

Descriptive Statistics

Fundamental methods to summarize and describe the overall characteristics of collected data.

- **Central Tendency:** Indicates where the data is concentrated.
 - **Mean:** The arithmetic average calculated by dividing the sum of all values by the count. Sensitive to extreme values (outliers).
 - **Median:** The middle value when data is arranged in order. Less distorted by outliers.
 - **Mode:** The most frequently occurring value in the dataset.
- **Dispersion:** Measures how widely the data is spread out.
 - **Range:** The difference between the maximum and minimum values.
 - **Variance:** The average of the squared distances (deviations) of each observation from the mean.
 - **Standard Deviation:** The square root of the variance; intuitively shows data variability in the same units as the original data.
 - **Quartiles:** Points that divide the data into four equal parts (Q1, Q2, Q3). **IQR (Interquartile Range) = Q3 - Q1** represents the spread of the middle 50% of the data.

Inferential Statistics

Methods for deducing characteristics and drawing conclusions about an uncollected Population based on an analysis of Sample data.

- **Hypothesis Testing:**
 - **Null Hypothesis (H0):** A hypothesis representing the 'status quo' or 'existing fact' that there is no difference or effect.
 - **Alternative Hypothesis (H1):** The researcher's claim that 'there is a difference' or effect to be proven.
 - **p-value (Significance Probability):** The probability of observing the current data (or more extreme data) by chance, assuming the null hypothesis is true.
 - **Significance Level (α):** The threshold for deciding whether to reject the null hypothe-

sis, commonly set at 0.05 (5%). Reject H0 if $p < \alpha$.

- **Confidence Interval:** A statistical range expected to contain the actual population parameter (e.g., mean). (Example: 95% Confidence Interval)

Major Probability Distributions

- **Normal Distribution:** A bell-shaped symmetric distribution centered around the mean; the most representative distribution for explaining natural and social phenomena.
- **Binomial Distribution:** The distribution of successes in a fixed number of independent Bernoulli trials (where results are binary, like success/failure).
- **Poisson Distribution:** The distribution of the number of rare events occurring within a fixed unit of time or space.
- **t-Distribution:** Similar in shape to the normal distribution but with thicker tails. Mainly used for mean inference when sample sizes are small.
- **Chi-squared Distribution:** Widely used in categorical data analysis, such as tests of independence or goodness-of-fit.

Major Statistical Testing Methods

- **t-test:** Checks if the difference between the means of two groups is statistically significant.
 - **Independent Samples t-test:** Compares means of two unrelated groups. (e.g., comparing scores between men and women)
 - **Paired Samples t-test:** Compares pre- and post-changes in the same group. (e.g., comparing weights before and after a diet)
- **ANOVA (Analysis of Variance):** Used to compare mean differences among three or more groups simultaneously.
- **Chi-squared Test:** Tests the association or independence between two categorical variables.
- **Correlation Analysis:** Checks how strong the linear relationship between two variables is.

- **Pearson Correlation Coefficient (r):** Values range from -1 (perfect negative correlation) to 1 (perfect positive correlation).

Regression Analysis

Method for constructing functional models to predict the value of a dependent variable (Y) through one or more independent variables (X).

- **Simple Linear Regression:** Models a linear relationship between one independent variable and the dependent variable.
- **Multiple Linear Regression:** Uses multiple independent variables for more precise prediction of the dependent variable.
- **Logistic Regression:** Models the probability of a specific event occurring when the dependent variable is categorical (e.g., pass/fail).

Statistical Errors

- **Type I Error (α):** Accidentally rejecting a null hypothesis that is actually true. (False Positive)
- **Type II Error (β):** Failing to reject a null hypothesis that is actually false. (False Negative)