Z TEST - Numerical Data

A **Z-test** is a statistical test used to determine whether there is a significant difference between the **sample mean** of a column and the **population mean** of a column or between the means of **two columns** when the population variance is known, and the sample size is large.

Types	One-Sample	Two-Sample	Paired Samples
Definition	Tests whether the sample mean of a column is significantly different from a known or claimed population mean of the same column.	Used to compare the means of two independent columns, assuming equal population variances and large column sizes.	Compares means of two related groups of a column (before-after or matched pairs)
Example	Is the average 'shipping_cost' in the DataFrame different from $10?	Is there a difference in average 'revenue' between orders shipped via 'Express' and 'Standard'?	Is there a significant change in sales after implementing discount rates?
Sample Size	> 30	> 30	> 30
Standard Deviation	Known	Known	Known

Z Proportion TEST - Categorical Data

A **z-proportion test** is used when we want to compare **proportions** instead of means. It helps determine whether the proportion of success in a **single column** or between **two columns in a dataframe** is significantly different from a specific value or each other

Types	One-sample one-tailed (Left-tailed & Right-tailed)	One-sample two-tailed	Two-sample one-tailed (Left-tailed & Right-tailed)	Two-sample two-tailed
Definition	This test checks if the proportion in a single column of a DataFrame is significantly less than or greater than a given threshold.	This test checks if the proportion in a column is significantly different (either lower or higher) from a specific reference value.	This test checks if the proportion in one group (one row of a grouped DataFrame) is significantly greater or lesser than another.	This test checks if the proportions between two filtered groups in a DataFrame are significantly different.
Example	A hospital claims that 30% of patients in the 'diabetes' column have the condition. You want to check if the actual percentage from your DataFrame is less than 30% (left-tailed) or greater than 30% (right-tailed).	You want to test if the percentage of 'Positive' entries in the 'Sentimental Analysis' column is different from 60%.	You grouped your DataFrame by 'gender', and now want to check if the proportion of 'diabetes' cases in 'Male' rows is higher than in 'Female' rows.	Compare the proportion of positive reviews between two 'shipping_mode' categories in your DataFrame to check if they differ.
Sample Size	> 30	> 30	> 30	> 30
Standard Deviation	Known	Known	Known	Known