

How to Report a Mann-Whitney U Test: A Step-by-Step Guide with Example

Authored by
mohammed looti

January 8, 2026

RECOMMENDED CITATION

mohammed looti (2026). *How to Report a Mann-Whitney U Test: A Step-by-Step Guide with Example*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=124999>

The Mann-Whitney U test (often abbreviated as the M-W U test or Wilcoxon rank-sum test) is a crucial non-parametric statistical test used extensively across fields such as psychology, medicine, and engineering. It serves the fundamental purpose of determining whether there is a statistically significant difference in the distributions of a continuous dependent variable between two independent samples. Unlike its parametric counterpart, the independent samples t-test, the M-W U test does not require the assumption that the data is normally distributed, nor does it require homogeneity of variances. This robustness makes it the preferred tool when dealing with ordinal data or small sample sizes where distributional assumptions cannot be reliably met.

When generating a report based on the results of a Mann-Whitney U test, clarity and accuracy are paramount. The reporting standards typically require the inclusion of several key statistics: the sample size for each group (n_1 and n_2), the calculated U statistic, the associated standardized test statistic (often the Z-value), and the resulting p-value. A comprehensive report should contextualize these figures within the research question, explaining whether the observed difference between the groups is statistically significant based on a predetermined alpha level, usually $\alpha = 0.05$. For instance, reporting the finding concisely might look like: "A Mann-Whitney U test revealed a significant difference in test scores between the control and experimental groups ($U = 75$, $n_1 = 15$, $n_2 = 15$, $p < 0.05$)." This structured reporting ensures readers can quickly grasp the findings and evaluate the statistical conclusion.

The Mann-Whitney U test operates by ranking all observations from both groups together, calculating the sum of ranks for each group, and subsequently determining the U statistic based on these rank sums. This approach allows the test to focus on the median difference between the groups, rather than the mean difference, thereby mitigating the influence of outliers that often skew results in parametric tests. It is particularly useful for comparing the differences between two independent samples when the underlying distributions are non-normal, or when the sample sizes are relatively small, often cited as $n < 30$ per group, although its application is not strictly limited by this threshold if non-normality persists.

Prerequisites and Assumptions for Non-Parametric Testing

Before deciding to run a Mann-Whitney U test, researchers must first verify that their data meets several fundamental prerequisites. The most critical condition is the existence of two independent samples; that is, the observations in one group must not influence the observations in the other group. This independence is essential for the test's validity. Furthermore, the dependent variable must be measured on at least an ordinal scale, meaning the data can be meaningfully ranked. If the data is continuous (interval or ratio), it is automatically suitable for the ranking procedure inherent in the M-W U test.

The primary reason for selecting this non-parametric statistical test over an independent samples t-

test often revolves around the assumption of normality. If diagnostic tests (such as the Shapiro-Wilk test or Kolmogorov-Smirnov test) indicate that the dependent variable's scores are not normally distributed within one or both groups, or if the sample size is insufficient to rely on the Central Limit Theorem, the M-W U test provides a reliable alternative. It is important to note, however, that while the M-W U test does not assume normality, it does assume that the shapes of the distributions for the two groups are similar if the goal is to interpret the results as a difference in medians. If the shapes are dissimilar, the test is still valid, but the interpretation shifts to a comparison of stochastic dominance--that a randomly selected score from one group is likely to be higher than a randomly selected score from the second group.

A rigorous reporting process begins with clearly stating the rationale for choosing the Mann-Whitney U test. This involves documenting the failure of the normality assumption or explaining the nature of the data (e.g., ordinal scores, small sample size). By establishing this justification early, the researcher enhances the methodological rigor of the study. After confirming suitability, the researcher moves to the analytical phase, calculating the ranks and the U statistic, which forms the numerical backbone of the final report. This careful preliminary work ensures that the conclusions drawn from the test are appropriate for the data structure being analyzed.

Core Elements Required in the Statistical Report

When preparing documentation for academic publication or internal reports, specific statistical values must be included to ensure full transparency and replicability. The reporting of a Mann-Whitney U test should always encompass the descriptive statistics, such as the median and interquartile range (IQR) for both groups, as these non-parametric measures accurately describe the central tendency and spread of the non-normally distributed data. Reporting the means and standard deviations is usually inappropriate in this context.

The fundamental required statistical metrics are the U statistic, the sample sizes (n_1 and n_2), and the corresponding probability value (p-value). In many statistical packages, particularly when sample sizes are large (typically $n > 20$ for each group), the software provides an approximation of the distribution using the standard normal distribution, yielding a Z-value. This standardized Z-value is often reported instead of or in addition to the raw U statistic because it simplifies the comparison against critical values and provides a more standardized measure of effect direction and magnitude. The structure of the report hinges on presenting these numerical outcomes alongside a clear verbal explanation of their meaning relative to the null hypothesis.

Finally, a complete statistical report requires an assessment of effect size. While the M-W U test determines significance, it does not inherently quantify the practical importance of the difference. Common effect size measures used in conjunction with the M-W U test include Pearson's r (calculated from the Z statistic and total sample size) or Glass's rank biserial correlation coefficient.

Including the effect size allows the audience to gauge not just whether a difference exists, but how large that difference is in practical terms, moving beyond mere statistical significance.

Establishing the General Reporting Structure

Regardless of the specific research context, the successful reporting of a Mann-Whitney U test must follow a standardized narrative structure, typically adhering to APA (American Psychological Association) guidelines in academic contexts. This structure ensures that all necessary methodological and statistical details are presented logically and concisely. This process begins by explicitly stating the purpose of the analysis and identifying the variables involved.

When reporting the results of a Mann-Whitney U test, we always use the following general structure:

A clear statement defining the independent variable (the groups being compared) and the dependent variable (the outcome measure).

Presentation of the descriptive statistics, specifically the median and IQR, for both groups.

The calculated test statistics, including the overall Z-value of the test and the corresponding p-value, often accompanied by the degrees of freedom (or sample sizes).

A concluding statement interpreting the statistical significance based on the comparison of the p-value to the chosen alpha level (α).

The use of concise and formal language is critical. The following template illustrates the structure often employed for the statistical statement within the results section of a technical paper:

A Mann-Whitney U test was performed to compare in (Median = , IQR =) and (Median = , IQR =).

There a significant difference in between and ; $U =$, $Z =$, $p =$. .

This template ensures that all essential components--the test type, the variables, descriptive statistics, and inferential results--are captured in a single, coherent narrative. The next section will apply this structure to a practical research example.

Detailed Research Example: Fuel Efficiency Comparison

Consider a scenario where automotive researchers aim to investigate the effectiveness of a novel fuel additive. They hypothesize that this new fuel blend might influence car performance, specifically measured by miles per gallon (MPG). Due to logistical constraints and cost, the researchers are limited to a small pilot study, which immediately suggests that the assumption of a normally distributed MPG outcome might be violated, or at least unreliable given the sample size.

To execute the experiment, they recruit two groups of vehicles: a control group consisting of 12 cars running on standard fuel, and an experimental group comprising 12 cars utilizing the new fuel blend. The researchers meticulously measure the MPG for all 24 cars over a standardized testing period. Since the combined sample size ($N = 24$) is relatively small, and there is concern about the normality of fuel efficiency data (which can often be skewed by various unmeasured factors), they appropriately select the Mann-Whitney U test to compare the central tendencies of the two independent groups.

The primary objective is to determine if the distribution of MPG scores in the new fuel group stochastically dominates the distribution in the standard fuel group, thereby indicating a practical difference in efficiency. The null hypothesis (H_0) states that there is no difference in MPG between the two fuel types, while the alternative hypothesis (H_A) suggests that the new fuel does indeed lead to a difference in MPG. The following steps detail how the analysis is conducted and interpreted using standard statistical software.

Analyzing Statistical Software Output (SPSS)

To determine if there is a significant difference between the two groups, the researchers perform a Mann-Whitney U test using statistical software like SPSS. The software processes the raw MPG data, ranks the scores, calculates the U statistic, and then uses an approximation (especially relevant for samples approaching $N \geq 20$) to generate the Z-value and the corresponding two-tailed p-value. The resulting output table is critical for obtaining the necessary reporting figures.

The output often presents several statistics, but the most important figures for determining significance are the standardized test statistic and the associated probability level. The image below represents a typical output obtained from such an analysis:

➔ NPar Tests

Mann-Whitney Test

Ranks				
	group	N	Mean Rank	Sum of Ranks
mpg	.00	12	10.67	128.00
	1.00	12	14.33	172.00
	Total	24		

Test Statistics^a

	mpg
Mann-Whitney U	50.000
Wilcoxon W	128.000
Z	-1.279
Asymp. Sig. (2-tailed)	.201
Exact Sig. [2*(1-tailed Sig.)]	.219 ^b

a. Grouping Variable: group

b. Not corrected for ties.

From this output, we extract two pivotal numerical results: the standardized **Z test statistic** and the **Asymptotic 2-tailed p-value**, which measures the probability of observing data as extreme as, or more extreme than, the current results, assuming the null hypothesis is true.

Z test statistic: -1.279

p-value (Asymptotic Significance): .201

The critical step involves comparing the calculated p-value (.201) to the predefined significance level ($\alpha = 0.05$). Since the p-value (.201) is considerably greater than 0.05, we fail to reject the null hypothesis. We conclude that there is not sufficient statistical evidence, based on these small samples, to assert that the true distribution of MPG is different between the cars receiving the new fuel and those receiving regular fuel. The observed difference is likely attributable to random sampling variability.

Writing the Final Report Narrative (APA Style)

Once the decision regarding the null hypothesis has been made, the results must be translated into a formal, readable narrative. This written report must clearly link the statistical output to the

research question and must include all necessary statistics. Assuming that the descriptive statistics showed, for example, that the standard fuel group had a Median MPG of 35.0 (IQR = 4.5) and the new fuel group had a Median MPG of 37.0 (IQR = 5.0), the narrative is constructed as follows:

Here is how we would report the results of this Mann-Whitney U test, ensuring all data points are properly contextualized:

A Mann-Whitney U test was performed to compare the distribution of average miles per gallon (MPG) between cars that received a new fuel ($n = 12$, $Mdn = 37.0$) and cars that received regular fuel ($n = 12$, $Mdn = 35.0$).

The test indicated that there was not a statistically significant difference in average miles per gallon between the two independent groups; $Z = -1.279$, $p = .201$. Therefore, the null hypothesis of equal distributions was retained. The small observed difference in median MPG between the new fuel and regular fuel groups did not reach the threshold for statistical significance at the $\alpha = 0.05$ level.

Refining the Interpretation and Effect Size

While the core result--the decision to retain or reject the null hypothesis--is primary, effective reporting goes further by providing context and magnitude. Since the outcome was non-significant ($p > 0.05$), the researcher should emphasize that the data does not support the claim that the new fuel is effective, rather than stating definitively that the new fuel has no effect. Statistical non-significance simply means the evidence is insufficient to reject the null hypothesis.

To enhance the detail of the report, calculating and including the effect size is highly recommended. For the M-W U test, the effect size r can be derived from the Z statistic and the total sample size (N) using the formula: $r = Z / \sqrt{N}$. In this example, $N = 24$. Thus, $r = -1.279 / \sqrt{24} \approx -1.279 / 4.899 \approx -0.26$. This effect size ($r = -0.26$) is typically considered a medium effect, suggesting that although the finding was not statistically significant, there might be a moderate tendency for the distributions to differ, warranting further investigation with a larger sample size.

Important Note on Reporting Precision: As a standard rule in scientific reporting, particularly in social sciences, p-values are typically rounded to three decimal places (e.g., $p = .201$). However, if a p-value is extremely small (e.g., $p = 0.00001$), it is generally reported as $p < .001$. Consistency in rounding and adherence to the required style guide (e.g., APA, MLA, specific journal guidelines) are crucial elements of professional statistical reporting. Always verify the industry or journal standard practice for rounding statistics.

Summary of Best Practices for Reporting

To summarize the best practices for documenting the Mann-Whitney U test results, researchers should prioritize four key areas: transparency, specificity, context, and completeness. Transparency involves clearly stating the non-parametric choice due to violations of parametric assumptions or data type. Specificity demands accurate reporting of descriptive statistics (medians/IQRs) and the inferential statistics (U , Z , p).

Context requires interpreting the results back into the practical research setting, explaining what the acceptance or rejection of the null hypothesis means for the original research question regarding the fuel efficiency. Finally, completeness mandates the inclusion of effect size measures, which prevent the reader from mistaking statistical non-significance for practical non-existence. Adhering to these guidelines ensures the M-W U test report is both technically sound and easily understood by a diverse audience.

The following tutorials provide additional information about the Mann-Whitney U test and related non-parametric methods: