Estimasi Titik dan Selang Kepercayaan untuk μ

Tim Dosen Pengantar Statistika

- **CPMK**

Mahasiswa Program Studi Teknologi Sains Data akan dapat menerapkan konsep dan metode statistika sesuai dengan permasalahan dan karakteristik data dengan baik dan benar. (C3)

- **Sub-CPMK**

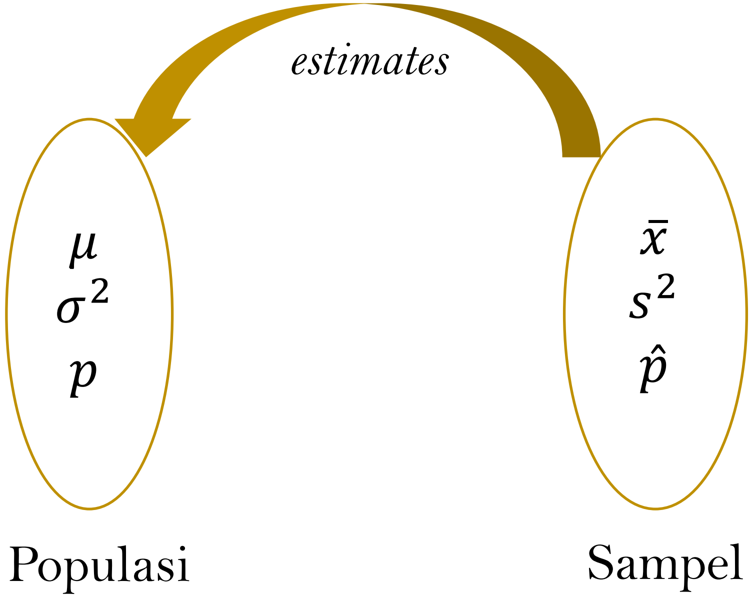
✓Mampu memformulasikan selang kepercayaan dengan baik dan benar (Sub-CPMK-6, C2),

✓Mampu menerapkan statistika inferensia dengan estimasi interval dan uji hipotesis dengan baik dan benar (Sub-CPMK-7, C3)

Outline

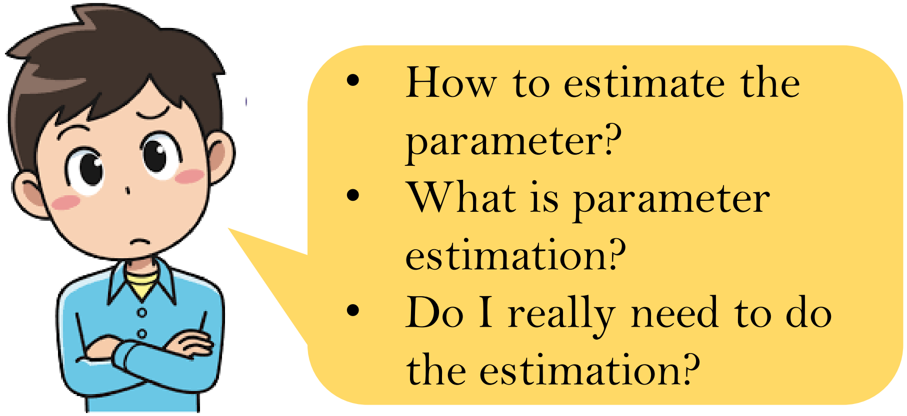
- Estimasi parameter
- Estimasi titik
- Estimasi interval / selang kepercayaan untuk μ dengan σ diketahui
- Tingkat kepercayaan
- Estimasi interval / selang kepercayaan untuk μ dengan σ tidak diketahui

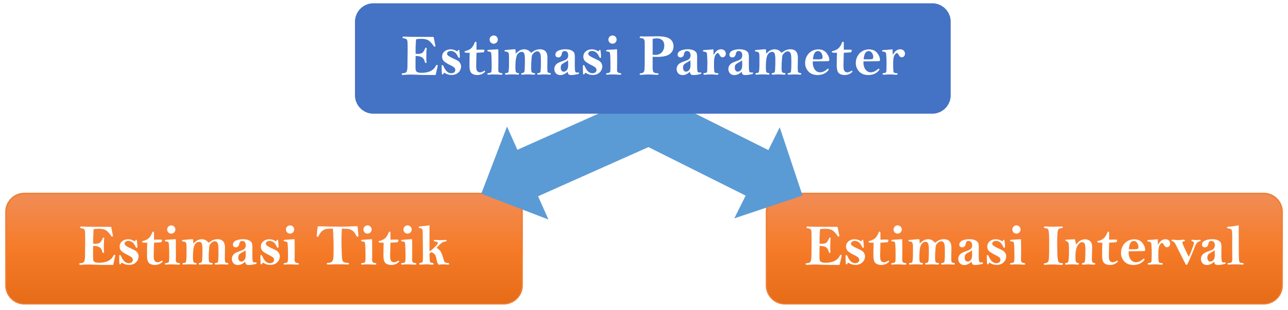
Estimasi Parameter



Estimasi parameter (penaksiran parameter) adalah pendugaan **karakteristik populasi (parameter)** dengan menggunakan **karakteristik sampel (statistik).**

Estimasi parameter diperlukan karena **nilai dari populasi (parameter) sulit diketahui** sehingga didekati dengan pengukuran melalui sampel.



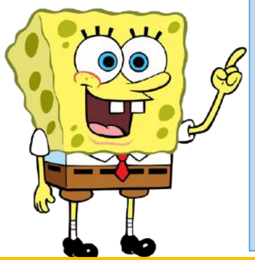


Estimasi Titik

- **Estimasi titik** adalah penaksiran karakteristik populasi dengan sebuah nilai karakteristik dari sampel.
- Estimasi titik **lebih mudah** dari segi **penghitungan**, **tetapi penaksirannya** sangat **diragukan.** Hal ini karena jarang ada nilai karakteristik populasi sama persis dengan nilai karakteristik sampel.

Estimasi Titik

	Parameter	Estimator
Rata-rata	$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
Variansi	$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2$	$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$
Proporsi	$p = \frac{X}{N}$	$\hat{p} = \frac{x}{n}$



Ingat! **Sifat estimator yang baik**:

1. *Unbiased*
2. *Efficient*
3. *Consistent*
4. *Sufficient*

Estimasi Interval

- Mengestimasi parameter populasi menggunakan statistik sampel cenderung sulit dilakukan sehingga pasti akan ada **error** dalam estimasinya.
- Error atau ketidakpastian dalam estimasi ini bisa dinyatakan menggunakan **estimasi interval**:

$$\boxed{point\ estimation \pm margin\ of\ error}$$

Estimasi Interval

- **Estimasi interval** adalah penaksiran populasi dengan nilai-nilai dalam suatu interval tertentu.
- Interval ini disebut **selang/interval kepercayaan (*confidence interval*).** Interval kepercayaan adalah rentang angka yang diyakini mencakup parameter populasi yang tidak diketahui.
- Dasar dari adanya estimasi interval adalah setiap penaksiran mengandung peluang kesalahan (error).
- Misalkan parameter yang akan diestimasi adalah θ dan d adalah *margin of error*. Estimasi intervalnya adalah $\hat{\theta} \pm d$ atau $\hat{\theta} - d < \theta < \hat{\theta} + d$.

Estimasi Interval

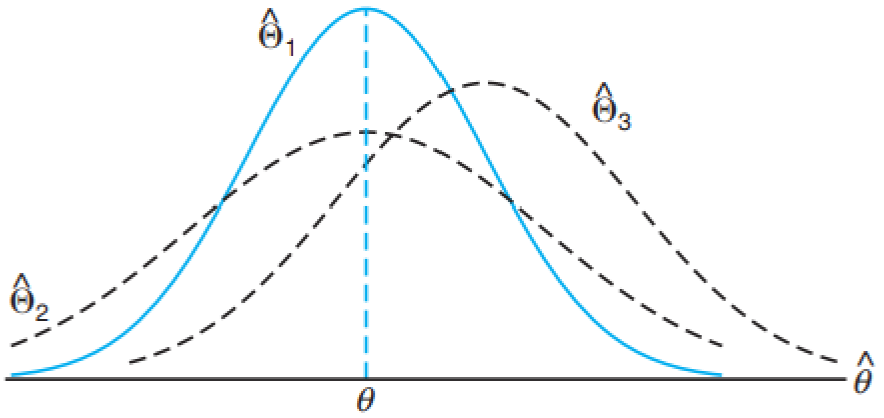
- Estimasi interval didasarkan pada suatu **distribusi peluang**, biasa yang digunakan adalah distribusi normal, distribusi student-t, distribusi F dan distribusi khi-kuadrat.
- Misalnya estimasi interval menggunakan distribusi normal, maka interval kepercayaannya:

$$\boxed{\hat{\theta} \pm Z_{\frac{\alpha}{2}} Se(\hat{\theta})}$$

$Z_{\frac{\alpha}{2}}$: Distribusi Normal Standar

$Se(\hat{\theta})$: *Standard Error*

$Z_{\frac{\alpha}{2}} Se(\hat{\theta})$: *Margin of error* atau *sampling error*



Distribusi sampling estimator berbeda dari θ

Estimasi Interval Rata-rata

1	Deviasi standar (σ) diketahui (sampel diambil dari populasi berdistribusi normal atau sampel besar)
	$\bar{x} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$
2	Deviasi standar (σ) tidak diketahui dengan sampel besar (n $\geq$ 30)
	$\bar{x} - Z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} < \mu < \bar{x} + Z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$
3	Deviasi standar (σ) tidak diketahui dengan sampel kecil (n < 30)
	$\bar{x} - t_{\frac{\alpha}{2},(n-1)} \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\frac{\alpha}{2},(n-1)} \frac{s}{\sqrt{n}}$

95% Probability Interval

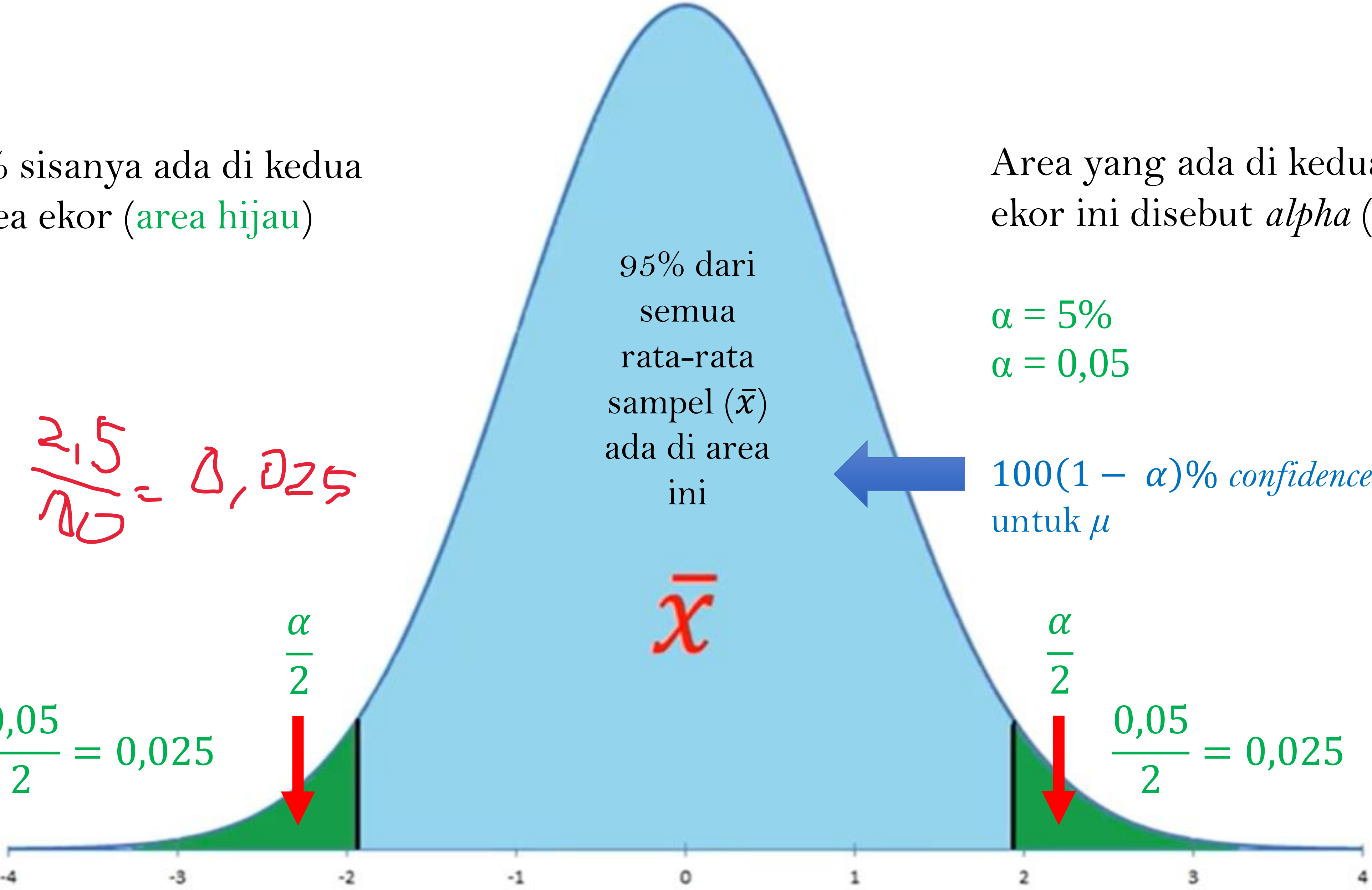
5% sisanya ada di kedua area ekor (area hijau)

$$\frac{2,5}{100} = 0,025$$

Area yang ada di kedua ekor ini disebut *alpha* (α)

α = 5%
α = 0,05

$100(1 - \alpha)\%$ *confidence interval* untuk μ



$$\frac{0,05}{2} = 0,025$$

$$\frac{0,05}{2} = 0,025$$

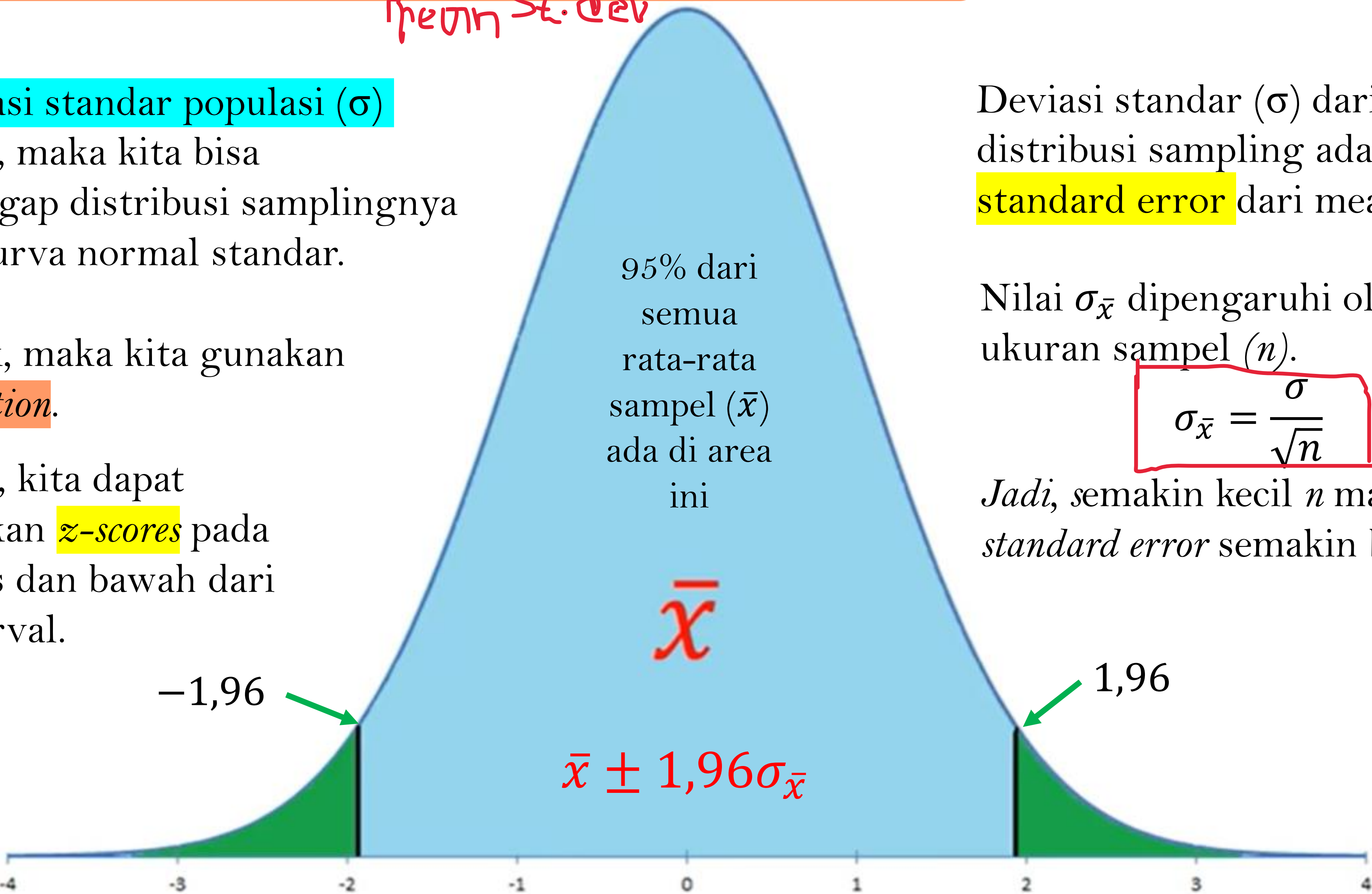
Estimasi Interval μ, σ diketahui

mean St.dev

Jika deviasi standar populasi (σ) diketahui, maka kita bisa menganggap distribusi samplingnya seperti kurva normal standar.

Jika tidak, maka kita gunakan *t-distribution*.

Sehingga, kita dapat menetapkan *z-scores* pada batas atas dan bawah dari 95% interval.



Deviasi standar (σ) dari distribusi sampling adalah standard error dari mean ($\sigma_{\bar{x}}$)

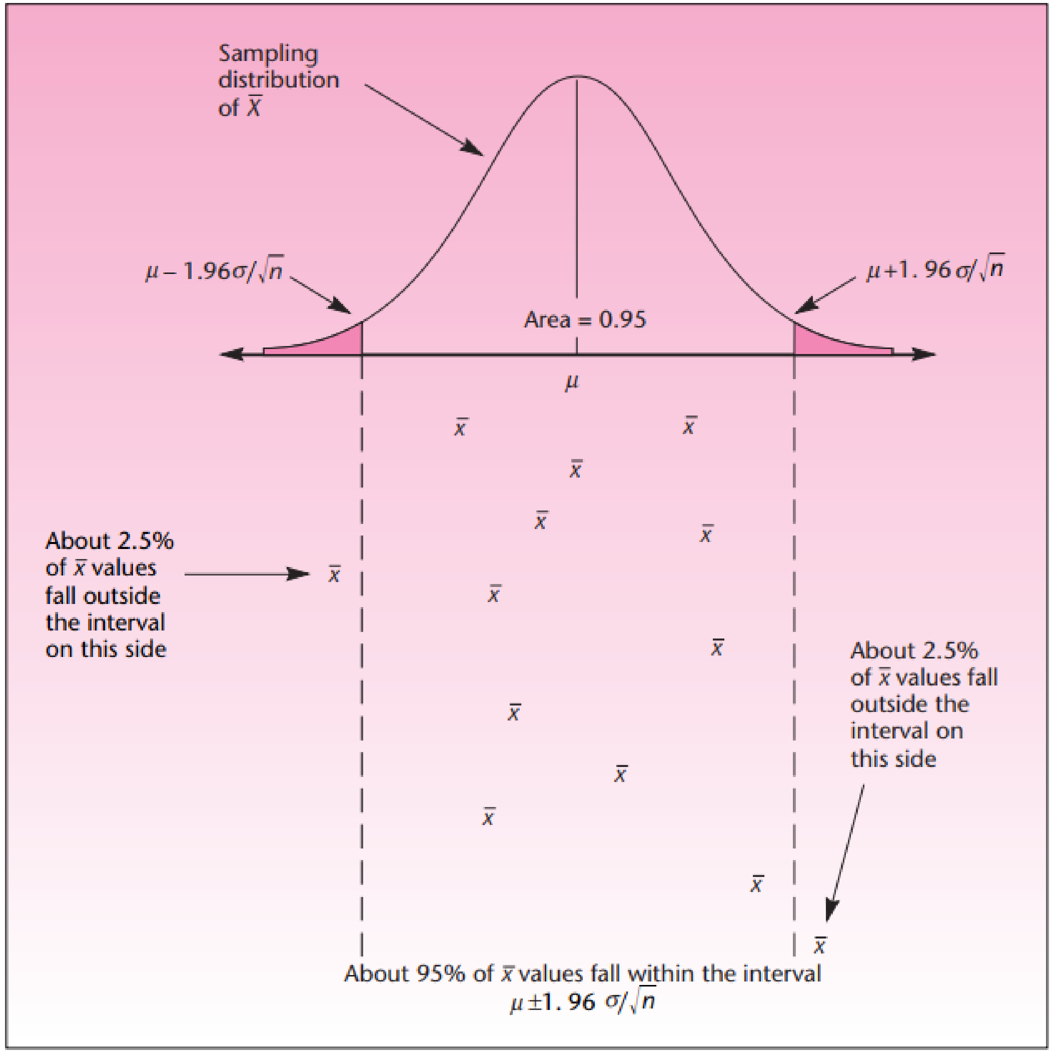
Nilai $\sigma_{\bar{x}}$ dipengaruhi oleh ukuran sampel *(n)*.

$$\boxed{\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}}$$

Jadi, semakin kecil n maka *standard error* semakin besar.

Estimasi Interval μ, σ diketahui

FIGURE 6–1 Probability Distribution of $\overline{X}$ and Some Resulting Values of the Statistic in Repeated Samplings

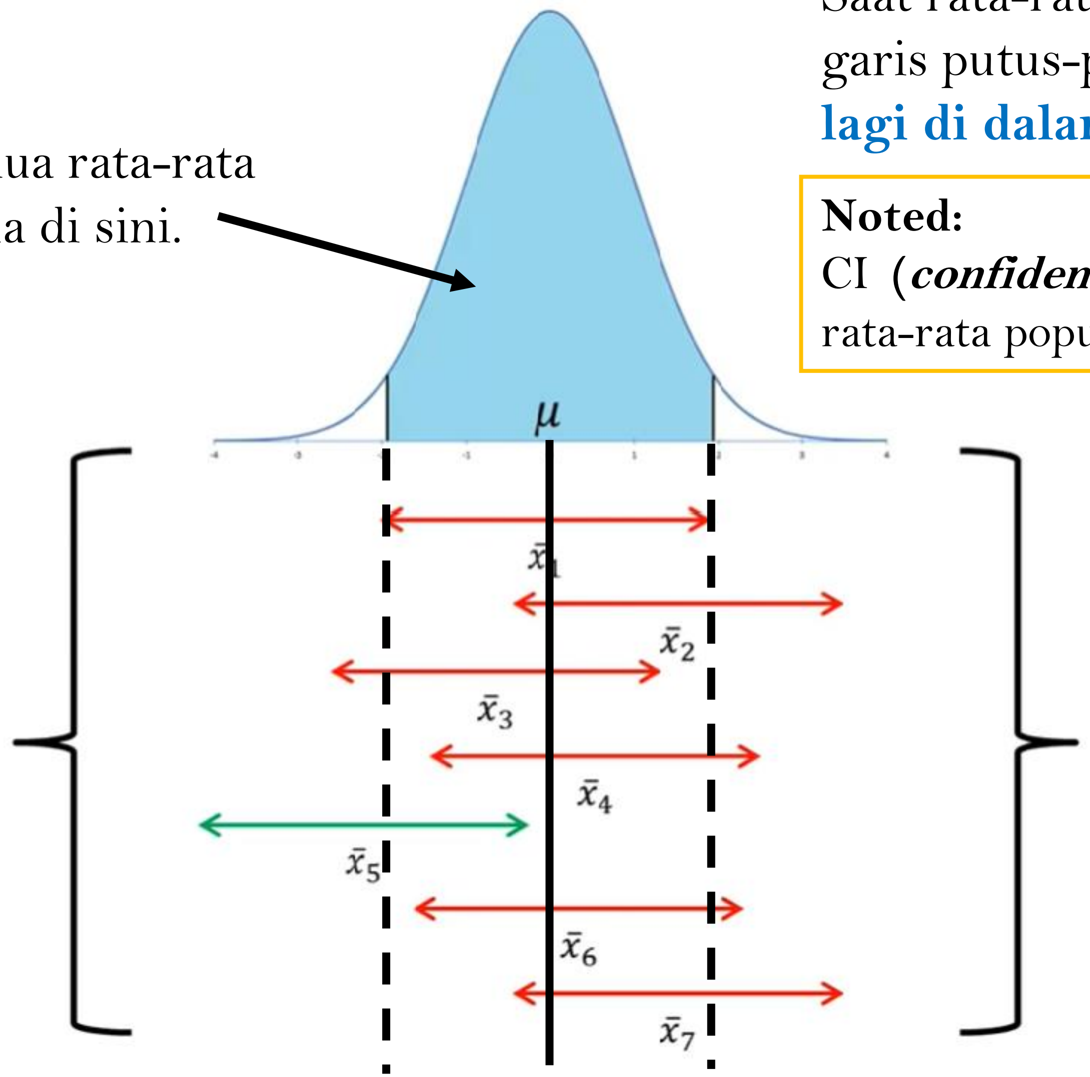


Estimasi Interval μ, σ diketahui

95% dari semua rata-rata sampel ($\bar{x}$) ada di sini.

Saat rata-rata sampel berada di luar batas garis putus-putus (yaitu $\bar{x}_5$), maka **μ tidak lagi di dalam interval.**

Noted:
CI (***confidence interval***) bukan *probability* bahwa rata-rata populasi (μ) terletak dalam interval.



Berbagai sampel dengan ukuran (n) yang sama

Sampel-sampel dengan n sama pasti memiliki ***standard error*** **($\sigma_{\bar{x}}$) yang sama**. Jadi, "lebar" dari interval 95% akan sama untuk sampel-sampel tersebut.

Estimasi Interval μ, σ diketahui

1000 $\rightarrow$ 25

$\bar{x} = 122\,m$ $\alpha = 5\%$

$\sigma = 20\,m$

Misal:

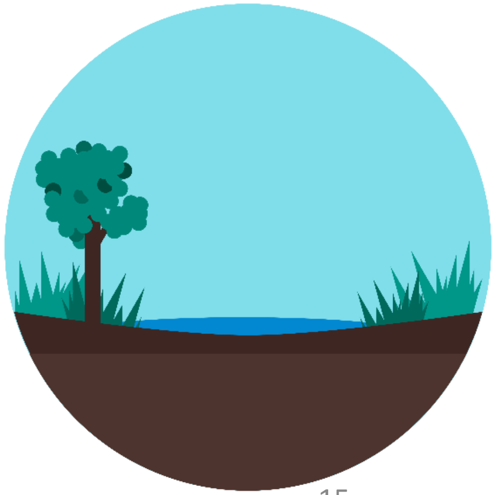
Kita ingin meneliti kedalaman danau-danau yang ada di Indonesia. Dari 1000 danau (N) yang ada, kita meneliti 25 danau (n) yang memiliki rata-rata kedalaman $\bar{x} = 122$ meter. Diketahui pula deviasi standar (σ) dari semua danau adalah 20 meter. Maka kita dapat menghitung *margin of error* dan *confidence interval* (CI) dengan level kepercayaan 95% $(\alpha = 5\%)$.

$$\bar{x} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = \bar{x} \pm 1{,}96\sigma_{\bar{x}}$$

$$= 122 \pm 1{,}96 \frac{20}{\sqrt{25}}$$

$$= 122 \pm 7{,}84$$

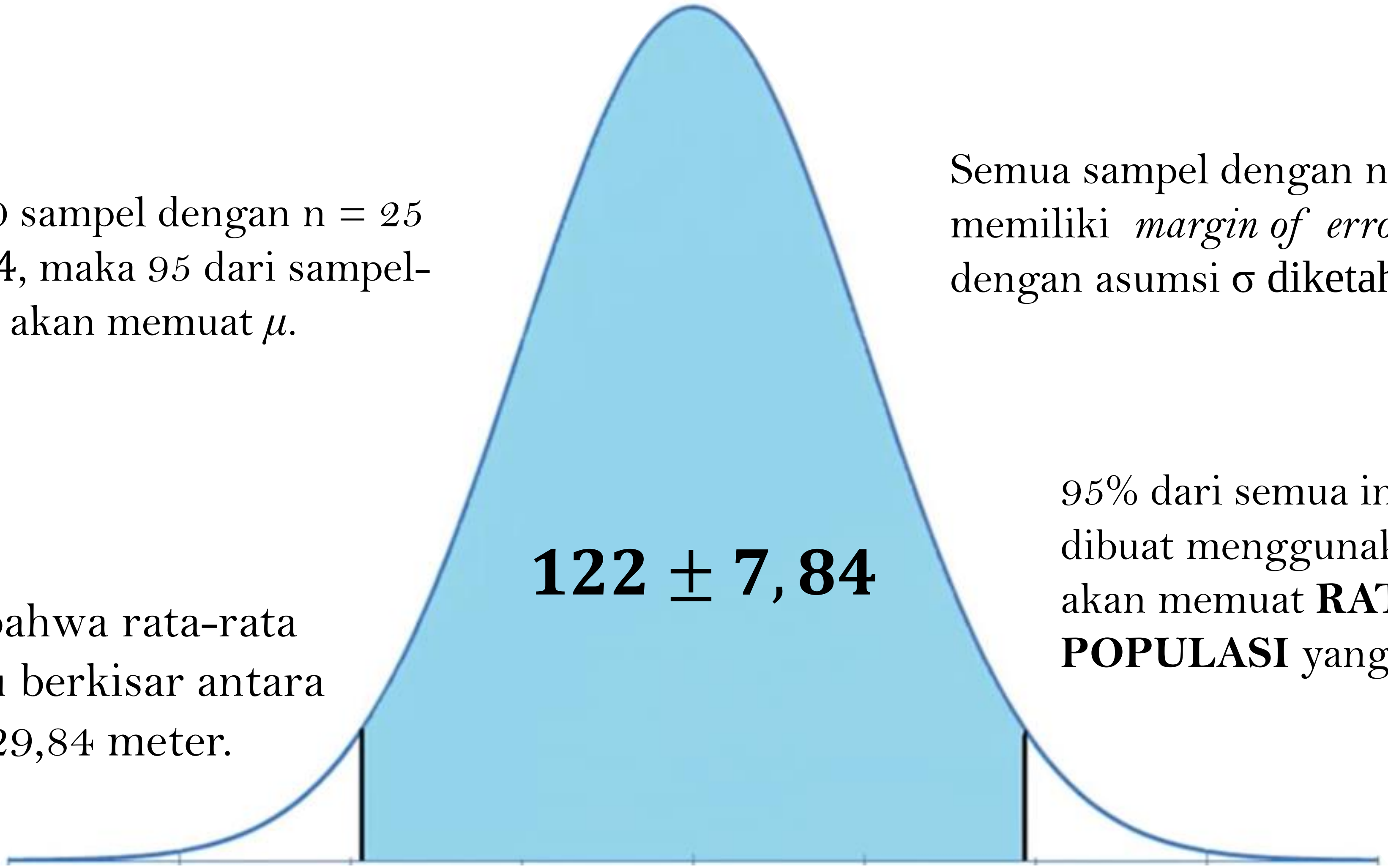
$$= [114{,}16;\ 129{,}84]$$



Estimasi Interval μ, σ diketahui

Jika diambil 100 sampel dengan n = 25 dengan $\bar{x} \pm 7{,}84$, maka 95 dari sampel-sampel tersebut akan memuat μ.

Semua sampel dengan n = 25 akan memiliki *margin of error* 7,84, dengan asumsi σ diketahui sebesar 20.



Interpretasi:
Kita 95% yakin bahwa rata-rata kedalaman danau berkisar antara 114,16 sampai 129,84 meter.

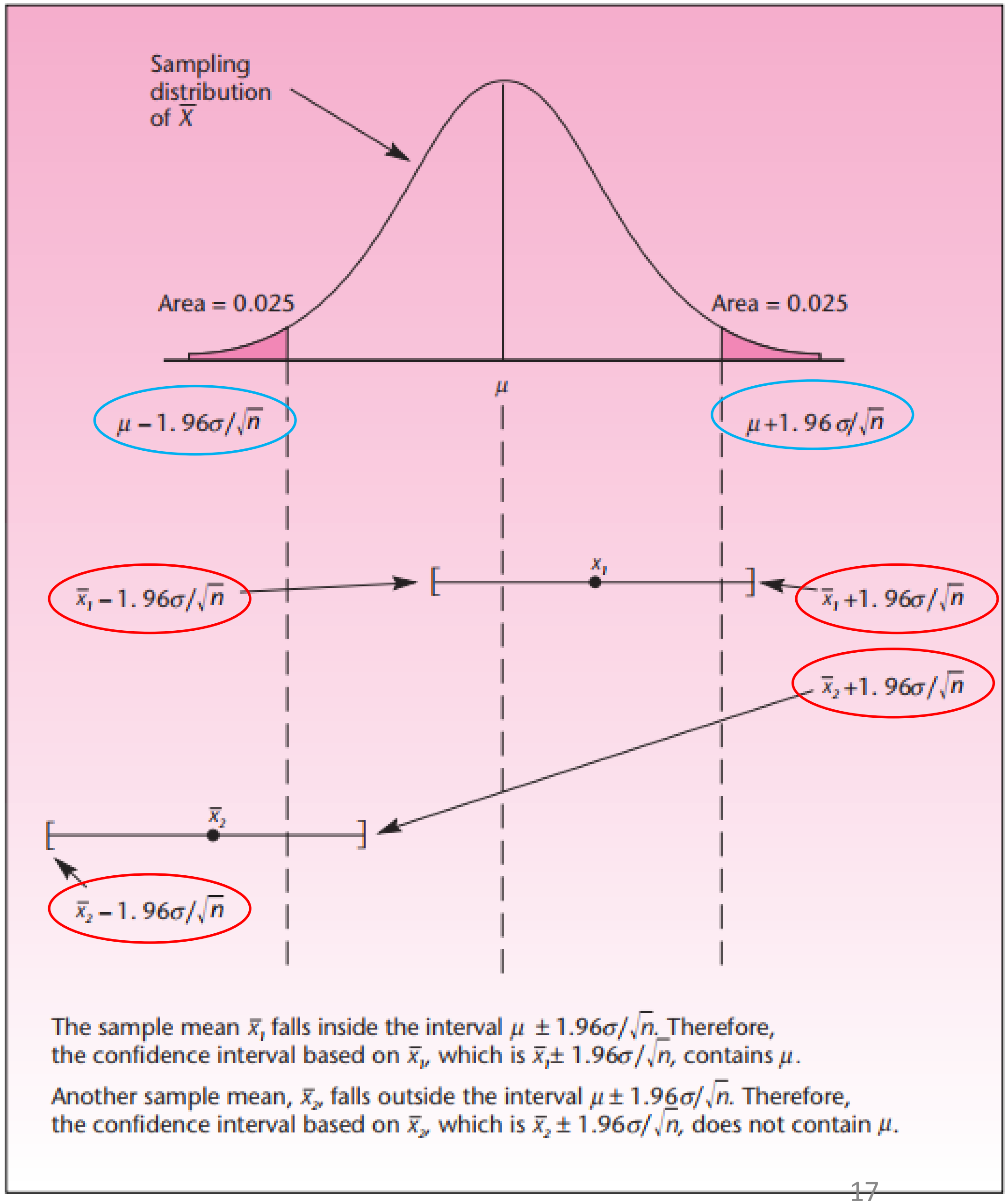
95% dari semua interval yang dibuat menggunakan $\bar{x} \pm 7{,}84$ akan memuat **RATA-RATA POPULASI** yang tidak diketahui.

$$\bar{x} \pm 1{,}96\sigma_{\bar{x}} = 122 \pm 7{,}84 = [114{,}16;\ 129{,}84]$$

Confidence level

- Definisikan $Z_{\frac{\alpha}{2}}$ sebagai nilai Z yang memotong area ekor bagian kanan *(right-tail)* dari nilai $\frac{\alpha}{2}$ di bawah kurva normal standar.
- Misal $Z_{\frac{\alpha}{2}} = 1{,}96$ dimana nilai $\frac{\alpha}{2} = 0{,}025$ (cek tabel distribusi normal).
- Nilai 1,96 adalah untuk *right-tail*, sedangkan *left-tail* sebesar -1,96.
- Sehingga area antara 1,96 dan -1,96 adalah $1 - \alpha = 1 - 2(0{,}025) = 0{,}95$ (untuk kedua ekor).
- Area $(1 - \alpha)$ ini disebut **koefisien kepercayaan**, sedangkan kombinasi dari area di kedua ekor (α) disebut ***error probability***.
- Jika koefisien kepercayaan kita kalikan 100 (persentase), maka nilai inilah yang disebut sebagai ***confidence level***.

FIGURE 6–2 Construction of a 95% Confidence Interval for the Population Mean μ



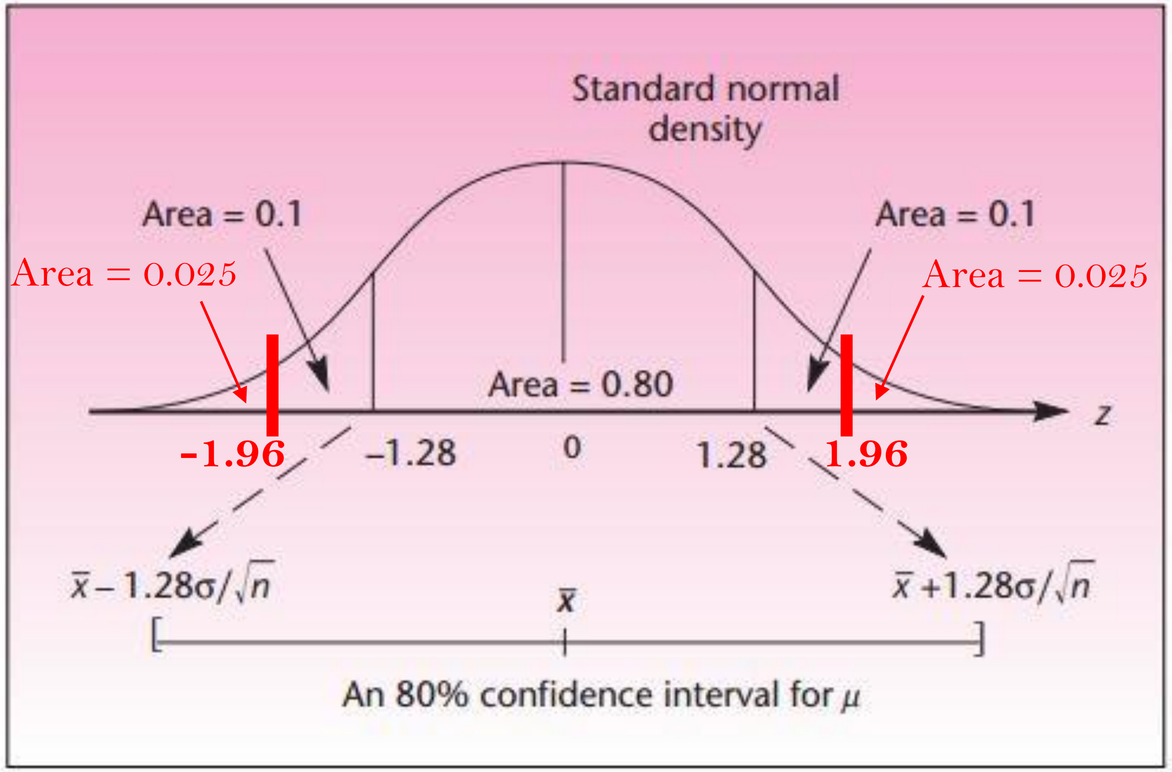
Confidence level

- *Confidence interval* memiliki level/tingkat kepercayaan (*confidence level*) yang biasanya bernilai 95% atau 99% atau 90%.
- Semakin tinggi *confidence level*, semakin lebar intervalnya.
- Contoh pada kasus sebelumnya tetapi kita ubah *confidence level* nya 80% ($\alpha = 20\%$).

$$\begin{aligned}\bar{x} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} &= 122 \pm Z_{\frac{0,2}{2}} \frac{20}{\sqrt{25}} \\ &= 122 \pm 1,28 \frac{20}{\sqrt{25}} \\ &= 122 \pm 5,12 \\ &= [116,88; 127,12]\end{aligned}$$

Interpretasi…?

Confidence level



Confidence level 95%	*Confidence level* 80%
$\bar{x} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$	$\bar{x} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$
$= 122 \pm Z_{\frac{0,05}{2}} \frac{20}{\sqrt{25}}$	$= 122 \pm Z_{\frac{0,2}{2}} \frac{20}{\sqrt{25}}$
$= 122 \pm 1,96 \frac{20}{\sqrt{25}}$	$= 122 \pm 1,28 \frac{20}{\sqrt{25}}$
$= 122 \pm 7,84$	$= 122 \pm 5,12$
$= [114,\underbrace{16;\ 129}_{},84]$	$= [116,\underbrace{88;\ 127}_{},12]$
$2 \times 7,84 = 15,68$	$2 \times 5,12 = 10,24$

Sehingga benar bahwa **semakin tinggi *confidence level*, semakin lebar intervalnya**, dengan catatan bahwa sampel diambil dari populasi yang sama dan ukuran sampel yang sama.

Confidence level

Perlu diingat bahwa interval yang lebar sebanding dengan *confidence level* yang tinggi. Jika ukuran sampel semakin besar dengan *confidence level* tetap, interval akan semakin kecil.

$n = 25$

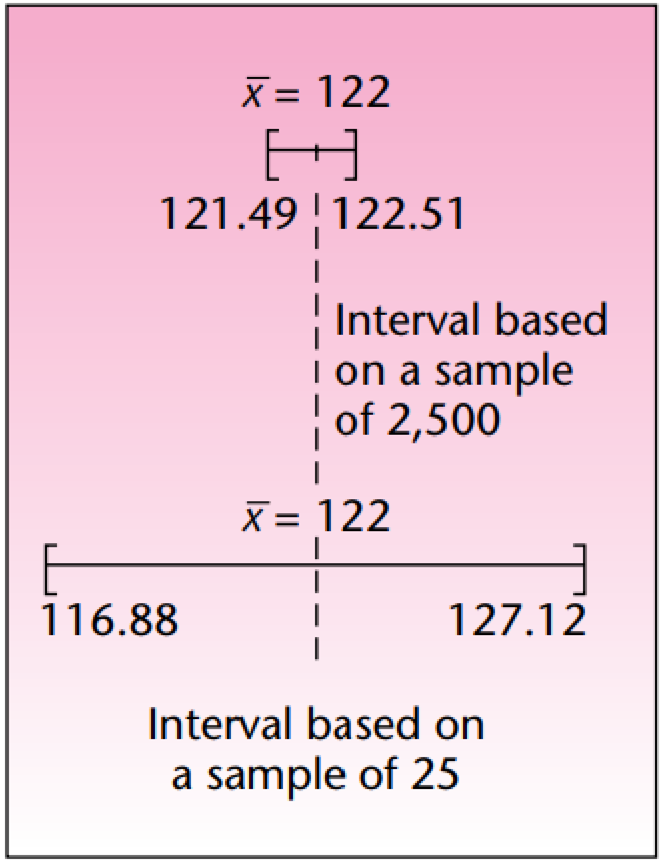
$$\bar{x} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$
$$= 122 \pm Z_{\frac{0,2}{2}} \frac{20}{\sqrt{25}}$$
$$= 122 \pm 1,28 \frac{20}{\sqrt{25}}$$
$$= 122 \pm 5,12$$
$$= [116,88; 127,12]$$

$2 \times 5,12 = 10,24$

$n = 2500$

$$\bar{x} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$
$$= 122 \pm Z_{\frac{0,2}{2}} \frac{20}{\sqrt{2500}}$$
$$= 122 \pm 1,28 \frac{20}{\sqrt{2500}}$$
$$= 122 \pm 0,512$$
$$= [121,49; 122,51]$$

$2 \times 0,512 = 1,024$



Estimasi Interval μ, σ tidak diketahui

Jika $\bar{x}$ dan s adalah mean dan deviasi standar sampel acak dari populasi normal dengan σ yang tidak diketahui, interval kepercayaan untuk μ adalah

$$\bar{x} - t_{\frac{\alpha}{2},(n-1)} \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\frac{\alpha}{2},(n-1)} \frac{s}{\sqrt{n}}$$

Ketika normalitas tidak dapat diasumsikan, σ tidak diketahui, dan $n \geq 30$, s dapat menggantikan σ dan interval kepercayaan adalah

$$\bar{x} - Z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} < \mu < \bar{x} + Z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}.$$

Ini sering disebut sebagai interval kepercayaan sampel-besar.

Estimasi Interval μ, σ tidak diketahui

100
$\bar{x} = 357,6$
$\sigma = 140$
CI 95%

Contoh sampel besar:

Seorang ekonom ingin mengestimasi rata-rata nominal cek oleh para nasabah di bank A. 100 akun nasabah dipilih secara acak dan diketahui nilai rata-ratanya \$357,6 dan deviasi standarnya \$140. Jika diinginkan *confidence level* 95%, hitung *confidence interval* untuk rata-rata nominal cek nasabah di bank A.

$$\bar{x} \pm Z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$
$$= 357,6 \pm Z_{\frac{0,05}{2}} \frac{140}{\sqrt{100}}$$
$$= 357,6 \pm 1,96 \frac{140}{\sqrt{100}}$$
$$= [330,16; 385,04]$$

Interpretasi…?

Latihan Soal

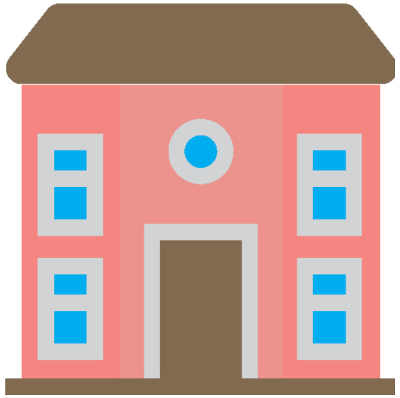
16

$\sigma = 75$ jt

$\bar{x} = 1345$ jt

α 5%

1. Agen *real estate* perlu memperkirakan nilai rata-rata properti hunian dengan ukuran tertentu di area tertentu. Agen real estat percaya bahwa standar deviasi dari nilai properti adalah $\sigma = 75$ juta dan bahwa nilai properti terdistribusi secara normal. Sebuah sampel acak dari 16 unit memberikan rata-rata sampel adalah 1.345 juta. Berikan interval kepercayaan 95% untuk nilai rata-rata semua properti jenis ini! Misalkan diperlukan selang kepercayaan 99%, hitung interval baru dan bandingkan dengan interval kepercayaan 95% yang sudah Anda hitung sebelumnya!



(i) $1345 \pm 1{,}96 \cdot \frac{75}{\sqrt{16}} = [1308{,}25 ; 1381{,}75]$

(ii) $1345 \pm 2{,}575 \cdot \frac{75}{\sqrt{16}} = [1296{,}71 ; 1393{,}28]$

$\therefore$ CI $\uparrow$ Interval $\uparrow$

Latihan Soal

32

$\bar{X} = 30{,}57$

2. Sebuah pabrik ban ingin memperkirakan rata-rata jumlah mil yang dapat ditempuh pada ban jenis tertentu sebelum ban tersebut aus. Sebuah sampel acak dari 32 ban dipilih; ban dikendarai sampai aus, dan jumlah mil yang ditempuh pada setiap ban dicatat. Data, dalam ribuan mil,

   32, 33, 28, 37, 29, 30, 25, 27,

   39, 40, 26, 26, 27, 30, 25, 30,

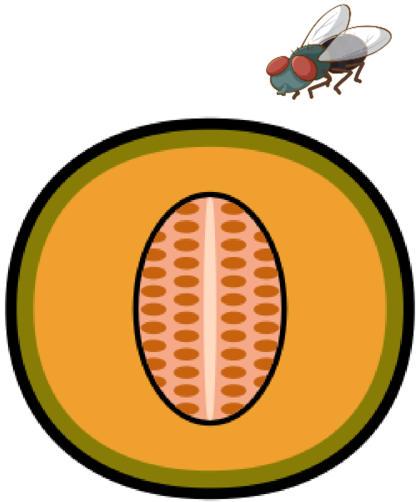
   31, 29, 24, 36, 25, 37, 37, 20,

   22, 35, 23, 28, 30, 36, 40, 41.

   a. Berikan interval kepercayaan 90% untuk jumlah rata-rata mil yang dapat ditempuh dengan ban jenis ini!
   b. Jika ukuran sampel menjadi 50, hitung interval kepercayaan 90% rata-rata jumlah mil yang ditempuh. Data tambahan sebagai berikut.
      28, 30, 32, 26, 40, 35, 23, 32, 38
      34, 37, 24, 28, 31, 30, 36, 40, 22

Latihan Soal

120



3. Seorang ahli entomologi menyemprotkan 120 lalat Melon dewasa dengan konsentrasi malathion yang rendah dan mengamati waktu kelangsungan hidup (*survival time*) mereka. Rata-rata dan standar deviasi dari hasil penyemprotan tersebut masing-masing adalah 18,3 dan 5,2 hari. Gunakan data ini untuk membangun interval kepercayaan 99% untuk waktu kelangsungan hidup rata-rata yang sebenarnya!
4. Sebuah sampel acak dari 60 batang energi cokelat dari merek tertentu memiliki rata-rata 230 kalori per batang, dengan standar deviasi 15 kalori. Buat interval kepercayaan 95% untuk kandungan kalori rata-rata sebenarnya dari bar energi merek ini. Asumsikan bahwa distribusi kandungan kalori kira-kira normal.

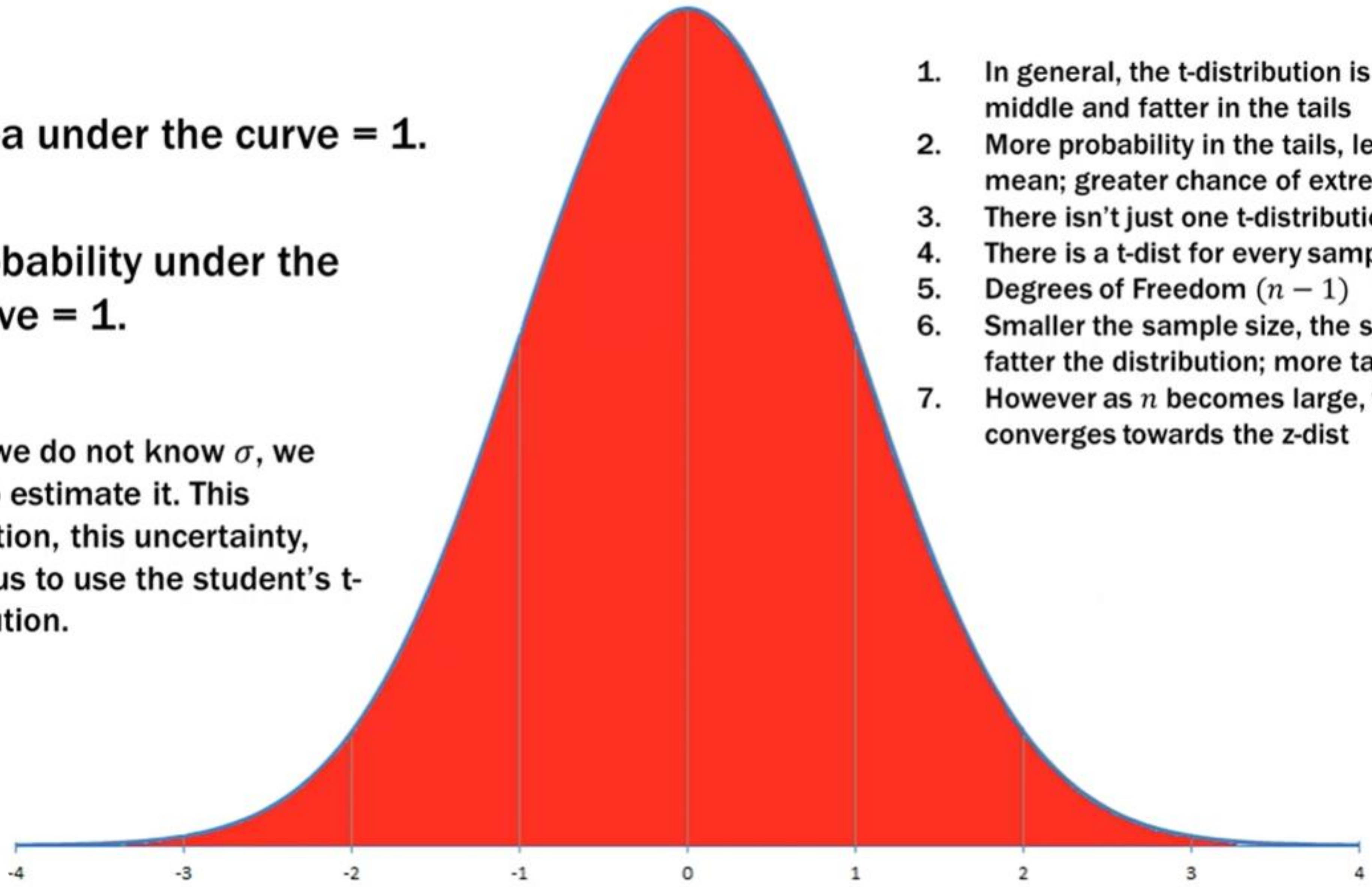
The Student's T-distribution

Area under the curve = 1.

Probability under the curve = 1.

When we do not know σ, we have to estimate it. This estimation, this uncertainty, forces us to use the student's t-distribution.

1. In general, the t-distribution is shorter in the middle and fatter in the tails
2. More probability in the tails, less near the mean; greater chance of extreme values
3. There isn't just one t-distribution
4. There is a t-dist for every sample size
5. Degrees of Freedom $(n-1)$
6. Smaller the sample size, the shorter and fatter the distribution; more tail probability
7. However as n becomes large, the t-dist converges towards the z-dist



NEXT WEEK

Student's T 95% Probability Interval

If we DO NOT KNOW the population standard deviation σ we treat the sampling distribution as a t-distribution.

By doing so, we can assign t-scores to the upper and lower boundary of the 95% interval *for each sample size.*

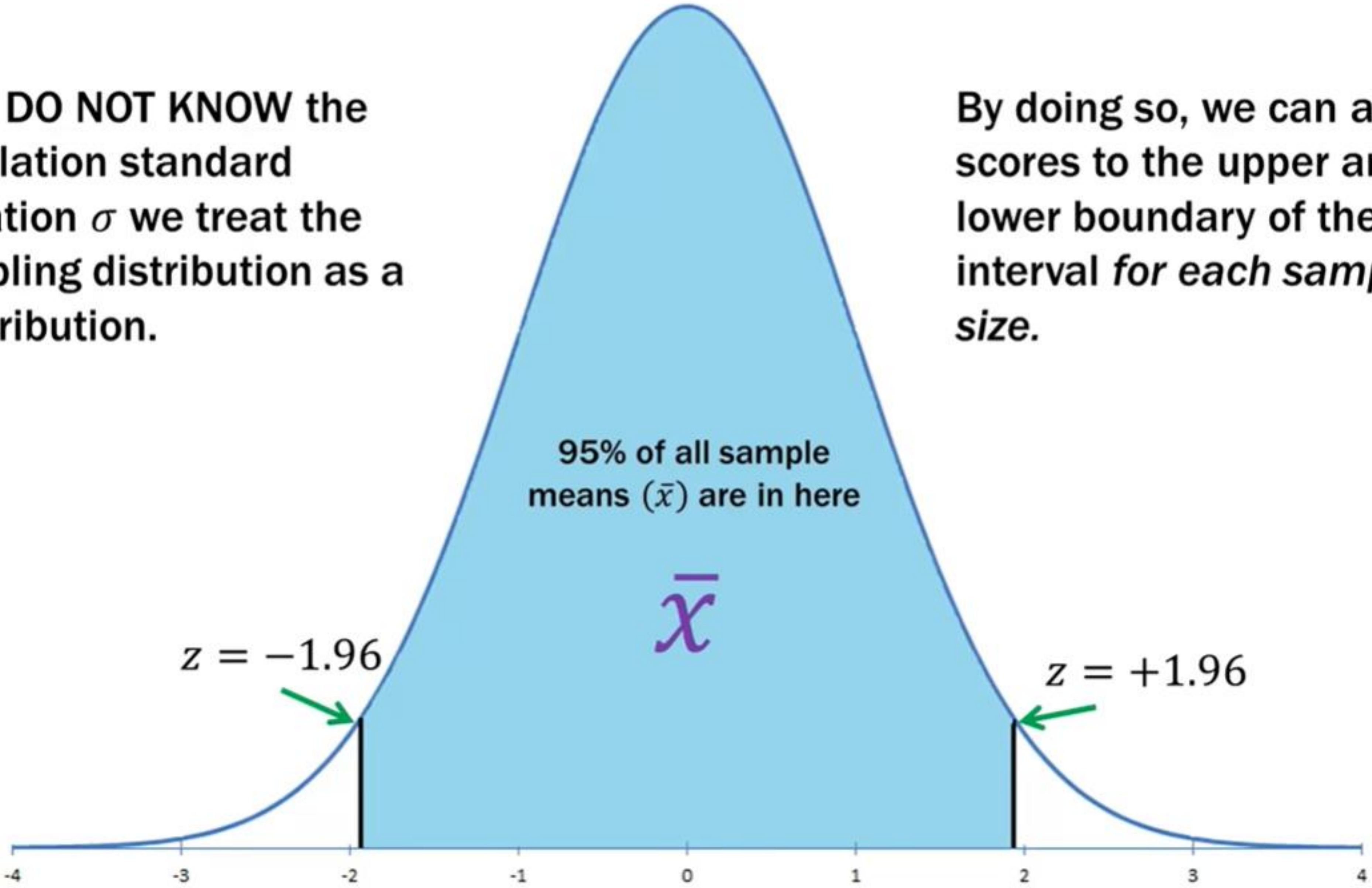


TABLE 6–1 Values and Probabilities of t Distributions

Degrees of Freedom	$t_{0.100}$	$t_{0.050}$	$t_{0.025}$	$t_{0.010}$	$t_{0.005}$
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
60	1.296	1.671	2.000	2.390	2.660
120	1.289	1.658	1.980	2.358	2.617
∞	1.282	1.645	1.960	2.326	2.576

NEXT WEEK

Contoh:

Seorang analis pasar saham ingin mengestimasi rata-rata return dari saham tertentu. Diambil sampel acak selama 15 hari dan didapatkan rata-rata return 10,37% serta deviasi standar 3,5%. Dengan *confidence level* 95%, berapa *confidence interval* untuk rata-rata return saham tersebut?

n = 15 ~ *t-distribution*

df = n-1 = 15-1 = 14

α = 5%

$t_{\frac{0,05}{2},14} = t_{0,025,14} = 2,145$

$$\bar{x} \pm t_{\frac{\alpha}{2},(n-1)} \frac{s}{\sqrt{n}}$$

$$= 10,37 \pm 2,145 \frac{3,5}{\sqrt{15}}$$

$$= 10,37 \pm 1,94$$

$$= [8,43; 12,31]$$

Interpretasi…?

Sumber

- Aczel, A. and Sounderpandian, J., 2008, *Complete Business Statistics*, 7th *Edition*, McGraw-Hill/Irwin, USA.
- Richard, A.J. and Bhattacharyya, G.K., 2019, *Statistics: Principles and Methods*, 8th *Edition*, John Wiley and Sons, USA.
- Walpole, R.E., Myers, R.H., Myers, S.L., and Ye, K.E., 2016, *Probability and Statistics for Engineers and Scientists*, 9th *Edition*, *Global Edition*, Pearson Education Limited, USA.
- https://youtu.be/BQ88ni4bJNA

Estimasi Titik dan Selang Kepercayaan untuk μ

Tim Dosen Pengantar Statistika