Statistika Inferensial

- Part 2 –

Distribusi Sampel

Muhammad Afif Hendrawan, S.Kom., M.T.

Outlines

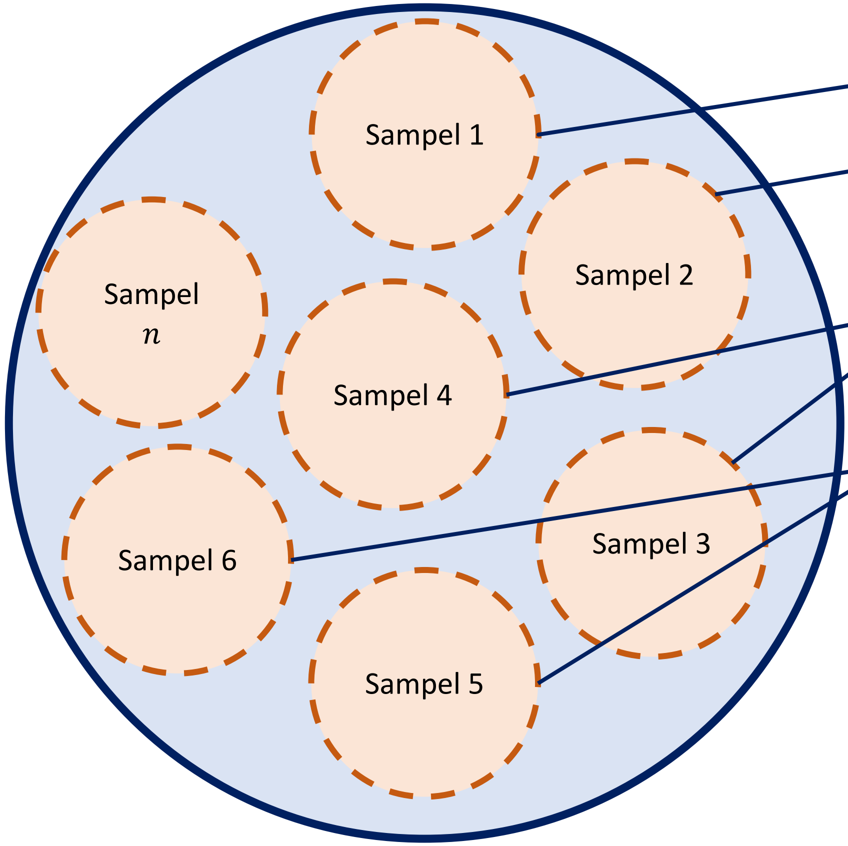
Distribusi Sampel

Distribusi t

Distribusi Sampel

Nilai yang mengambarkan sampel-sampel

Distribusi Rata-Rata Sampel #1



Populasi berukuran terhingga N
dengan parameter rata-rata μ

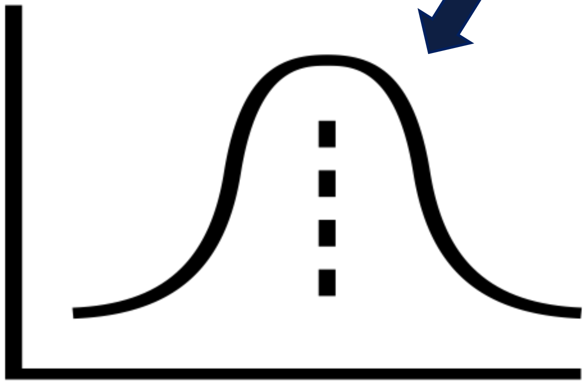
$$\left.\begin{aligned}
\bar{x}_{s_1} &= 9.6\\
\bar{x}_{s_2} &= 11.2\\
\bar{x}_{s_3} &= 10.4\\
\bar{x}_{s_4} &= 9.8\\
\bar{x}_{s_5} &= 10.8\\
\bar{x}_{s_6} &= 11.4\\
&\ \vdots\\
\bar{x}_{s_n} &= \cdots
\end{aligned}\right\}$$

DATA BARU
(data rata-rata sampel)

Sampling distribution of the sample mean

Jika diambil rata-ratanya (lagi)

Mean of sampling distribution of the sample mean



Akan terdistribusi secara normal

Distribusi Rata-Rata Sampel #2

Mean of sampling distribution of the mean (rata-rata dari distribusi rata-rata sampel) → $\mu_{\bar{x}}$

Nilai $\mu_{\bar{x}}$ akan sama dengan rata-rata populasi (μ), jika **distribusi rata-rata sampel terdistribusi secara normal**

Distribusi Rata-Rata Sampel #3

Apa hubungannya dengan distribusi normal?

Distribusi rata-rata sampel **akan terdistrubusi secara normal jika populasi juga terdistribusi secara normal**

Jika populasi tidak terdistrubusi secara normal (atau kita tidak tahu) → untuk membuat distribusi rata-rata sampel menjadi normal → **jumlah data dalam sampel $\geq$ 30**

CENTRAL LIMIT THEOREM (CLT)

Sehingga, kita dapat menggunakan sifat-sifat khusus yang ada pada distribusi normal (probabilitas)

JIKA

$$\mu_{\bar{x}} = \mu$$

Maka,

Variance dan STD Dari distribusi sampel

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$$

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

populasi

Bagaimana jika,

Kita juga tidak tau nilai asli dari populasinya?

$$S_{\bar{x}}^2 = \frac{S^2}{n}$$

$$\boxed{S_{\bar{x}} = \frac{S}{\sqrt{n}}}$$

Galat baku rata-rata/ Standard error of mean/ Standard error

Kita gunakan nilai sampelnya untuk melakukan estimasi

Distribusi Rata-Rata Sampel #4

Bagaimana jika data sampel besar?

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} \qquad S_{\bar{x}}^2 = \frac{S^2}{n}$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \qquad S_{\bar{x}} = \frac{S}{\sqrt{n}}$$

Untuk kondisi apabila sampelnya diambil dari populasi tak terhingga dengan ukuran sampel $\frac{n}{N} \leq 5\%$ atau data dapat dipilih kembali pada sampel

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} * \frac{N-n}{N-1} \qquad S_{\bar{x}}^2 = \frac{S^2}{n} * \frac{N-n}{N-1}$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} * \sqrt{\frac{N-n}{N-1}} \qquad S_{\bar{x}} = \frac{S}{\sqrt{n}} * \sqrt{\frac{N-n}{N-1}}$$

Untuk kondisi apabila sampelnya diambil dari populasi terhingga dengan ukuran sampel $> 5\%$ dari populasinya atau data **tidak dapat** dipilih kembali pada sampel

Distribusi Rata-Rata Sampel #5 - Contoh

Suatu populasi dengan jumlah N = 10 yang memiliki data sebagai berikut. 98, 99, 97, 98, 99, 98, 97, 97, 98, dan 99. Apabila dari populasi ini diambil sampel acak secara berulang-ulang dengan ukuran n = 2 maka diperoleh 45 buah sampel yang kemudian setiap sampel dihitung rata-ratanya. Data dari setiap sampel tersebut sebagai berikut.

Tabel 5.2. Nilai Rata-rata Sampel

No	Sampel	Rata-rata	No	Sampel	Rata-rata	No.	Sampel	Rata-rata
1	98, 99	98,5	16	99, 98	98,5	31	99, 98	98,5
2	98, 97	97,5	17	99, 99	99	32	99, 97	98
3	98, 98	98	18	97, 98	97,5	33	99, 97	98
4	98, 99	98,5	19	97, 99	98	34	99, 98	98,5
5	98, 98	98	20	97, 98	97,5	35	99, 99	99
6	98, 97	97,5	21	97, 97	97	36	98, 97	97,5
7	98, 97	97,5	22	97, 97	97	37	98, 97	97,5
8	98, 98	98	23	97, 98	97,5	38	98, 98	98
9	98, 99	98,5	24	97, 99	98	39	98, 99	98,5
10	99, 97	98	25	98, 99	98,5	40	97, 97	97
11	99, 98	98,5	26	98, 98	98	41	97, 98	97,5
12	99, 99	99	27	98, 97	97,5	42	97, 99	98
13	99, 98	98,5	28	98, 97	97,5	43	97, 98	97,5
14	99, 97	98	29	98, 98	98	44	97, 99	98
15	99, 97	98	30	98, 99	98,5	45	98, 99	98,5

Dari jumlah sampel sebanyak 45 buah rata-rata tersebut apabila dijumlahkan diperoleh sebesar 4410. Rata-rata dari rata-rata sampel tersebut adalah

$\mu_{\bar{X}} = \frac{4410}{45} = 98$. Rata-rata ini sama dengan rata-rata populasinya. Sedangkan simpangan baku populasinya dari perhitungan diperoleh $\sigma = 0{,}775$ (perhitungan ini dapat dilihat pada pokok pembahasan simpangan baku untuk populasi). Berdasarkan kedua perhitungan ini, maka dapat dihitung galat baku rata-rata, yaitu:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

$$= \frac{0{,}775}{\sqrt{2}} \sqrt{\frac{10-2}{10-1}} = 0{,}52$$

Sumber: Statistika Terapan, Mikha Agus Widiyanto (2013)

Syarat inferensi dengan Distribusi Rata-Rata Sampel #1

Random → Data sampel yang diambil harus bersifat acak (random)

Normal → Sampel terdisitribusi secara normal → Ukuran sampel cukup besar → Memenuhi kaidah CLT, ≥ 30

Independen *(aturan 10%)* →

- Ukuran sampel paling $\leq 10\%$ dari populasi
- Kondisi ini tidak masalah jika pengambilan sampel dengan teknik pengembalian
- Namun perlu diperhatikan, jika pengambilan sampel tanpa pengembalian

Syarat inferensi dengan Distribusi Rata-Rata Sampel #2

Contoh Kasus

Sebuah produsen bola diperkirakan memproduksi bola sebanyak 1.000.000 bola setiap tahun, dan tekanan tiap bola yang dihasilkan terdistribusi secara normal dengan **mean 8.7 PSI** dan simpangan bakunya **0.4 PSI**. Dipilih secara **acak 25 bola (tanpa pengembalian)** untuk mengetahui tekanannya. **Cari probabilitas dari mean sampel $\bar{x}$ berjarak 0.2 PSI dari mean populasi**

Cek Kondisi

Apakah kondisi sudah memenuhi syarat inferensi dengan distribusi rata-rata sampel?

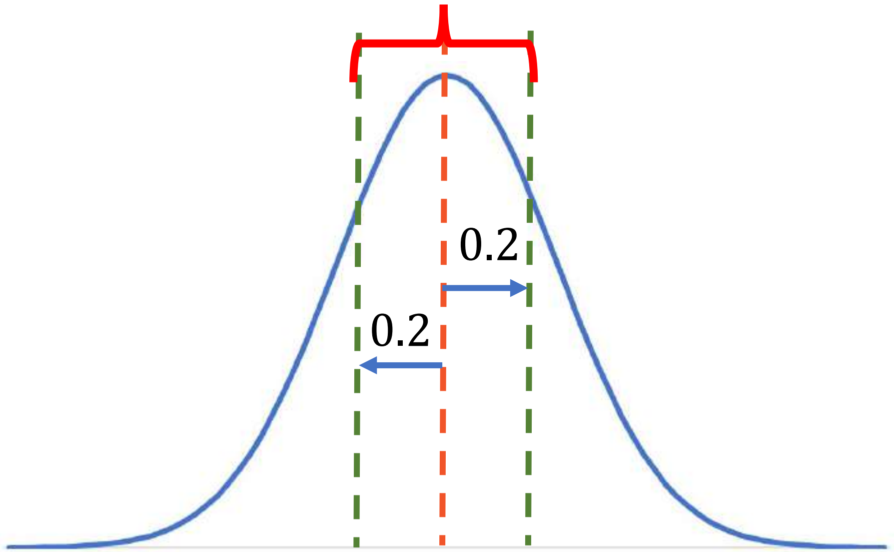
- Dari kasus, sudah dijelaskan bahwa, sampel di pilih secara acak → Asumsi random ✓
- Dari kasus, juga sudah dijelaskan bahwa tekanan bola terdistribusi secara normal → Asumsi normal ✓
- Apakah sampel $\leq 10\%$? Karena 25 bola sangat jauh dibandingkan 1.000.000 bola maka syarat independen (aturan 10%) juga terpenuhi ✓

Syarat inferensi dengan Distribusi Rata-Rata Sampel #2

Penyelesaian Kasus

Apa yang sebetulnya kita cari? “*Sampel mean berapa diantara 0.2 dari mean populasi*”
Berarti, kita menjadi **peluang nilai sampel mean berjarak 0.2 dari mean populasi**

Area yang di cari
Jadi, kita perlu nilai z



$\mu_{\bar{x}} = \mu \rightarrow CLT$

- Berdasarkan ”dalil” CLT, $\mu_{\bar{x}} = \mu$, maka nilai $\mu_{\bar{x}} = 8.7\ PSI$
- Nilai *standard error* / simpangan baku distribusi rata-rata sampel adalah,

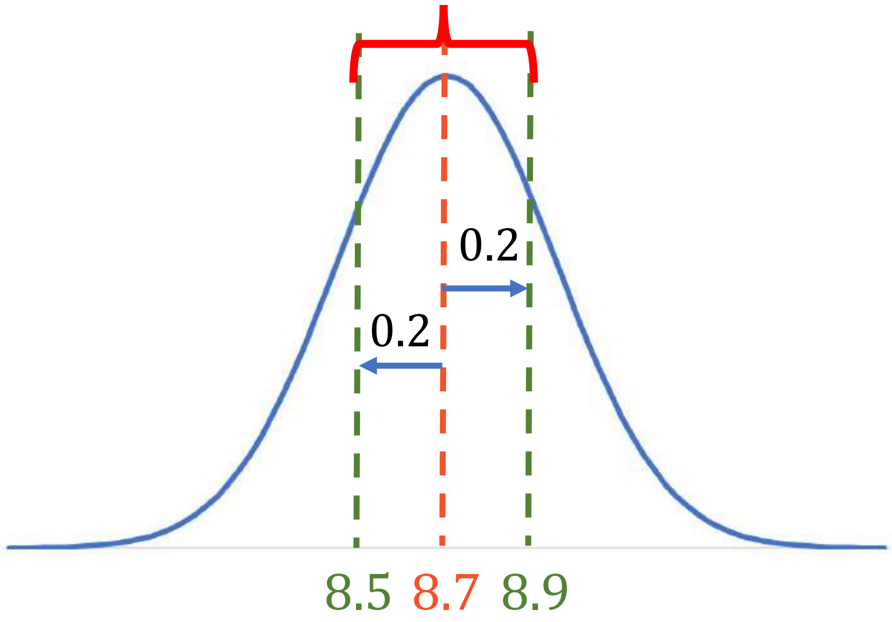
$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.4}{\sqrt{25}} = \frac{0.4}{5} = 0.08$$

- Selanjutnya, kita menjadi area sekitar 0.2 dari mean. Artinya, kita perlu $\mu_{\bar{x}} + 0.2$ dan $\mu_{\bar{x}} - 0.2$

Syarat inferensi dengan Distribusi Rata-Rata Sampel #3

Penyelesaian Kasus

Area yang di cari
Jadi, kita perlu nilai z



- Probabilitas sampel mean pada jarak 0.2 dapat dicari dengan nilai z,

$$z_{\bar{x}} = \frac{\bar{x}-\mu}{\sigma_{\bar{x}}} = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$$

- Maka,

$$P(8.5 < \bar{x} < 8.9 = P\left(\frac{8.5-8.7}{0.08} < z_{\bar{x}} < \frac{8.9-8.7}{0.08}\right)$$

$$= P\left(\frac{-0.2}{0.08} < z_{\bar{x}} < \frac{0.2}{0.08}\right)$$

$$= P(-2.50 < z_{\bar{x}} < 2.50)$$

Syarat inferensi dengan Distribusi Rata-Rata Sampel #3

Penyelesaian Kasus

Cek tabel Z

Kita cek nilai untuk $P(-2.50) = 0.0062$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
-2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
-2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064

Kita cek nilai untuk $P(2.50) = 0.9938$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964

Maka $P(-2.50 < z_{\bar{x}} < 0.25) = 0.9938 - 0.0062 = 0.9876 \approx 0.988 \approx 98.8\%$

Sehinga, peluang sampel 25 bola memiliki mean tekanan berada pada sekitar 0.2 dari mean populias adalah 98.8%

Distribusi t #1

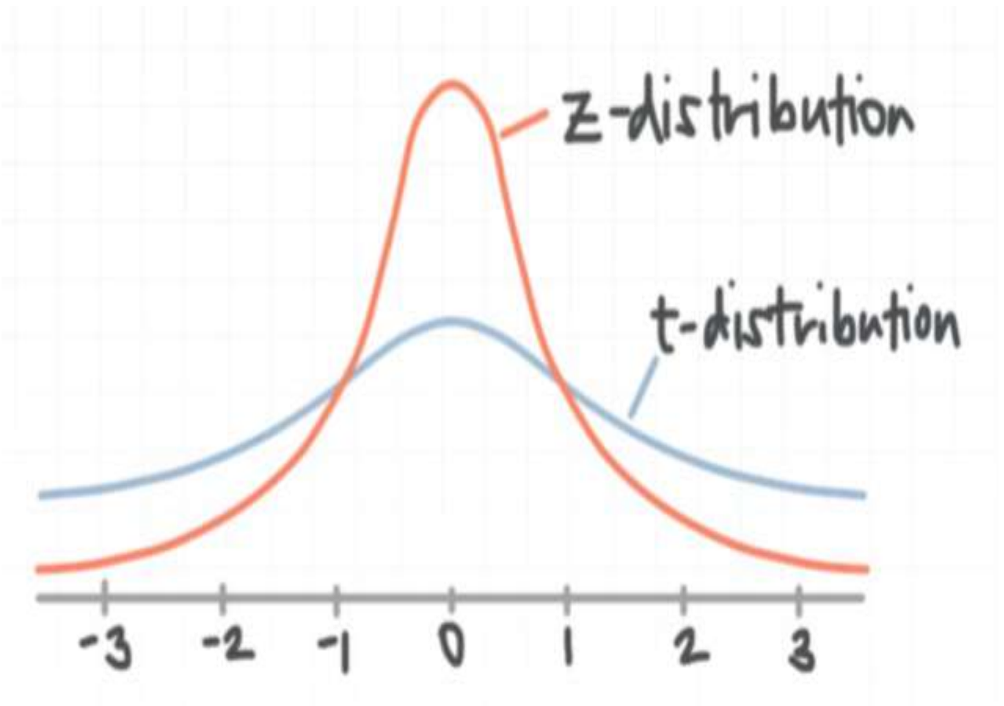
Secara bentuk, mirip dengan distribusi z, namun bentuk dari distribusi t lebih landai (*flat*) → **Simpangan baku lebih besar**

Bentuk dari distribusi t dan nilai dari distribusi t ($t-score$) tergantung dari ***degree of freedom (df)***

Nilai df adalah $n-1$, dimana n adalah ukuran sampel

Jika ukuran sampel $n = 30$ (CLT), maka distribusi t akan serupa dengan distribusi z

Best practice → Jika $n \geq 30$ gunakan z, jika tidak gunakan t



$$t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$$

Distribusi t #2

Nilai t $(t - score)$ dapat dicari dengan tabel t

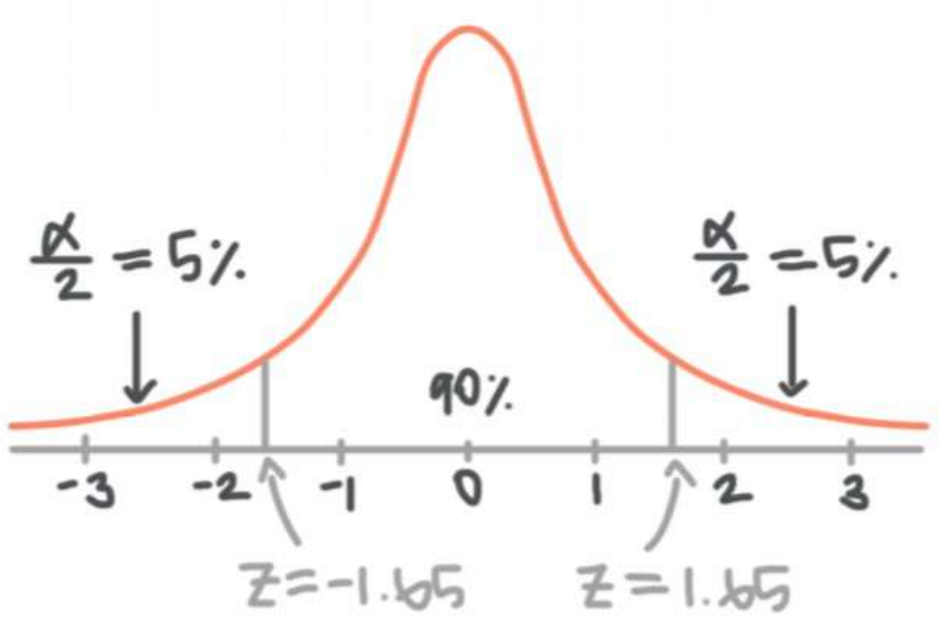
Jangan lupa, nilai n pada distribusi t adalah $n - 1$ atau nilai df

	Upper-tail probability p									
df	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.765	0.987	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence level C									

Confidence Interval #1

Bagaimana jika satu nilai dari sampel tidak bisa menggambarkan parameter populasi dengan tepat?

- Nilai yang kita cari sebelumnya, misal $\mu_{\bar{x}}$, adalah nilai *single point*.
- Nilai tersebut dapat benar dapat salah ketika digunakan untuk pengestimasi parameter μ dari populasi.
- Maka, kita bisa menggunakan estimasi dalam bentuk rentang (a, b) untuk meningkatkan kepecayaan kita dalam melakukan estimasi
- Tingkat kepecayaan disebut dengan *confidence level* atau *level of significance* di simbolkan dengan α
- Tingkat kepercayaan ini kita yang menentukan, tapi terdapat *trade-off* yang harus ditoleransi
- α yang tinggi berakibat pada interval (a, b) yang lebar



Contoh $\alpha = 90\%$

Confidence Interval #2

Nilai kepercayaan / signifikansi yang sering digunakan adalah,

- $\alpha \rightarrow 90\% \ \rightarrow z = \pm 1.65$
- $\alpha \rightarrow 95\% \ \rightarrow z = \pm 1.96$
- $\alpha \rightarrow 99\% \ \rightarrow z = \pm 2.58$

Confidence Interval #3

Nilai Confidence Interval

- $(a, b) = \bar{x} \pm z^* \sigma_{\bar{x}}, dimana\ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ →untuk sampel ≥ 30
- $(a, b) = \bar{x} \pm t^* S_{\bar{x}}, dimana\ S_{\bar{x}} = \frac{S}{\sqrt{n}}$ →untuk sampel < 30
- Nilai $z^* \sigma_{\bar{x}}$ atau $t^* S_{\bar{x}}$ disebut juga sebagai *margin of error* (*ME*)

Bagaimana jika sampel diambil dari populasi terhingga / data sampel tidak dikembalikan?

- $ME = z^* \sigma_{\bar{x}} * \sqrt{\frac{N-n}{N-1}}$ untuk sampel ≥ 30
- $ME = t^* S_{\bar{x}} * \sqrt{\frac{N-n}{N-1}}$ untuk sampel < 30

Confidence Interval #4

Nilai Confidence Interval

- $(a, b) = \bar{x} \pm z^{*}\sigma_{\bar{x}}, dimana\ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ →untuk sampel ≥ 30
- $(a, b) = \bar{x} \pm t^{*}S_{\bar{x}}, dimana\ S_{\bar{x}} = \frac{S}{\sqrt{n}}$ →untuk sampel < 30
- Nilai $z^{*}\sigma_{\bar{x}}$ atau $t^{*}S_{\bar{x}}$ disebut juga sebagai *margin of error* (*ME*)

Bagaimana jika sampel diambil dari populasi terhingga / data sampel tidak dikembalikan?

- $ME = z^{*}\sigma_{\bar{x}} * \sqrt{\frac{N-n}{N-1}}$ untuk sampel ≥ 30
- $ME = t^{*}S_{\bar{x}} * \sqrt{\frac{N-n}{N-1}}$ untuk sampel < 30

Confidence Interval #5

Contoh Kasus

Nilai mean ujian dari data **sampel terdiri dari 10 orang** mahasiswa yang dipilih secara acak adalah 86.7 dengan simpangan baku sampel adalah 5.72. Tentukan *confidence interval* dari nilai mean dengan nilai kepercayaan adalah 99%

Analsis Soal

- Dikarenakan sampel < 30 dan kita tidak tahu simpangan baku polulasi,
- Maka kita gunakan distribusi t. Nilai n adalah *degree of freedom* (df).
- Sehingga, $df = n - 1 = 10 - 1 = 9$.
- Cek nilai t dengan $df = 9$ dan $confidence\ level$ 99%

	Upper-tail probability p									
df	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence level C									

Nilai $t = 3.250$, maka nilai confidence interval,

$$(a, b) = \bar{x} \pm t^* \frac{S}{\sqrt{n}} = 86.7 \pm 3.250 * \frac{5.72}{\sqrt{10}}$$

$$(a, b) \approx 86.72 \pm 5.88 \rightarrow (80.82, 92.58)$$

Latihan

- Sebuah mesin pengisi botol minuman memiliki simpangan baku 1 oz. Kita akan mengambil sampel sejumlah 100 botol untuk mengetahui apakah rata-rata botol tersisi sebanyak 16 oz. Berapa confidence interval dari rata-rata jika kita menggunakan tingkat signifikansi 90%?


thank you
danke
謝謝
ngiyabonga
teşekkür ederim
gracias
спасибо
dank je
tapadh leat
bedankt
dziękuje
mochchakkeram
go raibh maith agat
obrigado
sukriya
kop khun krap
grazie
arigatō
takk
dakujem
merci
terima kasih
감사합니다
ありがとう
мерси
хвала
Баярлалаа
рахмат
mauruuru
hvala
sagolun
faafetai lava
kiitos
dankie
blagodaram
misaotra
matondo
paldies
mahalo
asante
manana
tenki
murakoze
mamnun
djiere dieuf
chnorakaloutioun
gracies
sulpay
taiku
trugarez
shukriya
dhanyavadagalu
diolch
xiexie
tanemirt
rahmet
najis tuke
kam sah hamnida
didi madloba
mesi
sobodi
dekuji
enkosi
bayarlalaa
gracie
koszonom
nandri
nanni
dhanyavad
vinaka
mersi
barka
welalin
tack
spas
kia ora
akun
dankon
aciu
ευχαριστώ
merca
tau