ESTIMASI PARAMETER

Prof. Dr. Arna Fariza, S.Kom., M.Kom.

TOPIK BAHASAN

1. Recap: Distribusi Sampling & CLT
2. Pengertian Estimasi Parameter
3. Estimasi Titik (Point Estimation)
4. Estimasi Interval (Interval Estimation)
5. Interval Kepercayaan (Confidence Interval)
6. Interval Kepercayaan Mean dengan Distribusi Z
7. Interval Kepercayaan Mean dengan Distribusi t
8. Aplikasi di Bidang IT
9. Latihan Soal & Kesimpulan

RECAP - DISTRIBUSI SAMPLING & CLT

Mengingat Kembali:
Kita menggunakan **sampel** untuk mengestimasi **populasi** karena meneliti seluruh populasi seringkali tidak mungkin.

- **Populasi:** Rata-rata sebenarnya (μ), Standar Deviasi (σ).
- **Sampel:** Rata-rata sampel ($\bar{x}$), Standar Deviasi sampel (s).

Teorema Limit Pusat (CLT):
Jika ukuran sampel $n \geq 30$, distribusi rata-rata sampel ($\bar{x}$) akan mendekati **Normal**.

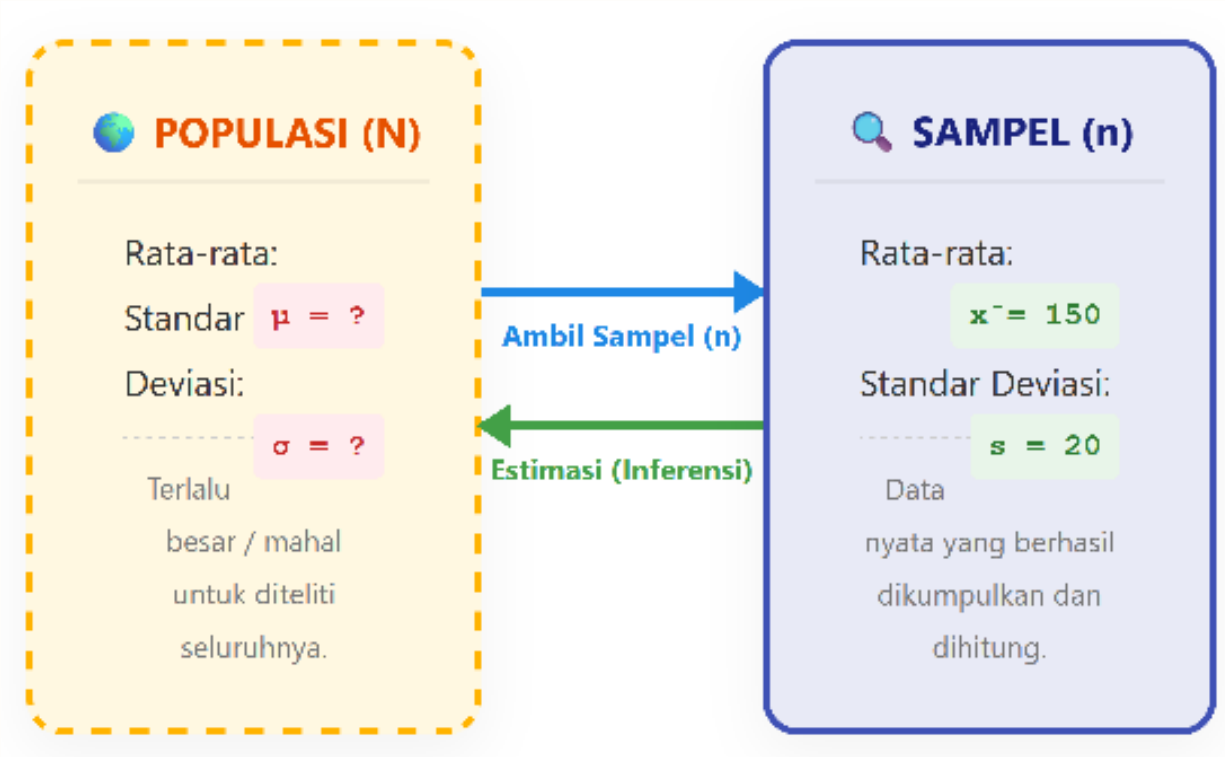
- Mean distribusi sampling: $\mu_{\bar{x}} = \mu$
- Standard Error: $SE = \sigma/\sqrt{n}$

Pertanyaan Kunci:
Jika kita hanya punya satu nilai $\bar{x}$ dari sampel, seberapa yakin kita bahwa nilai tersebut mendekati μ yang sebenarnya? Di sinilah **Estimasi Parameter** berperan.

PENGERTIAN ESTIMASI PARAMETER

Definisi:
Estimasi parameter adalah proses menggunakan data sampel (statistik) untuk menebak atau memperkirakan nilai parameter populasi yang tidak diketahui.

Ilustrasi:



Dua Jenis Estimasi:
1. **Estimasi Titik (Point Estimation)**
2. **Estimasi Interval (Interval Estimation)**

ESTIMASI TITIK (POINT ESTIMATION)

Definisi:
Estimasi titik adalah menggunakan **satu nilai tunggal** dari statistik sampel untuk menebak parameter populasi.

Contoh:
- Rata-rata waktu respons sampel ($\bar{x}$) = 150 ms digunakan untuk mengestimasi rata-rata waktu respons populasi (μ).
- Proporsi bug sampel ($\hat{p}$) = 5% digunakan untuk mengestimasi proporsi bug populasi (p).

Kelebihan:
- Sederhana, mudah dihitung, dan mudah dipahami.

Kelemahan (Sangat Penting):
- Hampir mustahil nilai sampel **tepat sama** dengan nilai populasi.
- Kita **tidak tahu seberapa akurat** tebakan tersebut (tidak ada ukuran tingkat keyakinan).

Solusi: Kita membutuhkan **Estimasi Interval**.

ESTIMASI INTERVAL (INTERVAL ESTIMATION)

Definisi:
Estimasi interval adalah menggunakan **rentang nilai** (interval) untuk mengestimasi parameter populasi, disertai dengan **tingkat kepercayaan** (confidence level) tertentu.

Contoh:
"Kita yakin **95%** bahwa rata-rata waktu respons server populasi berada di antara **145 ms hingga 155 ms**.“

Mengapa Lebih Baik?
- Memberikan rentang yang lebih realistis.
- Menyertakan ukuran presisi (melalui lebar interval).
- Menyertakan tingkat keyakinan (misal: 90%, 95%, 99%).

Istilah Penting: Hasil dari estimasi interval disebut **Interval Kepercayaan (Confidence Interval / CI)**.

INTERVAL KEPERCAYAAN (CI) - KONSEP

Apa arti "95% Yakin"?
Ini **bukan** berarti peluang parameter populasi berada di interval tersebut adalah 95%. Parameter populasi (μ) adalah nilai tetap, bukan variabel acak.

Interpretasi yang Benar:
Jika kita mengambil 100 sampel acak berbeda dan membuat interval kepercayaan 95% untuk masing-masing sampel, maka sekitar **95 dari 100 interval** tersebut akan berisi nilai rata-rata populasi (μ) yang sebenarnya.



TINGKAT KEPERCAYAAN & NILAI KRITIS (Z)

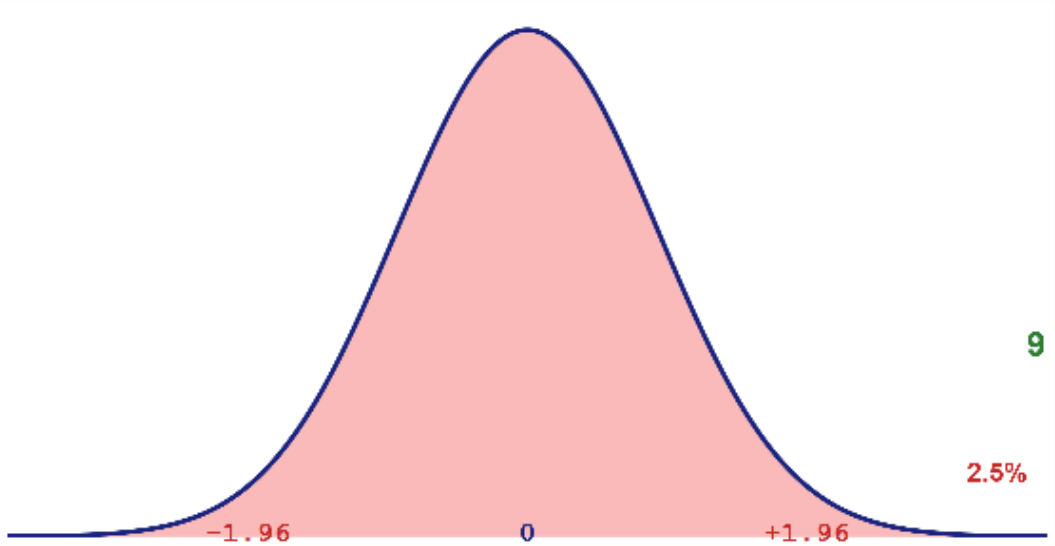
Tingkat Kepercayaan (Confidence Level):
Seberapa besar keyakinan kita (biasanya 90%, 95%, atau 99%).

Tingkat Signifikansi (α):
$\alpha = 1 -$ Confidence Level.

- Jika CI = 95%, maka $\alpha = 0.05$.
- Jika CI = 90%, maka $\alpha = 0.10$.

Nilai Kritis Z ($Z_{\alpha/2}$):
Nilai Z yang membatasi area di ekor distribusi normal.
(Catatan: Untuk 95%, area di setiap ekor adalah 2.5%, sehingga $P(Z < 1.96) = 0.9750$)

Tingkat Kepercayaan	α	$Z_{\alpha/2}$(dari Tabel Z)
90%	0.10	1.645
95%	0.05	1.96
99%	0.01	2.576



INTERVAL KEPERCAYAAN MEAN - DISTRIBUSI Z

Kapan Menggunakan Z?

1. Standar deviasi populasi (σ) **DIKETAHUI**.
2. ATAU Ukuran sampel besar ($n \geq 30$) sehingga standar deviasi sampel (s) dapat menggantikan σ (berdasarkan CLT).

Rumus Interval Kepercayaan:

$$\bar{x} - Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) \leq \mu \leq \bar{x} + Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

Atau disingkat:

$$\bar{x} \pm Z_{\alpha/2} \cdot SE$$

Keterangan:

- $\bar{x}$= Rata-rata sampel
- $Z_{\alpha/2}$= Nilai kritis Z
- σ= Standar deviasi populasi (atau s jika $n \geq 30$)
- n= Ukuran sampel
- SE= Standard Error = $\sigma/\sqrt{n}$

CONTOH SOAL - INTERVAL Z

Kasus: Evaluasi Kinerja Server Web

Sebuah tim DevOps mengukur waktu respons dari 64 request ke server web. Diketahui rata-rata waktu respons sampel ($\bar{x}$) adalah **150 ms**. Dari dokumentasi sistem, standar deviasi populasi (σ) diketahui sebesar **16 ms**.

Pertanyaan:

Buatlah interval kepercayaan **95%** untuk rata-rata waktu respons populasi!

Penyelesaian:

1. **Identifikasi Data:**
   $\bar{x} = 150$, $\sigma = 16$, $n = 64$.
   Tingkat kepercayaan 95% → $Z_{0.025} = 1.96$.
2. **Hitung Standard Error (SE):**
   $SE = \frac{\sigma}{\sqrt{n}} = \frac{16}{\sqrt{64}} = \frac{16}{8} = 2$ms.
3. **Hitung Margin of Error (MoE):**
   $MoE = Z \cdot SE = 1.96 \times 2 = 3.92$ms.
4. **Tentukan Interval:**
   Batas Bawah: $150 - 3.92 = \mathbf{146.08}$
   Batas Atas: $150 + 3.92 = \mathbf{153.92}$

Kesimpulan:

Kita yakin **95%** bahwa rata-rata waktu respons populasi sebenarnya berada di antara **146.08 ms hingga 153.92 ms**.

INTERVAL KEPERCAYAAN MEAN - DISTRIBUSI T

Kapan Menggunakan Distribusi t?

1. Standar deviasi populasi (σ) **TIDAK DIKETAHUI**.
2. Ukuran sampel **KECIL** ($n < 30$).
3. Populasi diasumsikan berdistribusi normal (atau mendekati normal).

Mengapa t, bukan Z?

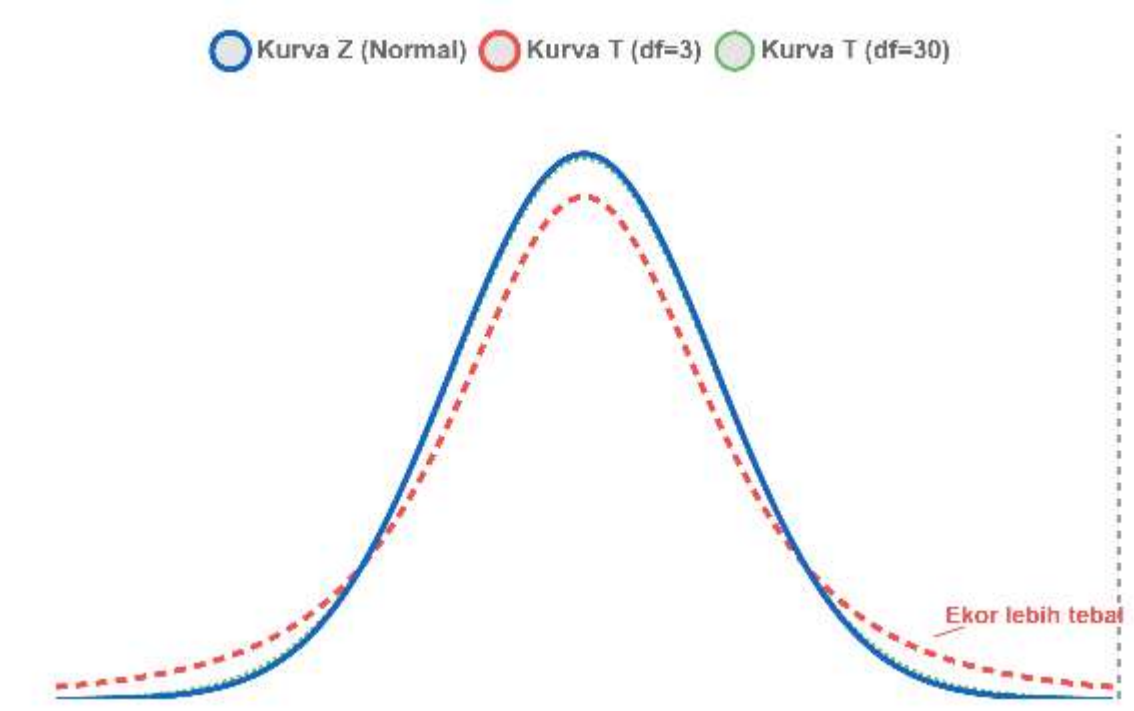
Jika σtidak diketahui dan $n$kecil, kita menggunakan standar deviasi sampel (s). Ketidakpastian ini membuat distribusi lebih lebar (ekor lebih tebal) dibandingkan Z. Ini disebut **Distribusi t-Student**.

Rumus:

$$\bar{x} \pm t_{\alpha/2,\, df} \left(\frac{s}{\sqrt{n}} \right)$$

- df(Degrees of Freedom / Derajat Kebebasan) = $n - 1$.
- Nilai $t$dicari dari **Tabel Distribusi t** berdasarkan dfdan $\alpha/2$.

PERBANDINGAN KURVA Z DAN T



- Semakin besar df (semakin besar n), distribusi t akan **semakin mirip** dengan distribusi normal Z.
- Jika $n \geq 30$, nilai t biasanya sudah mendekati nilai Z (bisa menggunakan Z sebagai pendekatan jika tabel t tidak tersedia).

CONTOH SOAL - INTERVAL T

Kasus: Pengujian Waktu Eksekusi Algoritma
Seorang developer menguji waktu eksekusi algoritma baru pada sampel berukuran kecil, yaitu $n = 16$ **kali percobaan**. Didapatkan rata-rata waktu eksekusi sampel ($\bar{x}$) = **25 detik** dan standar deviasi sampel (s) = **4 detik**.

Pertanyaan:
Buatlah interval kepercayaan **90%** untuk rata-rata waktu eksekusi populasi

CONTOH SOAL - INTERVAL T

Penyelesaian:

1. **Identifikasi Data:**
   $\bar{x} = 25$, $s = 4$, $n = 16$.
   $df = n - 1 = 15$.
   CI = 90%, $\alpha = 0.10$, $\alpha/2 = 0.05$.
   Dari Tabel t (df=15, $\alpha/2 = 0.05$), didapat $t = 1.753$.
2. **Hitung Standard Error (SE):**
   $SE = \frac{s}{\sqrt{n}} = \frac{4}{\sqrt{16}} = \frac{4}{4} = 1$detik.
3. **Hitung Margin of Error (MoE):**
   $MoE = t \cdot SE = 1.753 \times 1 = 1.753$detik.
4. **Tentukan Interval:**
   Batas Bawah: $25 - 1.753 = \mathbf{23.247}$
   Batas Atas: $25 + 1.753 = \mathbf{26.753}$

Kesimpulan:

Kita yakin **90%** bahwa rata-rata waktu eksekusi algoritma populasi sebenarnya berada di antara **23.25 detik hingga 26.75 detik**.

APLIKASI DI BIDANG TEKNIK INFORMATIKA

1. SLA (Service Level Agreement) Jaringan
Menentukan batas bawah dan batas atas waktu respons server. Jika CI 95% waktu respons adalah 100-120 ms, maka manajemen bisa berjanji ke klien dengan SLA maksimal 120 ms.

2. A/B Testing UI/UX
Mengestimasi selisih konversi antara dua desain antarmuka. Alih-alih hanya melihat nilai rata-rata, perusahaan melihat *confidence interval* untuk memastikan peningkatan konversi bukan sekadar kebetulan.

3. Quality Assurance (QA) Software
Mengestimasi rata-rata jumlah bug per modul. Memudahkan manajer proyek mengalokasikan waktu *debugging* dan sumber daya.

4. Latensi Jaringan
Mengestimasi rata-rata *ping* atau latensi. Digunakan oleh ISP untuk mengukur kualitas layanan mereka di berbagai wilayah.

LATIHAN SOAL

Soal 1 (Distribusi Z):
Sebuah sistem database memiliki waktu query rata-rata dari sampel sebanyak $n = 100$query sebesar $\bar{x} = 50$ ms. Diketahui standar deviasi populasi $\sigma = 10$ ms. Hitunglah interval kepercayaan 99% untuk rata-rata waktu query populasi!

Soal 2 (Distribusi t):
Sebuah tim QA menguji kecepatan loading sebuah fitur baru pada 9 perangkat. Didapatkan rata-rata waktu loading $s = 3.2$ detik dan standar deviasi sampel $s = 0.6$ detik. Hitunglah interval kepercayaan 95% untuk rata-rata waktu loading populasi! (t untuk df=8, $\alpha/2 = 0.025$ adalah **2.306**).

LATIHAN SOAL

Jawaban Soal 1:

- Diketahui: $\bar{x} = 50$, $\sigma = 10$, $n = 100$, CI = 99%.
- $Z_{0.005} = 2.576$(dari Tabel Z).
- $SE = 10/\sqrt{100} = 10/10 = 1$.
- $MoE = 2.576 \times 1 = 2.576$.
- Interval: 50 ± 2.576 →**(47.424 ms hingga 52.576 ms)**.

Jawaban Soal 2:

- Diketahui: $\bar{x} = 3.2$, $s = 0.6$, $n = 9$, df=8, $t_{0.025} = 2.306$.
- $SE = 0.6/\sqrt{9} = 0.6/3 = 0.2$.
- $MoE = 2.306 \times 0.2 = 0.4612$.
- Interval: 3.2 ± 0.4612 →**(2.7388 detik hingga 3.6612 detik)**.

KESIMPULAN

1. **Estimasi Parameter** digunakan untuk menebak parameter populasi menggunakan data sampel.
2. **Estimasi Titik** memberikan satu nilai tunggal, tetapi tidak memiliki ukuran akurasi.
3. **Estimasi Interval** memberikan rentang nilai dan lebih andal.
4. **Interval Kepercayaan (CI)** menunjukkan seberapa yakin kita bahwa interval tersebut memuat parameter populasi.
5. **Gunakan Distribusi Z** jika σdiketahui atau sampel besar ($n \geq 30$).
6. **Gunakan Distribusi t** jika σtidak diketahui dan sampel kecil ($n < 30$).
7. **Margin of Error (MoE)** adalah batas penyimpangan yang diizinkan; semakin besar sampel atau kecil standar deviasi, semakin kecil MoE.

TUGAS

Soal 1:
Sebuah aplikasi mobile memiliki waktu startup rata-rata dari 49 sampel sebesar 2.5 detik. Diketahui dari pengujian sebelumnya bahwa standar deviasi populasi adalah 0.5 detik. Buatlah interval kepercayaan 95% untuk rata-rata waktu startup populasi!

Soal 2:
Seorang data analyst mengukur penggunaan CPU dari 15 server. Didapatkan rata-rata penggunaan CPU sampel sebesar 65% dengan standar deviasi sampel 8%. Asumsikan data berdistribusi normal, buatlah interval kepercayaan 90% untuk rata-rata penggunaan CPU populasi! (t untuk df=14, $\alpha/2 = 0.05$adalah **1.761**).