Statistical Tests Cheat Sheet

This cheat sheet provides a quick reference for selecting appropriate statistical tests based on data type, research question, and assumptions. It's tailored for data science applications, including hypothesis testing, A/B testing, and exploratory data analysis.

1. Comparing Means

Tests to compare means across groups or against a population value.

Test	Use Case	Data Type	Assumptions	Example
One-sample t-test	Compare sample mean to a known population mean	Continuous, one sample	Normal distribution, independent observations	Test if average customer rating (e.g., 3.5/5) matches a target (e.g., 4/5)
Independent t-test/Two-sample t-test	Compare means of two independent groups	Continuous, two groups	Normal distribution, equal variances, independent observations	Compare average sales between two stores
Paired t-test	Compare means of two related groups (e.g., before/after)	Continuous, paired samples	Normal differences, paired observations	Compare user engagement before and after a website redesign

One-way ANOVA	Compare means of 3+ groups	Continuous, multiple groups	Normal distribution, equal variances, independent observations	Compare average revenue across multiple marketing campaigns
Welch's t-test	Compare means of two groups (unequal variances)	Continuous, two groups	Normal distribution, independent observations	Compare test scores between two classes with different variances
Mann-Whitney U Test	Non-parametric alternative to independent t-test	Continuous/ordinal, two groups	Non-normal data, independent observations	Compare user satisfaction scores between two products (non-normal data)
Wilcoxon Signed-Rank Test	Non-parametric alternative to paired t-test	Continuous/ordinal, paired samples	Non-normal differences, paired observations	Compare pre- and post-training performance scores
Kruskal-Wallis Test	Non-parametric alternative to one-way ANOVA	Continuous/ordinal, 3+ groups	Non-normal data, independent observations	Compare customer ratings across multiple product versions

2. Comparing Proportions

Tests for comparing proportions or frequencies across groups.

Test	Use Case	Data Type	Assumptions	Example
Z-test for Proportions	Compare proportion of one group to a known value	Categorical, one sample	Large sample size, independent observations	Test if click-through rate (CTR) matches an expected value (e.g., 10%)
Two-sample Z-test for Proportions (one-sided/two-sided)	Compare proportions between two groups	Categorical, two groups	Large sample size, independent observations	Compare conversion rates between two ad campaigns
Chi-Square Test of Independence	Test relationship between two categorical variables	Categorical, contingency table	Expected frequencies ≥5, independent observations	Test if customer gender affects product preference
Fisher’s Exact Test	Test relationship between two categorical variables (small samples)	Categorical, 2x2 table	Small sample size, independent observations	Test if treatment affects outcome in a small clinical trial

3. Comparing Variances(Homoscedasticity)

Tests to compare variability across groups.

Test	Use Case	Data Type	Assumptions	Example
F-test	Compare variances of two groups	Continuous, two groups	Normal distribution, independent observations	Compare variance in stock returns between two companies
Levene's Test	Compare variances of 2+ groups (robust to non-normality)	Continuous, multiple groups	Independent observations	Compare variance in response times across multiple servers
Bartlett's Test	Compare variances of 2+ groups	Continuous, multiple groups	Normal distribution, independent observations	Compare variance in test scores across multiple classes

4. Testing Relationships (Correlations)

Tests to assess relationships/correlations between variables.

Test	Use Case	Data Type	Assumptions	Example
Pearson Correlation	Measure linear relationship between two continuous variables	Continuous	Linear relationship, normal distribution, homoscedasticity	Test correlation between ad spend and revenue

Spearman's Rank Correlation	Non-parametric measure of monotonic relationship	Continuous/ordinal	Monotonic relationship, non-normal data	Test correlation between user ratings and time spent on app
Kendall Tau	Non-parametric measure of ordinal association	Ordinal	Ordinal data, small samples	Test association between ranked preferences and customer age

5. Testing Distributions

Tests to assess whether data follows a specific distribution or matches another dataset.

Test	Use Case	Data Type	Assumptions	Example
Shapiro-Wilk Test	Test if data is normally distributed	Continuous	Small to moderate sample size	Test if customer purchase amounts are normally distributed
Kolmogorov-Smirnov Test	Compare sample distribution to a reference or two samples	Continuous	Continuous data	Compare user session times to a normal distribution
Anderson-Darling Test	Test if data fits a specific distribution	Continuous	Continuous data	Test if sales data fits a Poisson distribution
Visualizations	Histogram, boxplots, Q-Q plots.	Continuous	Continuous data	Check if sales data has normal distribution.

6. Regression and Model-Based Tests

Tests used in regression analysis or model evaluation.

Test	Use Case	Data Type	Assumptions	Example
F-test (Regression)	Test significance of overall regression model	Continuous (dependent), mixed (independent)	Linear relationship, normal residuals, homoscedasticity	Test if a linear regression model predicts sales effectively
t-test (Regression Coefficients)	Test significance of individual predictors in regression	Continuous (dependent), mixed (independent)	Normal residuals, independent observations	Test if ad spend significantly predicts revenue
Breusch-Pagan Test	Test for heteroscedasticity in regression residuals	Continuous	Linear model fitted	Test if variance of residuals is constant in a revenue model

7. Other Non-Parametric Tests

Tests for non-normal data or small samples.

Test	Use Case	Data Type	Assumptions	Example
Sign Test	Non-parametric test for median difference in paired data	Continuous/ordinal, paired	Paired observations	Test if median user ratings changed after an update

Friedman Test	Non-parametric alternative to repeated-measures ANOVA	Continuous/ordinal, 3+ related groups	Related observations	Compare user satisfaction across multiple app versions
Run Test	Test randomness of a sequence	Binary/ordinal	Independent observations	Test if customer complaints occur randomly over time

8. Errors

Error	Use Case
Type I **False Positive**	occurs in statistical hypothesis testing when a null hypothesis that is actually true is rejected.
Type II **False Negative**	occurs in statistical hypothesis testing when a null hypothesis that is actually false is not rejected.

Factors Affecting Type I and Type II Errors

- **Sample Size:** In statistical hypothesis testing, larger sample sizes generally reduce the probability of both Type I and Type II errors. With larger samples, the estimates tend to be more precise, resulting in more accurate conclusions.
- **Significance Level:** The significance level (α) in hypothesis testing determines the probability of committing a Type I error. Choosing a lower significance level reduces the risk of Type I error but increases the risk of Type II error, and vice versa.

- **Effect Size:** The magnitude of the effect or difference being tested influences the probability of Type II error. Smaller effect sizes are more challenging to detect, increasing the likelihood of failing to reject the null hypothesis when it's false. If the effect size increases, it's easier to detect a true effect, and the probability of a type II error decreases. In practice, we wouldn't know the true effect size because we're only working with the sample data. Instead, we can determine the minimal important difference (MID), the smallest difference in a measured outcome that is considered to be meaningful or clinically relevant. We can set the effect size equal to the MID.
- **Data Variability :** Across different significance levels, the probability of type I and II error would decrease as the standard deviation decreases.
- **Statistical Power:** The power of Statistics (1 – β) dictates that the opportunity for rejecting a wrong null hypothesis is based on the inverse of the chance of committing a Type II error. The power level of the test rises, thus a chance of the Type II error dropping.
- **Power analysis** → "Is my sample size large enough to reliably detect the kind of difference I care about?" It is mainly useful for:
  Pre-study planning — determining the needed sample size.
  Post hoc power analysis — confirming whether a non-significant result was due to low power.

Relationships	Significance Level (α)	Sample Size	Data Variability	Effect Size
P(Type I Error)	Positive	Negative	Positive	N/A
P(Type II Error)	Negative	Negative	Positive	Negative

Sources:

- Grok
- https://towardsdatascience.com/hypothesis-testing-for-data-scientists-everything-you-need-to-know-8c36ddde4cd2/
- https://towardsdatascience.com/type-i-ii-errors-and-sample-size-calculation-in-hypothesis-testing-760dae42a065/
- https://stats.oarc.ucla.edu/spss/whatstat/what-statistical-analysis-should-i-usestatistical-analyses-using-spss/