



- Jenis-jenis analisa statistik
- Syarat-syarat Data yang baik
- Data Berdistribusi Normal
- Fungsi Perintah Data Statistik dengan Python
- Praktek Data Statistik dengan Jupyter Notebook.

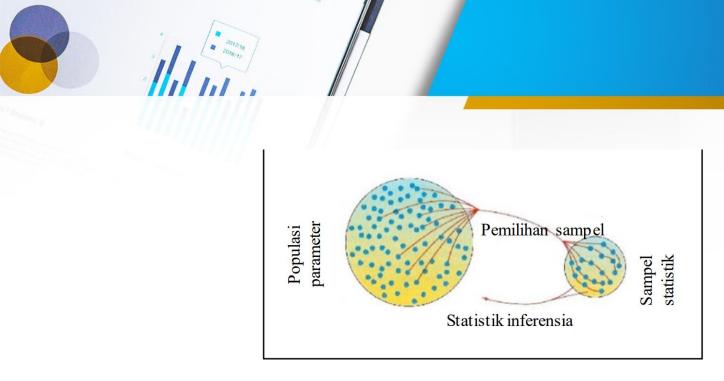


# Analytic Approach

 Sebagai Data Scientist kita perlu mempresentasikan bagian Analisis Data kepada audience [team Tech dan Business Owner/Decision] supaya memahami kualitas dari hasil data kita dan dapat dipertanggungjawabkan hasil kinerja dari data nya.



- Metoda Statistika merupakan prosedur yang digunakan untuk mengumpulkan, menyajikan, menganalisis, dan menginterpretasikan data. Para Peneliti dan ilmuwan menggunakan statistika sebagai alat bantu untuk membuat keputusan ilmiah yang dapat dipertanggungjawabkan secara logis
- Populasi adalah keseluruhan elemen atau unsur yang akan kita teliti tersebut. Sampel adalah sebagian anggota populasi dimana pengukuran yang akan dilakukan.



Gambar . Hubungan Populasi Parameter dengan Sampel dan Populasi



#### Statistika dan Data

Syarat-syarat Data yang baik adalah:

- 1. Realibel (dapat diandalkan), yang meliputi:
- a. Data harus objektif (sesuai keadaan yang benar). Data yang tersedia dapat menggambarkan keadaan yang sesungguhnya dari objek yang diamati. Misalnya, produksi ikan menurun dalam lima tahun terakhir dilaporkan meningkat, walaupun data tertera dalam statistika produksi; tetap saja tidak objektif.
- b. Data harus dapat mewakili (representative). Data yang tersedia mampu mewakili populasi (lingkup yang lebih besar) yang dimaksud. Misalnya : keadaan ekonomi nelayan "tidak miskin". Kesimpulan ini diambil berdasarkan survei yang respondennya pemilik kapal saja.
- c. Kesalahan baku (*Standard error*) kecil. Data dapat dikatakan baik apabila kesalahan bakunya kecil.

#### 2. Bermanfaat dan berguna, yang meliputi:

- a. Data relevan. Data yang dikumpulkan harus ada hubungannya dengan masalah yang akan dipecahkan. Misalnya, pemerintah akan meningkatkan pendapatan nelayan. Untuk memecahkan persoalan tersebut maka perlu diketahui tentang kepemilikan alat tangkap dan potensi sumber daya laut untuk mata pencaharian alternative.
- b. Data harus tepat waktu (*up to date*). Data yang dikumpulkan bersifat kekinian sehingga jika diambil sebagai bahan untuk pengambilan keputusan/perencanaan akan tepat keputusannya.





#### Jenis Analisa Statistik

- 1. Statistik Deskriptif
- a. Bertujuan untuk mendapatkan ringkasan umum dari data yang dimiliki
- b. Bisa dilakukan dengan membuat visulisasi data atau tabel ringkasan data [Insights and Ideas]
- 2. Statistik Inferensial
- Bertujuan untuk membuat kesimpulan dari data yang dimiliki apakah data mewakili populasi yang sesungguhnya sehingga kesimpulannya bias digeneralisasi.
- b. [Conclusions and Predictions]



# Statistika Deskriptif

Statistika Deskriptif atau ukuran pemusatan data adalah nilai-nilai yang dapat menggambarkan posisi pusat data. Biasanya yang digunakan adalah mean, median, modus, dll.

1. Rata-rata = pusat massa (centroid) dari suatu data sehingga simpangan (deviasi) kiri akan sama besar dengan simpangan kanannya. Rata-rata ini hanya berlaku untuk data kuantitatif (numerik), tidak bias untuk data kualitatif atau kategorik. Rata-rata memiliki sifat yaitu sangat sensitive terhadap pencilan (outlier). Artinya, jika ada pencilan maka rata-rata dapat berubah drastis. Rumusnya:

Rata-Rata = 
$$\frac{\text{Jumlah semua nilai}}{\text{Banyaknya nilai}}$$
 atau  $\overline{x} = \frac{x_1 + x_2 + x_3 + ... + x_n}{n} = \frac{1}{n} \sum_{i=1}^{n} x_i$ .

anna rount

Perhatikan data berikut.

Data I:

2, 3, 6, 9.

Rata-ratanya adalah:

$$\overline{x} = \frac{2+3+6+9}{4} = \frac{20}{4} = 5$$

Data II:

2, 3, 6, 25

Rata-ratanya adalah:

$$\overline{x} = \frac{2+3+6+25}{4} = \frac{36}{4} = 9$$



Modus adalah nilai data yang paling sering muncul atau nilai data yang mempunyai frekuensi terbesar.

Contoh: 4,5,5,5,6,7,8,8,9,10 Mempunyai modus nya adalah 5.

Median adalah suatu nilai yang membagi data yang telah diurutkan menjadi dua bagian yang sama banyak.

Contoh: 5,5,5,6,6,7,8,9,10,11,12. Mempunyai Median nya adalah 7.





### Statistik Inferensial

Statistika Inferensial (ukuran penyebaran data) menunjukkan variaasi atau keragaman data. Semakin besar nilainya menunjukkan bahwa data semakin menyebar. Ukuran yang dipakai yaitu Varians dan simpangan baku.

Varians dan simpangan baku merupakan ukuran keragaman yang telah memperhatikan keseluruhan data. Varians adalah jarak kuadrat setiap data terhadap rata-ratanya dibagi dengan banyaknya pengamatan. Sedangkan simpangan baku merupakan akar dari varians.



Jika data berasal dari data populasi maka:

Varians yang dinotasikan dengan  $\sigma^2$  dengan rumus  $\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$ .

Simpangan baku (standard deviation) yang dinotasikan dengan  $\sigma$  dengan rumus  $\sigma = \sqrt{\frac{\sum (x-\mu)^2}{N}}$ , di mana N adalah banyaknya data.

Jika data berasal dari data sampel maka:

Varians yang dinotasikan dengan s<sup>2</sup> dengan rumus:  $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2$ .

Simpangan baku (*standard deviation*) yang dinotasikan dengan s dengan rumus  $s = \sqrt{s^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}$ , di mana n adalah banyaknya data.

# 200718 200718

Contoh 1.10.

Dari data pada contoh sebelumnya: 2, 3, 6, 9. Telah diperoleh rata-ratanya adalah  $\bar{x} = 5$ . Jika dianggap data tersebut data populasi maka Variansnya adalah:

$$\sigma^2 = \frac{(2-5)^2 + (3-5)^2 + (6-5)^2 + (9-5)^2}{4} = 7,5.$$

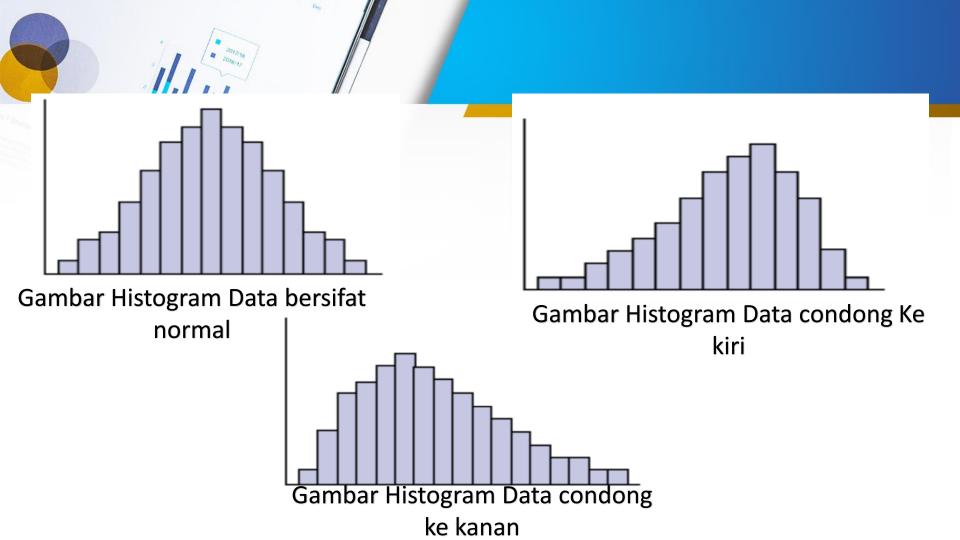
dan simpangan bakunya adalah:

$$\sigma = \sqrt{7.5} = 2.74.$$

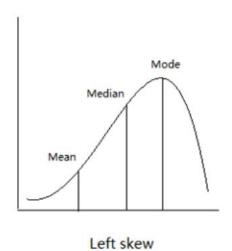


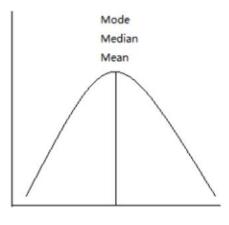
Sekumpulan data dikatakan normal jika histogramnya memiliki sifatsifat berikut :

- a. Tertinggi pada tengah-tengah interval
- Bergerak dari tengah-tengah interval ke masing-masing arah kanan dan kiri, tingginya menurun sehingga seperti bentuk lonceng
- c. Histogram adalah simetris di sekitar tengah intervalnya.

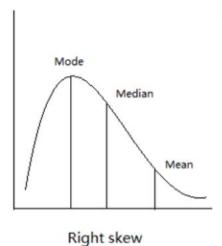








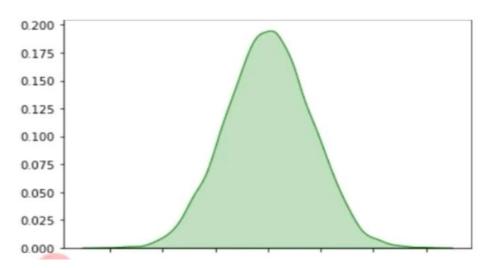
Normal Distribution





#### Distribusi Normal.

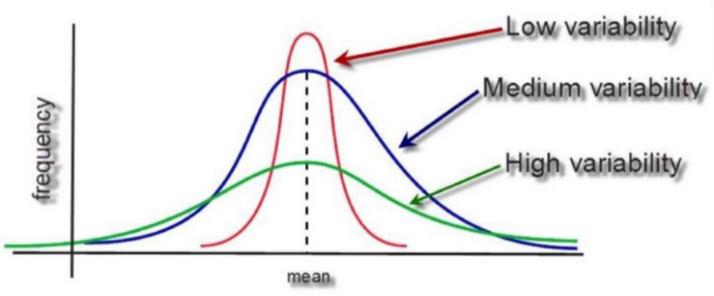
Misalkan kita ingin mengukur tinggi mahasiswa di suatu PTN tertentu, diketahui bahwa rata-rata tinggi mereka adalah 170 cm, Kemudian kita pilih mahasiswa secara random/acak, maka besar kemungkinan tinggi dari mahasiswa itu tersebut tidak jauh 170 cm, 169 cm, dan 171 cm, maka kemungkinan kecil Mahasiswa yang dipilih tadi itu memiliki tinggi yang berbeda ekstrim misalkan 300 cm, 50 cm, Maka apabila data seperti itu. bisa dikatakan datanva berdistribusi Normal.



Sumbu x = Random Variable Sumbu y = Probability Density Function (PDF)



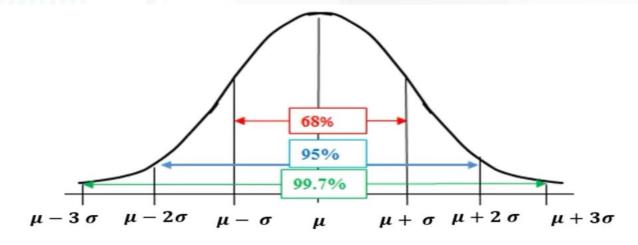
#### Kurva Distribusi Normal



Distribusi Normal : Semakin besar ukuran besar dari kurva nya, maka nilai standard deviasinya akan semakin besar  $(\sigma)$ . Maka semakin besar kurvanya, maka data yang diperoleh tidak normal/tidak baik.



$$\mu = 168 \text{ cm}$$
 $\sigma = 10 \text{ } cm$ 





### **Empirical Rule**

Jika sekumpulan data mendekati normal dengan  $\bar{x}$  dan s masing — masing adalah ratarata ( $\mu$ ) dan simpangan baku sampel ( $\sigma$ ),

#### maka berlaku bahwa :

- 1. Sekitar 68 % dari data tersebut terletak di dalam  $\bar{x} \pm s$ .
- 2. Sekitar 95 % dari data tersebut terletak di dalam  $\bar{x} \pm 2s$ .
- 3. Sekitar 99,7 % dari data tersebut terletak di dalam  $\bar{x} \pm 3s$ . (Outlier Data) Data yang tidak diperbolehkan untuk penelitian)





# Fungsi Statistika pada Python

- Import matplotlib.pyplot as plt
- Import numpy as np
- Import scipy.stats as stats
- Import seaborn as sns
- Sns.set()



# Let's Practice



## Challenge

#### Buatlah Resume yang berisi:

- Fungsi Statistik Deskriptif dan Inferensial menggunakan angka dan nilai tersendiri
- Buatlah deskripsi tentang statistik deskriptif dan inferensial menggunakan dataset yang ada di www.kaggle.com



#### Reference

Ratananingshi, M.Si & Dr. Sunaryo, Sony. Jakarta. *Statistika dan Ruang lingkupnya*: Jakarta: Universitas Terbuka Press

https://www.youtube.com/watch?v=BW9KLBNCmc&list=PLIeJsyt\_FUfKAB6K\_luK65

klldTZ0jlRA (Matematika Statistik dan Pendahuluan)

https://www.youtube.com/watch?v=b\_7o0ij49ok (Distribusi Normal dengan

Python)