

How to Perform McNemar's Test in Stata: A Step-by-Step Guide

Authored by
stats writer

March 8, 2026

RECOMMENDED CITATION

stats writer (2026). *How to Perform McNemar's Test in Stata: A Step-by-Step Guide*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=134676>

Understanding the Statistical Foundations of McNemar's Test

In the realm of biostatistics and social sciences, researchers frequently encounter scenarios where they must compare proportions across two related groups. **McNemar's test** is a non-parametric statistical method specifically designed for this purpose, particularly when dealing with paired data. Unlike the standard **Chi-squared test** for independence, which assumes that observations in each cell are independent of one another, this test accounts for the inherent dependency found in repeated measures or matched-pair designs. This makes it an essential tool for longitudinal studies where the same subjects are assessed before and after an intervention, or in case-control studies where subjects are matched based on specific demographic or clinical criteria.

The primary utility of this test lies in its ability to detect a change in a **binary outcome**--such as "pass/fail," "support/oppose," or "diseased/healthy"--over two distinct time points or conditions. When we analyze paired data, we are not simply looking at the total number of successes in each group; rather, we are focused on the "discordant pairs." These are the individuals who changed their status from the first observation to the second. By focusing on these transitions, the test provides a robust measure of whether the shift in one direction significantly outweighs the shift in the other, thereby indicating a statistically significant effect of the intervention or time variable.

To perform this analysis accurately within a software environment like **Stata**, it is crucial to understand that the data must be structured in a 2x2 **contingency table**. This table captures four types of pairs: those who remained positive in both instances, those who remained negative, and those who switched from positive to negative or vice versa. The statistical logic assumes that if there is no treatment effect, the number of individuals switching from "A to B" should be roughly equal to the number of individuals switching from "B to A." Any significant deviation from this balance suggests that the underlying proportions have shifted, leading us to reject the **null hypothesis**.

Theoretical Framework and the Null Hypothesis

The theoretical framework of **McNemar's test** is rooted in the concept of marginal homogeneity. In a 2x2 table, marginal homogeneity implies that the row totals are equal to the column totals, which signifies that the overall proportion of the outcome has not changed between the two measurement periods. The **null hypothesis** (H0) for this test states that the marginal probabilities for each outcome are the same, meaning the probability of a subject being in a specific category at time point one is equal to the probability of being in that same category at time point two. Conversely, the alternative hypothesis (H1) suggests that the marginal probabilities are different, indicating a shift in the population's response over time.

Mathematically, the test focuses on the off-diagonal elements of the 2x2 contingency table. If we

label the cells as A (both positive), B (first positive, second negative), C (first negative, second positive), and D (both negative), the test statistic is derived solely from cells B and C. These are the discordant cells that represent change. The **Chi-squared distribution** is used to determine if the difference between B and C is larger than what would be expected by random chance. Because the test relies on these specific transitions, it is often more powerful than other tests when the goal is to identify changes within subjects rather than differences between independent groups.

When applying this test, the researcher must ensure that the data meets specific assumptions. First, the **binary outcome** variable must be mutually exclusive and exhaustive. Second, the pairs must be independent of each other; that is, the response of one pair should not influence the response of another. Finally, the sample size must be sufficient for the **Chi-squared distribution** approximation to be valid. In cases where the number of discordant pairs is very small (typically less than 25), statisticians often recommend using the binomial distribution or applying an **continuity correction** to maintain the test's validity and prevent Type I errors.

Navigating the Stata Environment for Paired Proportions

Stata offers a streamlined approach to conducting **McNemar's test** through its built-in suite of epidemiological and statistical commands. Depending on how your data is stored, you will likely use one of two primary commands: **mcc** or **mcci**. The **mcc** command is utilized when you have a raw dataset where each observation represents a pair and variables indicate the status at each time point. This is the standard procedure for researchers working with large datasets where manual counting is impractical. The command automatically generates a 2x2 table and calculates the relevant statistics, including the **odds ratio** and confidence intervals.

For scenarios where you already have the summary counts from a published table or manual tally, **Stata** provides the "immediate" version of the command, known as **mcci**. The "i" at the end of the command stands for "immediate," allowing you to input the four cell values directly into the command line without needing an active dataset in memory. This is particularly useful for quick calculations or for verifying results found in existing literature. Both commands provide the same rigorous output, ensuring that the researcher has access to the **Chi-squared distribution** statistic and the **p-value** necessary for hypothesis testing.

When executing these commands, **Stata** assumes a specific order for the inputs. In the **mcci** command, the values must be entered in the order of the cells in a standard 2x2 table: A, B, C, and D. Specifically, these represent:

Cell A: Subjects who were positive at both time points (concordant).

Cell B: Subjects who were positive at time one but negative at time two (discordant).

Cell C: Subjects who were negative at time one but positive at time two (discordant).

Cell D: Subjects who were negative at both time points (concordant).

Correctly identifying these values is paramount, as the calculation of the **Chi-squared distribution** statistic relies entirely on the relationship between B and C.

A Practical Case Study: Analyzing Marketing Impact

To illustrate the application of this test, let us consider a practical example involving marketing research and public opinion. Suppose a team of researchers wants to evaluate whether a specific marketing video is effective at shifting public support for a newly proposed law. To test this, they design a pre-post study involving 100 participants. Each participant is surveyed twice: once before watching the video and once after. The **binary outcome** is "Support" or "Do Not Support." This design is a classic example of paired data because the same 100 individuals are providing two data points each, making their responses dependent.

The results of the survey are compiled into the following **contingency table**, which summarizes the shifts in opinion among the participants:

Before Marketing Video		
After Marketing Video	Support	Do not support
Support	30	40
Do not Support	12	18

In this dataset, we can see that 30 people supported the law both before and after the video, while 18 people remained consistently opposed. The interesting data points--the discordant pairs--are the 40 people who did not support the law initially but changed their minds after the video, and the 12 people who initially supported the law but withdrew their support after viewing the video. Because 40 is substantially different from 12, we might intuitively suspect that the video had a significant impact, but we must use **McNemar's test** to confirm this statistically.

By performing the test, we move beyond anecdotal observation to statistical **significance**. We are testing whether the 40-to-12 split is far enough from a 50/50 split to reject the **null hypothesis**. If the video had no effect, we would expect roughly equal numbers of people to change their minds in both directions. The fact that many more people shifted toward support than away from it suggests a systematic influence from the marketing material, which **McNemar's test** will quantify using the **Chi-squared distribution**.

Executing the Test Using the mcci Command

To analyze the marketing data in **Stata**, we use the **mcci** command. This immediate command allows us to input the raw counts from our table directly into the terminal. We enter the values following the top-left to bottom-right convention. In our case, the counts are 30 (Support/Support),

40 (Support After/Do Not Support Before), 12 (Do Not Support After/Support Before), and 18 (Do Not Support/Do Not Support). Note that the command structure in **Stata** expects the rows to represent the "cases" or the second measurement and columns to represent the "controls" or first measurement, though the symmetry of the formula for the test statistic makes the labels interchangeable as long as the discordant cells are correctly placed.

The syntax for this operation is as follows:

mcci 30 40 12 18

```
. mcci 30 40 12 18
```

Cases	Controls		Total
	Exposed	Unexposed	
Exposed	30	40	70
Unexposed	12	18	30
Total	42	58	100

```
McNemar's chi2(1) = 15.08 Prob > chi2 = 0.0001
Exact McNemar significance probability = 0.0001
```

Proportion with factor

Cases	.7			
Controls	.42			[95% Conf. Interval]
difference	.28	.1397545	.4202455	
ratio	1.666667	1.284237	2.162979	
rel. diff.	.4827586	.3075044	.6580128	
odds ratio	3.333333	1.71442	6.97982	(exact)

Once the command is executed, **Stata** generates a comprehensive output window. This output includes a reconstructed 2x2 table, marginal totals, and the percentages for each cell. Most importantly, the bottom of the output contains the **Chi-squared** statistic and the **p-value**. The output also typically includes an **odds ratio** with its corresponding confidence interval, which provides a measure of the effect size, indicating how much more likely a person is to support the law after the video compared to before.

Reviewing the output carefully is essential for accurate reporting. **Stata** will provide both the standard **McNemar's test** result and often an exact version based on the binomial distribution if the counts are low. For our marketing example, the software calculates the **Chi-squared** statistic as 15.08. This value is derived using the squares of the differences between the discordant pairs,

divided by their sum. A higher value here indicates a greater discrepancy between the two directions of change, leading to a smaller **p-value** and higher statistical **significance**.

Detailed Interpretation of Stata Output

Interpreting the **Stata** output requires focusing on three key metrics: the **Chi-squared** value, the **degrees of freedom**, and the **p-value**. In the provided example, the **Chi-squared** value is 15.08 with 1 degree of freedom. This **degree of freedom** is constant for all 2x2 tables in this test because the test is essentially evaluating a single comparison between two proportions. The formula used by **Stata** is $(B-C)^2 / (B+C)$. For our data, this calculation looks like: $(40-12)^2 / (40+12) = 784 / 52 = 15.0769$, which rounds to 15.08.

The **p-value** associated with this test is 0.0001. In the context of hypothesis testing, the **p-value** represents the probability of observing a difference as large as the one found in our sample if the **null hypothesis** were true. Since 0.0001 is significantly lower than the standard alpha level of 0.05, we have strong evidence to reject the **null hypothesis**. This leads to the conclusion that the change in the proportion of people supporting the law after watching the marketing video is statistically significant and unlikely to have occurred by random chance.

Additionally, **Stata** provides information on the **odds ratio**. In this context, the **odds ratio** is calculated as B/C ($40/12 \approx 3.33$). This tells us that the odds of switching from "Do Not Support" to "Support" are more than three times higher than the odds of switching in the opposite direction. Such measures of effect size are invaluable for understanding the practical implications of the results, going beyond a simple "yes/no" answer regarding significance to show the magnitude of the video's influence.

The Role of Continuity Corrections in Small Samples

While the standard **McNemar's test** is effective for large samples, statisticians often raise concerns about its accuracy when the number of discordant pairs is small. In such instances, the **Chi-squared distribution**, which is a continuous distribution, may not perfectly approximate the discrete nature of the counts in the table. To address this, many researchers apply the **continuity correction** (often called the Edwards correction). This involves subtracting 0.5 or 1 from the absolute difference between the discordant pairs before squaring it, which yields a more conservative **p-value** and reduces the risk of false positives.

The modified formula with the **continuity correction** is $(|B-C| - 1)^2 / (B+C)$. Applying this to our marketing example, the calculation would be $(|40-12| - 1)^2 / (40+12) = 27^2 / 52 = 729 / 52 = 14.019$. While the resulting **Chi-squared** value is slightly lower than the uncorrected 15.08, the **p-value** would still remain well below the 0.05 threshold. In practice, most statisticians recommend using this correction if any of the discordant cell counts (B or C) are less than 5, or if the total number of

discordant pairs (B+C) is less than 25.

When reporting results, it is considered best practice to specify whether a **continuity correction** was applied. **Stata** helps in this regard by providing multiple outputs, including the exact binomial test, which is the gold standard for very small samples. By choosing the most appropriate version of the test for your specific data distribution, you ensure the integrity of your statistical conclusions and adhere to the rigorous standards of **academic writing** and research.

Assumptions and Best Practices for Implementation

To ensure the validity of **McNemar's test**, researchers must strictly adhere to several fundamental assumptions. The most critical is that the data must be paired. This means that for every "before" measurement, there must be a corresponding "after" measurement from the same subject or a matched counterpart. If the groups were independent--for example, if you surveyed 100 people in one city and a different 100 people in another--this test would be inappropriate, and a standard **Chi-squared test** or a two-proportion z-test should be used instead.

Another important consideration is the nature of the variables. The test is designed for **binary data**. If your outcome variable has more than two categories (e.g., "Support," "Oppose," and "Neutral"), you should use the Stuart-Maxwell test, which is a generalized version of **McNemar's test** for larger tables. Furthermore, while the test is robust, it does not account for **confounding** variables. If you suspect that other factors (like age or income) are influencing the change in support, you might need to employ more complex models, such as **conditional logistic regression**.

In summary, performing **McNemar's test** in **Stata** is a powerful way to analyze changes in proportions within paired samples. By utilizing the **mcci** command, researchers can quickly determine the **statistical significance** of their findings and provide clear, quantifiable evidence for their hypotheses. Whether you are evaluating a marketing campaign, a medical treatment, or a policy change, understanding the mechanics and interpretation of this test is a vital skill for any data-driven professional. Always remember to consult with a **statistician** or reference official **Stata documentation** when dealing with complex datasets to ensure the highest level of accuracy in your analysis.