Common Analyses & Frequent Mistakes

Quick Answers to Student Questions

StudentStats.net

A practical reference for choosing the right test,
avoiding common pitfalls, and interpreting results correctly.

Checking Normality

Q: "How do I check if my data is normally distributed?"

Why it matters: Many statistical tests (t-tests, ANOVA, Pearson correlation, linear regression) assume that your data follow a roughly normal (bell-shaped) distribution. Violating this assumption can lead to unreliable p-values and misleading conclusions.

Visual Methods

- **Histogram:** Plot your data and look for the classic bell shape. A clearly skewed or multimodal histogram signals non-normality.
- **Q-Q Plot (Quantile-Quantile):** Plot your data quantiles against theoretical normal quantiles. If points fall roughly along the diagonal line, the distribution is approximately normal. Systematic curves away from the line indicate skewness or heavy tails.

Statistical Tests

- **Shapiro-Wilk test:** The preferred test for samples up to about 50 observations. It has strong statistical power for detecting departures from normality. A p-value below 0.05 suggests non-normal data.
- **Kolmogorov-Smirnov test:** Suitable for larger samples. It compares the empirical distribution against a reference normal distribution. Less powerful than Shapiro-Wilk for small samples.

Practical tip: Always start with visual inspection. With very large samples (n > 300), even tiny deviations from normality become statistically significant, so the tests may reject normality even when the data are "close enough" for parametric methods.

What if my data are NOT normal?

Switch to a nonparametric alternative. The table below shows the most common swaps:

Parametric Test	Nonparametric Alternative	When to Switch
Independent t-test	Mann-Whitney U	2 independent groups, non-normal
Paired t-test	Wilcoxon signed-rank	2 related measurements, non-normal
One-way ANOVA	Kruskal-Wallis H	3+ independent groups, non-normal
Repeated-measures ANOVA	Friedman test	3+ related measurements, non-normal
Pearson correlation	Spearman correlation	Non-linear or ordinal data

Small Sample Sizes

Q: "What do I do if I have few respondents?"

Small samples are one of the most common challenges in student research. A study with 15 participants is not automatically invalid, but it does require extra care.

Minimum Sample Sizes for Common Tests

Test	Recommended Minimum n	Notes
Independent t-test	15-20 per group	Larger if variance is unequal
Paired t-test	15-20 pairs	Each subject measured twice
Chi-square test	5+ expected per cell	Merge small categories if needed
One-way ANOVA	15-20 per group	Balanced groups preferred
Pearson / Spearman	25-30 total	Power depends on effect size
Simple regression	20+ per predictor	Rule of thumb: 10-20 per variable
Cronbach's alpha	30+ respondents	At least 5x the number of items

When Small Samples Are Acceptable

- Pilot studies and exploratory research where the goal is hypothesis generation, not confirmation.
- Rare populations (e.g., patients with an uncommon condition) where large samples are impractical.
- Qualitative or mixed-methods designs where depth matters more than breadth.

Use Nonparametric Tests

When n is small and normality cannot be verified, nonparametric tests (Mann-Whitney, Wilcoxon, Kruskal-Wallis) make fewer distributional assumptions and are safer choices.

Report Effect Sizes Alongside p-values

A small sample may fail to reach statistical significance ($p < 0.05$) even when there is a meaningful real-world effect. Always report an effect size measure (Cohen's d, eta-squared, r) so readers can judge practical significance independently of sample size.

Tips for Maximizing Statistical Power

- Use within-subjects (repeated-measures) designs when possible -- they need fewer participants.
- Reduce measurement error by using validated instruments and standardized procedures.

- Conduct a power analysis before data collection to determine the sample size you actually need.
- If your sample turns out small, acknowledge the limitation honestly in your discussion section.

Choosing Between Correlation Types

Q: "Pearson or Spearman -- which do I use?"

Both coefficients measure the strength and direction of a relationship between two variables, but they differ in their assumptions and appropriate use cases.

Pearson Correlation (r)

- Measures the **linear** relationship between two continuous variables.
- Assumes both variables are approximately normally distributed.
- Sensitive to outliers -- a single extreme value can inflate or deflate the coefficient.
- **Example:** Correlation between hours studied and exam score (both continuous, roughly normal).

Spearman Correlation (rho)

- Measures the **monotonic** (consistently increasing or decreasing) relationship.
- Works with ordinal data (e.g., Likert scales ranked 1-5) and non-normal continuous data.
- More robust to outliers than Pearson.
- **Example:** Correlation between satisfaction rating (1-5 scale) and frequency of use.

Side-by-Side Comparison

Feature	Pearson (r)	Spearman (rho)
Data type	Continuous (interval/ratio)	Ordinal or continuous
Relationship type	Linear only	Any monotonic relationship
Normality required?	Yes	No
Outlier sensitivity	High	Low
Typical use	Measured variables, large n	Likert scales, small n, skewed data

Decision Rule

Start with Pearson if your data are continuous, roughly normal, and the scatter plot shows a linear trend. If any of these conditions fail -- ordinal data, skewed distribution, curved relationship, or notable outliers -- use Spearman instead.

When to Use Logistic Regression

Q: "When should I use logistic regression instead of linear?"

The choice depends entirely on the type of **dependent variable** (outcome) you are trying to predict.

Linear Regression

- Use when your dependent variable is **continuous** (e.g., income, test score, blood pressure).
- Predicts the expected value of the outcome for given predictor values.
- **Example:** Predicting GPA (continuous, 0-4.0) from study hours and attendance rate.

Key assumptions: linear relationship between predictors and outcome, normally distributed residuals, homoscedasticity (constant variance of residuals), independence of observations.

Logistic Regression

- Use when your dependent variable is **binary / categorical** (e.g., pass/fail, yes/no, diagnosed/healthy).
- Predicts the probability of belonging to a particular category.
- **Example:** Predicting whether a student passes (1) or fails (0) based on attendance and prior grades.

Key assumptions: binary or ordinal outcome, independence of observations, little multicollinearity among predictors, linearity between predictors and the log-odds of the outcome, sufficiently large sample (at least 10 events per predictor).

Quick Comparison

Feature	Linear Regression	Logistic Regression
Dependent variable	Continuous	Binary / categorical
Output	Predicted value	Predicted probability (0-1)
Typical research Q	How much does X affect Y?	Does X increase the odds of Y?
Coefficient meaning	Change in Y per unit X	Change in log-odds per unit X
Goodness of fit	R-squared	Nagelkerke R-squared, AUC

Common mistake: Using linear regression for a yes/no outcome. This can produce predicted values outside the 0-1 range and violates fundamental assumptions. Always check your dependent variable type before choosing the model.

Reliability Analysis

Q: "Do I need Cronbach's alpha?"

When to use it: Cronbach's alpha measures **internal consistency** -- how closely related a set of items are as a group. You need it whenever you combine multiple questionnaire items into a single scale score.

- Questionnaires using Likert scales (e.g., 1 = Strongly Disagree to 5 = Strongly Agree).
- Any composite score built from multiple items intended to measure the same construct.
- You do NOT need it for single-item measures, objective tests with correct/wrong answers, or demographic variables.

What Values Are Acceptable?

Alpha Range	Interpretation	Action
> 0.90	Excellent	Proceed with confidence
0.80 - 0.90	Good	Acceptable for most research
0.70 - 0.80	Acceptable	Common threshold in social sciences
0.60 - 0.70	Questionable	Consider revising items
< 0.60	Poor	Do not use the scale as-is

What to Do if Alpha Is Too Low

- Check the "Alpha if item deleted" column -- removing a weak item may improve the overall alpha.
- Review item wording for ambiguity or double-barreled questions.
- Examine inter-item correlations; items below 0.3 may not belong in the scale.
- Consider whether the construct is truly unidimensional -- factor analysis can reveal multiple sub-dimensions.

Multiple Comparisons Problem

Q: "I ran many tests -- is that a problem?"

Yes, it can be. Every time you run a statistical test at the 0.05 significance level, there is a 5% chance of a false positive (Type I error). When you run many tests, these chances accumulate.

The Math Behind It

If you run 20 independent tests at alpha = 0.05, the probability of at least one false positive is:

$$1 - (1 - 0.05)^{20} = 0.64 \text{ (64\%)}$$

That means roughly two-thirds of the time you will find at least one "significant" result purely by chance.

Bonferroni Correction

The simplest fix: divide your significance level by the number of tests. If you run 10 tests, use alpha = 0.05 / 10 = 0.005 as your threshold for each individual test. This is conservative but effective.

When to Worry About It

- Running post-hoc pairwise comparisons after ANOVA (most software applies corrections automatically -- check your output).
- Testing many correlations in a correlation matrix.
- Comparing the same outcome across multiple subgroups.

When it is less of a concern: Pre-planned (a priori) comparisons based on specific hypotheses, or exploratory analyses clearly labeled as such in your write-up.

Top 10 Statistical Mistakes Students Make

1. Confusing correlation with causation. Finding that two variables are correlated does not mean one causes the other. A third variable may drive both, or the relationship may be coincidental.

2. Ignoring assumptions before running a test. Every statistical test has assumptions (normality, equal variances, independence). Skipping these checks can invalidate your results entirely.

3. Using the wrong test for the data type. Applying a parametric test to ordinal Likert data, or using a chi-square test when expected cell counts are below 5. Always match the test to your variable types.

4. Treating p > 0.05 as proof of no effect. A non-significant result means you failed to detect an effect, not that no effect exists. The study may simply lack statistical power.

5. Not reporting effect sizes. Statistical significance depends on sample size; effect sizes tell you whether the finding matters in practice. Always report both.

6. Running too many tests without correction. Each test at alpha = 0.05 carries a 5% false-positive risk. Multiple tests compound this risk (see Section 6).

7. Cherry-picking significant results. Running dozens of analyses and reporting only the ones that "worked" is a form of p-hacking. Report all analyses, including non-significant ones.

8. Using Cronbach's alpha when it is not needed. Alpha is for multi-item scales measuring a single construct, not for any set of questions grouped together. Misapplied alpha values are meaningless.

9. Deleting outliers without justification. Outliers should only be removed if there is a clear reason (data entry error, equipment malfunction). Removing them just to improve results is bad practice.

10. Copying SPSS output without interpretation. Pasting tables of numbers into your paper is not analysis. Explain what each result means in plain language, relate it to your hypotheses, and discuss practical implications.

Need More Help With Your Analysis?

Visit **StudentStats.net** for free guides, tutorials, and practical resources designed specifically for students working on theses, dissertations, and research projects.

www.studentstats.net

Free guides | Step-by-step tutorials | Statistical consulting

Helping students succeed with data since 2024