I[ntroduction to Statistics and Probability](https://microsoft.github.io/Data-Science-For-Beginners/" \l "/1-Introduction/04-stats-and-probability/README?id=a-brief-introduction-to-statistics-and-probability)

### Probability and Random Variables

* **Probability**: A number between 0 and 1 representing how likely an event is.
* **Random Variables**: Variables representing outcomes; can be discrete (countable) or continuous (range of real numbers).
* **Sample Space**: Set of all possible outcomes.

### Probability Distribution

* **Discrete Distribution**: Described by a function P(X) giving the probability of each outcome.
* **Continuous Distribution**: Described by a probability density function p(x).

### Mean, Variance, and Standard Deviation

* **Mean (Expectation)**: The average value of a dataset.
* **Variance**: Measures the spread of data points around the mean.
* **Standard Deviation**: The square root of variance, indicating the extent of deviation from the mean.

### Mode, Median, and Quartiles

* **Mode**: The most frequently occurring value in a dataset.
* **Median**: The middle value dividing the dataset into two equal halves.
* **Quartiles**: Values dividing data into four parts; Q1 (25th percentile), Q3 (75th percentile).
* **Interquartile Range (IQR)**: The difference between Q3 and Q1.

### Normal Distribution

* **Normal Distribution**: A bell-shaped curve representing data distribution with a specific mean and standard deviation.
* **Central Limit Theorem**: The mean of a large number of independent, identically distributed variables approximates a normal distribution.

### Confidence Intervals

* **Confidence Interval**: An estimated range of values for a population parameter, calculated from a sample.

### Hypothesis Testing

* **Hypothesis**: A statement to be tested statistically.
* **t-Test**: A statistical test comparing the means of two groups to determine if they are significantly different.

### Covariance and Correlation

* **Covariance**: Measures how two variables change together.
* **Correlation**: Normalized covariance, indicating the strength and direction of the relationship between two variables.