

Pengenalan Statistika Dasar

Jakarta, 10 Juli 2024

Anggota Tim Kerja Praktek :



Komang Ryandhi Suandita (5002211109)



Fadhlans Ferdinand (5002211121)



Maulana Yusuf Ikhwan Robbani (5002211167)

Daftar Isi

- 1 Konsep Dasar Statistika**
- 2 Distribusi Normal**
- 3 Korelasi**
- 4 Regresi**
- 5 Statistika dengan Bahasa Python**

Statistika adalah ilmu yang mempelajari tentang **pengumpulan**, penyajian dan **interpretasi** data.

Pentingnya mempelajari ilmu statistika:

Kaitannya dengan penyajian dan penafsiran kejadian yang bersifat “peluang” yang terjadi dalam suatu pengamatan / penelitian ilmiah.

Data:

Kumpulan fakta tiap individu

Agar data dapat menjadi informasi, perlu dikomunikasikan dengan **statistika**

Sebelum Masuk Materi, kita pahami data kita terlebih dahulu !

Terdapat 100 data nilai ujian matematika

Anak ke	1	2	97	98	99	100
Nilai anak	65	65	85	85	85	85
Nilai Anak setelah terurut	65	65	95	95	95

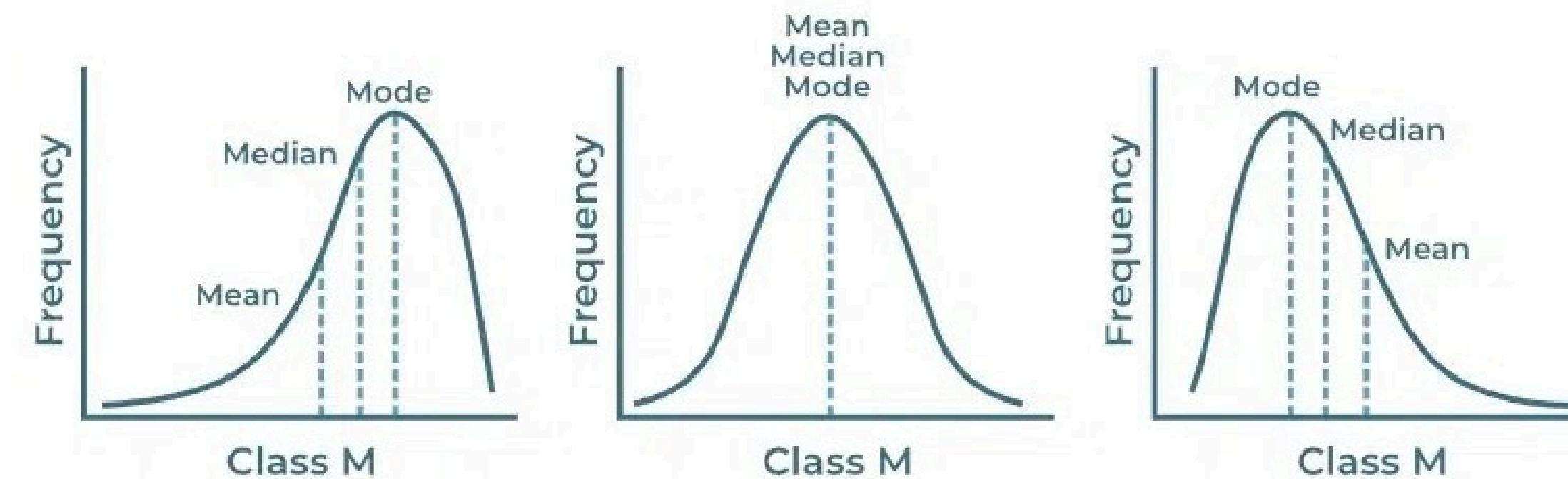
Data tersebut dimulai dari nilai 65 - 85.

Tiap nilai memiliki interval sebesar 5.

Data tersebut terdistribusi secara normal.

Ukuran Pemusatan Data

Berguna untuk mengetahui apakah data kita condong ke kiri atau kanan.
Untuk mengukurnya, perlu membandingkan **mean, median, modus**



- **Gambar 1.** Perhitungan Pemusatan Data (Sumber: [geeksforgeeks.com](https://www.geeksforgeeks.com/statistics-measures-of-central-tendency/))

Ukuran Pemusatan Data

Mean adalah suatu rata-rata dari deretan data yang kita miliki.

Cara menghitungnya:

1. Jumlahkan semua data
2. Bagi dengan banyaknya data

$$x = \frac{\text{data}_1 + \text{data}_2 + \dots + \text{data}_n}{n}$$

Karena data kita ada 100 jumlah
maka cara menghitungnya

$$x = \frac{65+65+\dots+95+95}{100}$$

Ukuran Pemusatan Data

Median adalah nilai tengah dari deretan data yang kita miliki.

Cara menghitungnya:

1. Ambil nilai tengah dari data kita miliki
2. Jika ada 2 nilai tengah, jumlahkan dan bagi 2

Karena data kita ada 100 jumlah
maka nilai tengah berada di 50
& 51

$$\begin{aligned}x &= \frac{\text{Nilai ke } 50 + \text{Nilai ke } 51}{2} \\x &= \frac{80+80}{2} = 80\end{aligned}$$

Ukuran Pemusatan Data

Modus adalah nilai yang sering muncul dari deretan data yang kita miliki.

Anak ke	1	2
Nilai anak	65	65	
Nilai Anak setelah terurut	65	65

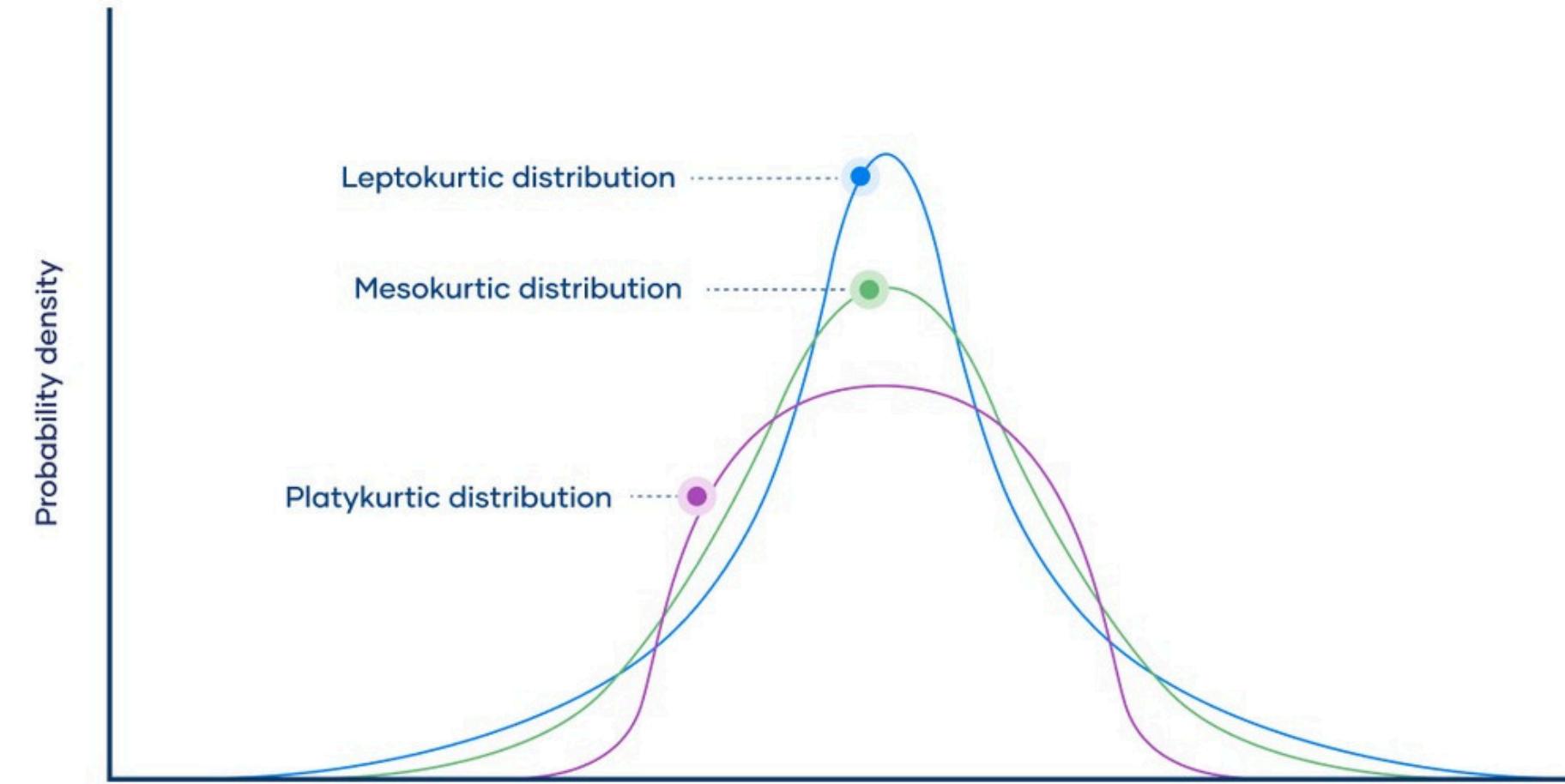
97	98	99	100
85	85	85	85
95	95	95	95

Karena data kita yang sering muncul adalah 80 dengan muncul sebanyak 30 kali

Ukuran Penyebaran Data

Dengan mengetahui penyebaran data, kita dapat mengetahui seberapa lancip atau landai distribusi data kita. Untuk mengetahui lancip/landai data kita, kita perlu mengetahui mean(rata-rata) dan variansi

Ketika variansi bertambah, kurva akan melandai menjadi platykurtic. Dan sebaliknya



- Gambar 2. Perhitungan Ketinggian distribusi Data (Sumber: analyticsvidhya.com)

Ukuran Penyebaran Data

Variansi adalah pengukuran untuk melihat seberapa jauh 1 data dengan rata-rata

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N} =$$

$$\frac{(\text{data}_1 - \text{rata rata})^2}{N} + \frac{(\text{data}_2 - \text{rata rata})^2}{N} + \dots + \frac{(\text{data}_n - \text{rata rata})^2}{N}$$

Karena data kita berjumlah 100, rata-rata nya 80, maka

$$\frac{(65 - 80)^2}{100} + \frac{(65 - 80)^2}{100} + \dots + \frac{(85 - 80)^2}{100} + \frac{(85 - 80)^2}{100} = 52.5$$

Sebanyak 100 penjumlahan

Ukuran Penyebaran Data

Simpangan baku adalah akar dari variansi

Akar 52,5 adalah ~7,2

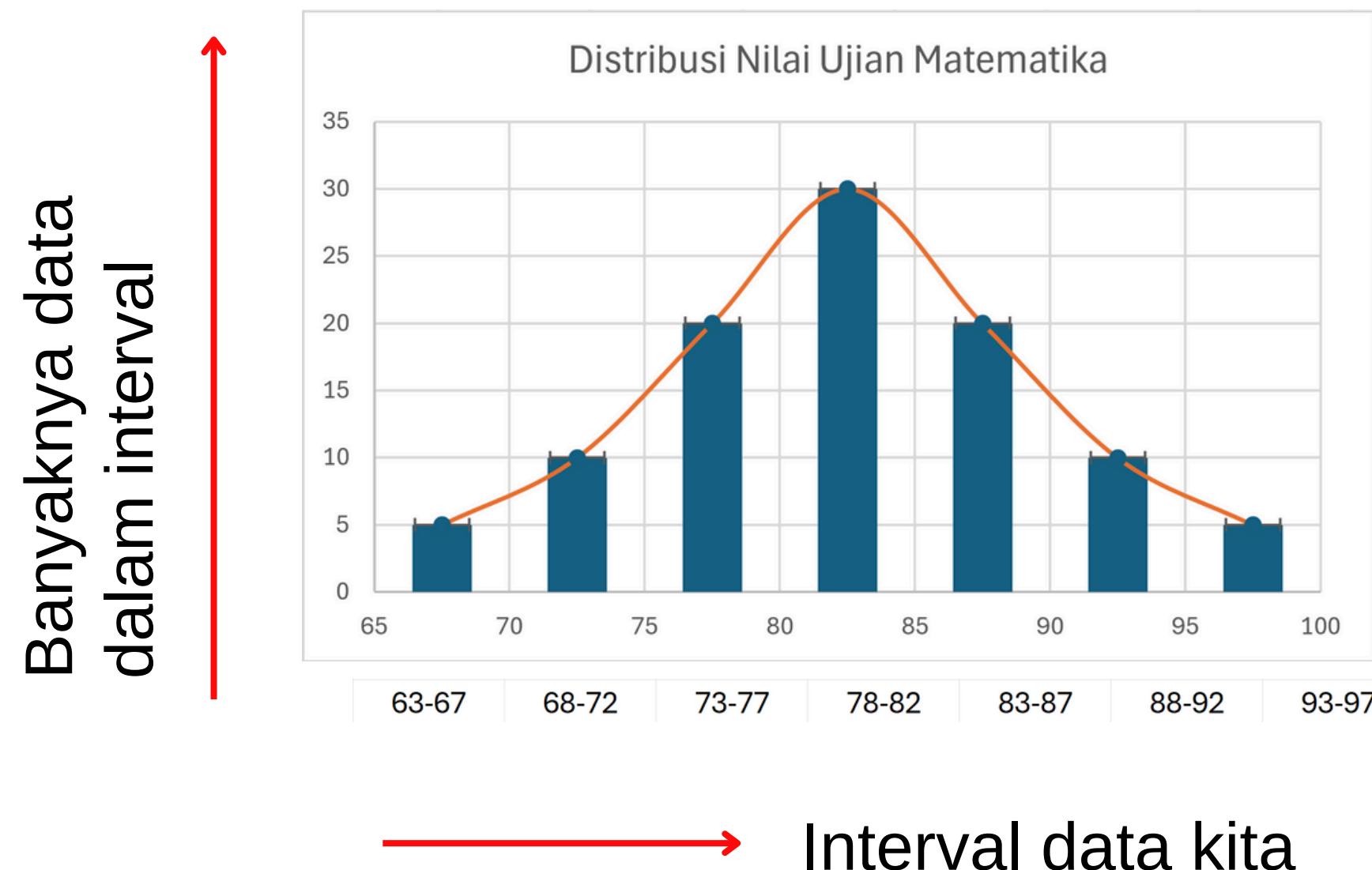
Simpangan rata-rata adalah variansi yang tidak dikuadrat

$$SR = \frac{\sum |x_i - \bar{x}|}{n} =$$
$$\frac{(\text{data}_1 - \text{rata_rata})}{N} + \frac{(\text{data}_2 - \text{rata_rata})}{N} + \dots + \frac{(\text{data}_n - \text{rata_rata})}{N}$$
$$\frac{(65 - 80)}{100} + \frac{(65 - 80)}{100} + \dots + \frac{(85 - 80)}{100} + \frac{(85 - 80)}{100} = 5.5$$

Umumnya untuk mengukur penyebaran data menggunakan variansi

Histogram

Histogram adalah gambar yg menunjukan banyaknya data dalam suatu interval



Cara baca:

Ambil contoh kotak tengah.
Pada data yang bernilai antara **78-82** memiliki kotak dengan **tinggi 30**. Artinya, nilai antara **78-82** berjumlah **30 data**

Cara Membuat Histogram

1. Data yang dimiliki **diurutkan**
2. **Buat interval.** Karena jarak antara satu data dengan data lain beda 5 maka buat 1 interval yang mengandung jarak sebesar 5. (Diambil 63-67, dst)
3. Masukan **semua angka-angka** yang dideretkan dalam **satu interval** yang sama. Contoh, karena 65 masuk interval 63-67 maka seluruh angka 65 dihitung jumlahnya.
4. Buat grafik dengan **sumbu x** adalah interval yang sudah dibuat dan sumbu y adalah **banyaknya data** dalam suatu interval.

Cara Membuat Histogram

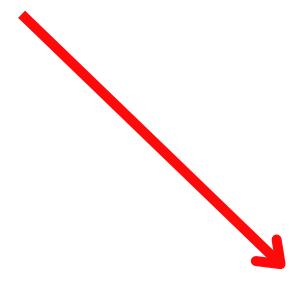
Data yang awalnya kita miliki

Anak ke	1	2
Nilai anak	65	65
Nilai Anak setelah terurut	65	65

97	98	99	100
85	85	85	85
95	95	95	95

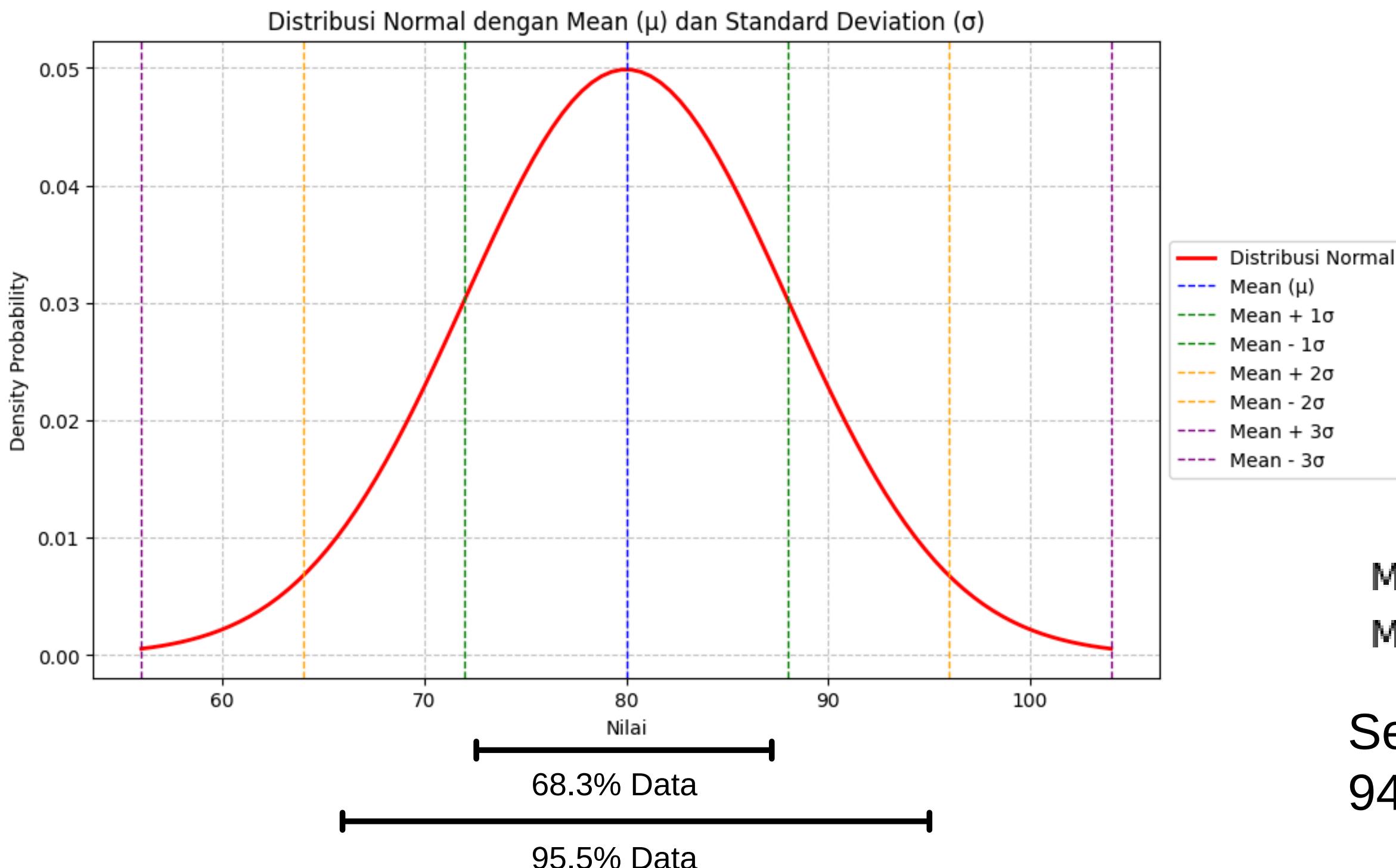
Cara Membuat Histogram

Diubah bentuknya jadi..



Jumlah	Gambaran Umumnya jadi				
30		80			
29		80			
28		80			
27		80			
26		80			
25		80			
24		80			
23		80			
22		80			
21		80			
20	75	80	85		
19	75	80	85		
18	75	80	85		
17	75	80	85		
16	75	80	85		
15	75	80	85		
14	75	80	85		
13	75	80	85		
12	75	80	85		
11	75	80	85		
10	70	75	80	85	90
9	70	75	80	85	90
8	70	75	80	85	90
7	70	75	80	85	90
6	70	75	80	85	90
5	65	70	75	80	85
4	65	70	75	80	85
3	65	70	75	80	85
2	65	70	75	80	85
1	65	70	75	80	85
	65	70	75	80	85
					90
					95
					Interval data

Standar deviasi



Gunanya apa?

1. Mengetahui lancip tidaknya kurva
2. Menentukan interval (contoh, 95% data berada dari rata-rata $- 2 \times$ standar deviasi)

Berdasarkan data kita

$$\begin{aligned} \text{Mean} + 2\sigma &= 80 + 2 \times \sigma = 65,5 \\ \text{Mean} - 2\sigma &= 80 - 2 \times \sigma = 94,4 \end{aligned}$$

Sehingga, data antara 65,5 sampai 94,4 sudah mewakili 95,5% data

Apa itu korelasi?

Korelasi adalah pengukuran statistik untuk mengetahui **pengaruh** antara 2 jenis data. Gampangnya, 2 kolom dikatakan **berkorelasi** ketika **satu nilai kolom naik, maka nilai kolom lain naik**

Misalkan terdapat data **nilai ujian matematika** dan **skor IQ** tiap murid

Nilai Ujian Matematika

no	data
1	60
2	60
3	65
4	65
5	65
6	65
7	65
8	65
9	70
10	70
11	70
12	70

86	90
87	90
88	90
89	90
90	90
91	90
92	90
93	90
94	95
95	95
96	95
97	95
98	95
99	100
100	100

Skor IQ

no	data	✓↑
1	91	
2	92	
3	92	
4	92	
5	93	
6	93	
7	93	
8	95	
9	95	
10	95	
11	96	
12	96	

86	117
87	117
88	118
89	118
90	118
91	118
92	118
93	119
94	119
95	119
96	120
97	120
98	120
99	120
100	120

Cara menghitung korelasi

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

r = Korelasi

x_i = Data IQ

\bar{x} = Rata-rata data IQ

y_i = Data Nilai Ujian

\bar{y} = Rata-rata data nilai ujian

Interval Koefisien	Keeratan Korelasi
0,00-0,20	Sangat Lemah
0,21-0,40	Lemah
0,41-0,70	Moderate / Sedang
0,71-0,90	Kuat
0,91-0,99	Sangat Kuat
1	Korelasi Sempurna

Nilai korelasi berada pada interval -1 sampai 1.

Jika nilai korelasi = +/- 1
Dikatakan korelasi kuat.

Dihitung, nilai korelasi antara skor IQ dan nilai ujian adalah = 0,937

Biasanya digunakan R^2 agar interval berada antara 0 sampai 1.

Sehingga korelasi kuadrat = 0,8787

Artinya, skor IQ berpengaruh besar pada nilai ujian

Apa itu Regresi?

Regresi dapat didefinisikan sebagai pengukuran yang digunakan untuk **mengukur seberapa besar perubahan** dalam satu variabel akan memengaruhi variabel lain. Gampangannya, kita **mencari garis lurus** yang jarak antara titik-titik dengan garisnya paling dekat. Sehingga, **terlihat tren atau perubahan** satu variabel terhadap variabel lain

Apa tujuan dari regresi?

1. Mengetahui **tren** dari data kita
2. Mengetahui **pengaruh suatu variabel** terhadap variabel lain
3. Memprediksi **nilai diluar data** yang kita miliki

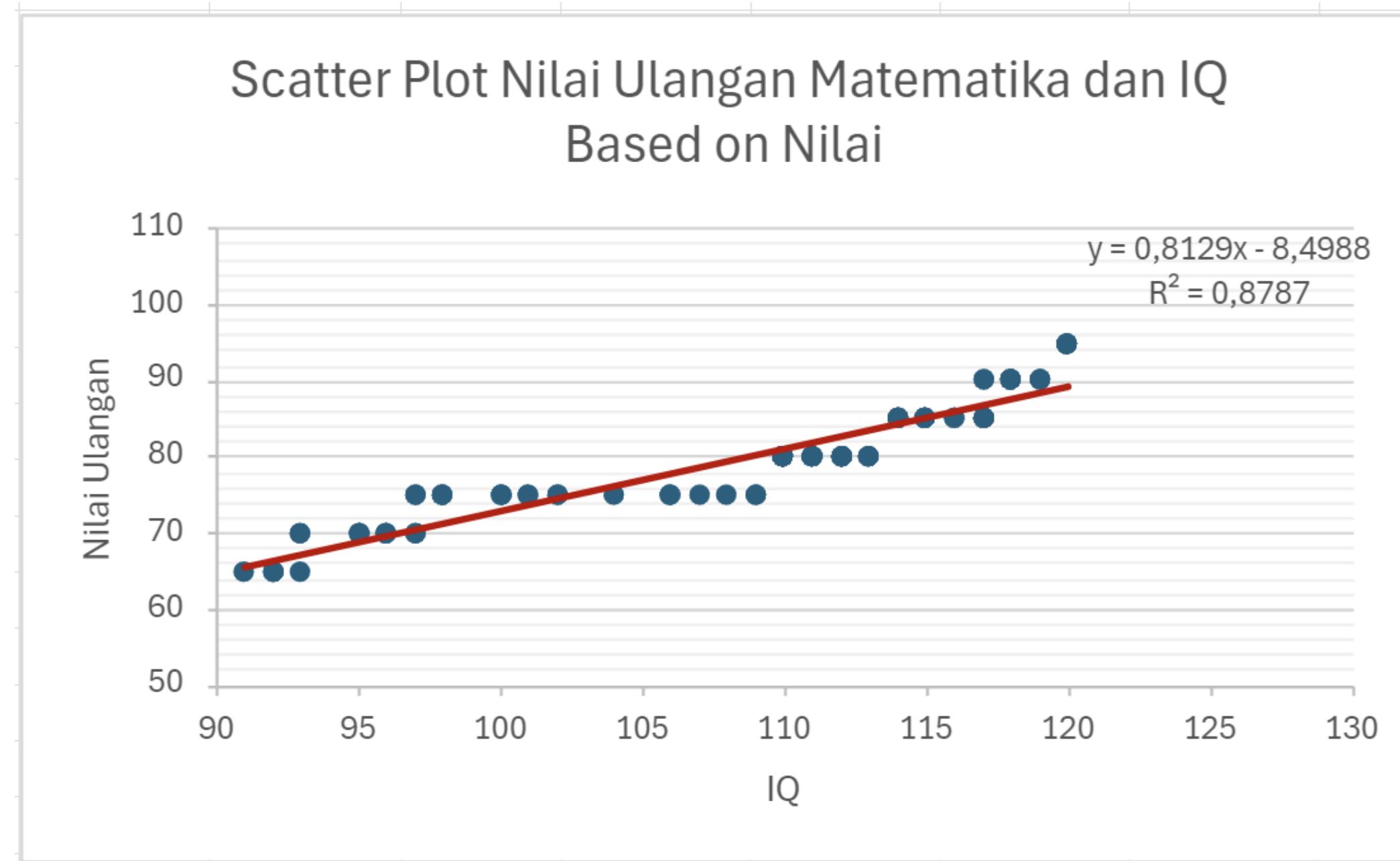
Proses melakukan regresi

1. Pilih 2 variabel yang diinginkan (Nilai Ujian & Skor IQ)
2. Petakan 2 variabel pada grafik. Variabel pertama di sumbu x (Independent), variabel kedua di sumbu y (Dependent)
3. Karena kita mencari suatu fungsi garis lurus $f(x) = ax + b$ dengan x adalah deretan data kita dan a dan b harus menjadikan eror garis lurus dengan data kita minimal. Sehingga cari a dan b dengan rumus:
4. Masukan a dan b ke fungsi $f(x)$ dan gambarkan pada grafik

$$a = \frac{\begin{vmatrix} \sum_{i=1}^n x_i Y_i & \sum_{i=1}^n x_i \\ \sum_{i=1}^n Y_i & n \end{vmatrix}}{\begin{vmatrix} \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & n \end{vmatrix}}$$

dan $b = \frac{\begin{vmatrix} \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i Y_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n Y_i \end{vmatrix}}{\begin{vmatrix} \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & n \end{vmatrix}}$

Hasil Visualisasi



Data yang digunakan

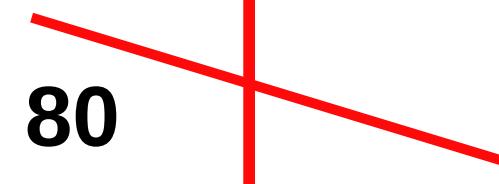
1. Nilai Ujian Matematika 100 Anak yang Terdistribusi Normal
2. Nilai Iq 100 Anak yang Mengikuti Ujian Matematika yang Terdistribusi Normal

Data 1 dan 2 memiliki :

Mean (Rata - Rata) : 80

Modus (Nilai yang Paling Sering Muncul) : 80

Median (Nilai Tengah) : 80

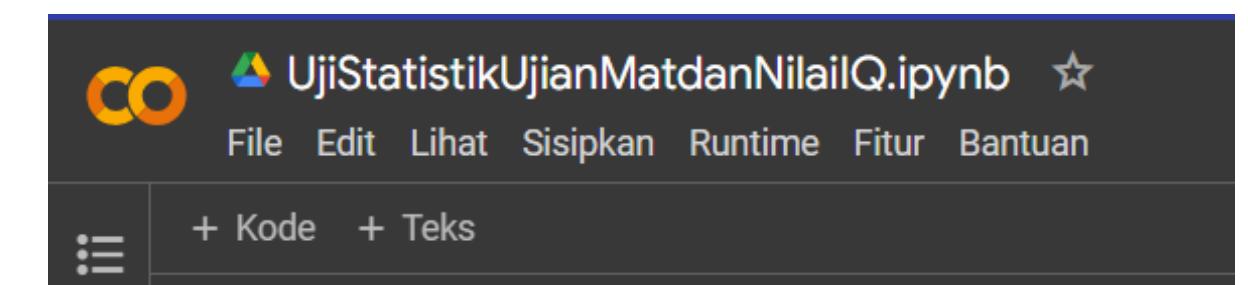


Akan Terdistribusi Normal

Tools yang digunakan :



[Link Google Colab](#)



*Penjelasan Detail ada di Google Collab

TERIMA KASIH

Disusun oleh :
Komang Ryandhi Suandita
Fadhlwan Ferdian
Maulana Yusuf Ikhsan Robbani