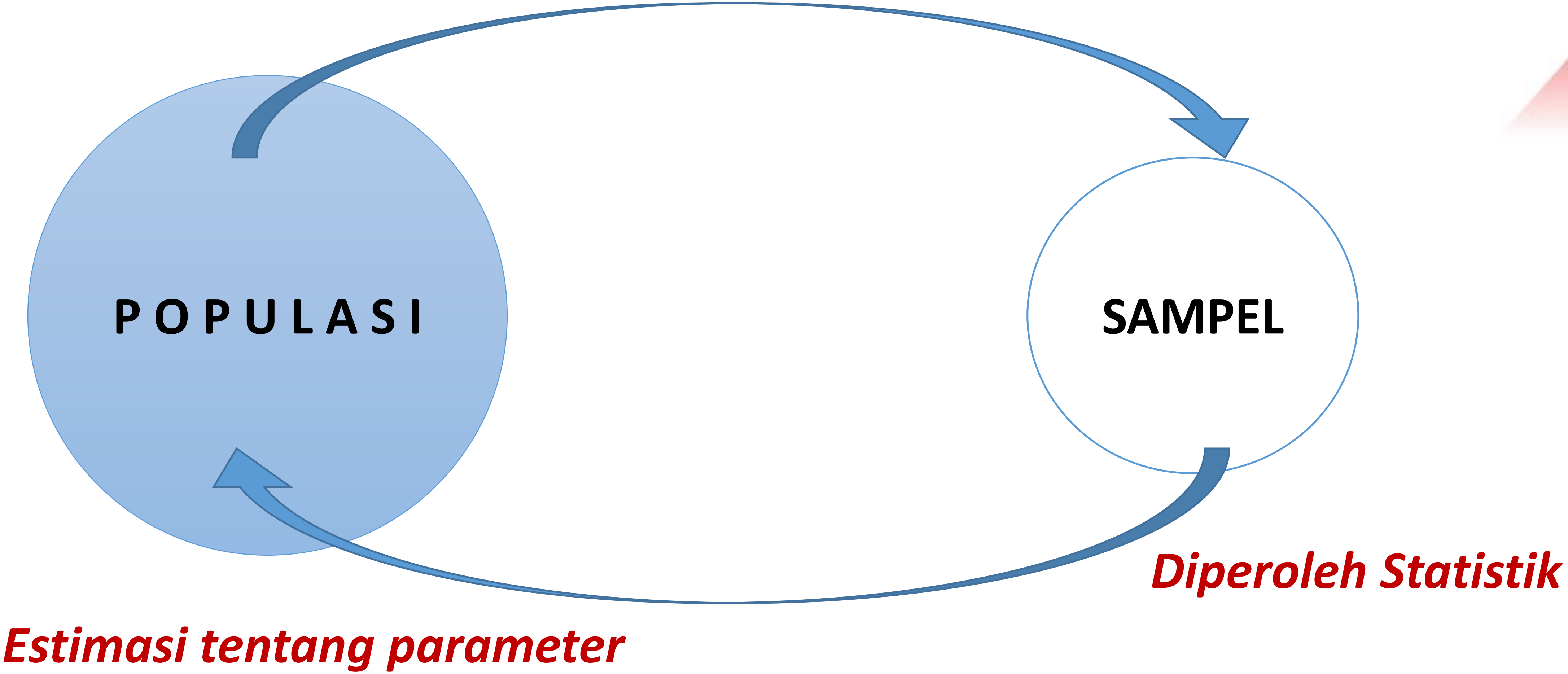
STATISTIKA DAN PROBABILITAS (CIV -110)

PERTEMUAN 10

Statistika Inferensial : Rentang Keyakinan

UNIVERSITAS PEMBANGUNAN JAYA
JL.CENDRAWASIH NO.1 SAWAH BARU, CIPUTAT 15413 TANGERANG SELATAN


POPULASI
SAMPEL
Diperoleh Statistik
Estimasi tentang parameter

Sampling dan Estimasi

ALASAN MENGAMBIL SAMPLING

1. Menghubungi keseluruhan populasi memakan waktu.
2. Meneliti seluruh item di dalam populasi seringkali terlalu mahal.
3. Hasil-hasil sampel umumnya memadai.

Dalam praktik riil, informasi mengenai **mean, standar deviasi,** atau bentuk populasi tidak bisa diketahui secara pasti.

Sampling Distribution also give us the foundation that allows us to take a sample and use it to estimate a population parameter.

PROSES ESTIMASI

POPULASI

SAMPEL ACAK

Mean, μ, tidak diketahui

Sampel

$$mean = \overline{X} = 50$$

Saya 95% yakin bahwa mean antara 40 dan 60

BAGAIMANA MENDUGA SUATU PARAMETER

Harga parameter dapat diestimasikan/ diduga dengan dua cara, yakni :

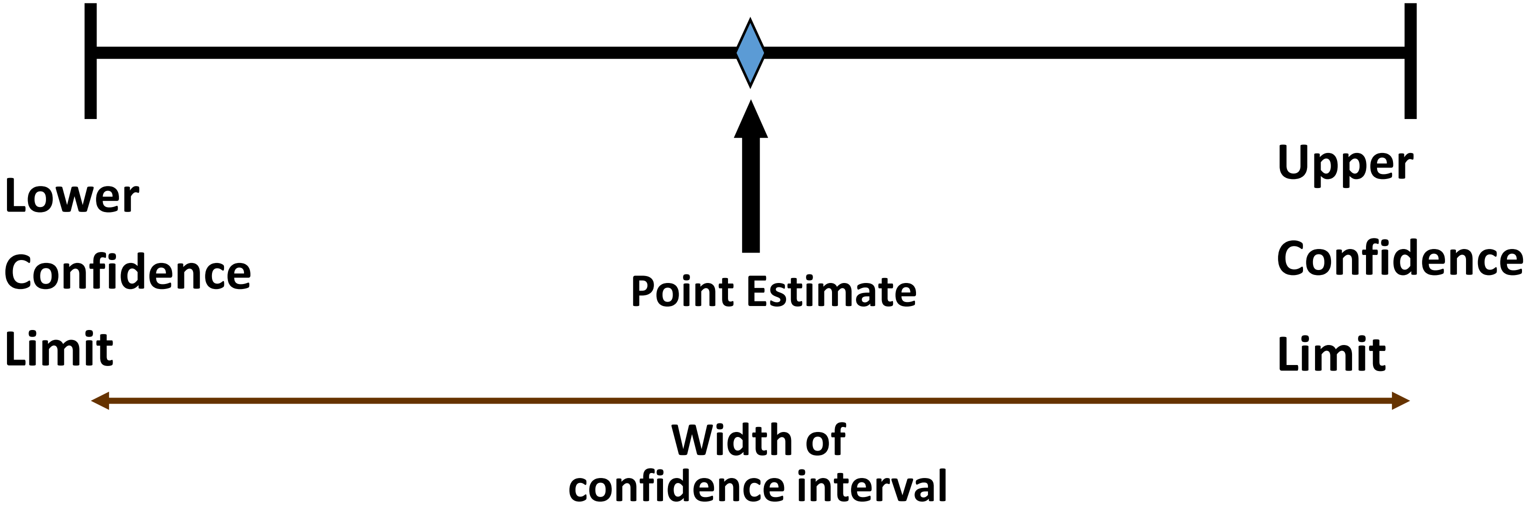
- *Point estimation* (Pendugaan Titik)
  merupakan nilai (suatu titik) yang digunakan untuk menduga suatu titik
- *Interval estimation* (Pendugaan Interval).
  suatu interval yang menyatakan **rentang** dimana suatu parameter populasi mungkin berada

Faktor-faktor yang menentukan luasnya Interval Kepercayaan

1) Besar sampel, **n**.
2) Keragaman populasi, biasanya **σ** diperkirakan oleh **s**.
3) Tingkat kepercayaan yang diinginkan

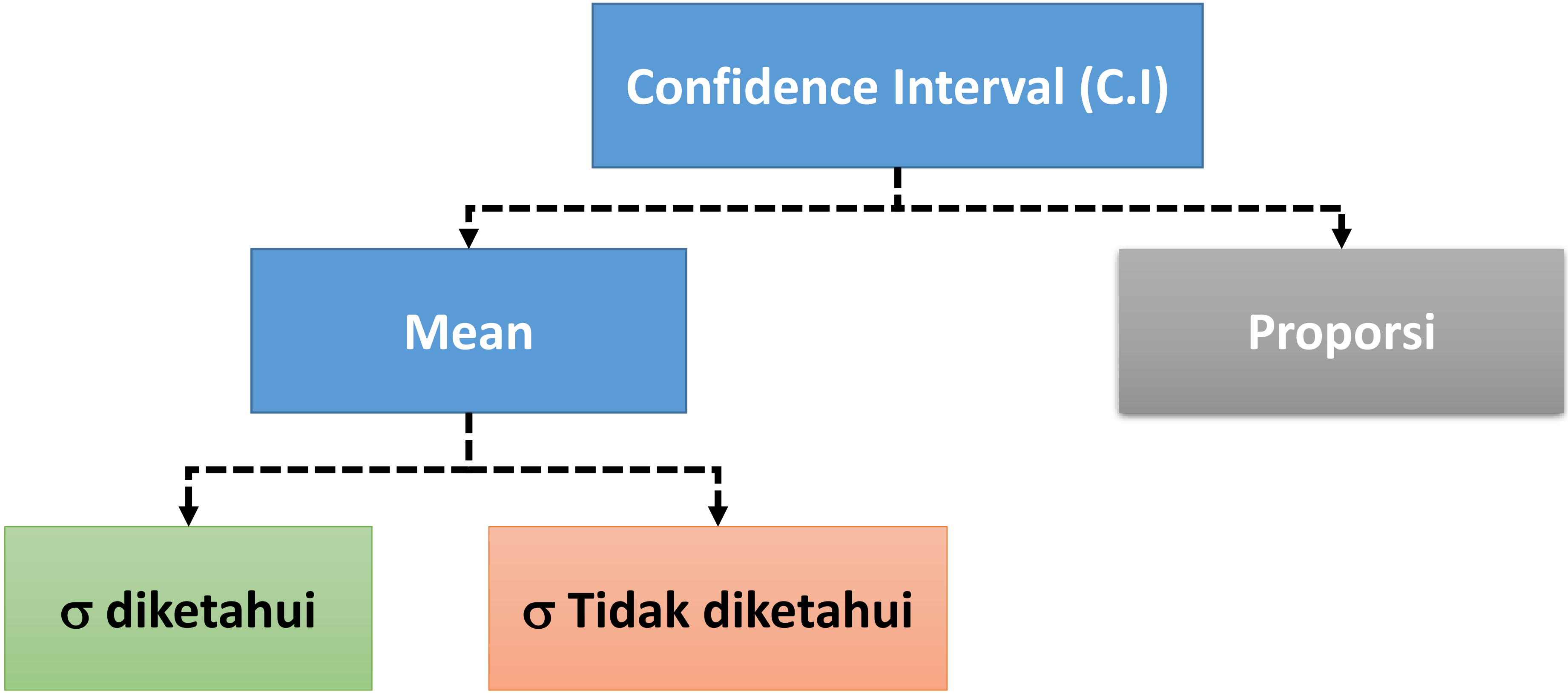
POINT ESTIMATION

Estimasi Parameter Populasi ...		Dengan statistik Sample
Mean	μ	$\bar{X}$
Proporsi	p	P_S
Varian	σ^2	S^2
Difference	$\mu_1 - \mu_2$	$\bar{X}_1 - \bar{X}_2$



Rumus umum untuk semua rentang keyakinan :

C.I = Point Estimate $\pm$ (Critical Value)(Standard Error)

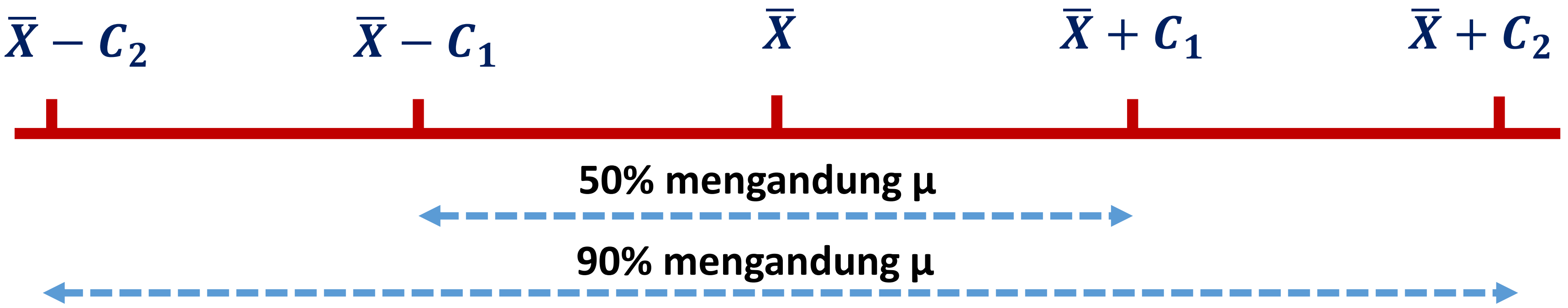

Confidence Interval (C.I)
Mean
Proporsi
σ diketahui
σ Tidak diketahui

Menghitung C.I apabila

Estimasi Parameter Populasi ...		Dengan statistik Sample
Mean	μ	$\bar{X}$
Proporsi	p	P_S
Varian	σ^2	S^2
Difference	$\mu_1 - \mu_2$	$\bar{X}_1 - \bar{X}_2$

ESTIMASI C.I UNTUK μ (σ diketahui)

- Apabila distribusi x diketahui, seberapa dekat μ tehadap x ?
- Ingin mengetahui probabilitas nilai sebenarnya μ terletak antara (x-c) dan (x+c)→ lihat gambar



Jika c tidak terbatas → probabilitas μ berada pada interval 100% **TIDAK BERGUNA**

Jika c dibuat lebih sempit dengan nilai c kecil → μ menjadi lebih tepat

Langkah kerjanya adalah sbb :

1) Tentukan probabilitas yang diinginkan (meletakkan μ pada interval) → umumnya 95%
2) Hitung berapa luas interval jika kemungkinan 95 % mengandung niai sebenarnya

Interval keyakinan (confidence interval)

95 = tingkat keyakinan (confidence level)

$$\boldsymbol{P(\bar{X} - c \leq \mu \leq \bar{X} + c) = 0.95}$$

3) Buatlah variable acak , yaitu Z : $Z = \frac{\bar{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}} = \sqrt{n}\,\frac{\bar{X}-\mu}{\sigma}$

Z = distribusi normal → mean = 0 dan varian = 1 sehingga persamaan :

$$P\left(\frac{-c\sqrt{n}}{\sigma} \le \sqrt{n}\,\frac{\bar{X}-\mu}{\sigma} \le \frac{c\sqrt{n}}{\sigma}\right) = 0.95 \;\Rightarrow\; P\left(\frac{-c\sqrt{n}}{\sigma} \le Z \le \frac{c\sqrt{n}}{\sigma}\right) = 0.95$$

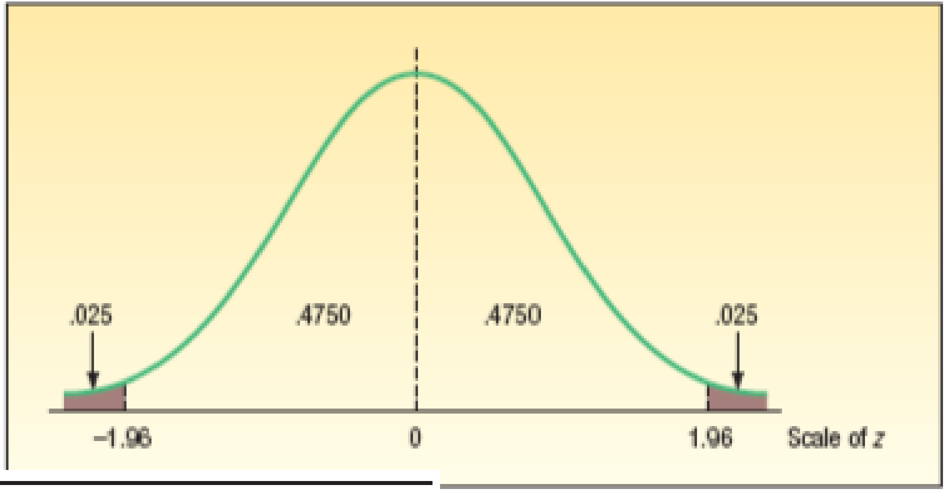
$$\boxed{\frac{-c\sqrt{n}}{\sigma} = \mathrm{a}}$$

Dari tabel dapat diketahui nilai **a = 1.96**

Terdapat 95% kemungkinan nilai sebenarnya μ akan berada pada diantara $\bar{x} - 1{,}96\,\sigma/\sqrt{n}$ dan $\bar{x} + 1{,}96\,\sigma/\sqrt{n}$

Confidence interval untuk rata-rata pupulasi bila σ diketahui	$\bar{X} \pm z\frac{\sigma}{\sqrt{n}}$

$\bar{x}$ = nilai rata-rata sampel
z = nilai untuk confidence level tertentu
σ = standar deviasi populasin
n = jumlah sampel yang diamati



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

Contoh Soal

American Management Association ingin memiliki informasi rata-rata pendapatan manajer toko di industri eceran. Sampel acak 256 manajer menyataka ratarata sampelnya $45.420. Standar deviasi populasi ini adalah $2.050

Berapa rata-rata populasinya?

Pada kasus ini, kita tidak tahu. Kita hanya tahu ratarata sampel $45.420. Dengan demikian, estimasi terbaik kita dari nilai populasi yang tidak dikeahui adalah angka sampel yang sesuai. Jadi, rata-rata sampel $45.420 merupakan estimasi titik dari ratarata populasi yang tidak diketahui

Berapa jangkauan nilai yang tepat untuk rata-rata populasinya?

$$\bar{X} \pm z\frac{\sigma}{\sqrt{n}} = \$45{,}420 \pm 1.96\frac{\$2{,}050}{\sqrt{256}} = \$45{,}420 \pm \$251$$

Cara biasanya adalah membulatkan titik ujung menjadi $45.169 dan $45.671, ini disebut batas kepercayaan. Tingkat kepercayaan adalah 95% yaitu $45.169-$45.671. $251 disebut sebagi batas kesalahan

Jika dipilih sampel 256 manajer, untuk masing2 sampel dihitung rata-ratanya dan membentuk interval kepercayaan 95%, kita dapat memperkirakan 95% interval kepercayaan ini memuat rata-rata populasi.

ESTIMASI C.I untuk μ (σ tidak diketahui) → t – distribution

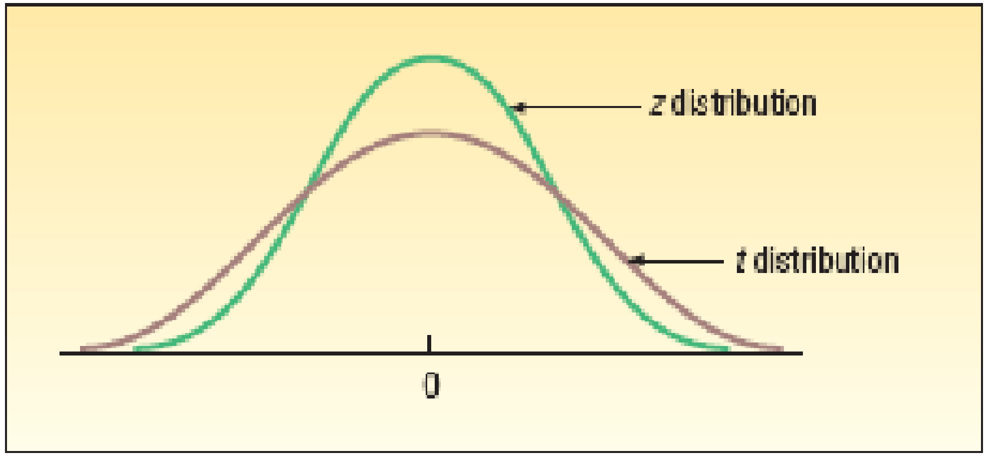
Pada kebanyakan situasi sampling standar deviasi (σ) tidak diketahui.

“Manager SDM PT.Anti Tekor ingin memperkirakan rata-rata jumlah jam bekerja pegawai paruh waktu yang bekerja tiap minggu. Ia memilih sampel 30 pegawai, menghubungi setiap pegawai dan bertanya berapa jam mereka bekerja minggu lalu”

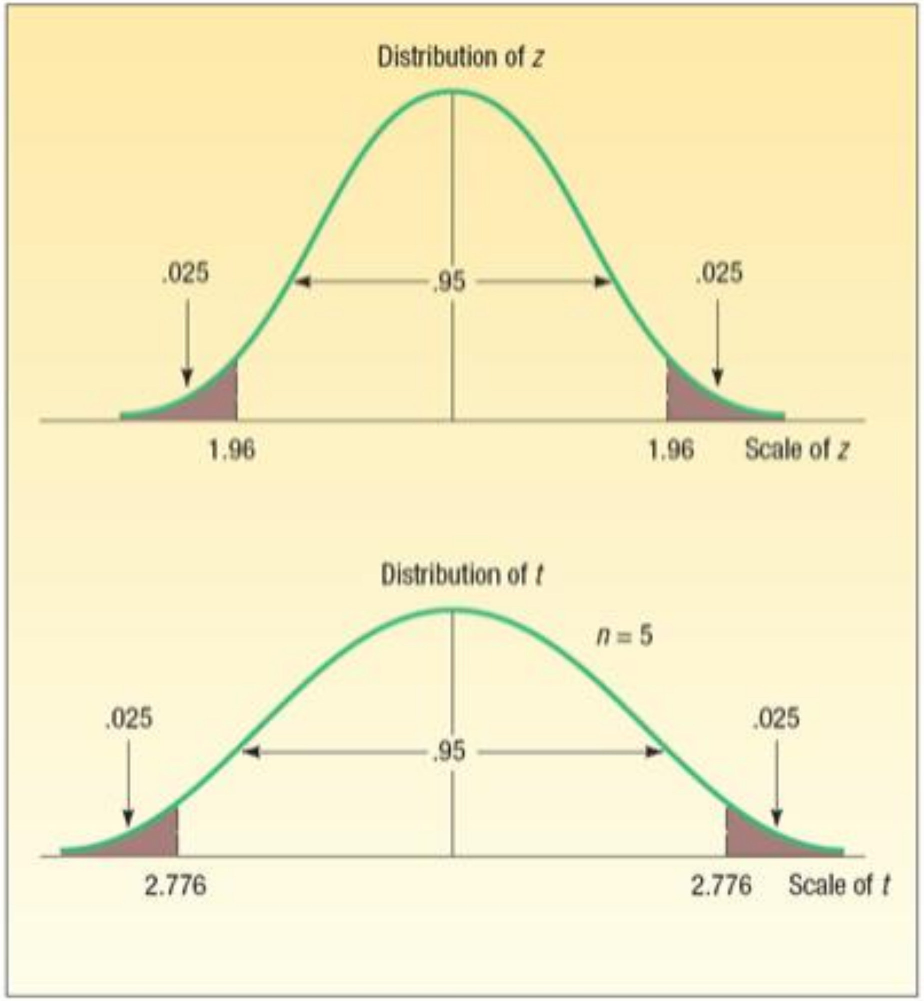
Standar deviasi(σ) populasi tidak memungkinkan diketahui

Karakteristik Distribusi t

1) Distribusinya, seperti distribusi z, merupakan distribusi kontinu.
2) Distribusinya, seperti distribusi z, berbentuk lonceng dan simetris.
3) Tidak hanya ada satu distribusi t, tetapi serumpun distribusi t. Seluruh distribusi t memiliki rata-rata 0, tetapi standar deviasinya berbeda-beda sesuai ukuran sampel, **n.**
4) Distribusi t lebih tersebar dan lebih landai di tengah daripada distribusi normal baku. Namun, semakin ukuran sampel bertambah, distribusi t mendekati distribusi normal baku karena kesalahan dalam penggunaan s untuk memperkirakan σ menurun dengan sampel yang lebih banyak.



Komparasi distribusi z dan t jika jumlah sampel (n) kecil



Menggunakan distribusi Z

Jika standar deviasi populasi diketahui atau jumlah sampel > 30

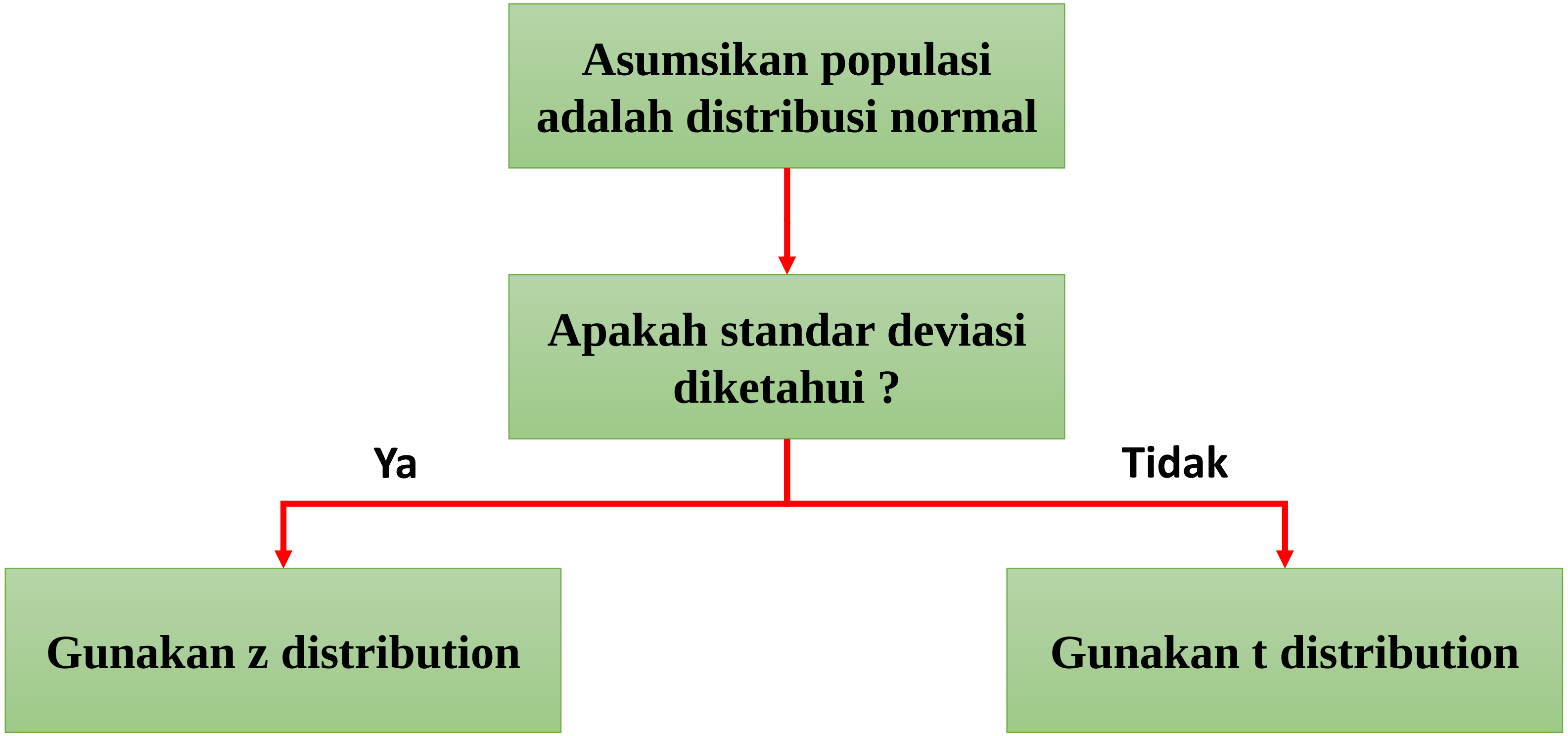
$$\bar{X} \pm z \frac{\sigma}{\sqrt{n}}$$

Menggunakan distribusi t

Jika standar deviasi populasi tidak diketahui atau jumlah sampel < 30

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

Kapan menentukan estimasi dengan distribusi Z atau distribusi t ????



Contoh soal

Pabrik ban ingin menyelidiki tebal jejak ban-ban produksinya. Sampel 10 ban yang menempuh 50.000 mil menyatakan rata-rata sampel jejak yang membekas adalah 0,32 inci dengan standar deviasi 0,09 inci. Gunakan interval kepercayaan 95% untuk rata-rata populasi.
Apakah tepat bagi pabrik untuk menyimpulkan bahwa setelah 50.000 mil rata-rata populasi jumlah jejak yang membekas adalah 0,03 inci?

Given in the problem:

$n = 10$

$\bar{x} = 0.32$

$s = 0.09$

Compute the C.I. using the t-dist. (since σ is unknown)

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

Compute the C.I.

using the t - dist. (since σ is unknown)

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

$$= \bar{X} \pm t_{.05/2,10-1} \frac{s}{\sqrt{n}}$$

$$= 0.32 \pm t_{.025,9} \frac{0.09}{\sqrt{10}}$$

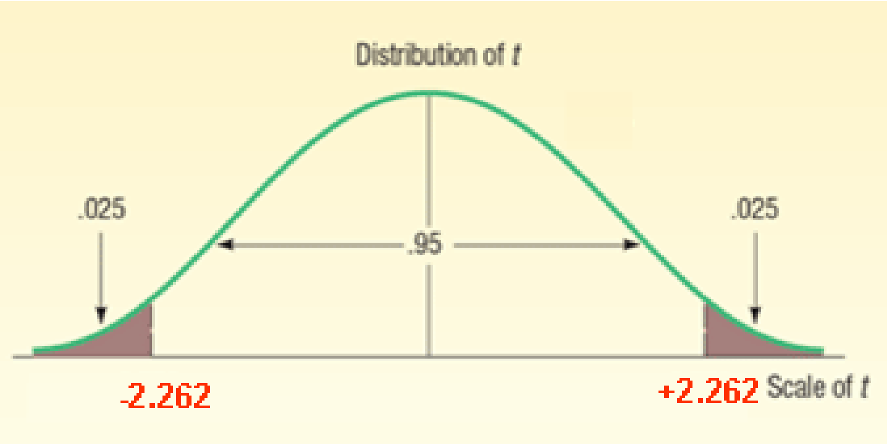
$$= 0.32 \pm 2.262 \frac{0.09}{\sqrt{10}}$$

$$= 0.32 \pm 0.064$$

$$= (0.256, 0.384)$$

Conclude : the manufacturer can be reasonably sure (95% confident) that the mean remaining tread depth is between 0.256 and 0.384 inches.

	Confidence Intervals				
	80%	90%	95%	98%	99%
	Level of Significance for One-Tailed Test				
df	0.100	0.050	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



ESTIMASI C.I untuk proporsi (π)

1. Perwakilan perusahaan menyatakan bahwa 45% penjualan Burger King dilakukan melalui *drive-through*.
2. Survei rumah-rumah di Chicago menunjukkan bahwa 85% bangunan memiliki ventilasi udara terpusat.
3. Survei terkini mengenai pria yang menikah di antara usia 35 dan 50 tahun menemukan bahwa 63% merasa bahwa kedua pasangan seharusnya mencari nafkah.

- Untuk membuat tingkat kepercayaan terhadap proporsi, kita perlu memenuhi asumsi berikut :
  a. Kondisi binomial telah terpenuhi. Kondisi tersebut antara lain:
     - Data sampel merupakan hasil penghitungan.
     - Hanya terdapat 2 kemungkinan hasil.
     - Probabilitas keberhasilan tetap sama dari satu percobaan ke percobaan berikutnya.
     - Percobaan-percobaannya saling beba. Ini berarti hasil pada suatu percobaan tidak mempengaruhi hasil percobaan lainnya.

  b. Nilai **n π** dan **n(1-π)** seharusnya lebih besar atau sama dengan 5. kondisi ini memungkinkan kita untuk menggunakan *teorema limit tengah* dan menerapkan distribusi normal baku, yakni, z, untuk mencapai suatu interval kemungkinan.

Sampel proporsi

$$\boldsymbol{p = \frac{X}{n}}$$

Confidence interval for a Population proportion

$$\boldsymbol{p \pm z\sqrt{\frac{p(1-p)}{n}}}$$

Contoh soal

Serikat perwakilan Bottle Blowers of America (BBA) sedang mempertimbangkan proposal untuk bergabung dengan Serikat Teamsters. Menurut anggaran rumah tangga serikat BBA, sedikitnya ¾ anggota serikat harus menyetujui merger apapun. Sampel acak 2.000 anggota BBA saat ini menunjukkan bahwa 1.600 diantaranya akan menyetujui proposal merger tersebut. Berapa estimasi proporsi dan populasinya?

Gunakan interval kepercayaan 95%. Dengan mendasarkan keputusan anda pada informasi sampel ini, dapatkah anda mengambil kesimpulan bahwa terdapat cukup proporsi dari populasi yang mendukung merger? Mengapa?

First, compute the sample proportion:

$$p = \frac{x}{n} = \frac{1{,}600}{2000} = 0.80$$

Compute the 95% C.I.

$$\text{C.I.} = p \pm z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

$$= 0.80 \pm 1.96\sqrt{\frac{.80(1-.80)}{2{,}000}} = .80 \pm .018$$

$$= (0.782,\ 0.818)$$

Conclude: The merger proposal will likely pass because the interval estimate includes values greater than 75 percent of the union membership.

Menentukan ukuran sampel yang tepat

Tiga faktor yang menentukan ukuran suatu sampel, walaupun tidak ada satupun yang berhubungan langsung dengan ukuran suatu populasi

1. Level of confidence yang diharapkan
2. Batas kesalahan (margin error) yang ditoleransi peneliti
3. Variasi dari populasi yang diamati

$$\boldsymbol{n} = \left(\frac{\boldsymbol{z}.\boldsymbol{\sigma}}{\boldsymbol{E}}\right)^{2}$$

n = jumlah sampel
z = nilai distribusi normal yang diharapkan pada tingkat keyakinan
σ = standar deviasi
E = maksimum error yang diijinkan

Contoh soal

A student in public administration wants to determine the mean amount members of city councils in large cities earn per month as remuneration for being a council member. The error in estimating the mean is to be less than $100 with a 95 percent level of confidence. The student found a report by the Department of Labor that estimated the standard deviation to be $1,000. What is the required sample size?

Given in the problem:

- *E*, the maximum allowable error, is $100
- The value of *z* for a 95 percent level of confidence is 1.96,
- The estimate of the standard deviation is $1,000.

$$n = \left(\frac{z \cdot \sigma}{E} \right)^2$$

$$= \left(\frac{(1.96)(\$1{,}000)}{\$100} \right)^2$$

$$= (19.6)^2$$

$$= 384.16$$

$$= 385$$

Menentukan ukuran sampel yang tepat dari proporsi

$$n = p(1-p)\left(\frac{Z}{E}\right)^2$$

n is the size of the sample
z is the standard normal value corresponding to the desired level of confidence
E is the maximum allowable error

The American Kennel Club wanted to estimate the proportion of children that have a dog as a pet. If the club wanted the estimate to be within 3% of the population proportion, how many children would they need to contact? Assume a 95% level of confidence and that the club estimated that 30% of the children have a dog as a pet.

$$n = (.30)(.70)\left(\frac{1.96}{.03}\right)^2 = 897$$

A study needs to estimate the proportion of cities that have private refuse collectors. The investigator wants the margin of error to be within .10 of the population proportion, the desired level of confidence is 90 percent, and no estimate is available for the population proportion. What is the required sample size?

$$n = (.5)(1-.5)\left(\frac{1.65}{.10}\right)^2 = 68.0625$$

$$n = 69 \text{ cities}$$