



FULLSTACK INTENSIVE BOOTCAMP

Data Analyst

MySkill

Let's Learn Basic Statistic

Oleh: Devita Eka Ratriningrum

JOIN THE BEST UPSKILLING COMMUNITY
WITH ME at myskill.id/bootcamp

MINI PORTOFOLIO

Course Summary



Statistik merupakan nilai-nilai ukuran data (fakta) yang mudah untuk dimengerti, sedangkan statistika adalah ilmu yang berkaitan dengan cara pengumpulan, pengolahan, analisis, dan penarikan kesimpulan atas data.

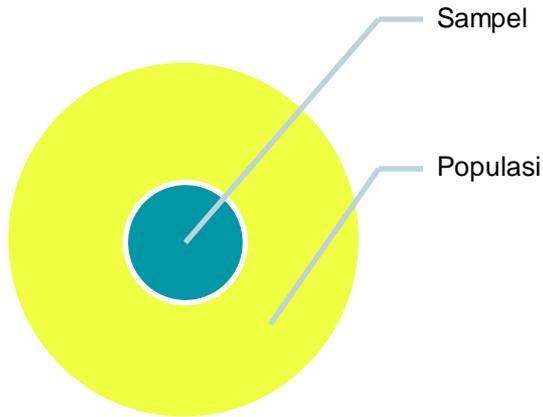
Terdapat 3 jenis statistik secara general :

1. Populasi dan Sampel
2. Statistik Deskriptif
3. Statistik Inferensial



1. Populasi dan Sampel

Populasi dan sampel merupakan sebuah metode untuk objek suatu penelitian yang akan menghasilkan suatu data yang diperoleh. Namun, jika ditelaah kembali, sampel merupakan bagian kecil dari populasi. Metode ini digunakan untuk mendapatkan data secara efektif dari segi tujuan dan juga efisien dari segi waktu dan biaya.



Populasi sendiri merupakan keseluruhan objek penelitian yang menjadi pusat perhatian dan menjadi sumber data penelitian.

Sampel merupakan bagian dari populasi yang dipilih menggunakan aturan tertentu yang akan digunakan nantinya untuk mengumpulkan informasi atau data yang menggambarkan sifat atau ciri yang dimiliki oleh populasi.

Jadi urutan dalam menggunakan metode populasi dan sampel adalah menentukan populasi yang ingin diuji, lalu mencari beberapa sampel dan data dari hasil pengukuran atau pengamatan akan diperoleh

1. Populasi dan Sampel

Ada beberapa tipe data yang digunakan dalam penelitian dengan metode populasi dan sampel:

1. Kualitatif (Kategorikal)

- *Nominal*, merupakan nilai yang tidak mengenal urutan dan operasi aritmatika atau tidak dapat diurutkan
Contoh: Married / Unmarried, Kewarganegaraan
- *Ordinal* atau Rank, merupakan nilai yang mengenal urutan dan operasi aritmatika atau dapat diurutkan
Contoh: Sangat Tidak Setuju (1), Setuju (2), Sangat Setuju (3)

2. Kuantitatif (Numerical)

- *Diskrit*, merupakan nilai yang berhubungan dengan proses menghitung dan pengamatan atas himpunan tertentu
Contoh: Banyaknya antrian bioskop, Kecepatan pesawat
- *Kontinu*, merupakan nilai yang didasarkan pada suatu selang atau interval sehingga meliputi semua bilangan riil.
Contoh: Suhu

1. Populasi dan Sampel

Dalam melakukan sampling, ada beberapa metode yang dapat kita terapkan, yaitu:

1. **Random Sampling**, merupakan proses dimana seluruh pengambilan sampel dilakukan secara independent dari anggota populasi lainnya
Contoh: mengambil 1 buah dalam sebuah keranjang yang berisikan berbagai macam buah
2. **Systematical Sampling**, merupakan proses dimana menentukan terlebih dahulu ukuran sampel kemudian menentukan elemen populasi dalam beberapa urutan dan akan dipilih secara berkala sampelnya dari daftar.
Contoh: mengambil setiap pelanggan dengan kelipatan 10 (pelanggan ke-10, pelanggan ke-20, pelanggan ke-30, dst.)
3. **Stratified Sampling**, merupakan metode sampling dengan cara membagi populasi menjadi beberapa karakteristik lalu diambil sampelnya secara acak dalam setiap kategori yang sudah di kategorikan.
Contoh: mengambil 50% data dari 40% data perempuan dan 50% data dari 60% data laki-laki

2. Statistik Deskriptif

Statistik deskriptif merupakan metode yang berhubungan dengan pengolahan data untuk mendapatkan suatu informasi. Dalam statistic deskriptif kita harus mengetahui macam-macam karakter distribusi guna memahami variabilitas data, mengetahui sejauh mana data tersebar, mengetahui pattern yang terbentuk hingga ketidakpastian atau risiko dari analisis tersebut.

Berikut macam-macam karakter-karakter distribusi:

1. **Ukuran Pemusatan**, merupakan statistik yang memberikan informasi dimana data terkumpul dengan ukuran atau jumlah tertentu

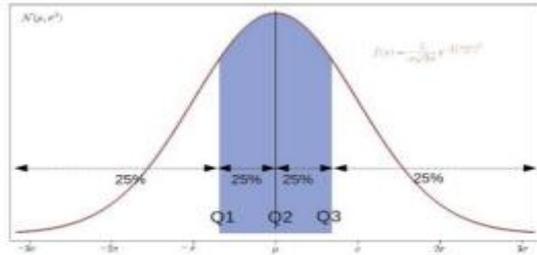
Untuk mengetahui seberapa terpusatnya data, kita dapat menggunakan beberapa parameter seperti:

- Mean, digunakan untuk mencari rata-rata suatu data
- Median, digunakan untuk mencari nilai tengah dari suatu data
- Modus, digunakan untuk mencari nilai yang paling sering muncul pada suatu data

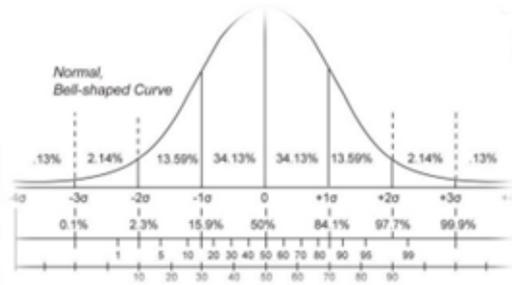


2. Statistik Deskriptif

- Kuartil, nilai yang terbagi menjadi sebuah dataset yang dibagi menjadi empat bagian sama besar.



- Percentile, nilai yang membagi sebuah dataset menjadi 10 dan 100 bagian yang sama besar



2. Statistik Deskriptif

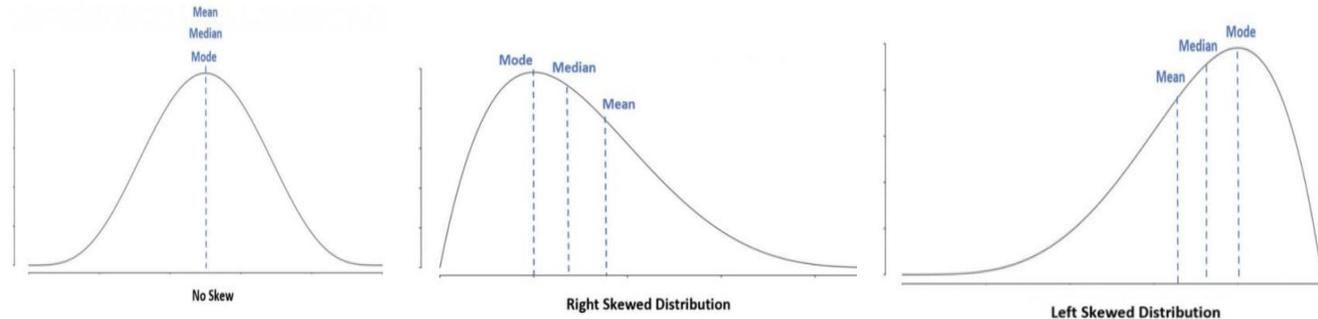
2. **Ukuran Penyebaran**, merupakan statistik yang memberikan informasi bagaimana data tersebar di sekitar pusat data

Untuk mengetahui seberapa tersebar data, kita dapat menggunakan beberapa parameter seperti:

- Range, digunakan untuk mengetahui nilai perbedaan nilai tertinggi dan nilai terendah
- Standard Deviasi (std.dev), digunakan untuk mengukur seberapa jauh rata-rata dari sekelompok data dari nilai – nilai dalam suatu kelompok data
- Variansi, digunakan untuk mengetahui nilai rata-rata dari perbedaan setiap data yang dihasilkan oleh mean

Penyebaran data memiliki karakter – karakter yang dapat digambarkan menjadi grafik

- Simetris
- Left Skewed
- Right Skewed
- Puncak



2. Statistik Deskriptif

3. **Kemiringan atau Skewness**, merupakan tingkat ketidaksimetrisan atau kejauhan simetri dari sebuah distribusi.

Terdapat beberapa kriteria untuk mengetahui model distribusi dari koefisien kemiringan sebagai berikut:

- Jika koefisien kemiringan < 0 , maka bentuk distribusinya negative (ekor bagian kiri lebih Panjang).
- Jika koefisien kemiringan $= 0$, maka bentuk distribusinya simetrik.
- Jika koefisien kemiringan > 0 , maka bentuk distribusinya positif (ekor bagian kanan lebih Panjang).

4. **Kelancipan atau Kurtosis**, merupakan tingkat kepuncakan dari sebuah distribusi yang biasanya diambil secara relatif terhadap suatu distribusi normal.

Terdapat beberapa kriteria untuk mengetahui model distribusi dari koefisien kurtosis sebagai berikut :

- Jika koefisien kurang dari 0,263 maka distribusinya adalah platikurtik
- Jika koefisien kurtosis sama dengan 0,263 maka distribusinya adalah mesokurtic
- Jika koefisien kurtosis lebih dari 0,263 maka distribusinya adalah leptokurtic.



3. Statistik Inferensial

Statistik inferensial merupakan metode yang menggunakan rumus statistic dan hasil perhitungannya dijadikan sebagai pedoman dalam pembuatan kesimpulan secara umum. Dengan menggunakan metode statistik inferensial, kita dapat membuat prediksi dari data yang dihasilkan.

Ada beberapa metode pengujian yang dapat digunakan dalam analisis inferensial, yaitu:

1. **Analisis regresi linier**, digunakan untuk memahami hubungan antara dua variabel dari kumpulan data, dimana salah satu variabelnya merupakan variabel dependen dan yang lainnya adalah variabel independen yang dapat digunakan lebih dari satu.
2. **Analisis varians**, digunakan untuk menguji dan menganalisis perbedaan antara dua atau lebih mean dari kumpulan data.
3. **Uji-T**, digunakan untuk membandingkan rata-rata dua kelompok dan memahami jika satu sama lain berbeda.
4. **Analisis korelasi**, digunakan untuk memahami sejauh mana dua variabel saling bergantung. Kita dapat melihat apakah variabel memiliki korelasi yang kuat atau lemah. Korelasi bisa negatif atau positif tergantung pada variabelnya.



Follow me!

LinkedIn : [linkedin.com/in/ratriiratri](https://www.linkedin.com/in/ratriiratri)

Bootcamp Data Analysis
by @myskill.id

