Common INFERENTIAL Statistical Tests Overview

1. **Key Points:**

   Data Type: Choose the test based on whether your data is continuous, categorical, or ordinal.

   Number of Groups/Conditions: Consider the number of groups or conditions being compared.

   Independence: Check whether the groups are independent or related (paired).

   Distribution Assumptions: Decide if you can assume normality or if you need a non-parametric test.

2. **REMEMBER:**

The tests listed in the table are primarily **inferential statistical tests** rather than descriptive analyses. Here's the distinction:

Descriptive Statistics: These are used to describe and summarize the main features of a dataset, providing simple summaries and visualizations. Common examples include measures of central tendency (mean, median, mode), measures of variability (standard deviation, variance), frequency distributions, and graphs such as histograms and box plots.

Inferential Statistics: These are used to make inferences or generalizations about a population based on a sample of data. The tests in the table (e.g., t-tests, ANOVA, chi-square tests) are inferential because they help determine if observed differences or relationships in the sample data are statistically significant and likely to be present in the larger population.

Descriptive statistics are generally the first step in data analysis, providing an overview of the data and identifying patterns or outliers. Inferential statistics are used to test hypotheses and make predictions based on the data.

The tests listed, such as t-tests, ANOVA, and chi-square tests, are tools for inferential analysis and are not typically used for descriptive purposes. Descriptive analysis would involve calculating basic statistics and creating visualizations to describe the dataset without making inferences beyond the data at hand.

Test	Purpose	Data Type	Key Assumptions	Example Scenario
Pearson Correlation	Measure the strength and direction of the relationship between two continuous variables	Continuous vs. Continuous	Linearity, normality of variables, homoscedasticity	Analyzing the relationship between hours of study and exam scores.
Spearman's Rank Correlation	Measure the strength and direction of the relationship between two ranked or ordinal variables	Ordinal or Continuous (non-normal) vs. Ordinal or Continuous	Non-parametric equivalent of Pearson correlation, does not assume normal distribution	Analyzing the relationship between customer satisfaction rankings and purchase frequency.
Chi-Square Test	Test association or independence between categorical variables	Categorical vs. Categorical	Expected frequency in each category should be at least 5; observations are independent	Analyzing the association between gender (male/female) and preference for a type of snack (sweet/savory).
Chi-Square Test for Goodness of Fit	Test if a sample distribution matches an expected distribution	Categorical	Large enough sample size, expected frequencies of at least 5 per category	Testing if a dice is fair based on the observed frequencies of each face.
t-Test (Independent)	Compare means between two independent groups	Continuous vs. Categorical (2 groups)	Data should be normally distributed; variances are equal; independent samples	Comparing average test scores between male and female students.
t-Test (Paired)	Compare means within the same group at two different times or under two conditions	Continuous (2 related measures)	Differences between pairs should be normally distributed	Measuring the effect of a diet on weight by comparing weights before and after the diet for the same individuals.

ANOVA (One-Way)	Compare means among three or more independent groups	Continuous vs. Categorical (3+ groups)	Data should be normally distributed; variances are equal; independent samples	Comparing the average effectiveness of three different teaching methods.
ANOVA (Repeated Measures)	Compare means across multiple measures or time points for the same subjects	Continuous (3+ related measures)	Sphericity assumption (variances of the differences are equal)	Measuring blood pressure at multiple time points for the same group of patients.
Factorial ANOVA	Compare means with two or more independent variables (factors)	Continuous vs. Categorical (multiple groups)	Similar assumptions to one-way ANOVA, plus no interaction effects between factors	Analyzing the impact of both study method and time of day on test performance.
MANOVA (Multivariate ANOVA)	Compare means across multiple dependent variables	Multiple Continuous vs. Categorical	Multivariate normality, homogeneity of variance-covariance matrices	Studying the effect of a treatment on both blood pressure and cholesterol levels.
Mann-Whitney U Test	Compare ranks between two independent groups	Ordinal or Continuous (non-normal) vs. Categorical (2 groups)	Non-parametric equivalent of independent t-test, does not assume normal distribution	Comparing customer satisfaction ratings between two stores.
Wilcoxon Signed-Rank Test	Compare ranks within the same group at two different times or under two conditions	Ordinal or Continuous (non-normal) (2 related measures)	Non-parametric equivalent of paired t-test, does not assume normal distribution	Comparing pre- and post-intervention anxiety levels in the same individuals.
Kruskal-Wallis Test	Compare ranks among three or more independent groups	Ordinal or Continuous (non-normal) vs. Categorical (3+ groups)	Non-parametric equivalent of one-way ANOVA, does not assume normal distribution	Comparing median test scores across different teaching methods.

Friedman Test	Compare ranks across multiple measures or time points for the same subjects	Ordinal or Continuous (non-normal) (3+ related measures)	Non-parametric equivalent of repeated measures ANOVA, does not assume normal distribution	Measuring the effect of different diets on weight across several time points for the same group.
Linear Regression	Predict the value of a dependent variable based on one or more independent variables	Continuous (dependent) vs. Continuous/Categorical (independent)	Linearity, independence, homoscedasticity, normality of residuals	Predicting house prices based on area, number of rooms, and location.
Logistic Regression	Predict the probability of a binary outcome based on one or more independent variables	Categorical (binary dependent) vs. Continuous/Categorical (independent)	Linearity of independent variables and log odds, independence of observations	Predicting whether a patient has a disease (yes/no) based on age, weight, and other factors.
Fisher's Exact Test	Test association between two categorical variables when sample sizes are small	Categorical	None, but typically used when expected frequencies are less than 5	Analyzing the relationship between two treatments and patient recovery in a small sample.
McNemar's Test	Test for changes in proportions on a dichotomous trait within paired samples	Categorical (binary)	Paired samples, dichotomous outcomes	Testing the effect of a drug on disease status (improved/not improved) in the same patients before and after treatment.
Cochran's Q Test	Compare proportions across multiple related groups with a binary outcome	Categorical (binary)	Similar to McNemar's test, but for more than two groups	Testing the effectiveness of three different treatments on disease status (cured/not cured) in the same patients.