What is the Success/Failure Condition in Statistics 3. The output of the file command is: /home/ccuser/workspace/learn-the-command-line/file ASCII text

Authored by
stats writer

December 21, 2025

The Success/Failure Condition in **statistical inference** is a crucial diagnostic tool used primarily to determine whether the sampling distribution of a proportion can be reasonably approximated by the **Normal Distribution**. This condition serves as a gatekeeper, ensuring that the theoretical assumptions underlying many common statistical procedures, such as constructing a **confidence interval** or performing a hypothesis test for a population proportion, are met. Without satisfying this condition, the results derived from the normal approximation may be inaccurate or unreliable, potentially leading to erroneous conclusions about the population being studied.

The core purpose of this check is to guarantee that the sample size is large enough relative to the population proportion to produce a sampling distribution that is sufficiently symmetric and bell-shaped. When the sample size is too small, or when the true proportion of success or failure is close to zero or one, the distribution remains highly skewed. By verifying the Success/Failure Condition, statisticians ensure that the model employed--often the standard normal model--provides an acceptable level of accuracy for estimating or testing the population parameter.

Establishing the Need for the Success/Failure Condition

The application of the Success/Failure Condition is deeply rooted in the framework of probability distributions that model dichotomous outcomes. Before we can use methods that rely on the smoothness of the normal curve, we must ensure that the underlying data-generating process is modeled accurately. Many real-world problems involve situations where an event either occurs ("success") or does not occur ("failure"), such as counting voters in favor, defective items, or successful medical treatments.

This condition, therefore, acts as a measurement that contributes significantly to the overall reliability of a chosen statistical method. It quantifies the degree to which a particular model produces results that match or exceed a predetermined level of accuracy associated with the normal distribution approximation. In essence, it is an essential evaluation of the quality and applicability of the statistical technique being used for inference concerning proportions.

The Foundation: Understanding Bernoulli Trials and Binomial Data

The data that necessitates the Success/Failure Condition originates from a series of independent Bernoulli trials. A **Bernoulli trial** is defined as an experiment with only two possible outcomes--conventionally labeled "success" or "failure"--where the probability of success, denoted by p, remains constant for every iteration of the experiment. The classic example of a Bernoulli trial is a single flip of a fair coin, where heads (success) and tails (failure) have a constant probability of 0.5.

When we perform a fixed number of **Bernoulli trials** (n) and count the total number of successes, the resulting probability distribution is known as the Binomial Distribution. This distribution is discrete, meaning it can only take on a finite number of values (0, 1, 2, ..., n). While the Binomial

Distribution is the exact model for the number of successes, calculating probabilities for large sample sizes (large *n*) using the binomial formula can become computationally intensive and complex.

Consequently, the need arises to simplify these calculations, which leads us directly to employing the **Normal Distribution** as a substitute. Understanding the relationship between these two distributions is paramount: the discrete nature of the Binomial Distribution must be handled appropriately when being approximated by the continuous **Normal Distribution**, and the Success/Failure Condition provides the necessary safety check for this transition.

Why We Approximate: Linking the Binomial to the Normal Distribution

When the sample size (*n*) is large, the shape of the **Binomial Distribution** starts to resemble the symmetrical, bell-shaped curve of the **Normal Distribution**. This phenomenon is supported by the **Central Limit Theorem**, which, in a broader sense, dictates that the sampling distribution of means or proportions tends toward normality as the sample size increases. However, the definition of "large enough" for *n* depends on the underlying probability *p*.

If the probability of success *p* is close to 0.5, the Binomial Distribution achieves symmetry relatively quickly, even with smaller sample sizes. Conversely, if *p* is very small (close to 0) or very large (close to 1), the distribution is highly skewed, and a much larger sample size is required before the distribution becomes sufficiently symmetric to justify using the normal approximation. The **Success/Failure Condition** provides a simple, quantifiable rule to standardize what constitutes a sufficiently large sample in various probabilistic scenarios.

When this approximation is valid, we can utilize the well-established properties and Z-tables of the **Normal Distribution** to calculate probabilities, critical values, and standard errors, simplifying the entire inference process. For example, calculating a **confidence interval** for a population proportion becomes straightforward when we can rely on the normal model to define the margin of error.

The Formal Success/Failure Condition Criteria

The **Success/Failure Condition** specifies the minimum required number of expected successes and expected failures in a sample. This expectation is calculated based on the sample size (*n*) and the hypothesized or estimated probability of success (*p*).

Success/Failure Condition: There should be at least 10 expected successes and 10 expected failures in a sample in order to use the normal distribution as an approximation for the binomial distribution.

Written formally using mathematical notation, we must verify that both of the following criteria are simultaneously satisfied:

Expected number of successes is at least 10: **np** ≥ **10**
Expected number of failures is at least 10: **n(1-p)** ≥ **10**

In these expressions, *n* represents the **sample size**, and *p* represents the true or estimated **probability of success** on a single trial. Meeting both inequalities ensures that the tails of the distribution are far enough away from 0 and *n*, preventing significant skewness and ensuring that the discrete bars of the binomial histogram are adequately covered by the smooth curve of the normal model.

It is important to acknowledge that some textbooks and introductory courses might suggest that only 5 expected successes and 5 expected failures are needed (i.e., np ≥ 5 and n(1-p) ≥ 5) to permit the use of the normal approximation. However, the criterion requiring 10 successes and 10 failures is generally regarded as a more **conservative** and robust standard, leading to more accurate results, especially near the critical boundaries. For expert statistical practice and adherence to common academic standards, the minimum threshold of 10 is strongly recommended and utilized throughout this discussion.

Case Study: Verifying the Condition for a Confidence Interval

To illustrate the application of this condition, consider a scenario where a political analyst wishes to estimate the proportion of residents in a large county who favor a specific newly proposed law. A random sample of 100 residents is selected, and their stance on the law is recorded. The resulting statistics are:

Sample size **n = 100**
Sample proportion in favor of the law **p? = 0.56**

The analyst intends to construct a **confidence interval** for the true population proportion (*p*). This interval calculation typically relies on the following standard formula, which involves a critical Z-value derived from the **Normal Distribution**:

Confidence Interval = p? +/- z*√p?(1-p?) / n

The components of this formula are defined as follows:

p?: The observed sample proportion, an estimate of the true population proportion.
z*: The critical z-value that corresponds to the desired level of confidence derived from the **Normal Distribution**.
n: The total **sample size**.

Since this calculation involves a z-value, it implicitly uses the **Normal Distribution** to approximate the true sampling distribution of the proportion, which is inherently binomial. Therefore, before we can calculate the interval accurately, we must verify that the **Success/Failure Condition** is met. We calculate the expected number of successes and failures based on the observed sample proportion, p? (since the true population proportion *p* is unknown):

Expected number of successes: n * p? = 100 * 0.56 = **56**

Expected number of failures: n * (1 - p?) = 100 * (1 - 0.56) = 100 * 0.44 = **44**

Since both the expected number of successes (56) and the expected number of failures (44) are greater than or equal to 10, the **Success/Failure Condition** is satisfied. The analyst is therefore justified in proceeding with the standard formula shown above to calculate the confidence interval, confident that the normal approximation will yield a reliable result.

A Necessary Partner: Adhering to the 10% Independence Rule

While the **Success/Failure Condition** addresses the shape (symmetry) of the sampling distribution, another crucial requirement must be met for the normal approximation to the Binomial Distribution to be valid: the **Independence Condition**, often quantified as the 10% Condition.

The formulas for the standard deviation (and thus the standard error used in confidence intervals) assume that the individual observations within the sample are statistically **independent**. However, in most practical sampling scenarios, particularly those involving finite populations (sampling without replacement), the trials are technically not independent; removing an item changes the probability of selecting the next item.

The 10% Condition provides a practical threshold to treat the trials as independent even when sampling without replacement. This condition states that the **sample size** (*n*) must not exceed 10% of the total population size (*N*). That is, $n \leq 0.10 * N$. If the sample takes up less than 10% of the population, the reduction in the population size has a negligible effect on the probabilities, and we can safely use the standard error formulas derived under the assumption of independence. Adhering to the 10% Condition ensures that the calculation of the standard error remains accurate.

Application to Complex Scenarios: Comparing Two Proportions

The principles established by the **Success/Failure Condition** extend logically to more complex statistical comparisons, such as estimating the difference between two population proportions. When working with two proportions--for instance, comparing the efficacy of two different treatments or the approval ratings in two different demographic groups--the requirement for the normal approximation becomes more stringent.

If you are creating a **confidence interval** or performing a hypothesis test for the difference between two proportions, you must verify that the required success/failure counts are met **in both samples independently**. This ensures that the sampling distributions for both sample proportions (p?1 and p?2) are individually well-approximated by the normal curve, which allows their combination (the difference, p?1 - p?2) to also be modeled accurately by the normal distribution.

Specifically, if you have Sample 1 (n1, p1) and Sample 2 (n2, p2), you must verify four separate inequalities:

Expected successes in Sample 1: **n1 * p?1** ≥ **10**
Expected failures in Sample 1: **n1 * (1 - p?1)** ≥ **10**
Expected successes in Sample 2: **n2 * p?2** ≥ **10**
Expected failures in Sample 2: **n2 * (1 - p?2)** ≥ **10**

Meeting all four conditions is absolutely necessary to ensure the validity and reliability of the resulting inference about the difference between the two population proportions. If any one of these four conditions is violated, alternative, non-parametric methods or exact binomial tests must be employed to maintain statistical integrity.