

NORMAL DISTRIBUTION



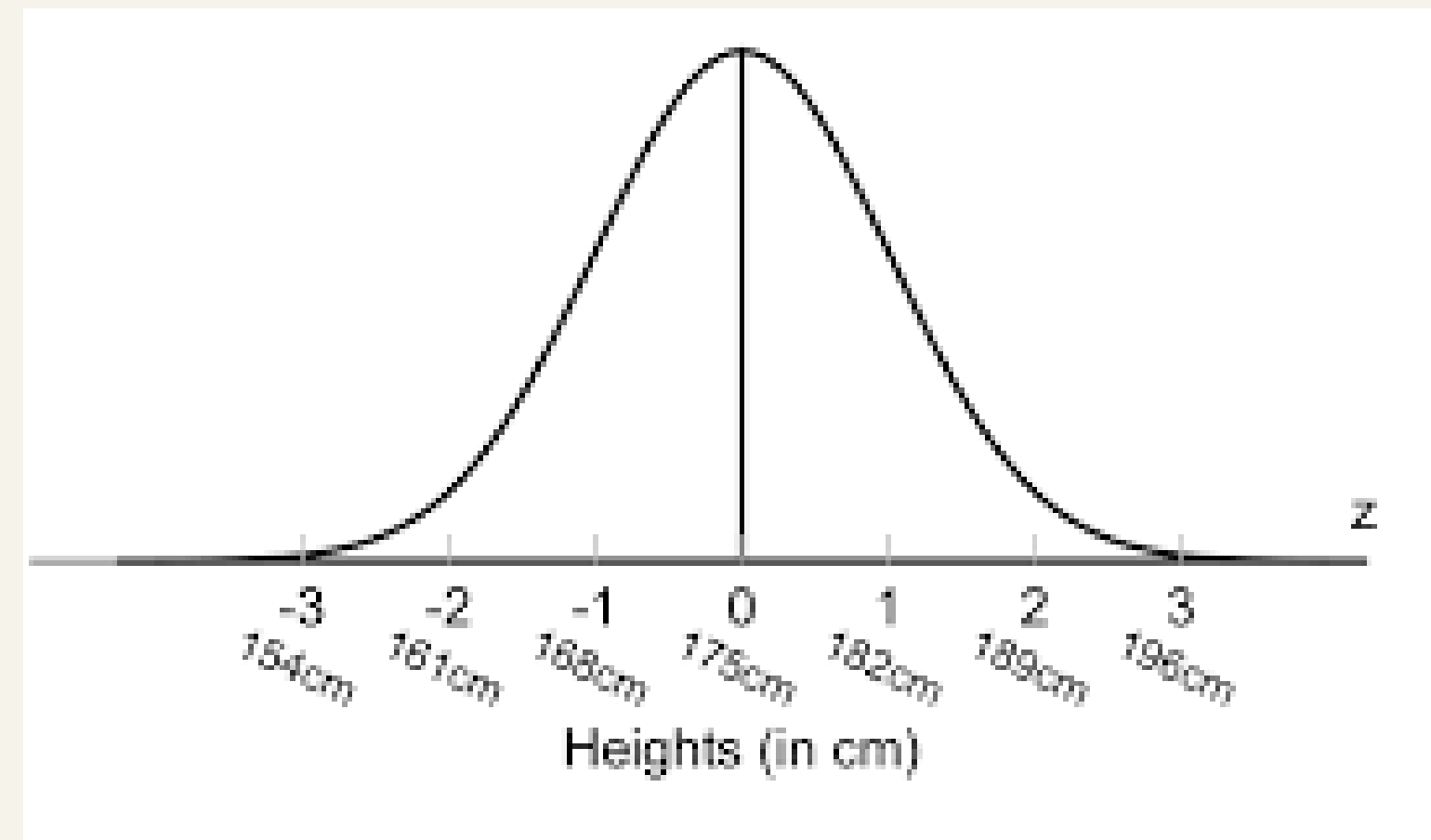
Pengantar Data Sains

Benhard Leroy Situmorang (Statistika/20244920002)

PENGERTIAN

2

Normal Distribution atau disebut juga dengan Distribusi Normal adalah sebuah distribusi probabilitas kontinu yang bentuknya itu digambarkan dengan kurva lonceng yang simetris. Distribusi ini biasanya digunakan untuk menggambarkan sebaran data yang ada dari variabel kontinu seperti tinggi badan, berat badan, atau tekanan darah.



SEJARAH

3

Sejarah pertama dalam Distribusi Normalitu dimulai pada tahun 1733 ketika ilmuwan De Moivre mengembangkan bentuk matematis dari kurva normal yang menjadi dasar statistika induksi. Lalu, distribusi itu dikembangkan oleh Carl Friedrich Gauss pada tahun 1794 dimana Gauss mengembangkan sebuah teori distribusi dengan fungsi eksponensial dua parameter.

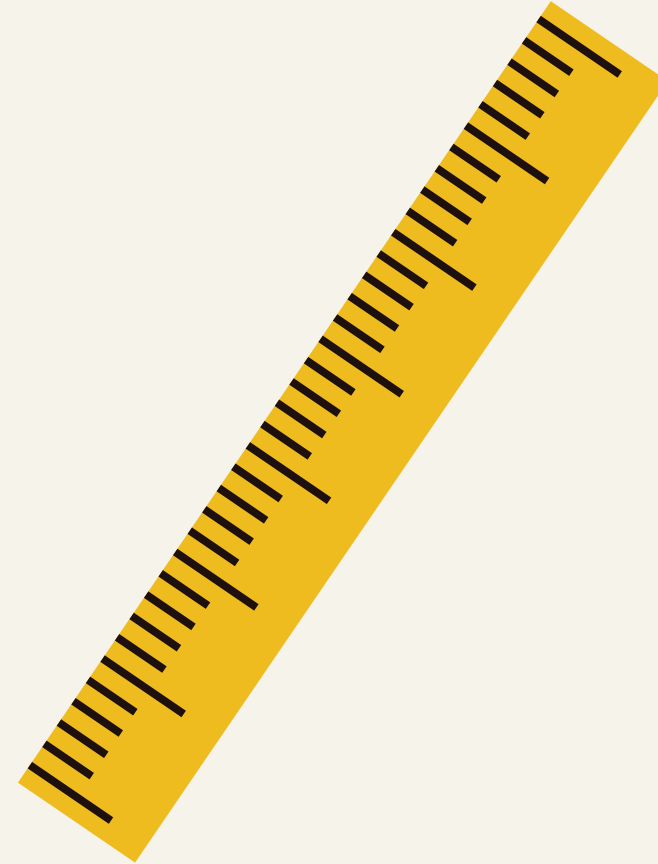


PARAMETER

4

Distribusi ini memiliki dua parameter yaitu:

1. Mean (μ) = Mean menentukan nilai penyebaran data berpusat dan berada di puncak kurva.
2. Standar deviasi (σ) = Standar deviasi mengukur nilai penyebaran data dari mean dan menunjukkan seberapa banyak data yang menyimpang dari mean. Semakin besar standar deviasi, maka data itu lebih tersebar, begitu sebaliknya.

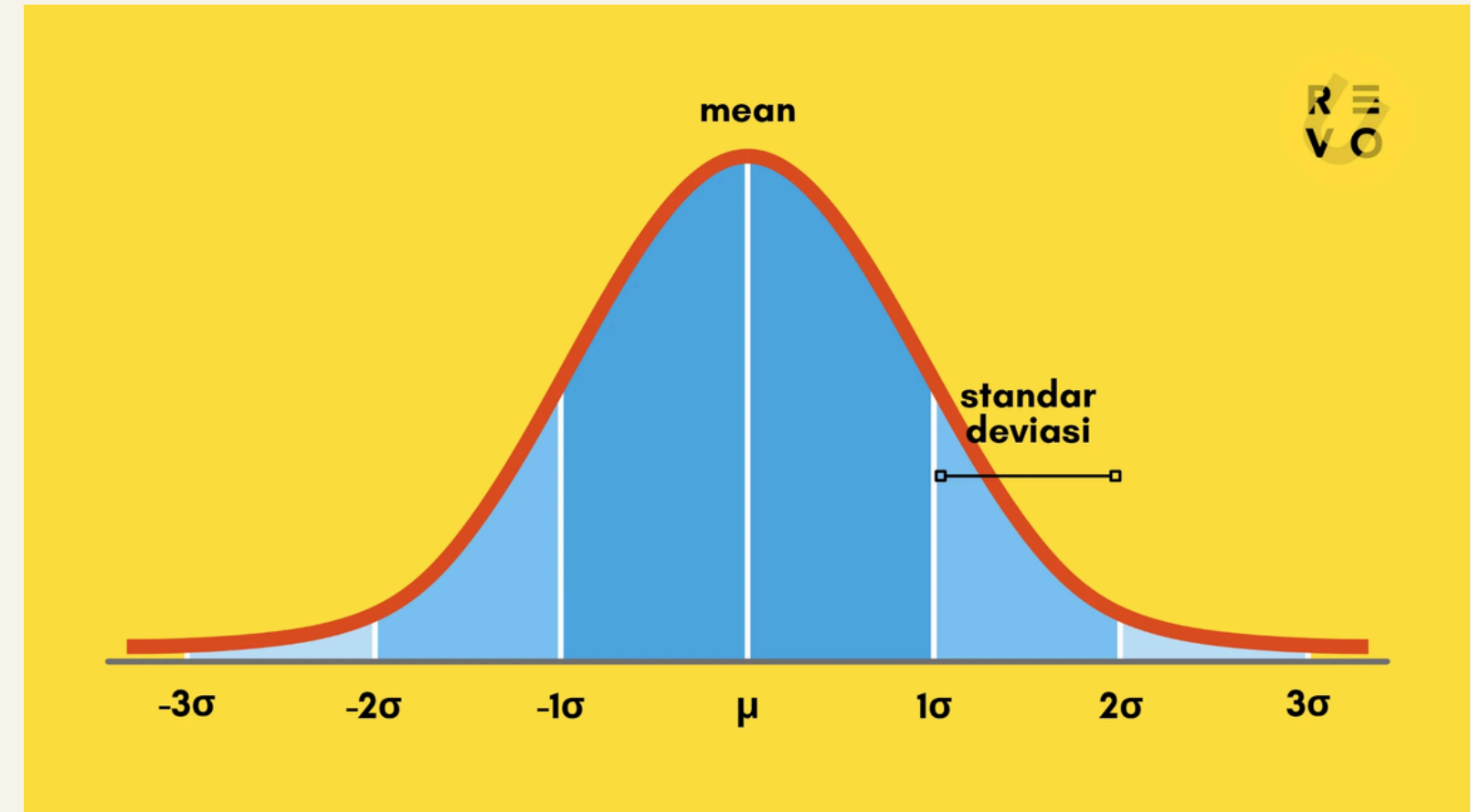


KARAKTERISTIK

5

Berikut adalah karakteristik dari Distribusi Normal:

1. Kurva berbentuk lonceng
2. Titik puncak berada di sekitar nilai mean
3. Nilai mean, median, dan modus yang sama
4. Tidak cocok untuk data dengan distribusi miring
5. Cocok dengan Variabel Kontinu



ATURAN EMPIRIS

6

Dalam sebuah data distribusi itu ada juga berbagai aturan yang harus ada untuk mendapatkan dan mengvisualisasikan sebuah distribusi. Dan di Distribusi Normal, terdapat juga aturan empiris yang muncul.

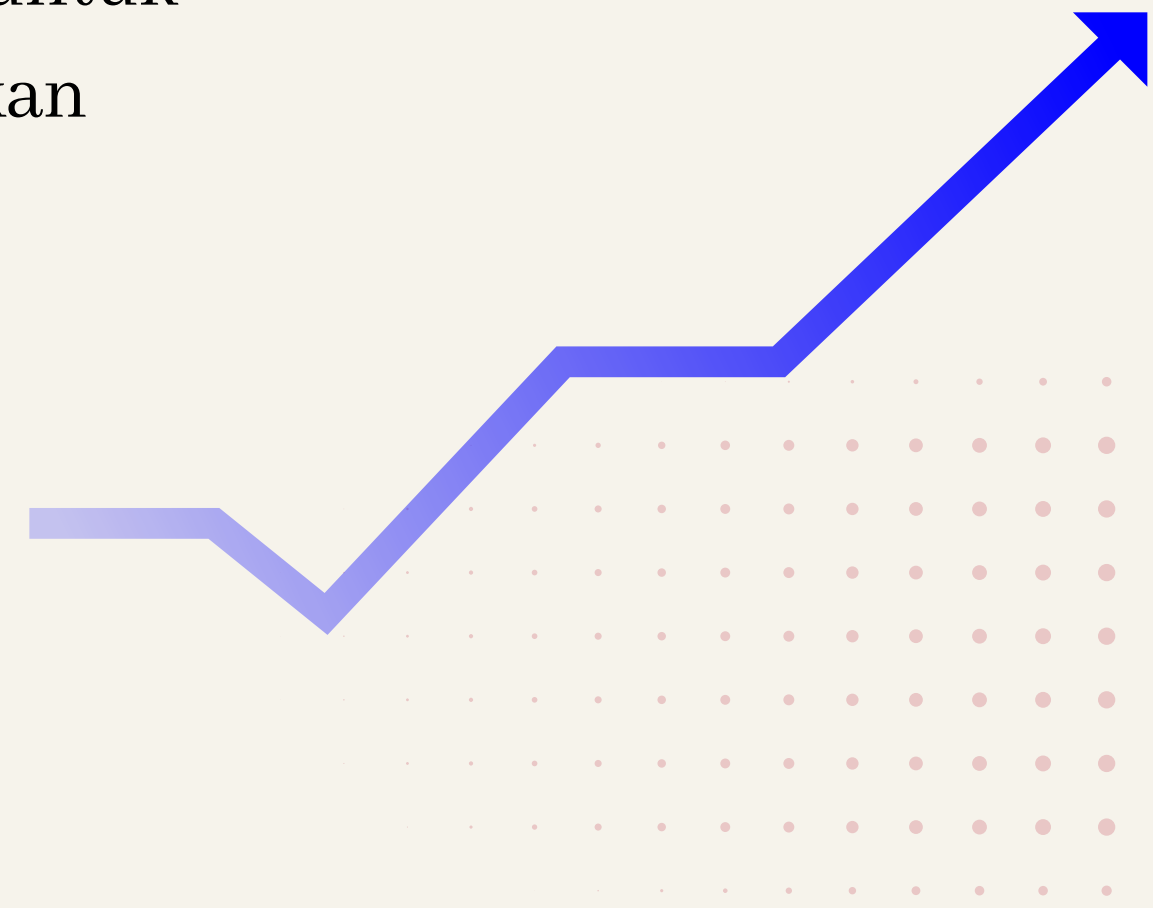
Aturan empiris atau aturan 68-95-99.7, adalah sebuah cara yang bermanfaat untuk memahami distribusi data dalam distribusi normal. Aturan ini dapat digunakan untuk memperkirakan probabilitas dan membuat prediksi berdasarkan karakteristik populasi.

Aturan:

Sekitar 68% dari data termasuk dalam satu standar deviasi dari rata-rata.

Sekitar 95% data termasuk dalam dua standar deviasi dari rata-rata.

Sekitar 99,7% data termasuk dalam tiga standar deviasi dari rata-rata.

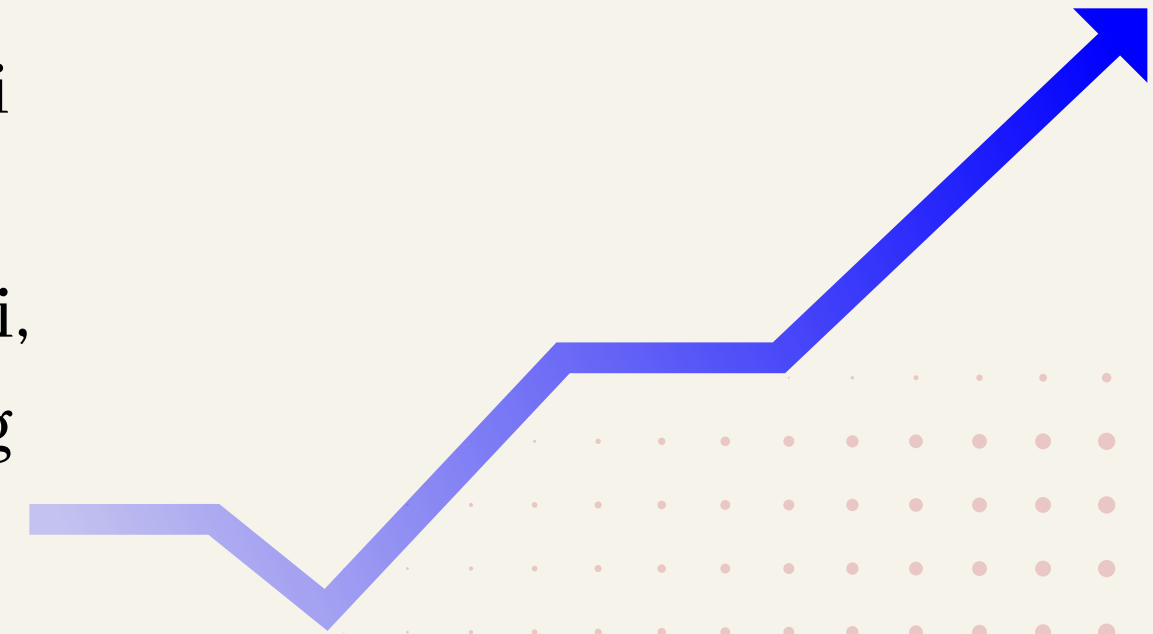


KEGUNAAN

7

Distribusi normal memiliki beberapa kegunaan untuk analisis data, yaitu:

1. Statistik infrensial: Jika terdapat data yang cukup besar, distribusi normal itu memudahkan peneliti untuk membuat inferensi atau prediksi tentang rata-rata dan varians populasi.
2. Pengujian hipotesis: Jika ukuran data cukup besar dan varians populasi diketahui, seseorang bisa menggunakan distribusi normal untuk menguji hipotesis tentang rata-rata populasi.
3. Analisis regresi: Distribusi normal itu juga berguna dalam analisis regresi, karena seseorang dapat membuat kesimpulan statistik yang valid tentang prediksi model.



PDF DAN CDF

8

PDF Distribusi Normal

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-1/2 \left(\frac{x-\mu}{\sigma} \right)^2}$$

μ : Rata-Rata

σ : Simpangan baku

e : Bilangan euler (2,71828)

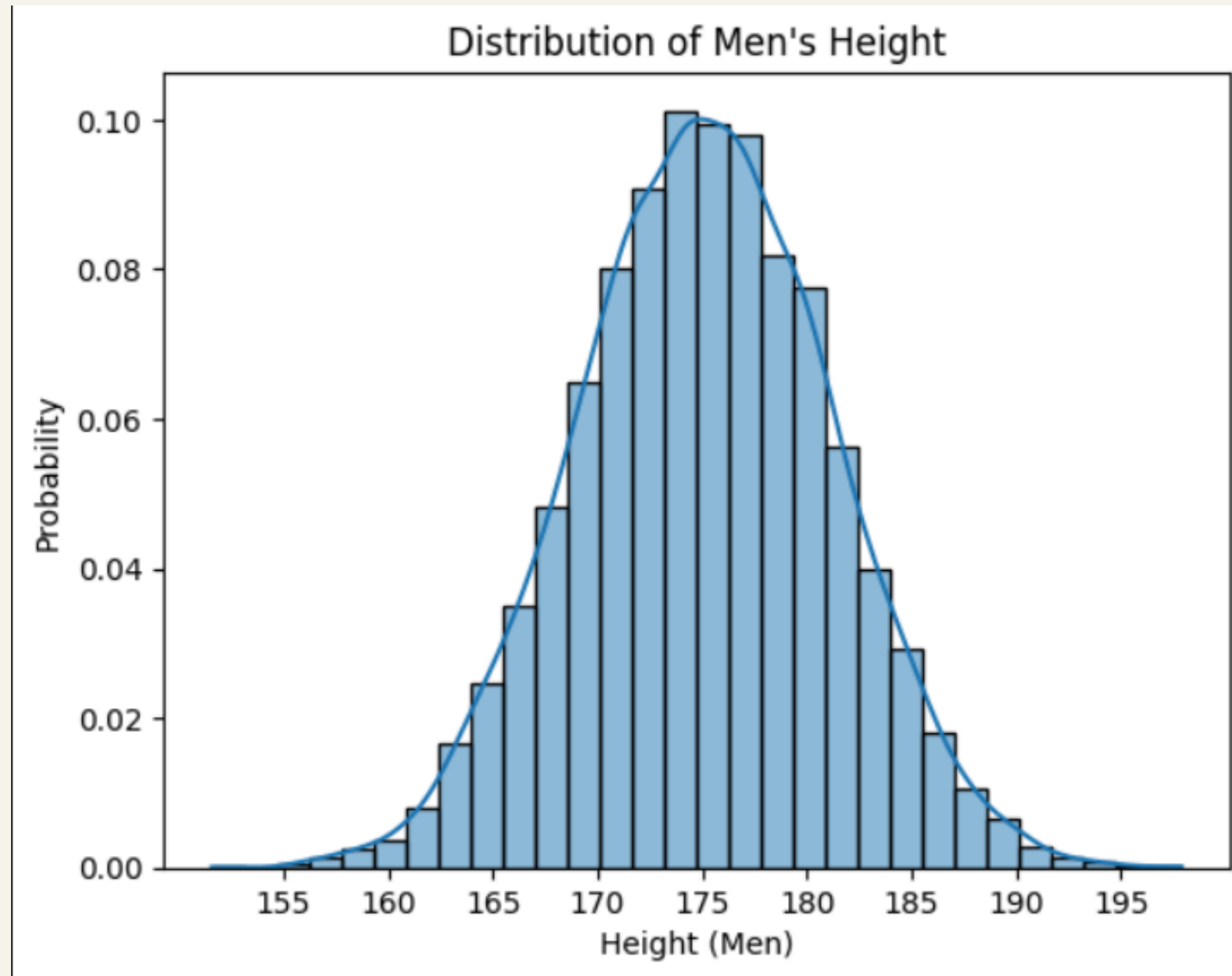
CDF Distribusi Normal

$$F(x) = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{x-\mu}{\sigma \sqrt{2}} \right) \right]$$

erf : error function

VISUALISASI DATA

9



Mean = 175 cm

Standar deviasi = 6

TABEL PERBANDINGAN

10

| Sample | Tinggi Badan | Z-Score |
|--------|--------------|-----------|
| 1 | 155 | -3,333333 |
| 2 | 160 | -2,5 |
| 3 | 165 | -1,666667 |
| 4 | 170 | -0,833333 |
| 5 | 175 | 0 |
| 6 | 180 | 0,8333333 |
| 7 | 185 | 1,6666667 |
| 8 | 190 | 2,5 |
| 9 | 195 | 3,3333333 |

Nilai $Z = \frac{x - \mu}{\sigma}$

x=Nilai Data

1. Distribusi Simetris

- Grafik berbentuk lonceng, maka data itu distribusi normal.
- Nilai rata-rata (μ) berada di tengah, yaitu sekitar 175 cm.

2. Mayoritas Data Dekat dengan Mean

- Sebagian besar data tinggi badan berada dalam rentang 165 cm - 185 cm. Ini sesuai dengan aturan empiris (68-95-99.7), di mana sekitar 68% data itu berada dalam 1 standar deviasi dari mean.

3. Terdapat Variasi dalam Tinggi Badan

- Meskipun sebagian besar data terkonsentrasi di sekitar nilai rata-rata, ada beberapa individu yang memiliki tinggi badan lebih rendah atau lebih tinggi dari rata-rata.

4. Model yang Baik untuk Populasi

- Jika distribusi tinggi badan mengikuti distribusi normal, model ini dapat digunakan untuk prediksi probabilitas tinggi badan tertentu dalam populasi.

SUMBER

1. Revoupedia
2. Medium
3. LMS-SPADA
4. Gramedia

TERIMA KASIH