

Statistical Methods

Statistical Method	Description	Formula	Real-Life Example
Mean (Average)	Measures the central tendency of a dataset.	$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$	Average income of employees in a company.
Median	The middle value of a dataset when ordered.		Median house price in a city.
Standard Deviation	Measures the spread of data points around the mean.	$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2}$	Variation in test scores among students.
Variance	The square of the standard deviation; measures data dispersion.	$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$	Variance in monthly sales figures.
T-Test	Compares the means of two groups to see if they are significantly different.	$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$	Comparing average salaries between two departments.
Chi-Square Test	Tests the association between categorical variables.	$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$	Testing if there is an association between gender and product preference.
ANOVA	Compares means	$F = \frac{MS_{\text{between}}}{MS_{\text{within}}}$	Comparing

(Analysis of Variance)	among three or more groups to see if at least one is different.		average customer satisfaction across different stores.
Correlation Coefficient	Measures the strength and direction of a linear relationship between two variables.	$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$	Correlation between hours studied and exam scores.
Linear Regression	Models the relationship between a dependent variable and one or more independent variables.	$y = \beta_0 + \beta_1 x + \epsilon$	Predicting house prices based on size and location.
Logistic Regression	Models the probability of a binary outcome based on one or more predictor variables.	$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x$	Predicting whether a customer will buy a product based on age and income.
Kruskal-Wallis Test	Non-parametric method for comparing medians of three or more groups.	N/A (Rank-based calculation)	Comparing median response times to different types of customer service queries.