

# Practical Statistics

- **Population**  
Keseluruhan data.
- **Sample**  
Bagian kecil/subset dari **population**.
- **Descriptive statistics**  
Suatu nilai yang merangkum data, tujuannya untuk menggambarkan keadaan data secara umum.
- **Exploratory Data Analysis (EDA)**  
Langkah awal dalam melakukan analisis data yang bertujuan untuk mengetahui karakteristik suatu data, biasanya dilakukan dengan membuat rangkuman **Descriptive statistics** dan visualisasi sederhana.
- **Inferential Statistics**  
Metode yang digunakan untuk menarik kesimpulan terhadap **population** dengan memanfaatkan informasi dari **sample**.
- **Measures of central tendency**  
Ukuran pemusatan data, menjelaskan titik sentral/pusat data.
- **Mean**  
Nilai rata-rata (total seluruh data dibagi dengan banyaknya data) dari data yang bertipe numerik/angka. Nilai rata-rata suatu **population** dinotasikan oleh  $\mu$ , sedangkan nilai rata-rata suatu **sample** dinotasikan oleh  $\bar{x}$ .
- **Quantile**  
Nilai yang membagi data terurut menjadi 4 bagian sama besar (—Q1—Q2—Q3—).
- **Median**  
Nilai yang membagi data terurut menjadi 2 bagian sama besar, sering disebut sebagai Q2.
- **Outlier**  
Data yang nilainya sangat ekstrim, sering disebut sebagai data yang anomali.
- **Trimmed mean**  
Nilai rata-rata yang diperoleh dari data terurut, yang sudah tidak terdapat **outlier** (**outlier** sudah dibuang).
- **Mode**  
Nilai yang paling sering muncul/keluar dari data.
- **Plot**  
Sebuah visualisasi yang menggambarkan/merepresentasikan suatu data, sering disebut sebagai **chart**.
- **Box plot**  
Plot yang digunakan untuk menggambarkan persebaran data yang bertipe numerik/angka.
- **Histogram**  
Plot yang digunakan untuk menggambarkan persebaran data yang bertipe numerik/angka.
- **Measures of spread**  
Ukuran penyebaran data, menjelaskan bagaimana persebaran suatu data.

- ***Variance***

Nilai yang menggambarkan seberapa bervariasi/beragamnya suatu data bertipe numerik/angka. Semakin besar nilai **variance** maka semakin beragam suatu data (heterogen), sedangkan semakin kecil nilai **variance** maka semakin sama/mirip setiap observasi pada data (homogen). Data yang observasinya bernilai sama, maka **variance** sama dengan 0.

- ***Standard deviation***

Rata-rata selisih/jarak setiap observasi dengan nilai **mean**, diperoleh dari akar kuadrat **variance**.

- ***Range***

Selisih/jarak antara observasi yang nilainya paling kecil (minimum) dengan yang nilainya paling besar (maksimum).

- ***IQR***

Selisih/jarak antara Q1 dan Q3.

- ***Standard error***

Nilai yang menggambarkan kedekatan antara **sample** dan **population**. Semakin kecil nilai **standard error** maka semakin dekat/representatif pula **sample** menggambarkan **population**, dan sebaliknya.

- ***Covariance***

Nilai yang menggambarkan hubungan (positif/negatif/tidak ada hubungan) antara dua variabel numerik. Namun **covariance** tidak dapat menggambarkan seberapa erat/kuat hubungan tersebut karena nilai **covariance** tidak memiliki batasan yang mutlak ( $-\infty$ ,  $+\infty$ ).

- ***Correlation***

Nilai yang menggambarkan keeratan hubungan (positif/negatif/tidak ada hubungan) antara dua variabel numerik. Nilai **correlation** mendekati 1 artinya kedua variabel berhubungan erat dan hubungannya positif, nilai **correlation** mendekati -1 artinya kedua variabel berhubungan erat dan hubungannya negatif, nilai **correlation** mendekati 0 artinya kedua variabel tidak saling berhubungan.

- ***Scatter plot***

Plot yang digunakan untuk menggambarkan persebaran dan hubungan (**correlation**) antara dua variabel bertipe numerik/angka.

- ***Data distribution***

Fungsi/bentuk yang menggambarkan persebaran data, sering disebut sebagai kumpulan nilai yang mungkin terjadi.

- ***Skewness***

Kecondongan dari suatu **data distribution** (data cenderung mengumpul di sisi bagian mana).

- ***Normal distribution***

Fungsi/bentuk yang menggambarkan persebaran data bertipe numerik/angka, bentuknya hampir menyerupai sebuah lonceng (simetris). Data yang memiliki distribusi normal cenderung mengelompok di sekitar **mean** (pusat lonceng).

- ***Central limit theorem***

Teorema ini menyatakan bahwa persebaran **sample** dengan distribusi tertentu yang diambil berulang kali dalam jumlah yang besar cenderung mengikuti **normal distribution**.

- ***Discrete variable***

Variabel yang berisi data bertipe numerik/angka bulat, contoh: jumlah siswa dalam satu kelas.

- ***Continuous variable***

Variabel yang berisi data bertipe numerik/angka desimal, contoh: tinggi badan.

- ***Probability mass function***

Peluang yang digunakan untuk menggambarkan kejadian pada **discrete variable**. Contoh: Dari total 50 siswa hanya 40 siswa yang hadir di kelas, maka peluang kehadiran siswa adalah  $40/50$  (0.8).

- ***Probability density function***

Peluang yang digunakan untuk menggambarkan kejadian pada **continuous variable**. Contoh: Peluang bertemu dengan orang yang memiliki tinggi badan minimal 165 cm di Algoritma.

- ***Standardization***

Proses untuk menyeragamkan skala data yang berbeda.

- ***Standard scores***

Nilai yang dihasilkan dari proses **Standardization**.

- ***z-score***

Standard score yang dihasilkan dari proses **standardization** dengan memanfaatkan **normal distribution**.

- ***Confidence interval***

Rentang nilai yang kemungkinan mengandung nilai parameter **population**, diperoleh dari informasi statistik **sample**.

- ***Margin of error***

Nilai yang menggambarkan besar kesalahan dari pengambilan **sample**.

- ***Significance level (alpha)***

Batas toleransi kesalahan/error yang diperbolehkan pada suatu pengujian.

- ***Hypothesis***

Dugaan sementara terhadap masalah pada **population** yang harus diuji kebenarannya.

- ***Hypothesis test***

Suatu metode yang digunakan untuk menarik kesimpulan/mengambil keputusan dari dua pernyataan/**hypothesis** yang saling bertolak belakang.

- ***Null hypothesis***

Menyatakan bahwa nilai parameter populasi sama dengan nilai yang ditentukan atau tidak memiliki pengaruh yang signifikan.

- ***Alternative hypothesis***

Menyatakan bahwa nilai parameter populasi tidak sama dengan nilai yang ditentukan atau memiliki pengaruh yang signifikan.

- ***z-test***

Salah satu uji statistik yang digunakan untuk menarik kesimpulan ketika melakukan **hypothesis test**, berdasarkan pada **normal distribution**.

- ***t-test***

Salah satu uji statistik yang digunakan untuk menarik kesimpulan ketika melakukan **hypothesis test**, berdasarkan pada **student's t distribution**.

- ***p-value***

Nilai/peluang kesalahan yang diperoleh dari hasil perhitungan statistik.