Peubah Acak Kontinu

Memodelkan probabilitas data pengukuran melalui Fungsi Padat Peluang (PDF), serta penerapan Distribusi Normal, Eksponensial, dan Uniform

Statistika dan Probabilitas · Pertemuan 06

Peta Pertemuan 06

1. **Konsep Peubah Acak Kontinu, PDF & CDF**

   Memahami variabel acak berbasis pengukuran dan bagaimana area di bawah kurva merepresentasikan probabilitas.

2. **Distribusi Uniform & Eksponensial**

   Memodelkan probabilitas kejadian yang merata (Uniform) dan waktu tunggu antar kejadian (Eksponensial).

3. **Distribusi Normal (Kurva Lonceng)**

   Mengenal distribusi paling fundamental di alam semesta dan data science, beserta karakteristik simetrisnya.

4. **Standarisasi & Z-Score**

   Mentransformasi berbagai skala Distribusi Normal ke dalam satu metrik standar baku (Normal Baku) untuk komparasi probabilitas.

Blok 1 — Konsep Dasar PDF & CDF

Transisi dari data hitungan diskret menuju data pengukuran berkelanjutan

Pemantik Diskusi: Paradoks Probabilitas Kontinu

PERTANYAAN PEMICU

- Mari bayangkan Anda sedang mengukur suhu ruangan saat ini dengan termometer digital yang **super presisi** tanpa batas.
- Berapakah probabilitas termometer tersebut menunjukkan angka **tepat secara absolut** di $25,00000000000\ldots^{\circ}C$ (dengan angka nol tak terhingga di belakang koma)?

Pemantik Diskusi: Paradoks Probabilitas Kontinu

PERTANYAAN PEMICU

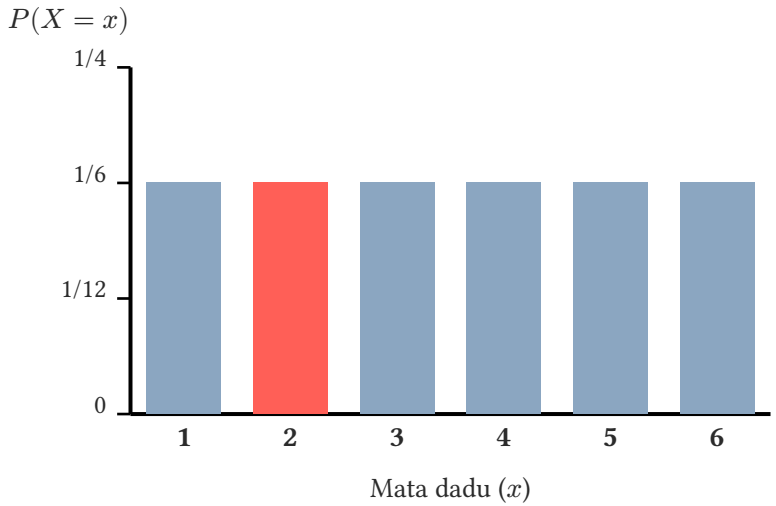
- Mari bayangkan Anda sedang mengukur suhu ruangan saat ini dengan termometer digital yang **super presisi** tanpa batas.
- Berapakah probabilitas termometer tersebut menunjukkan angka **tepat secara absolut** di $25,00000000000\ldots^{\circ}C$ (dengan angka nol tak terhingga di belakang koma)?

Jawabannya: Nol (0).

Skala kontinu (suhu, waktu, panjang, berat badan) dapat bernilai berapa pun dalam suatu rentang, termasuk desimal tak berujung. Probabilitas pada data kontinu diukur sebagai **luas di bawah kurva**, dan **satu titik tepat tidak memiliki lebar**, sehingga luasnya nol: $P(X = x) = 0$. Nol di sini bukan berarti mustahil, hanya saja peluang mengenai *satu nilai eksak* tidak bermakna. Yang bermakna adalah peluang pada **rentang**.

Analogi Visual 1: Probabilitas Diskret (Area Balok)

Untuk memahami paradoks probabilitas kontinu, kita ingat kembali bahwa **Probabilitas = Luas Area**. Pada peubah acak diskret (misal: mata dadu), setiap kejadian adalah titik bilangan bulat mandiri. Probabilitasnya divisualisasikan sebagai balok yang memiliki "lebar" logis.



Distribusi Diskret

Konsep Luas Area Diskret

Ambil contoh probabilitas mata dadu "2" (Balok Merah).

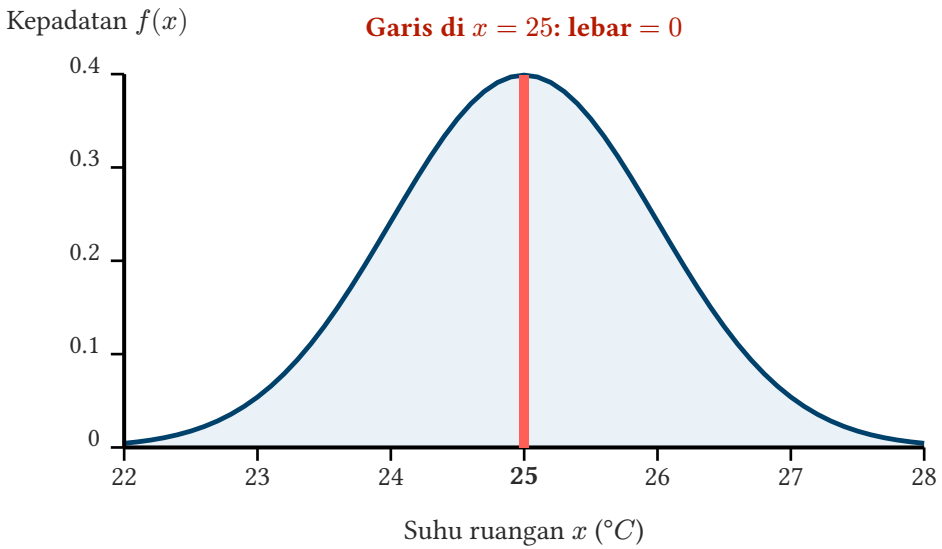
- **Tinggi Balok** = Besaran probabilitas (1/6)
- **Lebar Balok** = 1 (Mewakili 1 kejadian utuh)

Luas Area = Tinggi $\times$ Lebar = $\frac{1}{6}$

Karena lebarnya jelas ada, luas areanya bisa dihitung dan nilainya pasti lebih besar dari nol (> 0).

Analogi Visual 2: Titik Eksak pada Kontinu = Garis

Pada peubah acak kontinu (misal suhu ruangan), nilai yang mungkin menyebar mulus sehingga grafiknya berupa **kurva kepadatan** $f(x)$, dan probabilitas adalah **luas** di bawah kurva. Bagaimana jika kita bertanya tentang satu titik eksak, yaitu suhu **tepat** $25^\circ C$?



Kurva Kepadatan Kontinu & Satu Titik Eksak

Paradoks Titik Eksak = 0

Titik **tepat** $25^\circ C$ berupa **garis vertikal sangat tipis** (garis merah).

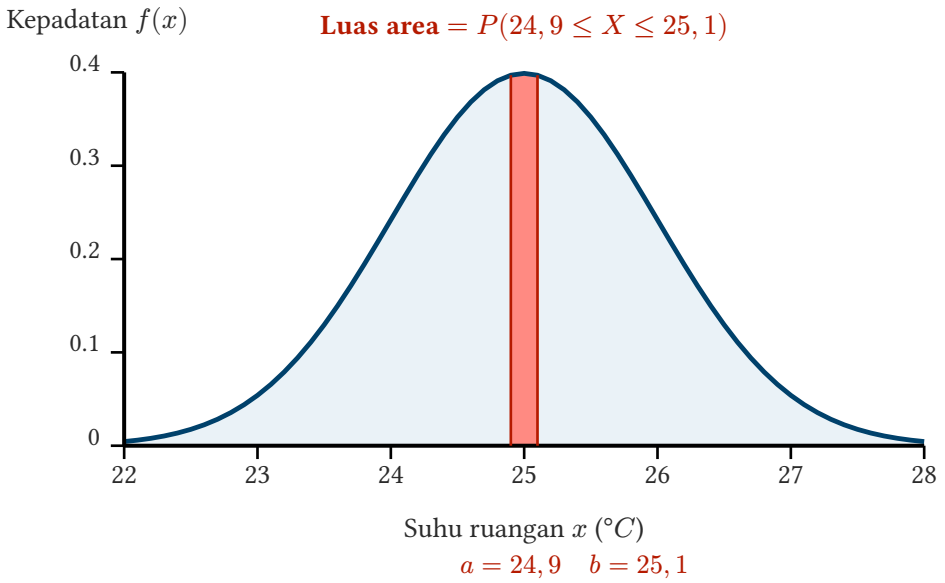
- **Tinggi garis** = kepadatan $f(25)$. Nilainya ada, tetapi ini **bukan** probabilitas.
- **Lebar garis** = $25 - 25 = 0$, karena satu titik tidak menempati rentang apa pun.

$$P(X = 25) = f(25) \times 0 = 0$$

Nol di sini bukan berarti suhu 25 mustahil muncul, melainkan peluang mengenai **satu nilai eksak** tidak memiliki luas. Peluang baru bermakna pada rentang.

Analogi Visual 3: Solusinya? Gunakan Rentang (Interval)

Agar luas probabilitasnya bermakna, kita ubah pertanyaan dari "satu titik" menjadi "**di antara dua titik**" (interval). Interval inilah yang menciptakan **lebar**, sehingga luas di bawah kurva menjadi lebih dari nol.



Probabilitas sebagai Luas di bawah Kurva pada Interval

Perhitungan Luas Rentang

Pertanyaan baru: **"Berapa probabilitas suhu berada di antara $24,9^{\circ}C$ dan $25,1^{\circ}C$?"**

- **Tinggi kurva** = kepadatan $f(x)$
- **Lebar rentang** = $25,1 - 24,9 = 0,2$

$$P(a \leq X \leq b) = \int_a^b f(x)dx > 0$$

Karena lebarnya 0,2 (bukan 0), luasnya positif. Untuk kurva ini, luasnya sekitar $0,2 \times f(25) \approx 0,08$, atau sekitar 8%.

Paradigma Baru: Probabilitas Berbasis Rentang (Interval)

Karena kita tidak bisa mencari peluang di **satu titik eksak**, kita mengubah paradigma dari menanyakan "satu titik" menjadi menanyakan **"area"** (rentang/interval).

Dari Titik ke Interval:

Pertanyaan Salah:

"Berapa peluang suhu **tepat** 25°C?"

Pertanyaan Benar:

"Berapa peluang suhu berada **di antara** 24,9°C hingga 25,1°C?"

Konsekuensi Matematis

> Pada Peubah Acak Kontinu, probabilitas **selalu** dievaluasi pada sebuah rentang interval, bukan titik tunggal.

$$P(a \leq X \leq b)$$

Catatan Penting:

Karena peluang batas eksak $P(X = a) = 0$ dan $P(X = b) = 0$, maka menggunakan tanda sama dengan $(\leq, \geq)$ ataupun tidak $(<, >)$ pada batas rentang **sama sekali tidak mengubah nilai probabilitasnya**:

$$P(a \leq X \leq b) = P(a < X < b)$$

Fungsi Padat Peluang (PDF) — Kepadatan Probabilitas

Jika probabilitas pada data kontinu bergantung pada **luas area**, maka kita memerlukan fungsi garis lengkung (kurva) sebagai "atap" pembentuk area tersebut. Fungsi pembuat atap inilah yang disebut **PDF** atau dilambangkan dengan $f(x)$.

Tiga Syarat Mutlak PDF:

1. Kurva tidak boleh berada di bawah nol (luas tidak mungkin negatif). Catatan: $f(x)$ adalah **kepadatan**, bukan probabilitas, sehingga nilainya boleh lebih dari 1; yang tidak boleh lebih dari 1 adalah **luasnya**:

$$f(x) \geq 0$$

2. Total keseluruhan luas area di bawah kurva (dari ujung negatif tak hingga sampai positif tak hingga) adalah mutlak 1 (atau 100%):

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

Perhitungan Peluang Interval

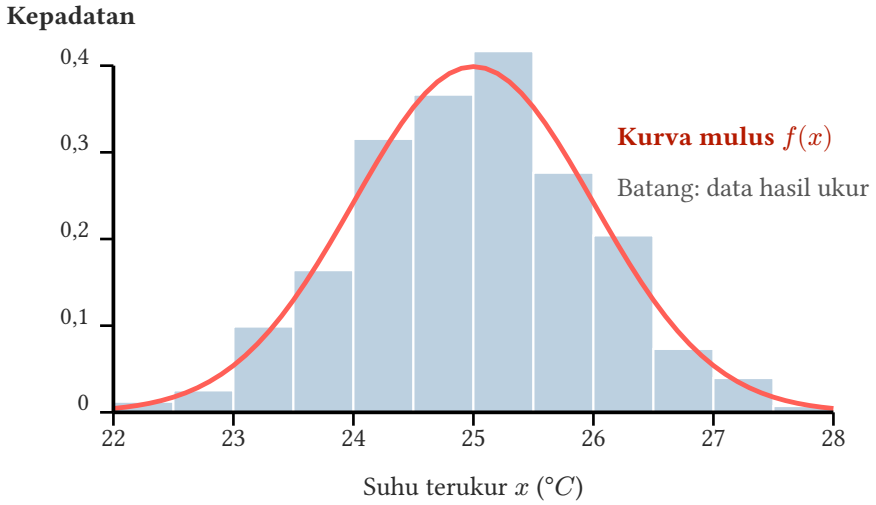
Syarat ketiga: Peluang variabel X berada di rentang $[a, b]$ dihitung dengan mencari **luas kurva integral** tepat di antara dua batas tersebut.

$$P(a \leq X \leq b) = \int_{a}^{b} f(x)dx$$

Dua pertanyaan lanjutan: (1) dari mana $f(x)$ untuk kasus nyata diperoleh? (2) adakah cara yang lebih efisien daripada mengintegral setiap kali? Kita jawab satu per satu.

Dari Mana $f(x)$ Berasal? Dari Data dan Model

Di soal buku, $f(x)$ biasanya diberikan. Di dunia nyata, $f(x)$ **dibangun** dari data pengukuran. Cara paling intuitif: ubah histogram menjadi kurva kepadatan.



Histogram (skala kepadatan) → kurva $f(x)$

Tiga Sumber $f(x)$

1. Dari data. Tinggi batang = frekuensi relatif ÷ lebar kelas, sehingga total luas = 1. Makin banyak data dan kelas makin sempit, batang makin menyerupai kurva mulus.

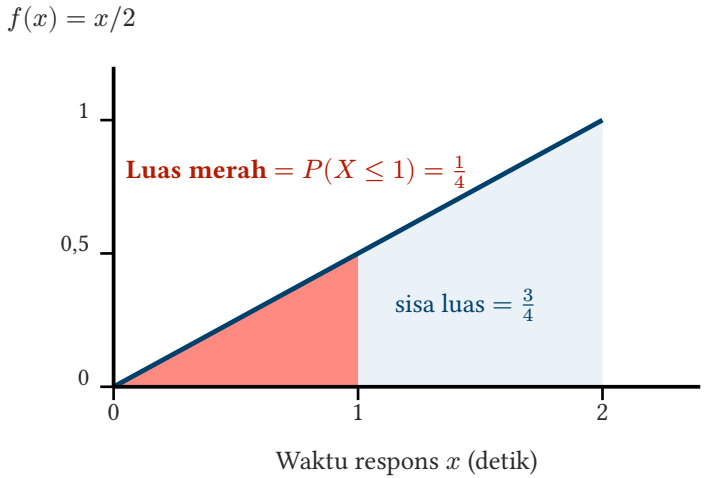
2. Dari model teori. Pilih keluarga distribusi yang cocok dengan proses fisiknya, lalu taksir parameternya dari data (misal $\mu \approx \bar{x}$, $\sigma \approx s$):

- nilai merata → Uniform
- waktu tunggu antar kejadian acak → Eksponensial
- hasil banyak pengaruh kecil → Normal

3. Dari syarat luas = 1. Bila bentuknya diketahui, konstanta ditentukan agar total luas = 1 (contoh di slide berikut).

Contoh: Menentukan $f(x)$ dari Syarat Luas = 1

Kasus. Waktu respons sebuah server X (detik) berada di $0 \le x \le 2$, dan makin lama makin mungkin: kepadatannya naik lurus, $f(x) = kx$. Berapa k, dan berapa peluang respons ≤ 1 detik?



Segitiga: total luas $= 1$

Ide Penyelesaian

Langkah 1. Total luas di bawah kurva harus 1. Dari sini kita cari konstanta k.

Langkah 2. Setelah k diketahui, tulis $f(x)$ lengkap.

Langkah 3. Peluang $P(X \le 1)$ adalah luas di kiri $x = 1$ (daerah merah).

Langkah 4. Cek dengan CDF $F(x)$.

Rincian hitungan setiap langkah ada di dua slide berikutnya.

Rincian 1: Mencari k (Luas Total = 1)

a. Tulis syaratnya. Luas total di bawah $f(x)$ harus 1:

$$\int_0^2 kx\,dx = 1$$

b. Keluarkan konstanta k dari integral:

$$k\int_0^2 x\,dx = 1$$

c. Cari antiturunan x. Aturan pangkat: $\int x^n dx = \frac{x^{n+1}}{n+1}$. Untuk $n = 1$:

$$\int x\,dx = \frac{x^2}{2}$$

d. Masukkan batas atas 2 dan bawah 0 (atas dikurangi bawah):

$$\left[\frac{x^2}{2}\right]_0^2 = \frac{2^2}{2} - \frac{0^2}{2} = \frac{4}{2} - 0 = 2$$

e. Selesaikan k:

$$k \cdot 2 = 1 \quad \Rightarrow \quad k = \tfrac{1}{2}$$

Cek dengan Geometri (tanpa integral)

Daerah di bawah $f(x) = kx$ adalah segitiga:

- alas $= 2$ (dari $x = 0$ sampai 2)
- tinggi $= f(2) = k \cdot 2 = 2k$

Luas segitiga $= \frac{1}{2} \cdot$ alas $\cdot$ tinggi:

$$\tfrac{1}{2} \cdot 2 \cdot 2k = 2k$$

Syarat $2k = 1$, jadi $k = \frac{1}{2}$. Hasilnya sama dengan integral.

Hasil: $f(x) = \frac{x}{2}$ untuk $0 \le x \le 2$, dan $f(x) = 0$ di luar itu.
Tingginya di $x = 2$ adalah $f(2) = 1$.

Rincian 2: Menghitung $P(X \leq 1)$ dan CDF

a. Peluang = luas di rentang. "≤ 1 detik" berarti dari batas bawah 0 sampai 1:

$$P(X \leq 1) = \int_0^1 \tfrac{x}{2}\, dx$$

b. Keluarkan faktor $\frac{1}{2}$:

$$= \tfrac{1}{2} \int_0^1 x\, dx = \tfrac{1}{2}\left[\tfrac{x^2}{2}\right]_0^1$$

c. Masukkan batas 1 dan 0:

$$= \tfrac{1}{2}\left(\tfrac{1^2}{2} - \tfrac{0^2}{2}\right) = \tfrac{1}{2} \cdot \tfrac{1}{2} = \tfrac{1}{4}$$

d. Cek geometri. Segitiga merah: alas 1, tinggi $f(1) = \frac{1}{2}$:

$$\tfrac{1}{2} \cdot 1 \cdot \tfrac{1}{2} = \tfrac{1}{4} \quad \checkmark$$

CDF: $F(x) = P(X \leq x)$

Ganti batas atas 1 dengan x (variabel di dalam integral kita namai t):

$$F(x) = \int_0^x \tfrac{t}{2}\, dt = \left[\tfrac{t^2}{4}\right]_0^x = \tfrac{x^2}{4}$$

Cek nilai:

- $F(1) = \frac{1}{4}$ cocok dengan $P(X \leq 1)$
- $F(2) = \frac{4}{4} = 1$ (total peluang)
- $F(0) = 0$

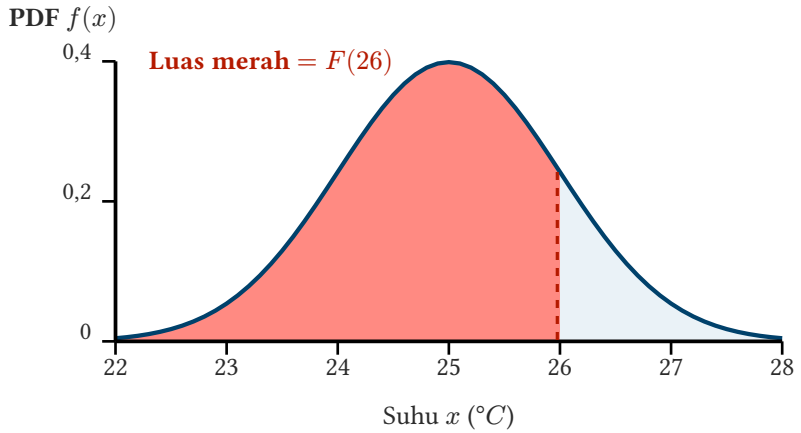
Turunan CDF adalah PDF:

$$F'(x) = \left(\tfrac{x^2}{4}\right)' = \tfrac{2x}{4} = \tfrac{x}{2} = f(x)$$

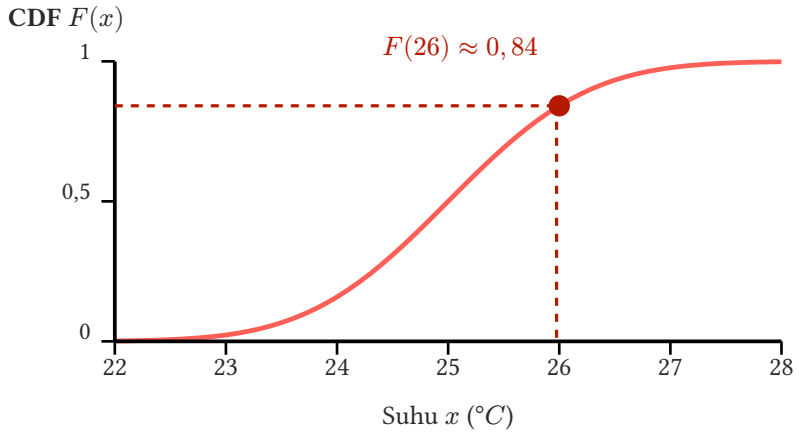
Dengan CDF, $P(a \leq X \leq b) = F(b) - F(a)$. Misalnya $P(1 \leq X \leq 2) = 1 - \frac{1}{4} = \frac{3}{4}$.

Fungsi Distribusi Kumulatif (CDF) — Akumulasi Area

Agar tidak menghitung integral berulang kali, kita pakai **CDF** $F(x)$: catatan *"running total"* yang merekap probabilitas yang sudah terkumpul **dari ujung kiri sampai titik** x.



(1) PDF: luas dari kiri sampai $x = 26$



(2) CDF: tinggi kurva di $x = 26$ = luas pada PDF

Definisi Formal CDF:

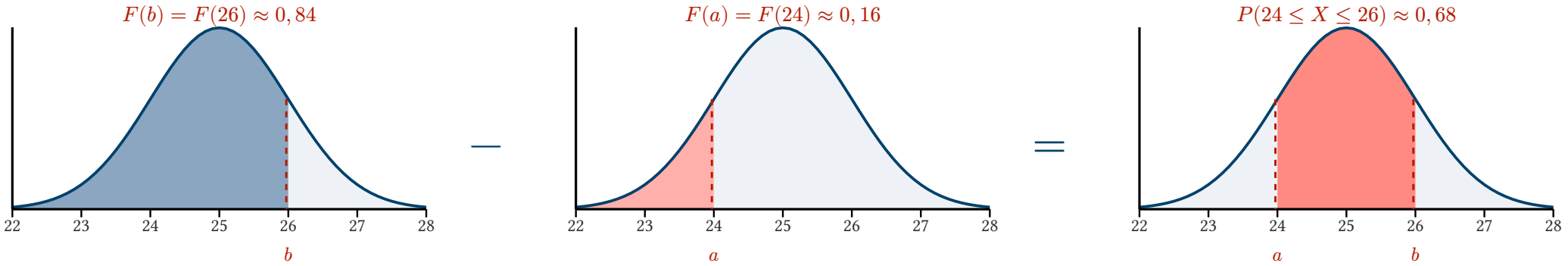
$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$$

Cara Membaca Grafik

- $F(x)$ **tidak pernah turun**; dimulai dari 0 di ujung kiri dan berakhir di 1 di ujung kanan.
- Semakin ke kanan, semakin banyak luas PDF yang terkumpul, sehingga CDF menanjak paling curam di sekitar puncak PDF.

Hubungan PDF & CDF: Jalan Pintas Menghitung Peluang

Dengan CDF, kita **tidak perlu mengintegral manual** setiap kali mencari peluang interval $[a, b]$. Cukup kurangkan dua luas akumulasi. Contoh: suhu X dengan $a = 24$ dan $b = 26$.



$$P(a \le X \le b) = F(b) - F(a)$$

Masukkan batas atas dan batas bawah ke dalam CDF yang sudah jadi.

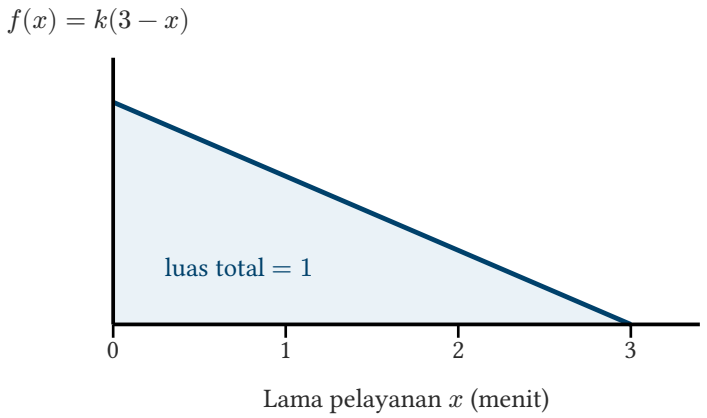
Logika Pengurangan Area

1. Ambil seluruh luas dari ujung kiri sampai $b \rightarrow F(b)$.
2. Buang luas dari ujung kiri sampai a yang tidak dibutuhkan $\rightarrow F(a)$.
3. Sisanya adalah luas tepat di antara a dan b: $0,84 - 0,16 = 0,68$ (sekitar 68%).

Soal Latihan: Lama Pelayanan di Loket

KASUS

Lama pelayanan seorang nasabah X (menit) di sebuah loket berada pada $0 \le x \le 3$. Makin lama makin jarang terjadi, dengan kepadatan yang turun lurus: $f(x) = k(3 - x)$ untuk $0 \le x \le 3$, dan $f(x) = 0$ di luar itu.



Pertanyaan

a. Tentukan konstanta k dan tulis $f(x)$ lengkap.

b. Tentukan CDF $F(x)$.

c. Hitung $P(X \le 1)$.

d. Hitung $P(1 \le X \le 2)$.

e. Hitung $P(X > 2)$ dan $P(X = 1)$.

Jawaban (a): Mencari k

1. Syarat luas total $= 1$:

$$\int_0^3 k(3-x)\,dx = 1$$

2. Keluarkan k dan cari antiturunan. $\int 3\,dx = 3x$ dan $\int x\,dx = \frac{x^2}{2}$:

$$k\left[3x - \tfrac{x^2}{2}\right]_0^3 = 1$$

3. Masukkan batas atas 3 dan bawah 0:

$$\left[3x - \tfrac{x^2}{2}\right]_0^3 = \left(9 - \tfrac{9}{2}\right) - (0-0) = \tfrac{9}{2}$$

4. Selesaikan k:

$$k \cdot \tfrac{9}{2} = 1 \quad \Rightarrow \quad k = \tfrac{2}{9}$$

Cek Geometri

Daerahnya segitiga: alas $= 3$, tinggi $= f(0) = 3k$.

$$\tfrac{1}{2} \cdot 3 \cdot 3k = 9\tfrac{k}{2} = 1 \quad \Rightarrow \quad k = \tfrac{2}{9}$$

Hasilnya sama dengan integral.

$f(x) = \frac{2}{9}(3-x)$ untuk $0 \le x \le 3$. Tinggi di $x = 0$: $f(0) = \frac{2}{3}$. Di $x = 3$: $f(3) = 0$.

Jawaban (b): Menentukan CDF $F(x)$

1. Definisi: $F(x) = P(X \leq x)$ adalah luas dari 0 sampai x. Variabel di dalam integral kita namai t:

$$F(x) = \int_0^x \tfrac{2}{9}(3 - t)\, dt$$

2. Keluarkan $\frac{2}{9}$, cari antiturunan:

$$F(x) = \tfrac{2}{9}\left[3t - \tfrac{t^2}{2}\right]_0^x$$

3. Masukkan batas x dan 0:

$$F(x) = \tfrac{2}{9}\left(3x - \tfrac{x^2}{2}\right) = \tfrac{6x - x^2}{9}$$

4. Bentuk lengkap:

$$F(x) = \begin{cases} 0 & \text{untuk } x < 0 \\ \frac{6x - x^2}{9} & \text{untuk } 0 \leq x \leq 3 \\ 1 & \text{untuk } x > 3 \end{cases}$$

Cek Hasil

- $F(0) = 0$ (belum ada luas terkumpul)
- $F(3) = \frac{18-9}{9} = 1$ (seluruh luas)
- Turunan harus kembali ke PDF:

$$F'(x) = \tfrac{6-2x}{9} = \tfrac{2}{9}(3 - x) = f(x) \quad \checkmark$$

Jawaban (c, d, e): Menghitung Peluang

c. $\boldsymbol{P(X \le 1) = F(1)}$

$$= \tfrac{6 \cdot 1 - 1^2}{9} = \tfrac{5}{9} \approx 0,556$$

d. $\boldsymbol{P(1 \le X \le 2) = F(2) - F(1)}$

$$F(2) = \tfrac{12-4}{9} = \tfrac{8}{9}$$

$$\tfrac{8}{9} - \tfrac{5}{9} = \tfrac{3}{9} = \tfrac{1}{3} \approx 0,333$$

e. $\boldsymbol{P(X > 2)}$**:** sisa luas di kanan 2 (aturan komplemen)

$$= 1 - F(2) = 1 - \tfrac{8}{9} = \tfrac{1}{9} \approx 0,111$$

$\boldsymbol{P(X = 1)}$**:** satu titik tidak punya lebar, jadi luasnya

$$P(X = 1) = \int_1^1 f(x)\,dx = 0$$

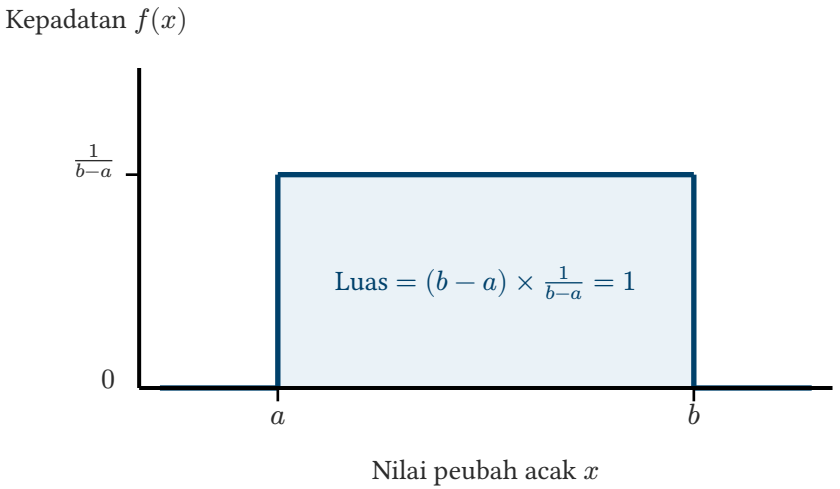
Cek: $\frac{5}{9} + \frac{3}{9} + \frac{1}{9} = \frac{9}{9} = 1$. Ketiga daerah $[0,1]$, $[1,2]$, $[2,3]$ saling lepas dan menutup seluruh rentang. Artinya, sekitar 56% nasabah dilayani dalam 1 menit atau kurang, dan hanya sekitar 11% yang butuh lebih dari 2 menit.

Blok 2 — Uniform & Eksponensial

Model probabilitas paling fundamental untuk probabilitas merata dan rentang waktu tunggu antar kejadian

Distribusi Uniform (Seragam) Kontinu

Distribusi Uniform Kontinu adalah model probabilitas yang paling sederhana. Ia terjadi ketika **setiap nilai di dalam sebuah rentang $[a, b]$ memiliki peluang (kepadatan) yang persis sama besar** untuk terpilih. Karena kepadatannya konstan secara absolut, kurva PDF-nya tidak melengkung, melainkan membentuk **persegi panjang datar**.



PDF Uniform: persegi panjang datar

Logika Tinggi "Persegi Panjang"

Ingat syarat mutlak PDF: **Total Luas Area harus = 1**.

Karena bentuk fungsinya adalah persegi panjang, maka: Luas = Panjang Alas × Tinggi = 1

- Panjang Alas = $(b - a)$
- Agar hasil perkaliannya pasti 1, maka **Tinggi kurva wajib diset ke kebalikannya**, yaitu $\frac{1}{b-a}$.

Parameter Matematis Distribusi Uniform

Secara matematis, profil dari Distribusi Uniform dilambangkan dengan $X \sim U(a, b)$, yang berarti peubah acak X menyebar seragam mulai dari batas bawah a hingga batas atas b.

Fungsi Kepadatan (PDF):

$f(x) = \frac{1}{b-a}$, untuk $a \leq x \leq b$

$f(x) = 0$, untuk nilai x di luar rentang.

Ukuran Pemusatan & Penyebaran

1. Nilai Harapan / Ekspektasi (μ)

Karena bentuknya simetris sempurna, rata-ratanya tepat berada di titik tengah antara a dan b.

$$\mu = \frac{a+b}{2}$$

2. Varians (σ^2)

Mengukur seberapa lebar atau fluktuatif rentang kejadian tersebut.

$$\sigma^2 = \frac{(b-a)^2}{12}$$

Studi Kasus: Kalkulasi Peluang Waktu Tunggu

Sebuah kereta bawah tanah dijadwalkan tiba setiap 20 menit. Seorang penumpang turun ke stasiun pada waktu yang sepenuhnya acak. Waktu tunggunya dapat dimodelkan berdistribusi seragam dari **0 hingga 20 menit**.

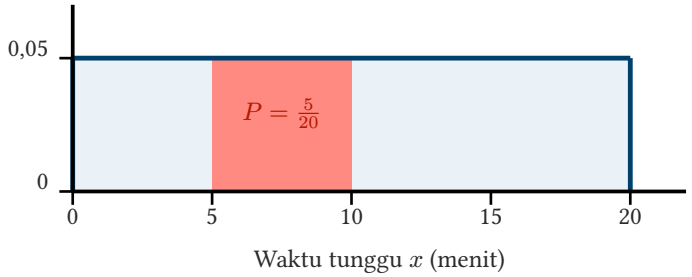
Pertanyaan:

Berapa probabilitas penumpang tersebut menunggu **di antara 5 hingga 10 menit**?

Studi Kasus: Kalkulasi Peluang Waktu Tunggu

1. Identifikasi Parameter

- Batas bawah rentang total (a) = 0
- Batas atas rentang total (b) = 20
- Tinggi Kurva PDF = $\frac{1}{20-0} = \frac{1}{20}$



2. Menghitung Luas Area Probabilitas

Probabilitas adalah luas persegi panjang dari interval target (yaitu 5 sampai 10):

$$P(5 \leq X \leq 10) = \text{Lebar Interval Target} \times \text{Tinggi PDF}$$

$$P(5 \leq X \leq 10) = (10 - 5) \times \frac{1}{20}$$

$$P(5 \leq X \leq 10) = 5 \times \frac{1}{20}$$

$$P = \frac{5}{20} = 0,25 \quad (\text{atau } 25\%)$$

Hubungan Simbiotis: Poisson vs Eksponensial

Distribusi Eksponensial adalah "saudara kembar" dari Distribusi Poisson yang kita pelajari minggu lalu. Keduanya memodelkan kejadian yang sama, namun dari **sudut pandang dimensi yang berlawanan**.

Sudut Pandang POISSON (Diskret)

Memodelkan **Jumlah Kejadian** dalam satu jendela waktu atau ruang yang sudah ditetapkan (tetap).

- Variabel (X): Jumlah (0, 1, 2, ...)
- Contoh: "Berapa banyak pelanggan yang datang ke bank antara jam 08:00 hingga 09:00?"

Sudut Pandang EKSPONENSIAL (Kontinu)

Memodelkan **Waktu Tunggu (Durasi)** atau jarak spasial di antara dua kejadian yang berurutan.

- Variabel (X): Waktu/Jarak ($X > 0$)
- Contoh: "Setelah pelanggan pertama masuk, berapa menit kita harus **menunggu** sampai pelanggan kedua masuk?"

Visualisasi Matematis & Parameter Eksponensial

Karena ini adalah pemodelan waktu tunggu, peluang untuk kejadian terjadi “segera” (waktu dekat) sangat padat, dan semakin lama kita menunggu, probabilitasnya semakin meluruh (menurun). Inilah yang membentuk kurva peluruhan eksponensial.

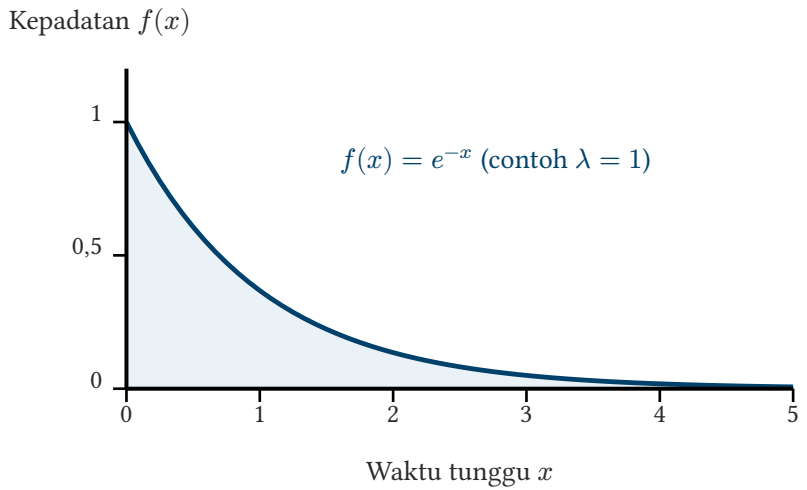
Fungsi Kepadatan (PDF):

$f(x) = \lambda e^{-\lambda x}$, untuk $x > 0$

Fungsi Kumulatif (CDF):

$F(x) = P(X \leq x) = 1 - e^{-\lambda x}$

Tips: Gunakan CDF untuk menghitung probabilitas waktu tunggu. Ini jauh lebih cepat daripada mengintegralkan PDF secara manual.



Parameter Statistik:

Ekspektasi (Rata-rata Waktu): $\mu = \frac{1}{\lambda}$

Sifat Unik: “Tanpa Ingatan” (*Memoryless Property*)

Distribusi Eksponensial adalah satu-satunya distribusi kontinu yang memiliki sifat **Memoryless** (Tidak punya ingatan). Artinya, **sejarah masa lalu tidak memengaruhi probabilitas di masa depan**.

$$P(X > s + t \mid X > s) = P(X > t)$$

Analogi Sifat Memoryless

Katakanlah Anda sedang menunggu bus yang jadwal kedatangannya berdistribusi eksponensial. Rata-rata waktu tunggunya adalah 15 menit.

Jika Anda sudah berdiri di halte selama 10 menit tanpa hasil ($s = 10$), probabilitas bus akan datang dalam 5 menit ke depan ($t = 5$) **SAMA PERSIS** dengan probabilitas seseorang yang baru saja tiba di halte dan berharap bus datang dalam 5 menit.

Sistem tidak “mengingat” bahwa Anda sudah lama menunggu. Probabilitasnya di-reset dari awal setiap saat.

Studi Kasus: Menghitung Probabilitas Waktu Tunggu

Sebuah komponen server API menerima permintaan masuk (*request*) dengan **tingkat rata-rata (λ) sebanyak 3 request per menit**.

Pertanyaan:

Berapa probabilitas waktu tunggu untuk **request berikutnya** akan datang dalam waktu **kurang dari 30 detik**?

Studi Kasus: Menghitung Probabilitas Waktu Tunggu

1. Sinkronisasi Satuan Waktu

- Tingkat kejadian (λ) = 3 request / menit
- Target waktu tunggu (x) = 30 detik.

Penting: Ubah x ke satuan yang sama dengan λ (menit). Maka $x = \frac{30}{60} = 0,5$ menit.

2. Eksekusi Rumus CDF

Karena kita mencari peluang menunggu **kurang dari** 0,5 menit, kita bisa langsung menggunakan rumus CDF:

$$P(X \leq 0,5) = 1 - e^{-\lambda \times x}$$

$$P(X \leq 0,5) = 1 - e^{-3 \times 0,5}$$

$$P(X \leq 0,5) = 1 - e^{-1,5}$$

Gunakan kalkulator untuk $e^{-1,5} \approx 0,2231$

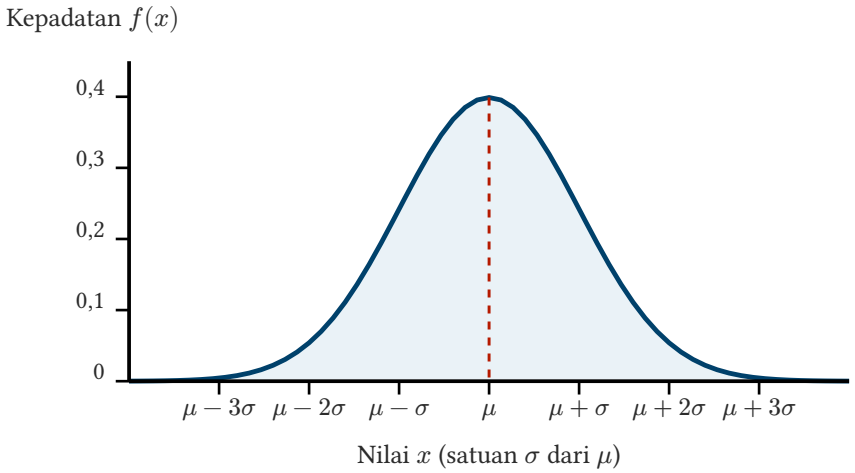
$$P = 1 - 0,2231 = 0,7769 \quad (77,69\%)$$

Blok 3 — Distribusi Normal & Z-Score

Memahami distribusi Gaussian yang menjadi pilar utama statistika inferensial dan sains data

Distribusi Normal (Kurva Lonceng)

Distribusi yang paling sering dijumpai di alam: tinggi badan, IQ, nilai ujian, galat pengukuran. Kurvanya simetris dan berbentuk lonceng. Bentuk dan posisinya ditentukan hanya oleh dua parameter: rata-rata μ dan simpangan baku σ.



Rumus Fungsi Kepadatan

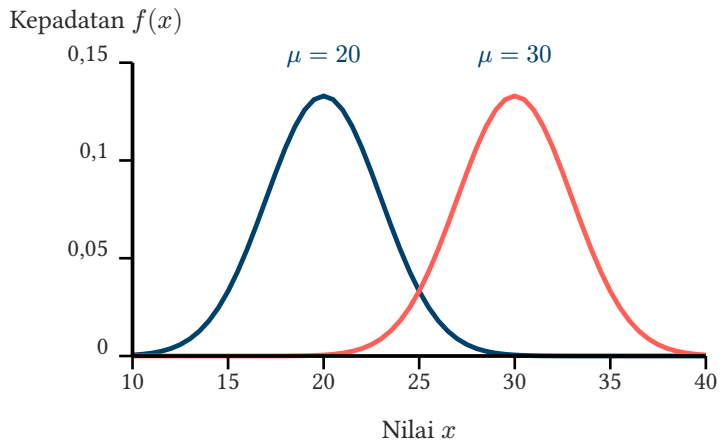
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- μ menentukan **letak** puncak (pusat kurva).
- σ menentukan **lebar** lonceng.
- Notasi: $X \sim N(\mu, \sigma^2)$.
- Rumus ini tidak perlu diintegralkan manual. Luasnya dibaca dari Tabel Z (dua slide berikutnya).

Sifat kurva: (1) simetris terhadap μ; (2) mean = median = modus = μ (di puncak); (3) luas total $= 1$, masing-masing sisi $0,5$; (4) ekor menjalar ke tak hingga dan tidak menyentuh sumbu x.

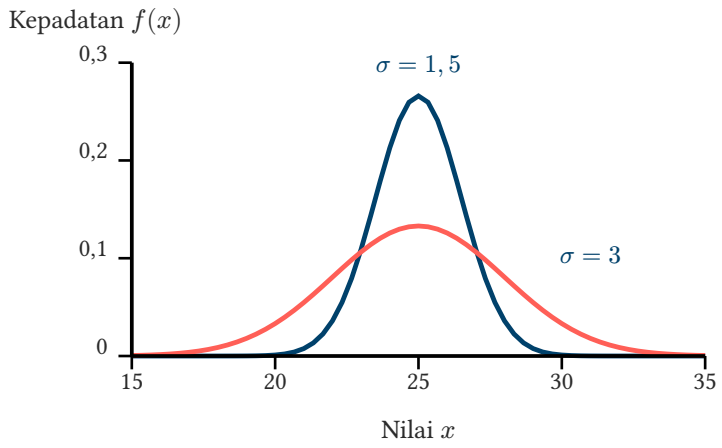
Pengaruh μ dan σ pada Bentuk Kurva

Dua parameter mengubah kurva dengan cara berbeda. Luas total di bawah setiap kurva tetap 1.



σ sama, μ berbeda

Kurva **bergeser** ke kanan atau kiri. Bentuknya tidak berubah.



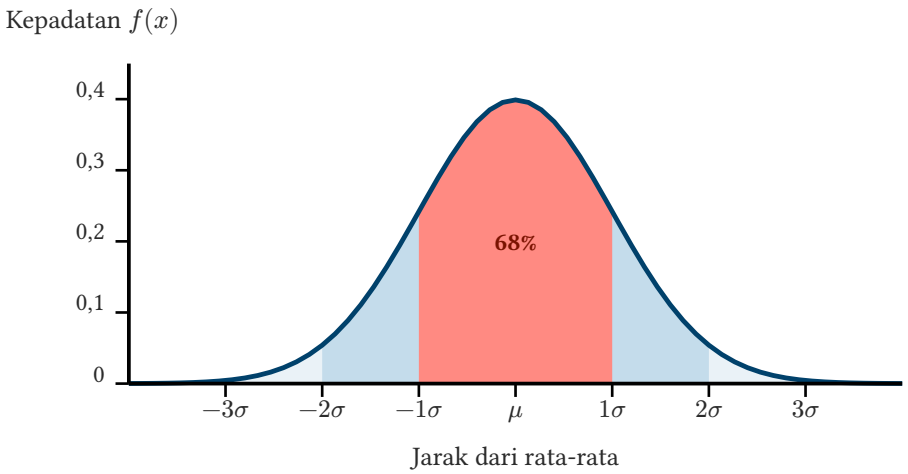
μ sama, σ berbeda

σ kecil: tinggi dan sempit. σ besar: pendek dan lebar.

Karena luas total harus tetap 1, kurva yang melebar otomatis menjadi lebih pendek. Itulah mengapa tinggi puncak $\frac{1}{\sigma\sqrt{2\pi}}$ makin kecil saat σ membesar.

Aturan Empiris 68 - 95 - 99,7

Untuk distribusi normal, proporsi data pada jarak tertentu dari rata-rata sudah **tetap**, berapa pun nilai μ dan σ.



Merah $\approx 68\%$, tambah biru sedang $\approx 95\%$, tambah biru muda $\approx 99,7\%$.

Cara Membaca

- $\mu \pm 1\sigma$: $\approx 68\%$ data
- $\mu \pm 2\sigma$: $\approx 95\%$ data
- $\mu \pm 3\sigma$: $\approx 99,7\%$ data

Luas per bagian (satu sisi): $\approx 34\%, 13,6\%, 2,1\%, 0,1\%$. Dua sisi dijumlahkan: $2 \times (34 + 13,6 + 2,1 + 0,1) \approx 99,7 + 0,3$ sisa ekor.

Contoh: IQ dengan $\mu = 100$, $\sigma = 15$. Sekitar 95% orang memiliki IQ antara $100 - 30 = 70$ dan $100 + 30 = 130$.

Data di luar $\pm 3\sigma$ (sisa $0,3\%$) sering dianggap **pencilan** (outlier).

Mengapa Perlu Standarisasi?

Setiap populasi punya skala berbeda (rupiah, cm, menit). Tabel luas hanya ada untuk **satu** kurva normal, yaitu yang berpusat di 0 dengan $\sigma = 1$. Jadi data apa pun kita ubah dulu ke skala itu.

Dua Langkah Mengubah Skala

1. Geser supaya pusatnya di 0: kurangi dengan rata-rata.

$$X - \mu$$

2. Skalakan supaya lebarnya jadi 1: bagi dengan simpangan baku.

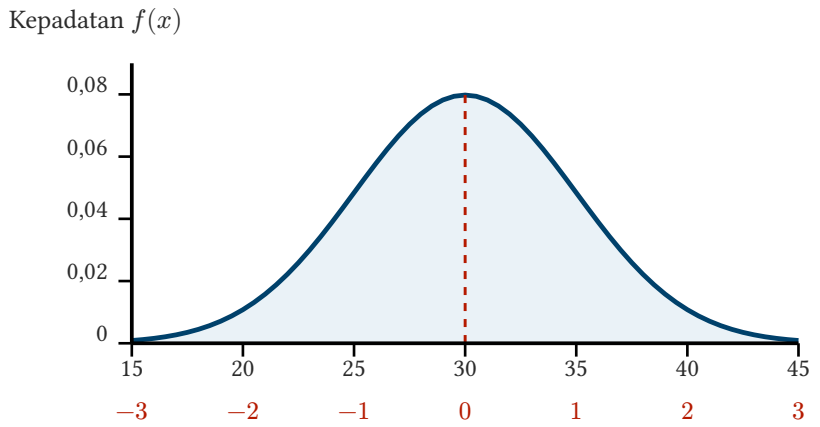
$$\frac{X-\mu}{\sigma}$$

Hasilnya disebut **Z-score**:

$$Z = \frac{X-\mu}{\sigma}$$

Z menyatakan **berapa simpangan baku** nilai X berada dari rata-rata. Distribusi Z disebut Normal Baku: $Z \sim N(0, 1)$.

Contoh: waktu tempuh, $\mu = 30$ menit, $\sigma = 5$ menit



Atas: waktu x (menit). Bawah (merah): Z

Cara Membaca Z-Score

Arti Tanda dan Nilai Z

- $\boldsymbol{Z = 0}$: nilai tepat sama dengan rata-rata.
- $\boldsymbol{Z > 0}$: nilai di **kanan** rata-rata. $Z = 1,5$ berarti $1,5$ simpangan baku di atas rata-rata.
- $\boldsymbol{Z < 0}$: nilai di **kiri** rata-rata. $Z = -2$ berarti 2 simpangan baku di bawah rata-rata.
- Makin besar $|Z|$, makin langka nilainya. $|Z| > 3$ sangat jarang.

Contoh: nilai ujian, $\mu = 70$, $\sigma = 10$

- $X = 85$: $Z = \frac{85-70}{10} = 1,5$
- $X = 55$: $Z = \frac{55-70}{10} = -1,5$
- $X = 70$: $Z = 0$

Kegunaan: Membandingkan Skala Berbeda

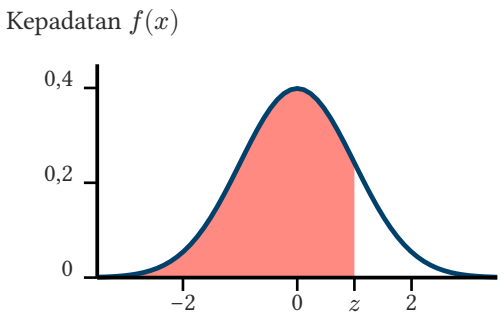
Budi mendapat 80 di Matematika dan 85 di Fisika. Mana yang lebih baik secara relatif?

Mata kuliah	μ	σ	Z
Matematika	70	5	$\frac{80-70}{5} = 2$
Fisika	75	10	$\frac{85-75}{10} = 1$

Nilai Fisika lebih tinggi (85), tetapi Z Matematika lebih besar ($2 > 1$). Budi **relatif lebih unggul di Matematika** karena nilainya lebih jauh di atas rata-rata kelasnya.

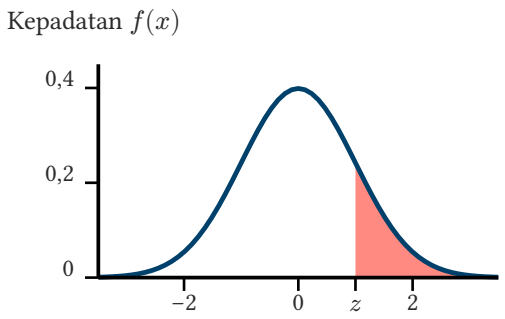
Membaca Tabel Z: Luas di Bawah Kurva Normal Baku

Tabel Z memberikan luas dari ujung kiri sampai z, yaitu $\Phi(z) = P(Z \le z)$. Semua jenis peluang disusun dari nilai ini.



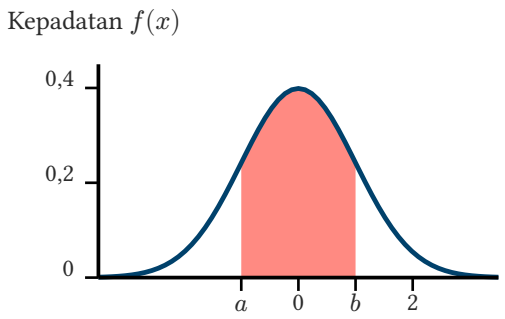
Luas kiri

$P(Z \le z) = \Phi(z)$ (langsung dari tabel)



Luas kanan

$P(Z > z) = 1 - \Phi(z)$



Antara dua titik

$P(a \le Z \le b) = \Phi(b) - \Phi(a)$

Sifat Simetri

Untuk z negatif: $\Phi(-z) = 1 - \Phi(z)$. Contoh: $\Phi(-1) = 1 - 0,8413 = 0,1587$.

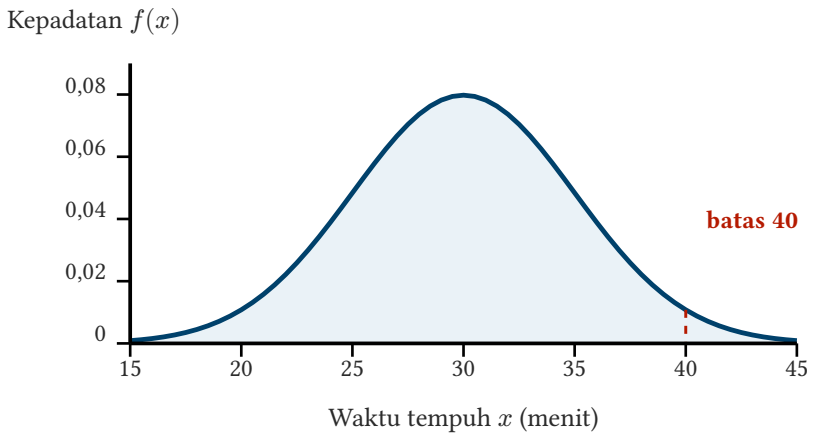
Nilai Tabel yang Sering Dipakai

$\Phi(1,00) = 0,8413$, $\Phi(1,50) = 0,9332$, $\Phi(1,96) = 0,9750$, $\Phi(2,00) = 0,9772$, $\Phi(3,00) = 0,9987$

Latihan: Implementasi Z-Score & Tabel Normal

KASUS

Waktu tempuh mahasiswa dari rumah ke kampus berdistribusi normal dengan rata-rata $\mu = 30$ menit dan simpangan baku $\sigma = 5$ menit. Seorang mahasiswa berangkat 40 menit sebelum kuliah dimulai.



Pertanyaan

1. Hitung Z-Score untuk waktu tempuh 40 menit.

2. Berapa peluang ia **terlambat**, yaitu waktu tempuh > 40 menit?

3. Berapa peluang waktu tempuhnya antara 25 dan 35 menit?

Pembahasan 1: Z-Score dan Peluang Terlambat

Langkah 1. Tulis yang diketahui: $\mu = 30$, $\sigma = 5$, batas $x = 40$.

Langkah 2. Standarisasi:

$$Z = \frac{40 - 30}{5} = \frac{10}{5} = 2,00$$

Batas 40 menit berada 2 simpangan baku di atas rata-rata.

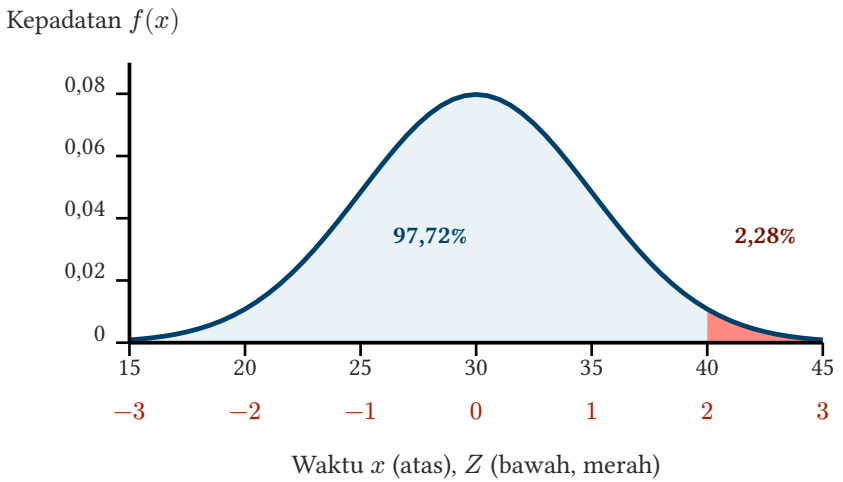
Langkah 3. Ubah ke bentuk Z:

$$P(X > 40) = P(Z > 2,00)$$

Langkah 4. Pakai luas kanan: tabel memberi luas kiri, jadi

$$P(Z > 2,00) = 1 - \Phi(2,00) = 1 - 0,9772$$

$$P = 0,0228 \textbf{ (2,28\%)}$$



Waktu x (atas), Z (bawah, merah)

Biru $= \Phi(2,00) = 0,9772$. Merah (ekor kanan) $= 1 - 0,9772 = 0,0228$.

Hanya sekitar $2,28\%$ kemungkinan terlambat. Sesuai aturan empiris: di luar 2σ ada sekitar 5%, dan setengahnya (ekor kanan) sekitar $2,5\%$.

Pembahasan 2: Peluang Antara 25 dan 35 Menit

Langkah 1. Standarisasi kedua batas:

$$z_1 = \frac{25-30}{5} = -1,00 \quad z_2 = \frac{35-30}{5} = 1,00$$

Langkah 2. Bentuk selisih luas kiri:

$$P(25 \le X \le 35) = \Phi(1,00) - \Phi(-1,00)$$

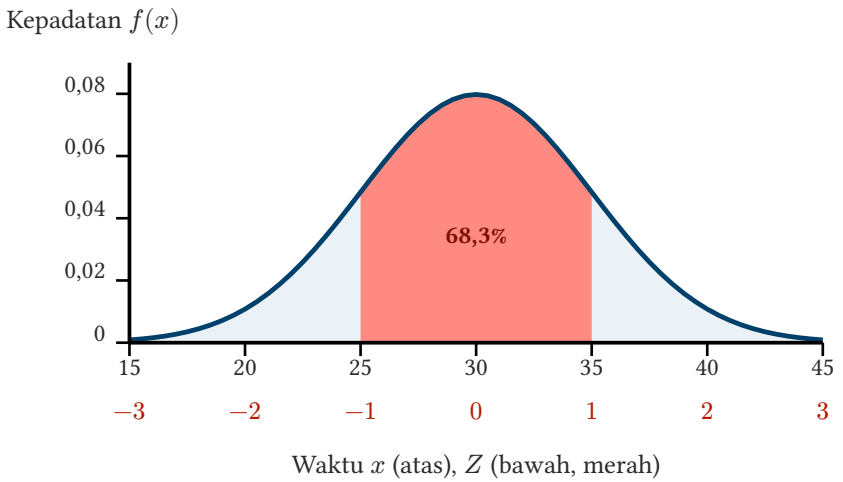
Langkah 3. Baca tabel: $\Phi(1,00) = 0,8413$ dan

$$\Phi(-1,00) = 1 - 0,8413 = 0,1587$$

Langkah 4. Kurangkan:

$$0,8413 - 0,1587 = 0,6826$$

$$P \approx 0,683 \textbf{ (68,3\%)}$$



Waktu x (atas), Z (bawah, merah)

Daerah merah: dari -1σ sampai $+1\sigma$.

Hasilnya cocok dengan aturan empiris: sekitar 68% data berada dalam $\mu \pm 1\sigma$. Jadi sekitar 68% mahasiswa menempuh perjalanan 25 sampai 35 menit.

Penutup & Evaluasi

Rekapitulasi distribusi kontinu dan persiapan teori inferensi

Rekapitulasi Modul Probabilitas

Pertemuan hari ini menutup paruh pertama perkuliahan kita. Kita telah menyelesaikan evolusi analisis data dari ranah diskret ke kontinu.

Peubah Acak Diskret (Review Pertemuan 5)

- Hitungan *countable* (0, 1, 2, ...).
- Peluang tepat di titik x dihitung dengan $P(X = x)$.
- Menggunakan Binomial (Sukses/Gagal n kali) dan Poisson (Kejadian per interval).

Peubah Acak Kontinu (Pertemuan 6)

- Pengukuran tanpa batas resolusi (suhu, durasi, bobot).
- Peluang titik eksak $P(X = x)$ selalu 0. Peluang dihitung via luasan area PDF (integral/rentang).
- Model: Uniform, Eksponensial, dan pilar utama Gaussian (Distribusi Normal).

Persiapan Materi Selanjutnya: Jembatan Inferensial

Tanya Jawab?

Mulai minggu depan, kita akan masuk ke Paruh Kedua silabus. Kita tidak lagi berbicara soal probabilitas data teoritis yang dikalkulasi secara manual dari persamaan. Kita akan mulai mempelajari data nyata. Bagaimana jika kita tidak punya akses ke seluruh Populasi?

Agenda Pertemuan 07 Berikutnya:

Sampling dan Distribusi Sampling: Mengenal konsep populasi vs sampel, metode pengambilan data (acak, berstrata), dan “Keajaiban” Teorema Limit Pusat (CLT) sebagai pondasi inferensial sebelum UTS.