How to Interpret Statistical Results

A Clear Guide for Academic Writing

StudentStats.net

Free resources to help students succeed in research

Table of Contents

1. Why Interpretation Matters

Running a statistical test is only half the work. The other half -- often the more important half -- is understanding what the numbers actually mean for your research question. Software will always produce output: p-values, test statistics, confidence intervals. But numbers without interpretation are just numbers.

The Interpretation Gap

Many students treat statistical analysis as a black box: data goes in, a p-value comes out, and the hypothesis is either 'supported' or 'not supported.' This approach misses the richness of what your data is telling you. A statistically significant result might be trivially small in practical terms, while a non-significant result might reflect insufficient sample size rather than the absence of an effect.

How Poor Interpretation Undermines Research

Misinterpreting results leads to overstated conclusions, misleading abstracts, and flawed recommendations. Common consequences include:

- Claiming a treatment 'works' based solely on $p < 0.05$, ignoring that the effect is negligibly small.
- Concluding that two groups are 'the same' because a test was not significant, when the study simply lacked statistical power.
- Presenting correlation results as evidence of causation.
- Selectively reporting only significant findings while ignoring non-significant ones.

Good interpretation connects statistical output back to the research question, considers effect size alongside significance, acknowledges limitations, and presents results honestly. This guide will help you develop those skills.

2. Understanding P-Values

What a P-Value Actually Means

A p-value is the probability of obtaining results at least as extreme as the ones observed, assuming the null hypothesis is true. It is **not** the probability that the null hypothesis is true, and it is **not** the probability that your results occurred by chance.

Think of it this way: if there truly were no difference between your groups (or no relationship between your variables), how surprising would your data be? A small p-value means 'very surprising under the null hypothesis,' which gives you reason to doubt that null hypothesis.

The 0.05 Threshold

The conventional alpha level of 0.05 means we accept a 5% risk of incorrectly rejecting the null hypothesis (a Type I error). This threshold is a convention, not a law of nature. It was popularized in the early 20th century and has remained the default in most social and behavioral sciences. Some fields use stricter thresholds (e.g., 0.01 in medical research, 0.001 in genetics).

What $p < 0.05$ DOES Mean

- The observed data would be unlikely if the null hypothesis were true.
- There is sufficient evidence to reject the null hypothesis at the chosen significance level.
- The result is 'statistically significant' by conventional standards.

What $p < 0.05$ Does NOT Mean

- It does NOT mean the effect is large or practically important.
- It does NOT mean the null hypothesis is false (it may still be true -- you accepted a 5% error rate).
- It does NOT mean your study is free from bias or methodological problems.
- It does NOT tell you the probability that the null hypothesis is true or false.

The 'No Effect' Misinterpretation

One of the most common errors is interpreting $p > 0.05$ as proof that no effect exists. A non-significant result means you failed to find sufficient evidence against the null hypothesis -- it does not confirm the null. The study may have been underpowered (too few participants), the measure may have been imprecise, or the effect may exist but be smaller than the study could detect. Always consider statistical power and confidence intervals when interpreting non-significant results.

Statistical Significance vs. Practical Significance

A statistically significant result is not necessarily a meaningful one. With a very large sample, even tiny, trivial differences can produce $p < 0.05$. For example, a study with 10,000 participants might find a statistically significant difference of 0.3 points on a 100-point scale. While technically significant, this difference has no practical value. Always report and interpret effect sizes alongside p-values to assess whether a result matters in the real world.

3. Reading Key Statistics

This section explains the most common statistics you will encounter in research, what each one tells you, and how to interpret it properly.

Mean and Standard Deviation

The **mean** (M) is the arithmetic average of a set of values. It tells you the central tendency of your data. The **standard deviation** (SD) tells you how spread out the values are around the mean. A small SD means data points cluster tightly around the mean; a large SD means they are more dispersed.

How to report: M = 4.52, SD = 1.23. Always report both together so readers can assess both the typical value and the variability in your data.

t-Statistic and Degrees of Freedom

The **t-statistic** measures how many standard errors the sample mean difference is from zero (or from a hypothesized value). Larger absolute values of t indicate greater evidence against the null hypothesis. **Degrees of freedom** (df) reflect the sample size and affect the shape of the t-distribution used to calculate the p-value.

How to report: $t(58) = 2.45$, $p = .017$. The number in parentheses is the degrees of freedom.

F-Statistic (ANOVA)

The **F-statistic** compares variance between groups to variance within groups. It is used in Analysis of Variance (ANOVA) to test whether three or more group means differ significantly. A larger F indicates greater differences between groups relative to within-group variability.

How to report: $F(2, 87) = 5.34$, $p = .007$. The first df value is the number of groups minus one; the second is the total sample size minus the number of groups.

Chi-Square Statistic

The **chi-square statistic** (X^2) tests whether observed frequencies in categorical data differ significantly from expected frequencies. It is used for tests of independence (are two categorical variables related?) and goodness-of-fit tests (does observed data match a theoretical distribution?).

How to report: $X^2(4, N = 150) = 11.28$, $p = .024$. Include degrees of freedom and sample size.

Correlation Coefficient (r)

Pearson's **r** measures the strength and direction of a linear relationship between two continuous variables. Values range from -1 to +1. A positive r indicates that as one variable increases, the other tends to increase. A negative r indicates an inverse relationship. General guidelines for interpreting strength:

- $|r|$ = .10 to .29: small (weak) relationship
- $|r|$ = .30 to .49: medium (moderate) relationship
- $|r|$ = .50 to 1.0: large (strong) relationship

How to report: $r(98) = .45$, $p < .001$. Include degrees of freedom (N - 2).

R-Squared (R^2)

R^2 represents the proportion of variance in the dependent variable that is explained by the independent variable(s). An R^2 of .36 means 36% of the variability in the outcome is accounted for by the predictor(s). It is commonly reported in regression analyses.

How to report: $R^2 = .36$, indicating that the model explained 36% of the variance in the outcome variable.

Confidence Intervals

A **95% confidence interval** provides a range of plausible values for a population parameter. If you were to repeat the study many times, approximately 95% of the calculated intervals would contain the true parameter. Wider intervals indicate less precision; narrower intervals indicate more precision.

How to report: M = 12.4, 95% CI [10.8, 14.0]. Always specify the confidence level (usually 95%).

Effect Sizes: Cohen's d and Eta-Squared

Cohen's d quantifies the difference between two group means in standard deviation units. Guidelines: d = 0.2 (small), d = 0.5 (medium), d = 0.8 (large).

Eta-squared (n^2) represents the proportion of total variance attributable to a factor in ANOVA. Guidelines: $n^2 = .01$ (small), $n^2 = .06$ (medium), $n^2 = .14$ (large).

Why effect sizes matter: They tell you the magnitude of an effect regardless of sample size. A result can be statistically significant but practically meaningless if the effect size is very small. Many journals now require effect sizes to be reported.

4. Writing the Results Section

Structure of a Results Paragraph

A well-written results paragraph follows a clear four-part structure:

- **Restate the hypothesis** or research question being tested.
- **Name the test** used and why it was appropriate.
- **Report the results** including test statistic, degrees of freedom, p-value, and effect size.
- **State the conclusion** in plain language -- was the hypothesis supported?

Template Sentences for Common Tests

Independent Samples t-Test

> *"An independent samples t-test was conducted to compare [DV] between [Group 1] and [Group 2]. There was a significant difference in scores for [Group 1] (M = __, SD = __) and [Group 2] (M = __, SD = __); t(__) = __, p = __, d = __."*

One-Way ANOVA

> *"A one-way ANOVA was conducted to examine the effect of [IV] on [DV]. The results showed a statistically significant difference between groups, F(__, __) = __, p = __,* n^2 *= __. Post hoc comparisons using Tukey HSD indicated that..."*

Pearson Correlation

> *"A Pearson product-moment correlation was computed to assess the relationship between [Variable 1] and [Variable 2]. There was a [positive/negative] correlation between the two variables, r(__) = __, p = __."*

Chi-Square Test of Independence

> *"A chi-square test of independence was performed to examine the relationship between [Variable 1] and [Variable 2]. The relation between these variables was significant,* X^2*(__, N = __) = __, p = __."*

Linear Regression

> *"A simple linear regression was calculated to predict [DV] based on [IV]. A significant regression equation was found, F(__, __) = __, p = __, with an* R^2 *of __. Participants' predicted [DV] is equal to __ + __([IV])."*

APA Formatting Basics for Statistics

- Italicize statistical symbols: *t*, *F*, *p*, *r*, *M*, *SD*, *N*, *n*, *df*.

- Do not use a leading zero for values that cannot exceed 1.0 (e.g., $p = .034$, not $p = 0.034$; $r = .52$, not $r = 0.52$).
- Report exact p-values to two or three decimal places (e.g., $p = .032$). Use $p < .001$ for very small values.
- Round most statistics to two decimal places.
- Place spaces around equals signs and comparison operators: $t(45) = 2.31$, $p = .025$.
- Always include degrees of freedom in parentheses immediately after the test statistic symbol.

5. Common Interpretation Mistakes

Even experienced researchers sometimes fall into interpretation traps. Being aware of these common mistakes will strengthen your results section and your overall research.

1. Confusing Correlation with Causation

Finding that two variables are correlated does not mean one causes the other. A significant correlation between ice cream sales and drowning rates does not mean ice cream causes drowning -- both are driven by hot weather. Unless your study uses an experimental design with random assignment, avoid causal language. Use phrases like 'is associated with' or 'is related to' rather than 'causes' or 'leads to.'

2. Ignoring Effect Size When p Is Significant

A significant p-value tells you an effect probably exists, but not how big it is. With large samples, even trivially small effects become statistically significant. Always report and discuss effect size (Cohen's d, eta-squared, R^2) to help readers understand the practical importance of your finding.

3. Cherry-Picking Significant Results

Running many tests and reporting only the significant ones inflates the false-positive rate. If you test 20 hypotheses at alpha = .05, you would expect one significant result by chance alone. Report all analyses you conducted, including non-significant results. If you ran multiple comparisons, apply appropriate corrections (e.g., Bonferroni).

4. Misinterpreting Non-Significant Results

A p-value above .05 does not prove the null hypothesis. It means the data did not provide strong enough evidence to reject it. The absence of evidence is not evidence of absence. Consider reporting confidence intervals, which can show whether the range of plausible effects includes meaningful values, even when $p > .05$.

5. Over-Interpreting Small Sample Results

Small samples produce unstable estimates. An effect size of $d = 1.2$ from a study with 10 participants per group should be interpreted cautiously -- the confidence interval around that estimate will be very wide. Acknowledge sample size limitations and avoid sweeping generalizations from small studies.

6. Reporting Only P-Values Without Descriptive Statistics

A p-value alone tells the reader almost nothing. Always include descriptive statistics (means, standard deviations, frequencies) so readers can understand the direction and magnitude of effects. A complete report includes: descriptive statistics, test statistic with degrees of freedom, exact p-value, effect size, and confidence intervals when possible.

6. Quick Reference Table

Statistic	What It Tells You	How to Report (APA)
Mean (M)	Central tendency of data	M = 4.52, SD = 1.23
Standard Deviation (SD)	Spread of data around the mean	Reported alongside the mean
t-statistic	Difference between two group means in SE units	t(58) = 2.45, p = .017, d = 0.64
F-statistic	Ratio of between-group to within-group variance	F(2, 87) = 5.34, p = .007
Chi-square (X^2)	Association between categorical variables	X^2(4, N = 150) = 11.28, p = .024
Correlation (r)	Strength and direction of linear relationship	r(98) = .45, p < .001
R^2	Proportion of variance explained by model	R^2 = .36
95% CI	Range of plausible values for parameter	95% CI [10.8, 14.0]
Cohen's d	Standardized mean difference (effect size)	d = 0.64 (medium effect)
Eta-squared (n^2)	Proportion of variance due to a factor	n^2 = .09 (medium effect)

StudentStats.net

Free tools and guides to help students master statistics

Visit us at:

www.studentstats.net

> Step-by-step statistical test calculators
> Free guides for SPSS, R, and Excel
> APA results templates you can copy and paste
> Sample size and power analysis tools

Helping students turn data into insight.