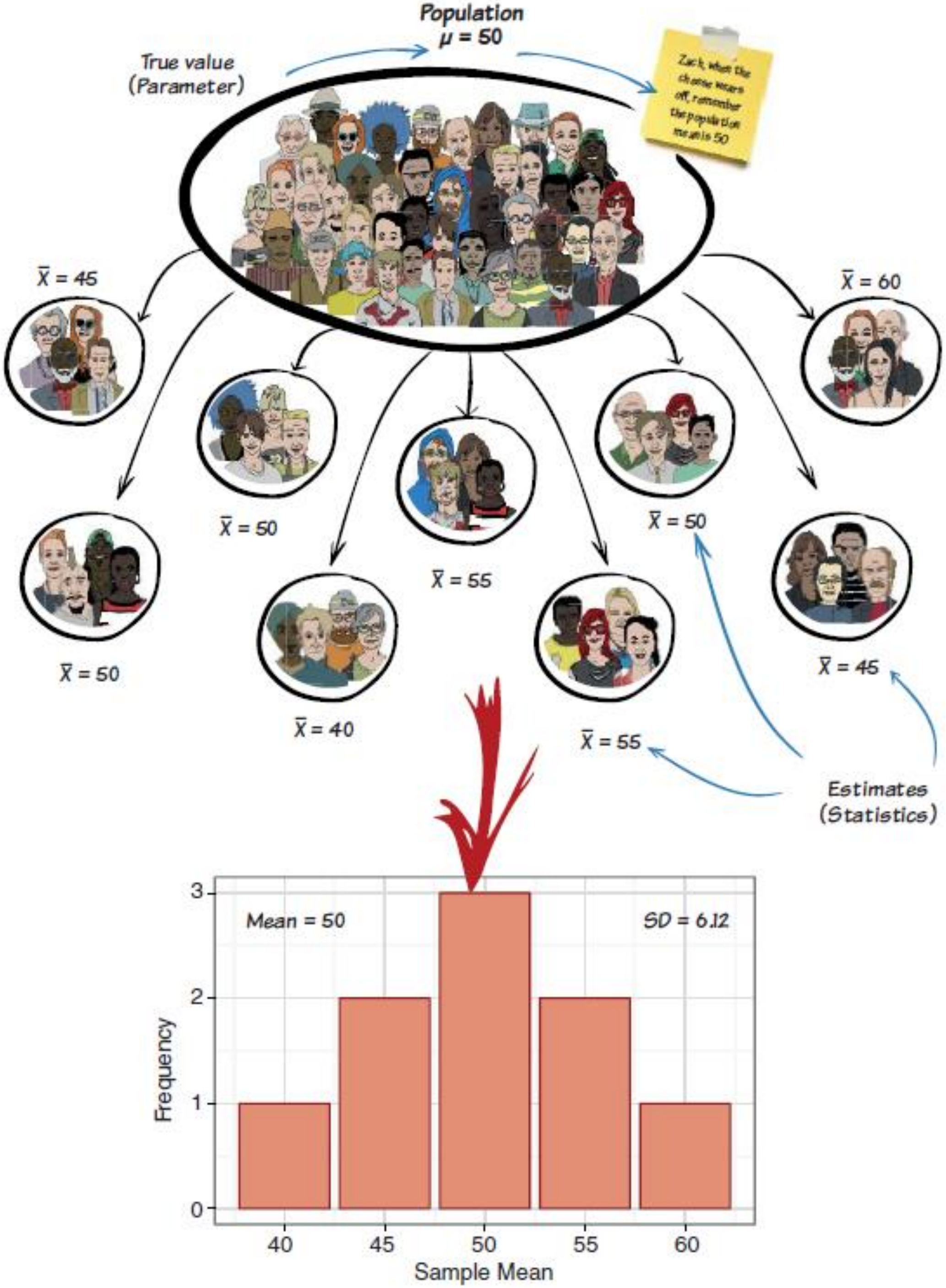
Confidence Interval

Tim Pengajar MK Data dan Pustaka

Pokok Bahasan

1. *Standard error*
2. *Sampling error*
3. Menghitung *confidence interval (Mean dan SD)*
4. Menginterpretasi *confidence interval*



Ilustrasi Standard Error

Bahan diskusi

Menurut anda, kenapa nilai rata-rata sampel yang diambil dari populasi ada kemungkinan akan berbeda dengan rata-rata populasinya? Perhatikan gambar disamping.

Referensi:

Field, A. (2016). *An Adventure in Statistics:* ***Page : 270****. The Reality Enigma.* London: SAGE Publications.

Ilustrasi Standard Error

Hasil menghitung rata-rata sampel yang diambil dari populasi sebesa $\bar{x}$ =45, padahal rata-rata populasi $\boldsymbol{\mu}$ = 50. Terdapat selisih 5.

Hasil menghitung rata-rata sampel yang diambil dari populasi sebesar $\bar{x}$ =60, padahal rata-rata populasi $\boldsymbol{\mu}$ = 50. Terdapat selisih 10.

Variasi yang terjadi pada distribusi sampling ini yang disebut sebagai **Standard Error (SE)**

Semakin kecil SE, maka estimasi rata-rata sampel terhadap rata-rata populasi menjadi semakin baik

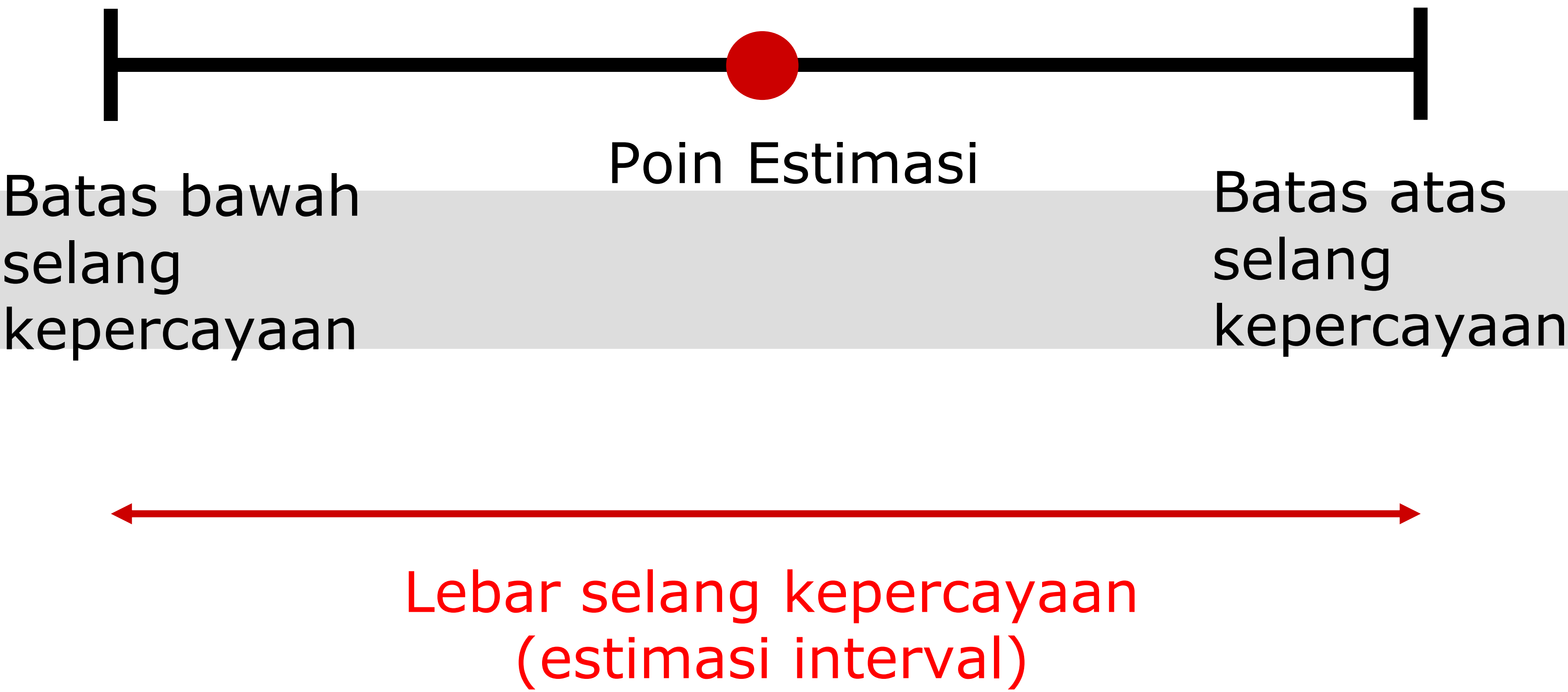
Estimasi

Confidence interval merupakan proses melakukan estimasi. Estimasi adalah proses memperkirakan nilai parameter (pada populasi) dengan mengunakan nilai statistik (pada sampel)

Ada 2 jenis estimasi :

- Estimasi titik (*point estimate*)
- Estimasi interval (*interval estimate*)

Estimasi titik dan interval



Estimator Yang Valid

1. Unbiased : apabila nilai estimator sama dengan nilai yang diestimasi
2. Efisien : apabila estimator memiliki varians yang kecil
3. Konsisten :
   a. Jika ukuran sampel semakin bertambah maka estimator akan mendekati parameternya
   b. Jika ukuran sampel bertambah tak berhingga maka distribusi sampling estimator akan mengecil

Estimasi Titik

Memperkirakan nilai parameter dengan menggunakan satu nilai statistik tertentu

<u>Contoh :</u>

- **μ diestimasi dengan $\bar{x}$**
- **σ^2 diestimasi dengan s^2**
- **P diestimasi dengan p**

Estimasi Interval

Memperkirakan nilai parameter dengan menggunakan 2 (dua) nilai atau dalam interval tertentu atau disebut juga Confidence Interval

Jenis Estimasi :

1). Estimasi untuk rata-rata

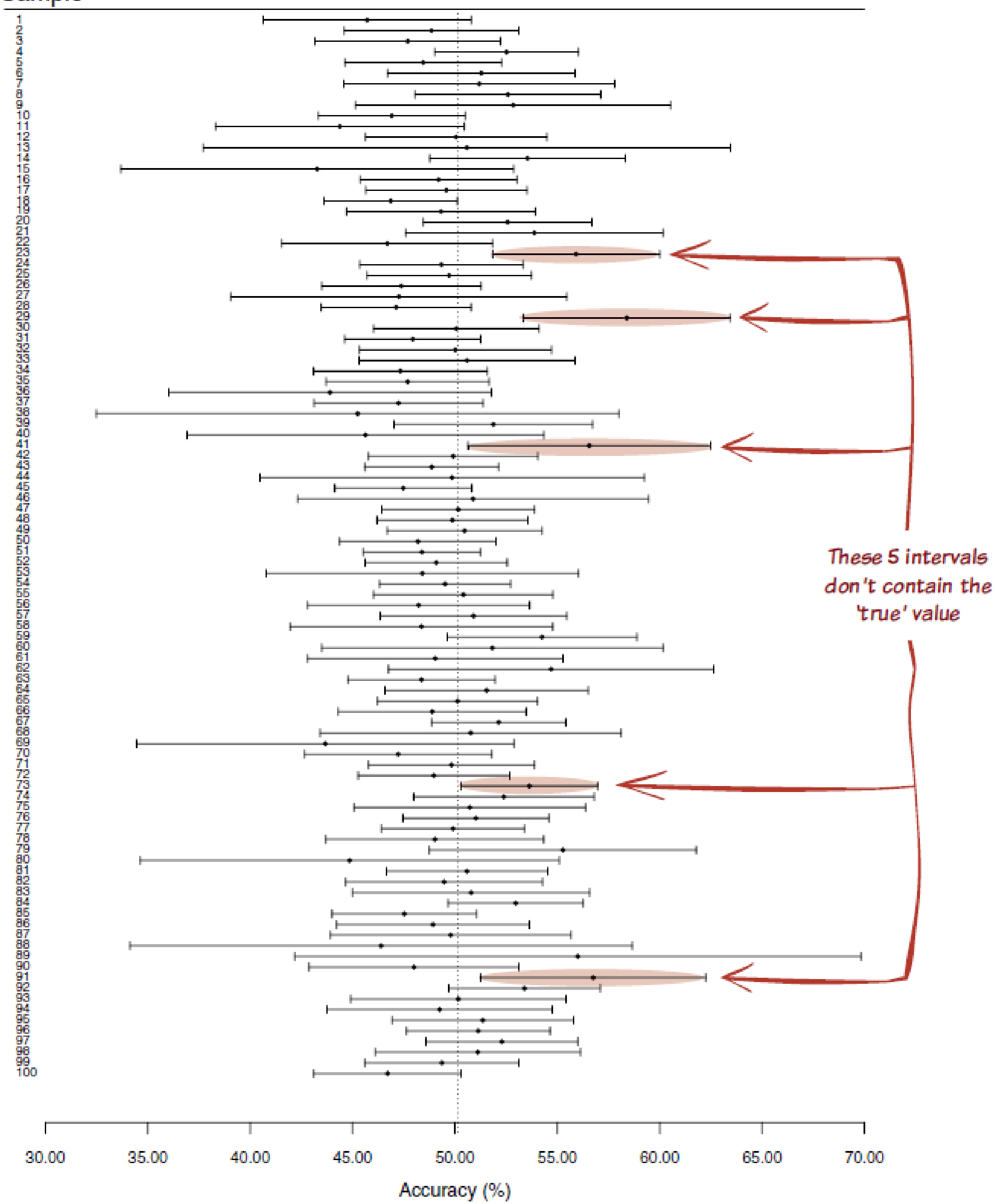
2). Estimasi untuk varians

3). Estimasi untuk proporsi

Definisi Confidence Interval

Perkiraan/ estimasi rentang nilai parameter populasi sebenarnya. Disebut juga selang kepercayaan

Tingkat kepercayaan (**level of confidence**) dalam confidence interval merupakan peluang bahwa selang **kepercayaan** yang terbentuk benar-benar memuat parameter populasi jika proses estimasinya dilakukan secara berulang-ulang.



Ilustrasi dalam Memaknai Confidence Interval (CI 95%)

Dilakukan 100 kali sampling pada populasi yang sama. Dan pada pada setiap sampel dihitung estimasi intervalnya.

Jika terdapat 5 estimasi interval tidak presisi, maka kondisi ini yang disebut Confidence Interval 95% atau dengan level of confidence $\alpha=5\%$

Referensi:
Field, A. (2016). *An Adventure in Statistics:* ***Page : 280.*** *The Reality Enigma.* London: SAGE Publications.

Confidence Interval 95%

Selang Kepercayaan 95% artinya kita percaya bahwa **95**% sample yang kita ambil akan memuat nilai parameter aslinya.

Formula Umum Estimasi Interval

Point Estimasi ± (Nilai Kritis) (Standar Eror)

Standard error (SE) adalah **standar** deviasi dari distribusi sampling suatu statistik. **Standard error** merujuk kepada perkiraan **standar** deviasi dari sampel tertentu yang digunakan untuk menghitung suatu nilai estimator.

Nilai Kritis adalah nilai batas penerimaan kesalahan yang diharapkan, bergantung dari distribusi sampel. Nilai ini dihasilkan dari table kritis pada distribusi tersebut.

Estimasi Rata-rata

Jika standar deviasi (σ) diketahui

$$\bar{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \hat{\mu} < \bar{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

dimana :

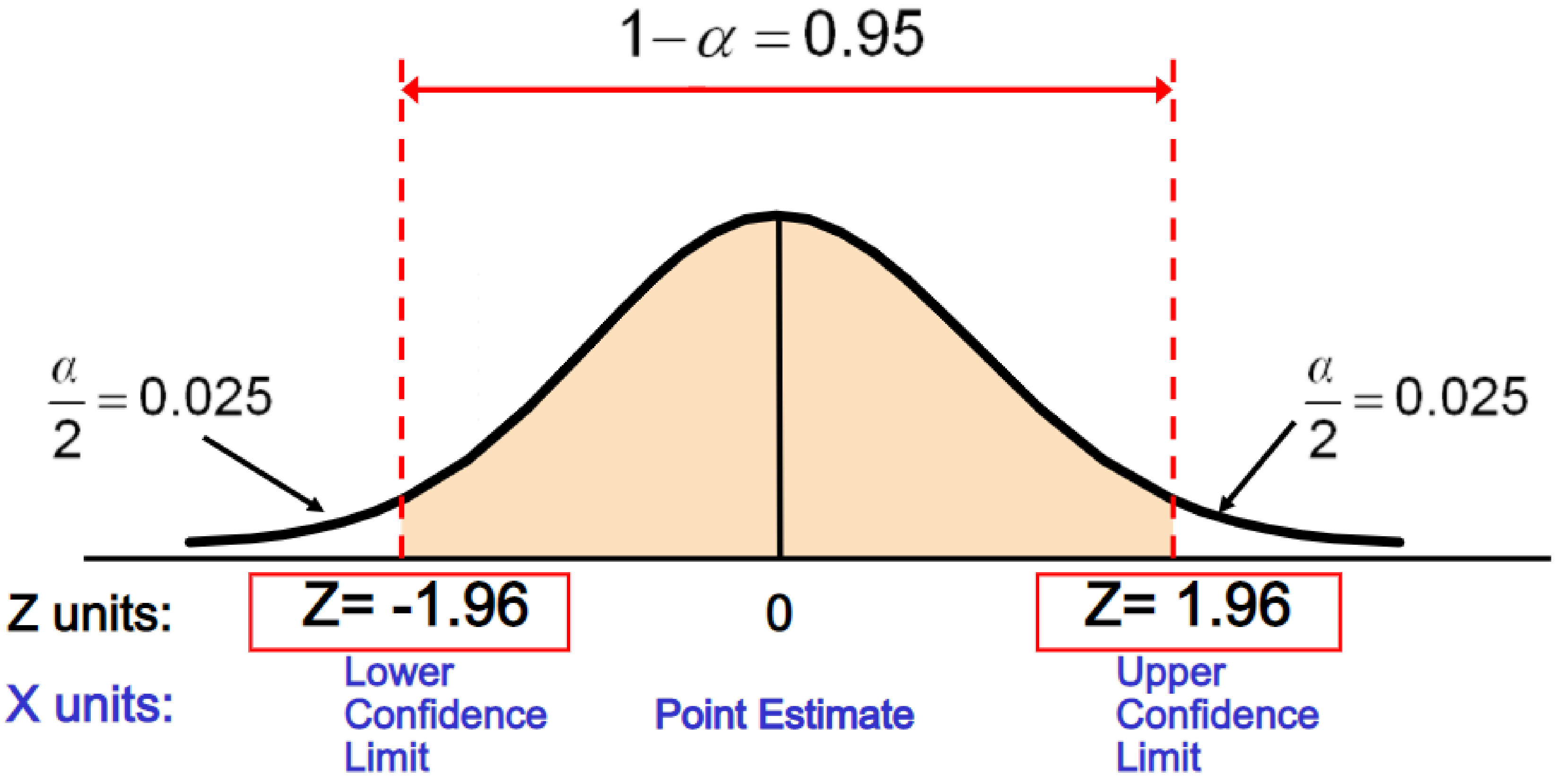
x = rata-rata dari sampel

Z = tabel Normal Standart

N = banyaknya data

Menentukan Nilai Kritis, Z (Distribusi Normal Standar)

Selang Kepercayaan 95% (Nilai kritis → Z = ±1.96)



Tabel Z (Nilai Kritis Z)

Pada umumnya menggunakan level kepercayaan 90%, 95%, 99%

Level Kepercayaan	(1 - α)	Nilai Z kritis
80%	0.80	1.28
90%	0.90	1.645
95%	0.95	1.96
98%	0.98	2.33
99%	0.99	2.58
99.9%	0.999	3.27

Nilai kritis Z (Tabel Normal Standard)

Contoh Kasus

Rata-rata nilai Statistika sampel random 36 mahasiswa adalah 2,6. Jika diketahui nilai standar deviasi populasi 0,3 maka hitunglah selang kepercayaan 95% untuk rata-rata nilai statistika semua mahasiswa?

$$\bar{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \hat{\mu} < \bar{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$2{,}6 - (1{,}96)\frac{0{,}3}{\sqrt{36}} < \hat{\mu} < 2{,}6 + (1{,}96)\frac{0{,}3}{\sqrt{36}}$$

$$2{,}5 < \hat{\mu} < 2{,}7$$

Kita percaya bahwa **95**%nilai $\hat{\mu}$ akan berada pada interval antara 2,5 sampai dengan 2,7

Estimasi Rata-rata

Jika standar deviasi (σ) tidak diketahui

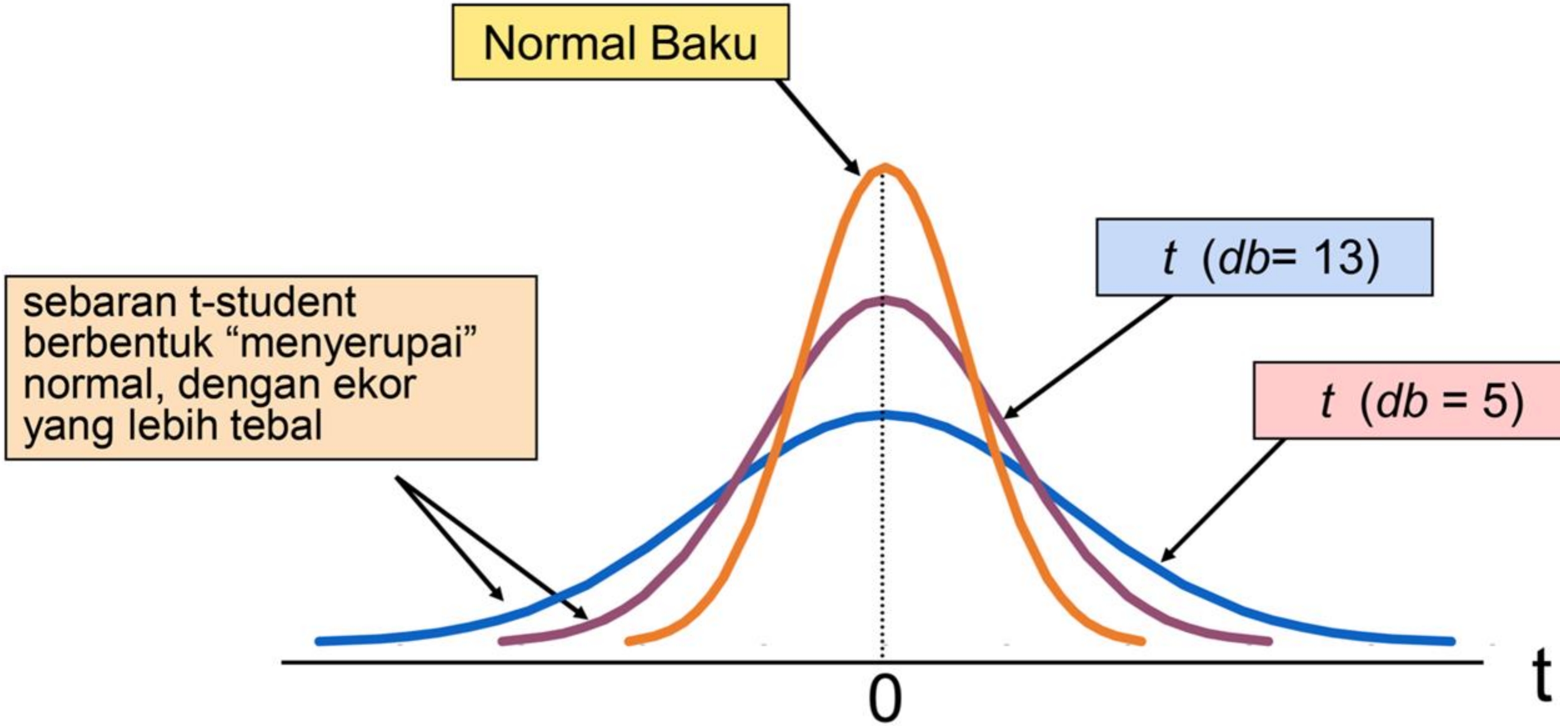
Karena standar deviasi di populasi tidak diketahui, maka digunakan nilai standar deviasi dari sampel (s), yaitu :

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

Maka rata-rata di populasi diestimasi dengan :

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

Sebaran t vs Sebaran Normal



semakin besar derajat bebas, sebaran t-student semakin mendekati sebaran normal baku

Tabel t (Nilai Kritis t)

df	Tingkat signifikansi uji satu arah					
	0,10	0,05	0,025	0,01	0,005	0,0005
	Tingkat signiflkansi uji dua arah					
	0,20	0,10	0,05	0,02	0,01	0T001
1	3,078	6,314	12,706	31,821	63,657	636,619
2	1,886	2,920	4,303	6,965	9,925	31,599
3	1,638	2,353	3,182	4,541	5,841	12,924
4	1,533	2,132	2,776	3,747	4,604	8.610
5	1,476	2,015	2,571	3,385	4,032	6,869
6	1,440	1,943	2,447	3,143	3,707	5,959
7	1,415	1,895	2,365	2,998	3,499	5,408
8	1,397	1,860	2,306	2,896	3,355	5,041

Nilai kritis *t* dengan jumlah n=6 pada tingkat kepercayaan **α**=5% ⇒ Nilai *t*(5 ; **α**=5%) = 2,571

Tabel t-student untuk Nilai kritis t

Contoh Kasus

Beda contoh ini dengan sebelumnya adalah nilai parameter **σ** populasi tidak diketahui. Sehingga estimasi interval menggunakan nilai kritis t

Ingin diketahui berat badan bayi yang lahir di Klinik Bersalin “Anak Impian”. Dilakukan pengambilan sampel terhadap bayi-bayi yang dilahirkan di klinik tersebut. Dengan menggunakan teknik sampling yang benar, diperoleh 15 bayi sebagai sampel penelitian, dan berat lahirnya (dalam kg) adalah sebagai berikut :

3.1	3.2	2.9	2.7	2.7	2.8	2.9	2.5
2.4	2.6	3.4	3.5	3.0	2.7	2.8	

Hitunglah selang kepercayaan 90% untuk rata-rata berat badan semua bayi yang lahir di Klinik Bersalin tersebut?

Penyelesaian

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

$$2{,}88 - 1{,}761\,\frac{0{,}3144}{\sqrt{15}} < \hat{\mu} < 2{,}88 + 1{,}761\,\frac{0{,}3144}{\sqrt{15}}$$

$$2{,}73705 < \hat{\mu} < 3{,}02295$$

Artinya: Kita percaya bahwa **90**% nilai $\hat{\mu}$ akan berada pada interval antara 2,74 sampai dengan 3,023

Estimasi Varians

Nilai σ^2 akan diestimasi dengan rumus :

$$\frac{(n-1)s^2}{\chi^2_{(n-1;\alpha/2)}} < \hat{\sigma}^2 < \frac{(n-1)s^2}{\chi^2_{(n-1;1-\alpha/2)}}$$

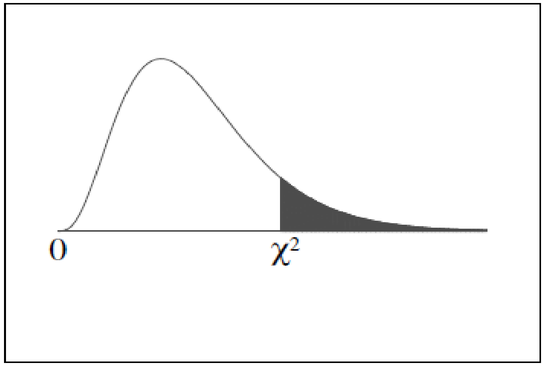
Dimana :

n = banyaknya data

s = standar deviasi data

χ^2 = tabel *chi-square* dengan derajat bebas (n –1) dan α tertentu

Tabel Chi-Square (Nilai Kritis χ^2)



The shaded area is equal to α for $\chi^2 = \chi^2_\alpha$.

df	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.950}$	$\chi^2_{.900}$	$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750

Nilai kritis χ^2 dengan jumlah n=6 pada tingkat kepercayaan **α**=5% ⇒ Nilai kritis $\chi^2_{(5\,;\,0{,}025)}$ = 12,833 dan $\chi^2_{(5\,;0{,}975)}$ = 0,831

Tabel Chi Square (Nilai Kritis Chi-Square)

Contoh Kasus

Dengan menggunakan data yang sama yaitu data tentang berat badan 15 bayi yang dilahirkan di Klinik Bersalin “Anak Impian” maka hitunglah estimasi interval untuk s^2 dengan CI 90%

$$\frac{(n-1)s^2}{\chi^2_{(n-1;\alpha/2)}} < \hat{\sigma}^2 < \frac{(n-1)s^2}{\chi^2_{(n-1;1-\alpha/2)}}$$

$$\frac{(15-1)0{,}0983}{23{,}68} < \hat{\sigma}^2 < \frac{(15-1)0{,}0983}{6{,}57}$$

$$0{,}0581 < \hat{\sigma}^2 < 0{,}2055$$

Estimasi Proporsi

Estimasi interval untuk nilai p :

$$p - Z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} < \hat{P} < p + Z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

Dimana :

p = proporsi di sampel

n = besar sampel

Z = tabel normal standar

Contoh Kasus

Pada suatu sampel random 500 keluarga yang memiliki pesawat televisi di Kota A, ditemukan 340 keluarga telah mengubah televisi analognya menjadi televisi digital. Hitunglah selang kepercayaan 95% untuk proporsi sesungguhnya dari keluarga yang telah memiliki televisi digital di Kota tersebut.

$$p - Z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} < \hat{P} < p + Z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

$$\frac{34}{50} - 1{,}96\sqrt{\frac{(34/50)(1-34/50)}{500}} < \hat{P} < \frac{34}{50} + 1{,}96\sqrt{\frac{(34/50)(1-34/50)}{500}}$$

$$0{,}64 < \hat{P} < 0{,}72$$

Menentukan Ukuran Sampel

Menentukan ukuran sampel berdasarkan selang kepercayaan rata-rata

$$\bar{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \hat{\mu} < \bar{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$\boxed{\bar{x} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}$$

Sampling error
(margin of error)

$$\boxed{e = Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}$$

Menentukan Ukuran Sampel

$$\boxed{e = Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}$$

Sampling error
(margin of error)

$$\boxed{n = \frac{Z^2\sigma^2}{e^2}}$$

Sampling Error : kesalahan yang ditimbulkan karena kita hanya mengamati sebagian saja (contoh), tidak semuanya (populasi).

Example: Sample Size for Mean

What sample size is needed to be 90% confident of being correct within ± 5? A pilot study suggested that the standard deviation is 45.

$$n = \frac{Z^2\sigma^2}{e^2} = \frac{(1.645^2)(45^2)}{5^2} = 219.2 \approx 220$$

Latihan

Contoh kasus pada slide 17. Jika ingin percaya 95% bahwa estimasi untuk $\boldsymbol{\mu}$ akan terjadi kesalahan sebesar kurang dari 0,05. Maka berapa jumlah sampel yang harus diambil?

See you next week