected students prefer football over basketball. If the classroom size (N) is 20, and the trials were hypothetically independent (sampling with replacement), the probability calculation would be straightforward, as the base probability ($10/20$) never changes.

When trials are independent, the probability that all 4 students prefer football is calculated simply by multiplying the constant probabilities: $P(\text{All 4 students prefer football}) = (10/20) \text{ times } (10/20) \text{ times } (10/20) \text{ times } (10/20) = \mathbf{.0625}$. However, if our trials are *not* independent (e.g., once we sample one student, they cannot be placed back in the classroom), the calculation must account for the diminishing population and changing proportions. If the classroom size is 20 and the sample size is 4, the sample size (4) is 20% of the population (20), thus violating the 10% Condition. In this non-independent scenario, the probability accounting for the dependency is: $P(\text{All 4 students prefer football}) = (10/20) \text{ times } (9/19) \text{ times } (8/18) \text{ times } (7/17) = \mathbf{.0433}$. These two probabilities are quite different, highlighting the necessity of adhering to the condition.

However, consider the following table that shows the probability that all 4 randomly selected students prefer football, based on increasing classroom size (N):

Classroom Size	Sample Size / Classroom Size	$P(\text{All 4 students prefer football})$ <i>Independent trials</i>	$P(\text{All 4 students prefer football})$ <i>Not Independent trials</i>
20	20.0%	$10/20 * 10/20 * 10/20 * 10/20 = .0625$	$10/20 * 9/19 * 8/18 * 7/17 = .0433$
40	10.0%	$20/40 * 20/40 * 20/40 * 20/40 = .0625$	$20/40 * 19/39 * 18/38 * 17/37 = .0530$
100	4.0%	$50/100 * 50/100 * 50/100 * 50/100 = .0625$	$50/100 * 49/99 * 48/98 * 47/97 = .0587$
1,000	0.4%	$500/1,000 * 500/1,000 * 500/1,000 * 500/1,000 = .0625$	$500/1,000 * 499/999 * 498/998 * 497/997 = .0621$

As the sample size relative to the population size (e.g., "classroom size" in this example) decreases, the calculated probability between independent trials and non-independent trials gets closer and closer. This convergence illustrates the diminishing impact of dependency as the ratio n/N falls. Note that when the sample size is exactly 10% of the population size ($n/N=0.1$), the difference between the probabilities of independent trials and non-independent trials is relatively similar, establishing this ratio as the working limit for acceptable error. This is the mathematical justification for the 10% threshold.

Practical Applications and Best Practices

The application of the 10% Condition is fundamental to the appropriate use of large sample statistical tools. The primary benefit is allowing the use of standard formulas for calculating the standard error of a sample proportion or mean, which assume independence. If the condition is met, the researcher avoids the complexity of incorporating the Finite Population Correction (FPC) factor into every calculation. Ignoring the FPC when n/N exceeds 10% results in standard errors that are too large, leading to confidence intervals that are overly wide and potentially incorrect conclusions in hypothesis testing.

Violating the 10% Condition leads to standard errors that are systematically underestimated, causing confidence intervals to be artificially narrow and p -values to be erroneously small, suggesting a level of precision that does not exist. For example, if a researcher draws a sample size of 300 from a population of 1000 (30% ratio), calculating the standard deviation using the standard formula would ignore the dependency, resulting in highly inflated confidence in the results and a substantial risk of committing a Type I error.

When designing a study, statisticians must prioritize keeping the sample size far below the 10% threshold whenever possible. While 10% is the acceptable maximum, optimal practice dictates utilizing a much smaller proportion--perhaps 5% or less--especially when dealing with small to moderately sized populations. The smaller the ratio n/N , the closer the calculated statistics will align with the ideal scenario of perfect independence, thereby enhancing the reliability and generalizability of the research findings. If the sample size must exceed 10% due to research constraints, the researcher must explicitly use the FPC to adjust for the lack of independence.

Conclusion

The 10% Condition is a fundamental and indispensable heuristic in statistical practice, serving as a boundary for acceptable dependency when sampling without replacement from a finite population. It articulates that our sample size should be less than or equal to 10% of the population size in order to safely make the assumption that a set of Bernoulli trials is independent for the purpose of applying approximations and standard formulas. This condition is rooted in minimizing

the impact of the finite population correction factor, ensuring that the error introduced by assuming independence is statistically negligible.

Adherence to this rule allows researchers to simplify their statistical modeling, relying on standard formulas for standard error and variance. Of course, it's best if our sample size is much less than 10% of the population size so that our inferences about the population are as accurate as possible. For example, we'd prefer that our sample size is only 5% of the population compared to 10%, as this further minimizes the risk of undermining the independence assumption. By respecting the 10% Condition, statisticians ensure that their conclusions are robust, reliable, and reflective of the true population parameters.

ARABPSYCHOLOGY.COM