Learning the Binomial Test in R: A Step-by-Step Guide

Authored by

Mohammed looti

November 8, 2025

Understanding the Binomial Test and Its Purpose

The Binomial Test is a fundamental statistical tool used to assess whether the proportion of successes observed in a sample significantly deviates from a specific, predetermined theoretical or hypothesized proportion. This test is applicable exclusively when dealing with data that follows a binomial distribution--meaning the outcomes of the trials are dichotomous (e.g., success or failure, yes or no) and independent of one another. The primary goal is to determine if the true underlying population proportion (pi) aligns with the expected value (p) based on the evidence collected from a limited number of trials (n).

Before employing this test, it is crucial to ensure that the data meets the necessary assumptions. Specifically, the trials must be fixed in number, each trial must be independent, and the probability of success (p) must remain constant across all trials. For instance, testing the effectiveness rate of a new manufacturing process or determining if a coin is fair are classic scenarios where the binomial test provides robust statistical inference. The test provides an exact probability (p-value), making it particularly valuable for smaller sample sizes where asymptotic tests might be less reliable.

In essence, the binomial test compares the number of observed "successes" (x) within a fixed number of trials (n) against what would be expected if the null condition (p) were true. A large disparity between the observed outcome and the expected outcome suggests that the true population proportion (pi) is likely different from the hypothesized value (p), prompting a careful examination of the resulting statistical output to draw meaningful conclusions.

Formulating Null and Alternative Hypotheses

Every statistical test begins with the formal establishment of the Null hypothesis (H_0) and the Alternative hypothesis (H_A). These statements represent the two competing views regarding the true state of the population proportion (pi). The Null hypothesis always represents the status quo--the assumption of no effect or no difference, asserting that the population proportion is exactly equal to the hypothesized value (p). Conversely, the Alternative hypothesis posits that the true population proportion is not equal to, or is directional relative to, the hypothesized value.

The standard formulation for the hypotheses in a two-tailed (non-directional) binomial test is as follows:

H_0: $pi = p$ (The population proportion pi is equal to the hypothesized proportion p.)

H_A: $pi neq p$ (The population proportion pi is not equal to the hypothesized proportion p.)

It is important to recognize that the Binomial Test offers flexibility in the Alternative hypothesis depending on the research question. If the researcher is specifically interested in whether the true proportion is larger than p (a right-tailed test) or smaller than p (a left-tailed test), the Alternative hypothesis is adjusted accordingly. This choice--whether to use a one-tailed or two-tailed test--must be made before data collection and statistical analysis to prevent biases in interpretation, fundamentally influencing how the resulting P-value is calculated and interpreted regarding statistical significance.

Executing the Binomial Test in R

The statistical programming environment R simplifies the execution of the binomial test through its built-in function, `binom.test()`. This function efficiently calculates the exact probability of observing a given number of successes or more extreme outcomes, assuming the null hypothesis is true. To utilize this function effectively, three primary arguments must be supplied, corresponding directly to the core components of the binomial experiment: the number of successes, the total number of trials, and the null probability of success.

The core syntax for the function is concise and powerful, allowing users to specify all necessary parameters within a single command. Understanding the role of each parameter is essential for accurate modeling and interpretation of the results:

binom.test(x, n, p)

Where the arguments are defined as:

x: Represents the specific number of successful outcomes observed in the sample.

n: Represents the total number of independent trials or observations conducted.

p: Represents the hypothesized probability of success on any single trial, corresponding directly to the value specified in the Null hypothesis (H_0).

Additionally, the function accepts an optional argument, `alternative`, which allows the user to specify the direction of the Alternative hypothesis. By default, R performs a two-tailed test (`alternative="two.sided"`), but this can be explicitly changed to `alternative="less"` for a left-tailed test or `alternative="greater"` for a right-tailed test. The following practical examples demonstrate how to apply this function across different hypothesis scenarios, ranging from non-directional inquiries to specific directional predictions.

Case Study 1: The Two-Tailed Hypothesis

Consider a classic probability investigation involving a standard six-sided die. We hypothesize that

the die is fair, meaning the true probability of rolling any specific number, such as "3," is $1/6$ or approximately 0.1667. To test this statistical significance, we roll the die 24 times and record the results. If the die lands on "3" a total of 9 times, we must determine if this observed frequency (9 out of 24) is significantly different from the expected frequency (4 out of 24, or $1/6 times 24$). The Null hypothesis states that the true probability (pi) equals $1/6$, while the Alternative hypothesis states that pi is simply not equal to $1/6$.

The observed number of successes is $x=9$, the number of trials is $n=24$, and the hypothesized probability is $p=1/6$. We input these values into the R function to execute the two-tailed binomial test:

#perform two-tailed Binomial test
binom.test(9, 24, 1/6)

#output
Exact binomial test

data: 9 and 24
number of successes = 9, number of trials = 24, p-value = 0.01176
alternative hypothesis: true probability of success is not equal to 0.1666667
95 percent confidence interval:
0.1879929 0.5940636
sample estimates:
probability of success
0.375

Upon reviewing the output, the calculated P-value is **0.01176**. Since this value is considerably less than the conventional level of significance, $alpha = 0.05$, we have sufficient evidence to reject the Null hypothesis (H_0). We conclude that the die does **not** land on the number "3" during $1/6$ of the rolls. The observed sample estimate (0.375) is much higher than the expected probability (0.167), suggesting that the die is biased toward rolling a "3," leading to a statistically significant outcome.

Case Study 2: Exploring One-Tailed Tests

One-tailed tests are employed when the research question suggests a specific direction for the deviation from the hypothesized proportion. This allows for increased statistical power in detecting effects that occur only in that predicted direction. We will examine both the left-tailed (less than) and the right-tailed (greater than) applications of the `binom.test()` function.

Left-Tailed Test (Testing for Lower Probability)

Imagine testing whether a coin is inherently biased toward landing on tails, meaning the probability of landing on heads (pi) is less than the expected 0.5. We flip the coin 30 times and observe 11 heads. The Null hypothesis is $H_0: pi = 0.5$, while the Alternative hypothesis is $H_A: pi < 0.5$. Since we are testing for a proportion that is less than the hypothesized value, we set the `alternative` argument to "less."

#perform left-tailed Binomial test
binom.test(11, 30, 0.5, alternative="less")

#output
Exact binomial test

data: 11 and 30
number of successes = 11, number of trials = 30, p-value = 0.1002
alternative hypothesis: true probability of success is less than 0.5
95 percent confidence interval:
0.0000000 0.5330863
sample estimates:
probability of success
0.3666667

The resulting P-value is **0.1002**. Because this P-value is greater than 0.05, we fail to reject the Null hypothesis. Although the sample proportion (0.367) is lower than 0.5, the difference is not statistically large enough to conclude, with 95% confidence, that the coin is genuinely biased against landing on heads. We do not have sufficient evidence to support the claim that the true probability of heads is less than 0.5.

Right-Tailed Test (Testing for Higher Probability)

Consider a manufacturing scenario where widgets historically have an effectiveness rate of 80% ($p=0.8$). A company implements a new production system intended to boost this rate. From a batch produced under the new system, 50 widgets are randomly sampled, and 46 are found to be effective. The question is whether this new observed rate (46/50 = 92%) provides sufficient evidence to conclude that the new system is superior. We set up the hypotheses as $H_0: pi = 0.8$ and $H_A: pi > 0.8$. We use the `alternative="greater"` option for this right-tailed test.

#perform right-tailed Binomial test
binom.test(46, 50, 0.8, alternative="greater")

```
#output
Exact binomial test

data: 46 and 50
number of successes = 46, number of trials = 50, p-value = 0.0185
alternative hypothesis: true probability of success is greater than 0.8
95 percent confidence interval:
0.8262088 1.0000000
sample estimates:
probability of success
0.92
```

In this instance, the P-value is **0.0185**. Since 0.0185 is less than the critical threshold of 0.05, we reject the Null hypothesis. This statistical result provides strong evidence that the new system has successfully increased the widget effectiveness rate beyond the initial 80%. The sample success rate of 92% is statistically significant, validating the effectiveness of the new manufacturing procedures.

Interpreting R Output and Statistical Significance

Interpreting the output generated by `binom.test()` requires attention to several key metrics beyond just the P-value. The output provides the specific data used (number of successes and trials), the full hypothesis being tested (in the "alternative hypothesis" line), the sample estimate, and, critically, the confidence interval. Together, these elements paint a complete picture of the statistical inference drawn from the sample.

The core decision criterion remains the P-value. The P-value represents the probability of observing the sampled data (or data more extreme) if the Null hypothesis were perfectly true. If this probability is very small--conventionally less than $alpha=0.05$--it suggests that the observed data is highly unlikely under the assumption of the Null, leading us to reject H_0 in favor of H_A. Conversely, a large P-value indicates that the observed sample data is consistent with the Null hypothesis, leading to a failure to reject H_0. Note that failing to reject the Null hypothesis does not prove its truth; it merely means there is insufficient evidence to conclude the Alternative hypothesis is true.

Furthermore, the output includes the **confidence interval**, typically set at 95%. This interval provides a range of plausible values for the true population proportion (pi). In the case of a two-tailed test, if the hypothesized proportion (p) falls outside the 95% confidence interval, the result is statistically significant at the 0.05 level, reinforcing the rejection of H_0. For one-tailed tests, the confidence interval displayed is adjusted, reflecting the directional nature of the test (e.g., the

right-tailed test above shows a lower bound greater than 0.826 and an upper bound of 1.000, clearly excluding the null value of 0.8). By systematically examining the P-value, the confidence interval, and the sample estimate, researchers can confidently communicate the results of their binomial analysis.