Estimasi dan Confidence Interval

Tjipto Juwono, Ph.D.

November 2017

SURYA UNIVERSITY

Point Estimate

Point Estimate:

Adalah suatu nilai tunggal (*point*) yang diperoleh dari sample yang digunakan untuk melakukan estimasi nilai yang bersesuaian dari populasi.

$$\begin{aligned} \bar{X} &\rightarrow \mu \\ s &\rightarrow \sigma \\ s^2 &\rightarrow \sigma^2 \end{aligned} \tag{1}$$

Confidence Interval Estimate

Confidence Interval Estimate:

Adalah suatu rentang nilai (*interval*) yang diperoleh dari sample sehingga parameter populasi berada pada rentang nilai itu dengan probabilitas tertentu. Probabilitas bahwa parameter populasi berada pada rentang tersebut disebut *level of confidence*.

$$\text{C.I.} = \text{point estimate} \pm \text{margin of error}$$

Confidence Interval Estimate

Lebar rentang *confidence interval* ditentukan oleh:

1. Ukuran sample, n.
2. Variasi dalam populasi yang diukur dengan σ dan diestimasi di dalam sample dengan s.
3. level of confidence yang diminta.

Confidence Interval for Population Mean - σ diketahui

Confidence Interval untuk Population Mean jika σ diketahui

$$\bar{X} \pm z \frac{\sigma}{\sqrt{n}} \tag{2}$$

Dengan:

$\bar{X}$: mean dari sample

z : nilai-z untuk confidence-level yang dipilih

σ : standard deviasi populasi

n : jumlah observasi di dalam sample

Confidence Interval for Population Mean - σ diketahui

Confidence Interval untuk Population Mean jika σ diketahui

$$\bar{X} \pm z\frac{\sigma}{\sqrt{n}} \tag{3}$$

Lebar Confidence Interval: ditentukan oleh confidence level dan besarnya standard error dari mean

Standard error dari mean: Ditentukan oleh dua parameter

a Standard Deviation, σ

b Jumlah observasi di dalam sample, n

Bagaimana Cara Menentukan z jika diberikan Confidence Level tertentu?

1. Misalkan diberikan CL=95%. Maka besar probabilitas yang akan kita cari di tabel adalah:

$$\frac{0.95}{2} = 0.4750 \tag{4}$$

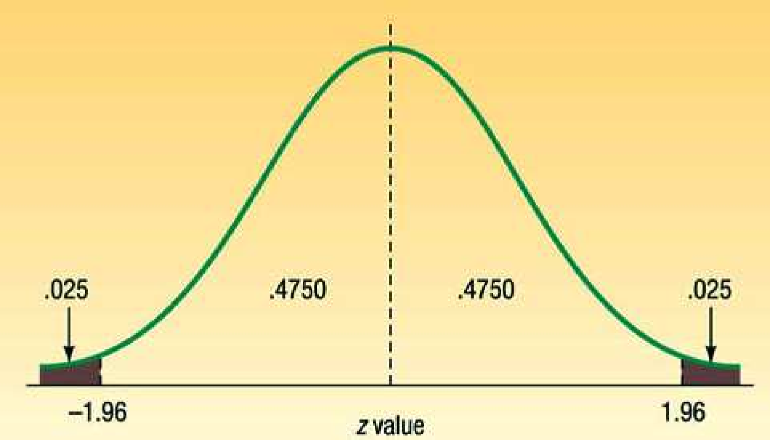
2. Carilah harga z di tabel, yang bersesuaian dengan nilai **0.4750**

Confidence Level

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936

Diperoleh: $\mathbf{z = 1.96}$

Confidence Level



Apa Arti Estimasi Confidence Interval?

Misalkan kita mempunyai Confidence Level 95%. Dan kemudian kita menghitung Confidence Interval berdasarkan data-data yang diperoleh. Apa makna dari Confidence Interval itu?

1. Misalkan dari populasi yang disediakan, kita mengambil 100 sample
2. Dari 100 sample itu kita memperoleh 100 Confidence Interval
3. Maka dari 100 Confidence Interval tersebut dapat diharapkan bahwa 95 di antaranya akan memuat parameter populasi yang sesuai.
4. Selain itu, 95% dari 100 sample mean yang kita peroleh akan berada pada rentang **1.96** dari parameter populasi yang kita hipotesakan.

Contoh

Contoh

Suatu survey ingin mengetahui mean income dari manager tingkat menengah pada suatu industri tertentu. Sample random dari 256 manager menunjukkan sample mean sebesar $\bar{X} = \$45420$. Standard deviasi dari populasinya adalah $\sigma = \$2050$.

Contoh

Contoh

Suatu survey ingin mengetahui mean income dari manager tingkat menengah pada suatu industri tertentu. Sample random dari 256 manager menunjukkan sample mean sebesar $\bar{X} = \$45420$. Standard deviasi dari populasinya adalah $\sigma = \$2050$.

1. Berapa mean dari populasi? (μ)
2. Dapatkan rentang nilai yang masuk akal dari mean dari populasi!
3. Apa makna dari hasil tersebut?

Contoh

1. Berapa mean dari populasi? (μ)
   Dalam kasus ini, kita tidak tahu berapa nilai dari μ.
   Yang kita miliki adalah sample dari mean $\bar{X}$=\$45420.
   Sample dari mean ini adalah point estimate dari mean dari populasi.
   $$\bar{X} \to \mu \tag{5}$$

Contoh

2. Dapatkan rentang nilai yang masuk akal dari mean dari populasi! Dengan 95% level of confidence:

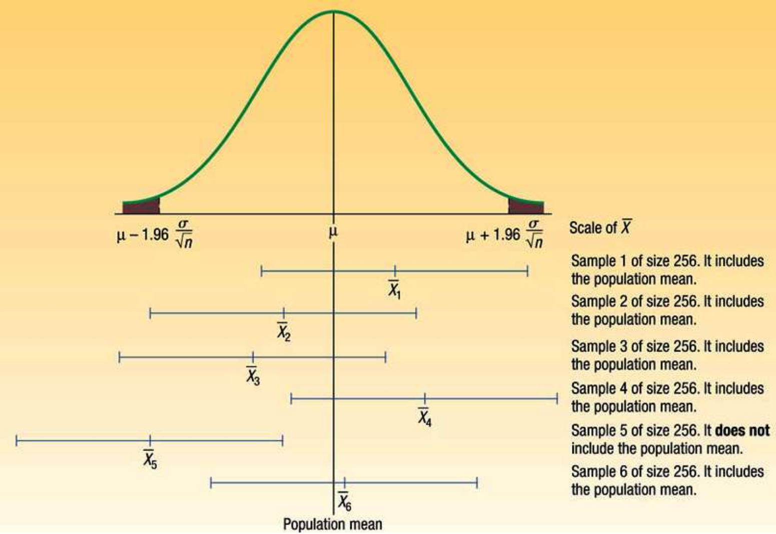
$$\begin{aligned} \bar{X} \pm z\frac{\sigma}{\sqrt{n}} &= \$45420 \pm 1.96\frac{\$2050}{\sqrt{256}} \\ &= \$45420 \pm \$251 \end{aligned} \tag{6}$$

Diperoleh rentang antara \$45169 s/d \$45671. Sedangkan $\pm$\$251 disebut sebagai margin of error.

Contoh

3. Apa makna dari hasil tersebut?
   Jika kita mengambil banyak sample (misalkan 100 sample) dari populasi, dengan masing-masing sample berukuran $n = 256$, dan kita menghitung mean dan confidence interval untuk masing-masing sample, maka kita dapat mengharapkan bahwa 95% dari semua confidence interval itu akan memuat mean dari populasi di dalamnya. Dapat juga kita mengatakan bahwa 5% dari semua confidence interval itu tidak memuat μ di dalamnya.

Contoh



$\mu - 1.96 \frac{\sigma}{\sqrt{n}}$
μ
$\mu + 1.96 \frac{\sigma}{\sqrt{n}}$
Scale of $\overline{X}$
Sample 1 of size 256. It includes the population mean.
Sample 2 of size 256. It includes the population mean.
Sample 3 of size 256. It includes the population mean.
Sample 4 of size 256. It includes the population mean.
Sample 5 of size 256. It **does not** include the population mean.
Sample 6 of size 256. It includes the population mean.
$\overline{X}_1$
$\overline{X}_2$
$\overline{X}_3$
$\overline{X}_4$
$\overline{X}_5$
$\overline{X}_6$
Population mean

Standard Deviasi Populasi (σ) Tidak Diketahui

Di dalam kebanyakan situasi, σ tidak diketahui. Berikut ini beberapa contoh di mana kemungkinan besar σ tidak diketahui.

1. Seorang mahasiswa ingin mengetahui mean dari jumlah jam yang dipakai mahasiswa untuk bekerja (dibayar) di luar kampus. Ia mewawancarai sample yang terdiri dari 30 mahasiswa.
2. Dilakukan survey terhadap 40 mahasiswa, dan diwawancarai berapa mean dari jarak tempat tinggal ke kampus.
3. Duapuluh mahasiswa yang hampir lulus ditanya berapa student loan yang harus dibayar.

σ Tidak Diketahui

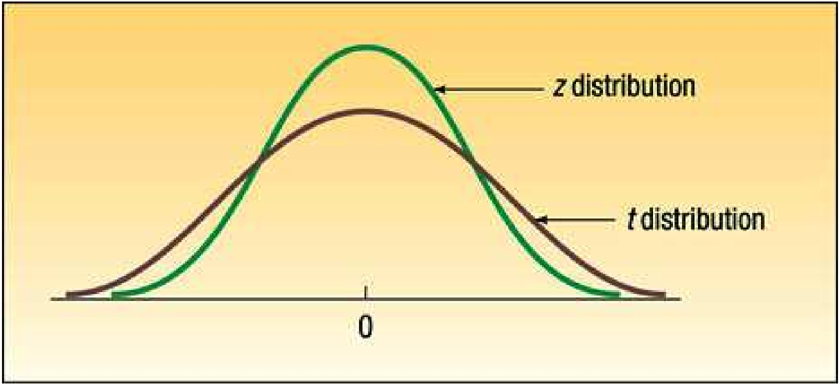
σ tidak diketahui

Jika σ tidak diketahui maka kita harus menggunakan t-distribution.

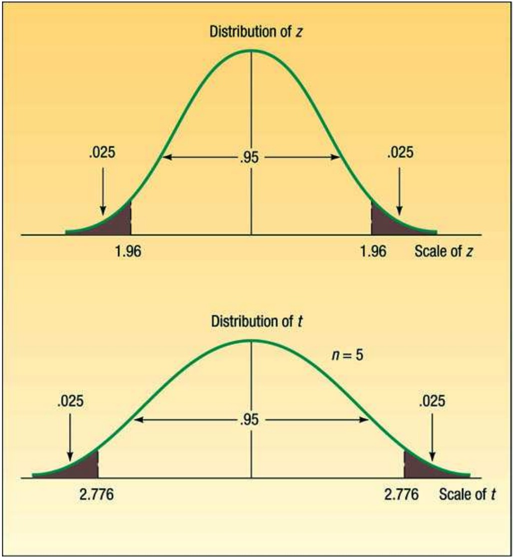
Karakteristik t-distribution

1. Distribusi-t adalah distribusi yang bersifat kontinu (sama seperti distribusi normal).
2. Distribusi-t berbentuk lonceng (bell-shaped) dan simetris (sama seperti distribusi normal).
3. Berbeda dengan distribusi normal yang hanya ada satu, distribusi-t terdiri dari banyak distribusi (”family of distribution”). Semua distribusi-t itu mempunyai mean nol, tetapi mempunyai standard deviasi yang besarnya bergantung pada ukuran sample n.
4. Distribusi-t lebih menyebar, dan lebih rata di pusat daripada distribusi normal. Namun, jika ukuran sample n ditingkatkan, distribusi-t akan semakin mendekati distribusi normal.

Perbandingan distribusi-t dengan distribusi normal



Perbandingan distribusi-t dengan distribusi normal



σ tidak diketahui

Confidence Interval untuk Population Mean jika σ tidak diketahui

$$\begin{aligned} &\bar{X} \pm t_{\alpha/2,df} \frac{s}{\sqrt{n}} \\ &= \bar{X} \pm t_{\alpha/2,n-1} \frac{s}{\sqrt{n}} \end{aligned} \tag{7}$$

Dengan:

$\bar{X}$: mean dari sample
t : nilai-t untuk confidence-level yang dipilih
s : standard deviasi sample
n : jumlah observasi di dalam sample
$df = n - 1$: degree of freedom
α : level of significance (1-α=level of confidence)

Contoh Distribusi-t

Sebuah perusahaan ban ingin menguji daya tahan ulir pada ban-bannya. Sebuah sample terdiri dari 10 ban yang telah menempuh 50ribu mil, menunjukkan sample mean ulir tersisa sebesar 0.32 inci, dengan standard deviasi 0.09 inci.

1. Susunlah 95% confidence interval untuk mean dari populasi!
2. Apakah masuk akal untuk menyimpulkan bahwa setelah menempuh 50ribu mil, mean dari populasi untuk besar ulir tersisa adalah 0.3 inci?

Contoh Distribusi-t

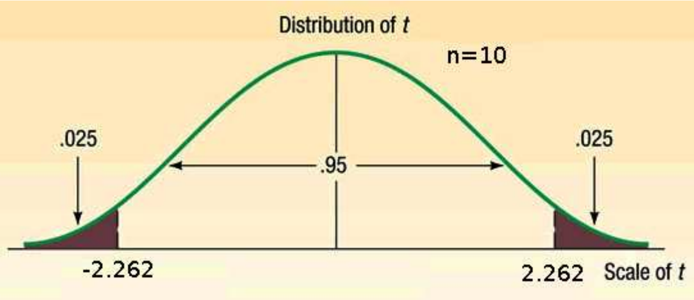
Diketahui dari soal:

$$\begin{aligned} n &= 10 \\ \bar{X} &= 0.32 \\ s &= 0.09 \end{aligned} \tag{8}$$

Hitung C.I. dengan menggunakan distribusi-t (karena σ tidak diketahui)

$$\bar{X} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} \tag{9}$$

	Confidence Intervals				
	80%	90%	95%	98%	99%
	Level of Significance for One-Tailed Test				
df	0.10	0.05	0.025	0.010	0.005
	Level of Significance for Two-Tailed Test				
	0.20	0.10	0.05	0.02	0.01
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	$\boxed{2.262}$	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169


Distribution of t
n=10
.025
.95
.025
-2.262
2.262
Scale of t

Contoh Distribusi-t

$$
\begin{aligned}
\bar{X} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} &= \bar{X} \pm t_{0.05/2,10-1}\frac{s}{\sqrt{n}} \\
&= 0.32 \pm t_{0.025,9}\frac{0.09}{\sqrt{10}} \\
&= 0.32 \pm 2.262\frac{0.09}{\sqrt{10}} \qquad (10) \\
&= 0.32 \pm 0.064 \\
&= (0.256, 0.384)
\end{aligned}
$$

Kesimpulan: Perusahaan ban cukup yakin (95% yakin) bahwa mean dari ulir tersisa adalah berada di antara 0.256 dan 0.384 inci. Cukup masuk akal untuk berasumsi bahwa mean dari populasi adalah 0.3 inci, karena 0.3 inci berada di antara 0.256 dan 0.384 inci.

Contoh 2

Manager dari sebuah mall di Florida ingin membuat estimasi rata-rata uang yang dibelanjakan. Sebuah sample terdiri dari costumer memberikan hasil sebagai berikut.

$48.16	$42.22	$46.82	$51.45	$23.78	$41.86	$54.86
37.92	52.64	48.59	50.82	46.94	61.83	61.69
49.17	61.46	51.35	52.68	58.84	43.88	

Manager ingin tahu apakah mean dari populasi lebih dekat ke $50, atau bisa lebih tinggi lagi (misalnya $60).

Contoh 2

$$\begin{aligned}
\bar{X} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} &= \bar{X} \pm t_{0.05/2,20-1}\frac{s}{\sqrt{n}} \\
&= 49.53 \pm t_{0.05/2,20-1}\frac{9.01}{\sqrt{20}} \\
&= 49.53 \pm 2.093\frac{9.01}{\sqrt{20}} \\
&= 49.53 \pm 4.22 \\
&= (45.13, 53.57)
\end{aligned} \tag{11}$$

Nilai \$60 tidak termasuk di dalam interval, sedangkan \$50 berada dalam interval. Jadi masuk akal kalau disimpulkan bahwa mean dari populasi adalah \$50.

Distribusi-z atau Distribusi-t

Asumsikan populasi mempunyai distribusi normal

Gunakan distribusi-z

1. σ diketahui, atau
2. $n \geq 30$

$$\bar{X} \pm z \frac{\sigma}{\sqrt{n}}$$

Gunakan distribusi-t

1. σ tidak diketahui, dan
2. $n < 30$

$$\bar{X} \pm t \frac{s}{\sqrt{n}}$$

Soal

Seorang peneliti mengamati sejenis burung, dan memperoleh data-data sebagai berikut.
Sampel yang terdiri dari 10 burung jantan mempunyai massa (dalam gram):
620, 600, 597, 520, 585, 570, 585, 540, 545, 570
Sampel kedua terdiri dari 9 burung betina mempunyai massa (dalam gram):
550, 540, 545, 510, 530, 585, 525, 520, 530
(a) Susun Confidence Interval (95%) untuk masing-masing sampel,
(b) apakah perbedaan rata-rata massa antara kedua sampel itu mempunyai arti.