

 ibimbing



 Kampus
Merdeka
INDONESIA JAYA

python data analyst

linkedin : <https://www.linkedin.com/in/ali-murtadho>

telegram : t.me/alimurtadho_id

github : [@alimurtadho](https://github.com/alimurtadho)

medium : medium.com/@dho_aldho



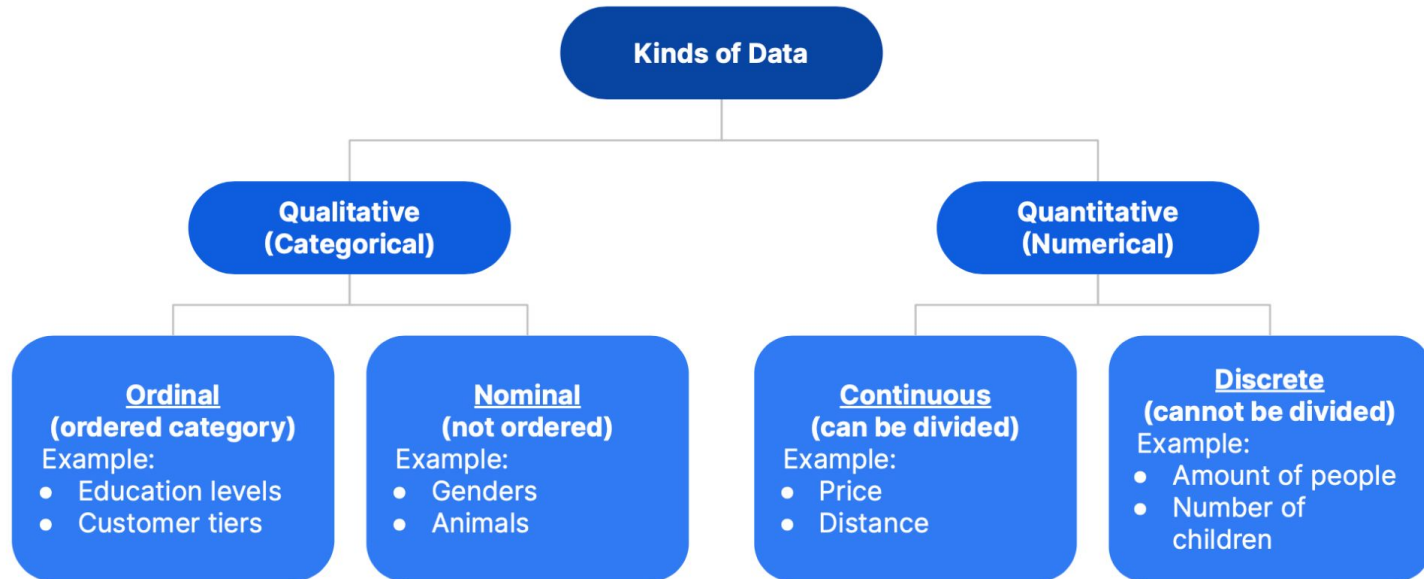
Introduction to Statistics

What is Statistics Descriptive?

- Descriptive statistics atau statistik deskriptif adalah cabang statistik yang berfokus pada peringkasan dan penggambaran karakteristik dari suatu kumpulan data. Ini melibatkan penggunaan berbagai teknik statistik dalam meringkas data yang bermakna.
- Tujuan dari statistik deskriptif untuk memberikan informasi tentang kecenderungan sentral (central tendency), variabilitas, dan distribusi data.

Kind of Data

Sebelum bahas metode statistik, kita perlu memahami jenis data yang dapat ditangani. Hal ini penting karena metode yang digunakan bergantung pada jenis datanya.



Measurement of Central Tendency

Mean, Median, Modus

- **Mean:** Mean atau rata-rata dihitung dengan menjumlahkan semua nilai dalam kumpulan data dan kemudian membaginya dengan jumlah total nilai. Rata-rata dipengaruhi oleh nilai ekstrim dan memberikan ukuran rata-rata aritmatik dari data. Rata-rata banyak digunakan dan cocok untuk kumpulan data dengan distribusi simetris (berbentuk normal).
- **Median:** Median merupakan nilai tengah dalam kumpulan data yang sudah diurutkan. Ini memisahkan setengah data yang lebih tinggi dari setengah data yang lebih rendah. Jika jumlah pengamatan genap, median adalah rata-rata dari dua nilai tengah. Median tidak terpengaruh oleh nilai ekstrim dan cocok untuk kumpulan data dengan distribusi miring atau outlier.
- **Mode:** Modus merupakan nilai yang paling sering muncul dalam kumpulan data. Ini bisa berupa nilai tunggal atau beberapa nilai jika terjadi frekuensi yang sama (seri). Modus berguna untuk data kategorikal atau diskrit, dan juga dapat diterapkan pada data kontinu dengan mengelompokkan nilai ke dalam interval (rentang).

Oleh karena itu, penting untuk memahami karakteristik dan keterbatasan dari masing-masing ukuran tendensi sentral (central tendency) sebelum memilih mana yang akan digunakan dalam analisis.

How Do We Know the Spreadness of the Data?

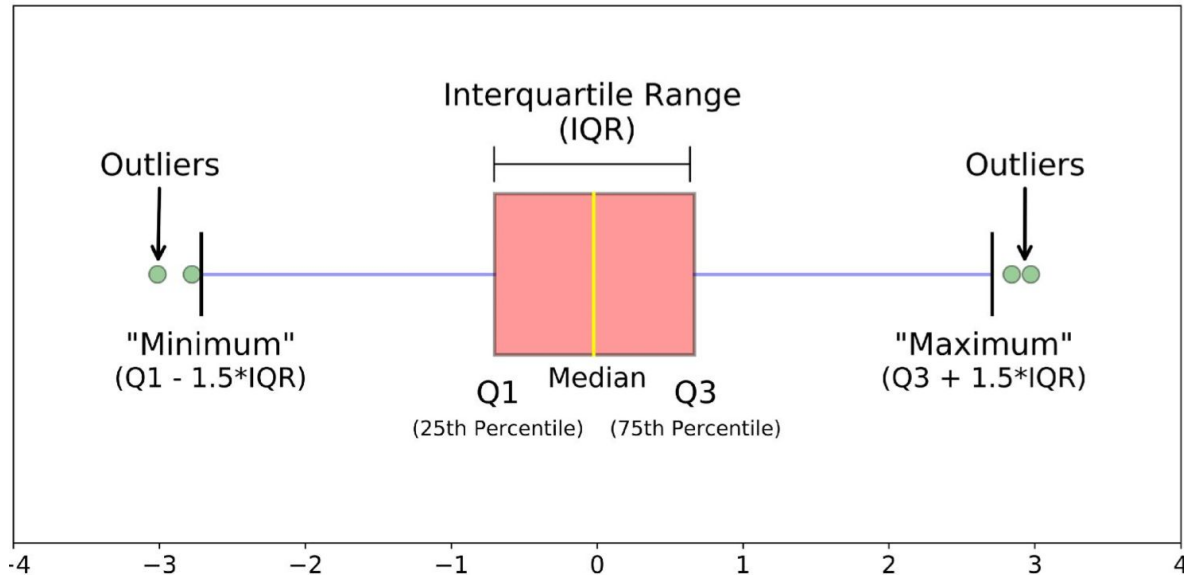
Untuk mengetahui tingkat "sebaran" atau "variabilitas" data, kita dapat menggunakan berbagai ukuran statistik. Berikut adalah beberapa metode umum:

- **Range:** Range atau rentang menyediakan cara sederhana untuk menilai penyebaran data. Dihitung dengan mengurangi nilai terkecil dari nilai terbesar dalam kumpulan data. Namun, rentang hanya mempertimbangkan nilai ekstrim dan mungkin tidak menangkap gambaran keseluruhan variabilitas.
- **Interquartile Range (IQR):** IQR mengukur penyebaran bagian tengah data. Ini adalah selisih antara kuartil ketiga (persentil ke-75) dan kuartil pertama (persentil ke-25) dari kumpulan data. Dengan berfokus pada 50% data di tengah, IQR kurang sensitif terhadap nilai ekstrim.
- **Variance:** Variansi menghitung deviasi kuadrat rata-rata dari setiap titik data dari mean. Ini memberikan ukuran penyebaran keseluruhan data, dengan mempertimbangkan semua nilai. Namun, variansi berada dalam satuan kuadrat, sehingga lebih sulit untuk ditafsirkan secara langsung.
- **Standard Deviation:** Deviasi standar adalah akar kuadrat dari variansi. Ini mewakili jumlah tipikal atau rata-rata penyimpangan titik data dari mean. Ini adalah ukuran penyebaran yang banyak digunakan karena berada dalam unit yang sama dengan data asli, sehingga membuatnya lebih mudah untuk ditafsirkan.

Measurement of Dispersion

Box and Whisker Plot

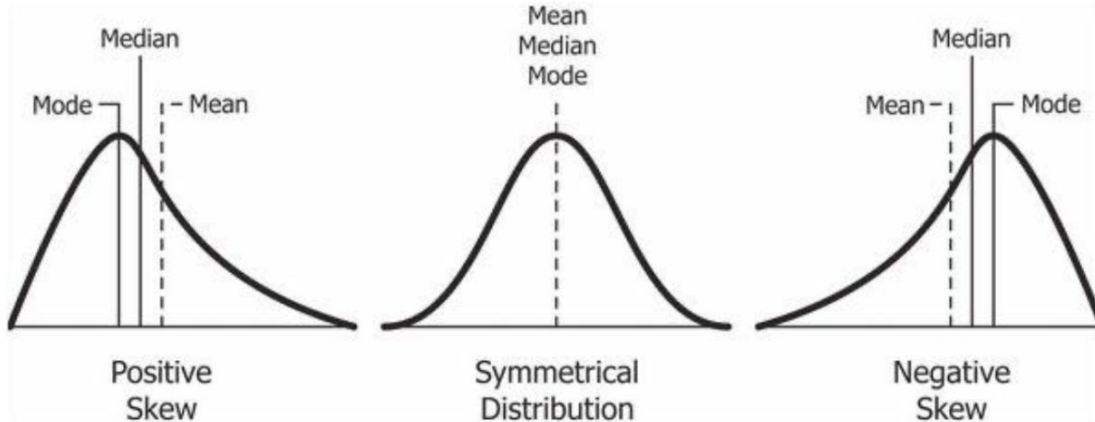
Whisker Plot dikenal sebagai Box Plot adalah representasi grafis dari distribusi sebuah kumpulan data. Plot ini memberikan ringkasan tentang penyebaran, pusat, dan adanya nilai outlier (nilai yang jauh berbeda dari data lainnya) dalam data tersebut. Plot ini terdiri dari lima komponen utama: the minimum value, the first quartile (Q1), the median (Q2), the third quartile (Q3), and the maximum value.



Normality Test - Skewness

Ada banyak metode statistik yang menggunakan asumsi bahwa data berdistribusi normal. Lalu, bagaimana cara kita mengetahui kenormalan data kita? Kita dapat melakukan uji normalitas pada data kita menggunakan metrik skewness dan kurtosis.

Skewness adalah metrik yang menggambarkan kesimetrisan data.



<Normal Distribution>

Jika skewness adalah 0, data tersebut simetris sempurna, meskipun ini sangat tidak mungkin untuk data dunia nyata.

- Jika skewness kurang dari -1 atau lebih besar dari 1, distribusinya sangat miring.
- Jika skewness antara -1 dan -0.5 atau antara 0.5 dan 1, distribusinya agak miring.
- Jika skewness antara -0.5 dan 0.5, distribusinya kira-kira simetris.

1. hands on python



Saatnya Quiz

Questions?