Which Statistical Test to Use (and When)

A Practical Guide for Students

www.studentstats.net

1. Introduction

Selecting the appropriate statistical test is one of the most important decisions you will make in any research project. Using the wrong test can lead to invalid conclusions, wasted effort, and results that do not hold up under scrutiny. Whether you are working on a university dissertation, a lab report, or an independent research paper, the test you choose must align with your research question, the type of data you have collected, and the assumptions your data can satisfy.

Many students feel overwhelmed when confronted with the long list of available tests. The good news is that the decision can be broken down into a handful of straightforward questions: What kind of variables are you working with? How many groups are you comparing? Does your data meet certain distributional requirements? Once you answer these questions, the correct test almost always becomes clear.

What This Guide Covers

This guide walks you through the fundamental concepts behind statistical testing, provides a decision table you can use to identify the right test quickly, and then explains each of the most commonly used tests in plain language. We also include a quick-reference summary table and a section on common mistakes to avoid. By the end, you should feel confident about choosing and justifying a statistical test for your own work.

2. Key Concepts Before You Choose

Types of Variables

Understanding the measurement level of your variables is the first step in choosing a test. There are four standard levels:

- **Nominal** - Categories with no natural order. Examples: gender, blood type, country of birth.
- **Ordinal** - Categories with a meaningful order but unequal intervals. Examples: satisfaction ratings (low / medium / high), education level.
- **Interval** - Numeric values with equal intervals but no true zero point. Examples: temperature in Celsius, calendar year.
- **Ratio** - Numeric values with equal intervals and a meaningful zero. Examples: weight, height, income, reaction time.

Interval and ratio variables are often grouped together as **continuous** (or scale) variables. Most parametric tests require at least interval-level measurement.

Dependent vs. Independent Variables

The **independent variable** (IV) is the factor you manipulate or use to define groups. The **dependent variable** (DV) is the outcome you measure. For example, if you are testing whether a new teaching method improves exam scores, the teaching method is the IV and the exam score is the DV. Correctly identifying these roles determines which statistical test fits your design.

Parametric vs. Nonparametric Tests

Parametric tests assume that your data follow a specific distribution (usually normal) and that variances across groups are roughly equal. They tend to be more powerful, meaning they are better at detecting a real effect when one exists. Common parametric tests include the t-test, ANOVA, and Pearson correlation.

Nonparametric tests make fewer assumptions about the shape of your data. They work with ranks rather than raw values, making them suitable for ordinal data or for continuous data that violate normality assumptions. Common nonparametric tests include Mann-Whitney U, Kruskal-Wallis, Wilcoxon signed-rank, and Spearman correlation.

A useful rule of thumb: if your sample is large (roughly $n > 30$ per group) and the data are not heavily skewed, parametric tests are usually appropriate. For small or skewed samples, consider a nonparametric alternative.

What Is a p-Value?

A p-value tells you the probability of obtaining results at least as extreme as yours if there were truly no effect or no difference in the population. A small p-value (commonly below 0.05) suggests that the observed result is unlikely to have occurred by chance alone, leading you to reject the null hypothesis.

Important: a p-value does **not** tell you the size or practical importance of an effect. Always pair your p-value with an effect size measure (such as Cohen's d or r) and consider whether the result is meaningful in the real world, not just statistically significant.

3. Choosing the Right Test: Decision Table

Use the table below to identify the recommended test based on your research goal, the types of variables involved, and your sample characteristics. The 'Parametric' column shows the test to use when standard assumptions are met; the 'Nonparametric' column shows the alternative when they are not.

Your Goal	IV / Groups	DV Type	Parametric Test	Nonparametric Alternative
Compare means of 2 independent groups	1 IV, 2 groups	Continuous	Independent samples t-test	Mann-Whitney U
Compare means of 3+ independent groups	1 IV, 3+ groups	Continuous	One-way ANOVA	Kruskal-Wallis
Compare before and after (paired)	1 IV, 2 related measures	Continuous	Paired samples t-test	Wilcoxon signed-rank
Test association between two categorical variables	2 categorical variables	Nominal / Ordinal	Chi-square test of independence	Fisher's exact test (small n)
Measure linear relationship between 2 continuous variables	2 continuous variables	Continuous	Pearson correlation (r)	Spearman correlation (rho)
Predict a continuous outcome from one or more predictors	1+ IVs	Continuous DV	Linear regression	Nonparametric regression
Predict a binary outcome (yes/no) from predictors	1+ IVs	Binary DV	Logistic regression	---

Tip: If you are unsure whether your data meet parametric assumptions, run a Shapiro-Wilk test for normality and Levene's test for equality of variances before choosing your primary analysis.

4. Common Tests Explained

Below you will find a plain-language explanation of each frequently used test, including what it does, when to use it, the key assumptions, and a concrete example scenario.

Independent Samples t-Test

What it does: Compares the means of a continuous variable between two separate (unrelated) groups to determine whether the difference is statistically significant.

When to use it: You have one categorical independent variable with exactly two levels and one continuous dependent variable. The participants in each group are different people.

Key assumptions: The dependent variable is approximately normally distributed within each group. The variances of the two groups are roughly equal (homogeneity of variance). Observations are independent of one another.

Example: A researcher wants to know whether students who study with background music score differently on a memory test than students who study in silence. Group A (n=35) studies with music; Group B (n=35) studies without. An independent t-test compares the mean test scores of the two groups.

Paired Samples t-Test

What it does: Compares the means of a continuous variable measured at two time points or under two conditions for the same participants.

When to use it: You have repeated measures on the same individuals (e.g., pre-test and post-test) or matched pairs, and the outcome variable is continuous.

Key assumptions: The differences between paired observations are approximately normally distributed. Each pair of observations is independent of other pairs.

Example: A psychologist measures anxiety levels in 40 patients before and after an 8-week therapy program. A paired t-test determines whether the mean anxiety score changed significantly from pre-treatment to post-treatment.

One-Way ANOVA

What it does: Tests whether the means of a continuous variable differ across three or more independent groups. It is an extension of the independent t-test to more than two groups.

When to use it: You have one categorical independent variable with three or more levels and a continuous dependent variable.

Key assumptions: The dependent variable is approximately normally distributed in each group. Variances are approximately equal across groups (Levene's test). Observations are independent.

Example: A nutritionist compares weight loss across three different diets (low-carb, Mediterranean, and intermittent fasting) after 12 weeks. One-way ANOVA tests whether at least one diet produces a significantly different mean weight loss. If significant, a post-hoc test (e.g., Tukey HSD) identifies which pairs differ.

Chi-Square Test of Independence

What it does: Assesses whether two categorical variables are associated or independent of each other. It compares observed frequencies in a contingency table to the frequencies expected if no relationship existed.

When to use it: Both variables are categorical (nominal or ordinal) and you want to know if there is a statistically significant association between them.

Key assumptions: Observations are independent. Expected frequencies in each cell of the contingency table should generally be 5 or more. For smaller expected counts, consider Fisher's exact test instead.

Example: A sociologist surveys 200 people and records their preferred news source (TV, online, newspaper) and age group (18-34, 35-54, 55+). A chi-square test reveals whether news preference is associated with age group.

Mann-Whitney U Test

What it does: A nonparametric alternative to the independent t-test. It compares the distributions of a continuous or ordinal variable between two independent groups using ranks rather than raw means.

When to use it: You want to compare two independent groups but your data are ordinal, or your continuous data violate the normality assumption.

Key assumptions: Observations are independent. The two distributions have the same shape (the test is sensitive to differences in central tendency, not shape, when this holds).

Example: A marketing team asks 50 customers and 50 non-customers to rate brand trust on a 7-point Likert scale. Because Likert data are ordinal, a Mann-Whitney U test is used to determine whether the two groups differ in their trust ratings.

Wilcoxon Signed-Rank Test

What it does: A nonparametric alternative to the paired t-test. It compares two related measurements by analyzing the signed ranks of the differences.

When to use it: You have paired or repeated-measures data that are ordinal, or continuous data where the distribution of differences is not normal.

Key assumptions: The differences between pairs are independent and come from a symmetric distribution (though the test is fairly robust to moderate violations of symmetry).

Example: A teacher records student confidence ratings (1-10 scale) before and after a public speaking workshop. Because the scale is ordinal, a Wilcoxon signed-rank test checks whether confidence changed significantly.

Pearson Correlation (r)

What it does: Measures the strength and direction of the linear relationship between two continuous variables. The coefficient r ranges from -1 (perfect negative) to +1 (perfect positive), with 0 indicating no linear association.

When to use it: Both variables are continuous (interval or ratio) and you expect a linear relationship.

Key assumptions: Both variables are approximately normally distributed (bivariate normality). The relationship is linear. There are no extreme outliers distorting the result.

Example: An economist examines whether hours of weekly study are linearly related to GPA in a sample of 120 university students. A Pearson correlation of $r = 0.58$ would indicate a moderately strong positive relationship.

Spearman Correlation (rho)

What it does: A nonparametric measure of the strength and direction of the monotonic relationship between two variables. It works on ranks, making it suitable for ordinal data or continuous data that are not normally distributed.

When to use it: At least one variable is ordinal, the relationship is monotonic but not necessarily linear, or the normality assumption for Pearson is violated.

Key assumptions: The relationship between the variables is monotonic (consistently increasing or decreasing, though not necessarily at a constant rate). Observations are independent.

Example: A researcher ranks 30 job applicants on interview performance and separately on years of experience. Spearman correlation tests whether higher experience ranks tend to go with higher performance ranks.

5. Quick Reference Table

The table below summarizes every test discussed in this guide. Keep it handy as a one-page cheat sheet when planning your analysis.

Test	Type	Purpose	Key Assumption
Independent t-test	Parametric	Compare means of 2 independent groups	Normal distribution, equal variances
Paired t-test	Parametric	Compare means of 2 related measures	Normal distribution of differences
One-way ANOVA	Parametric	Compare means of 3+ independent groups	Normal distribution, equal variances
Chi-square	Nonparametric	Test association between 2 categorical variables	Expected cell counts >= 5
Mann- Whitney U	Nonparametric	Compare 2 independent groups (ranks)	Independent observations
Wilcoxon signed-rank	Nonparametric	Compare 2 related measures (ranks)	Symmetric distribution of differences
Pearson r	Parametric	Linear relationship between 2 continuous variables	Bivariate normality, linearity
Spearman rho	Nonparametric	Monotonic relationship between 2 variables	Monotonic relationship

6. Common Mistakes to Avoid

Using a parametric test when assumptions are violated

Always check normality (e.g., Shapiro-Wilk test, Q-Q plots) and homogeneity of variance (Levene's test) before running parametric analyses. If assumptions fail and your sample is small, switch to the nonparametric alternative.

Running multiple t-tests instead of ANOVA

Comparing three or more groups with repeated t-tests inflates your Type I error rate. Use ANOVA first, then apply a post-hoc correction (such as Tukey or Bonferroni) to identify which specific pairs differ.

Confusing correlation with causation

A significant Pearson or Spearman correlation tells you that two variables move together. It does not prove that one causes the other. Always consider alternative explanations and confounding variables.

Ignoring effect size

Statistical significance depends heavily on sample size. A very large sample can produce a significant p-value for a trivially small effect. Always report an effect size measure (Cohen's d, eta-squared, r) alongside your p-value.

Using the wrong level of measurement

Treating ordinal Likert-scale data as continuous and applying parametric tests is a common shortcut. While some researchers argue it is acceptable with many scale points, it is safer to use nonparametric methods for clearly ordinal data, especially with small samples.

Forgetting to check for outliers

Outliers can dramatically affect means, standard deviations, and correlation coefficients. Inspect your data with box plots and z-score checks before analysis, and report how you handled any extreme values.

Not reporting full results

A complete report includes the test statistic, degrees of freedom (where applicable), the p-value, and an effect size. Saying only '$p < 0.05$' is not enough for your reader to evaluate the finding.

Misinterpreting a non-significant p-value

A p-value above 0.05 does not prove that no effect exists. It means the data did not provide strong enough evidence to reject the null hypothesis. Consider your sample size and whether you had enough statistical power to detect a meaningful effect.

Need help with your statistical analysis?

Visit **studentstats.net**

or email **office@studentstats.net**