Statistical Terms: Definitions, Formulas, and Analogies

Term	**Definition**	**Formula**	**Simple Analogy**
Mean			
The average of a set of numbers. It represents the central value.			
Mean = (xi) / n			
The average score of a student in multiple subjects.			

Median
The middle value in a sorted list of numbers. If the list is even, it's the average of the two middle numbers.
If odd: The middle number. If even: Median = (x(n/2) + x(n/2 + 1)) / 2
The middle point in a queue, where half are in front and half are behind.

Mode
The number that appears

most frequently in a data set.
No formula. It's the most frequent value.
The most common shoe size sold in a store.

Variance
A measure of how much the numbers in a data set differ from the mean.
Variance (2) = (xi -)2 / n
How spread out the students' test scores are from the average score.

Standard Deviation
The square root of variance, representing the average distance of each data point from the mean.
Standard Deviation () = ((xi -)2 / n)
How much, on average, each students score deviates from the average score.

Covariance
A measure of the relationship between two variables; indicates whether they tend to move together.
Cov(X, Y) = (xi - x)(yi -) / n
Whether taller people tend to have larger shoe sizes.

Correlation Coefficient (r)
A standardized measure of the strength and direction of the linear relationship between two variables.
r = Cov(X, Y) / (x)
How closely two variables are related, like height and weight.

Coefficient of Determination (R2)
The proportion of variance in the dependent variable that is predictable from the independent variable(s).
R2 = 1 - (SS_res / SS_tot)

The percentage of how well a regression model explains the data; like how well your study habits predict your grades.

Skewness
A measure of the asymmetry of the probability distribution of a real-valued random variable.
Skewness = (xi -)3 / (n3)
Whether a distribution of ages in a class is tilted to younger or older students.

Kurtosis
A measure of the 'tailedness' of the probability distribution.
Kurtosis = (xi -)4 / (n4) - 3
How extreme or common the outliers are in a distribution, like the presence of very high or very low test scores.

Confidence Interval
A range of values that is likely to contain the population parameter with a certain level of confidence.
CI = Mean (Z /n)
The margin of error in a poll, showing the range within which the true value is expected to lie.

P-value
The probability of obtaining test results at least as extreme as the observed results, assuming that the null hypothesis is true.
No simple formula; derived from statistical tests.
The likelihood that an observed effect is due to chance, like flipping a coin and getting heads 10 times in a row.

Null Hypothesis (H0)

A general statement that there is no relationship between two measured phenomena or no association among groups.
No formula; it's a statement tested in hypothesis testing.
Assuming a new drug has no effect, then testing to see if evidence suggests otherwise.

F-test
A statistical test to compare the variances of two populations to determine if they are significantly different.
F = Variance of Group 1 / Variance of Group 2
Comparing the variability of test scores between two different classes to see if one is more consistent.

t-test

A statistical test used to determine if there is a significant difference between the means of two groups.
t = Difference in Means / Standard Error
Comparing the average scores of two groups to see if they perform differently on a test.

Chi-square test
A statistical test used to determine if there is a significant association between two categorical variables.
2 = ((Oi - Ei)2 / Ei)
Checking if theres a relationship between gender and preference for a particular product in a survey.

ANOVA (Analysis of Variance)

A statistical test to compare the means of three or more groups to see if at least one is different.
No single formula; based on F-distribution.
Comparing the average heights of plants grown with different fertilizers.

Type I Error
The error of rejecting the null hypothesis when it is actually true.
No formula; occurs in hypothesis testing.
Convicting an innocent person (false positive).

Type II Error
The error of failing to reject the null hypothesis when it is actually false.
No formula; occurs in hypothesis testing.
Letting a guilty person go

free (false negative).

Power of a Test
The probability that a test correctly rejects a false null hypothesis (detects an effect when there is one).
Power = 1 -
The ability of a medical test to correctly identify a disease.

Regression Analysis
A statistical method for estimating the relationships among variables.
Linear Regression: y = mx + c
Predicting someones weight based on their height using a line of best fit.