Estimasi Titik dan Selang Kepercayaan untuk μ – Distribusi t & Estimasi Mean 2 Populasi

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)

Estimasi Mean

- Distribusi sampling $\bar{X}$ berpusat di μ, dan dalam kebanyakan aplikasi varians lebih kecil daripada estimator μ lainnya. Dengan demikian, mean sampel $\bar{x}$ akan digunakan sebagai estimasi titik untuk mean populasi μ. Ingat bahwa $\sigma_{\bar{X}}^2 = \sigma^2/n$, maka sampel yang besar akan menghasilkan nilai $\bar{X}$ yang berasal dari distribusi sampling dengan varians yang kecil. Oleh karena itu, $\bar{x}$ kemungkinan menjadi estimator yang sangat akurat untuk μ ketika n besar.
- Mari kita perhatikan estimasi interval dari μ. Jika sampel kita dipilih dari populasi normal atau, jika gagal, jika n cukup besar, kita dapat menetapkan selang kepercayaan untuk μ dengan mempertimbangkan distribusi sampling $\bar{X}$. Menurut Teorema Limit Pusat, kita dapat mengharapkan distribusi sampling $\bar{X}$ mendekati terdistribusi normal dengan mean $\mu_{\bar{X}} = \mu$ dan deviasi standar $\sigma_{\bar{X}} = \sigma/\sqrt{n}$

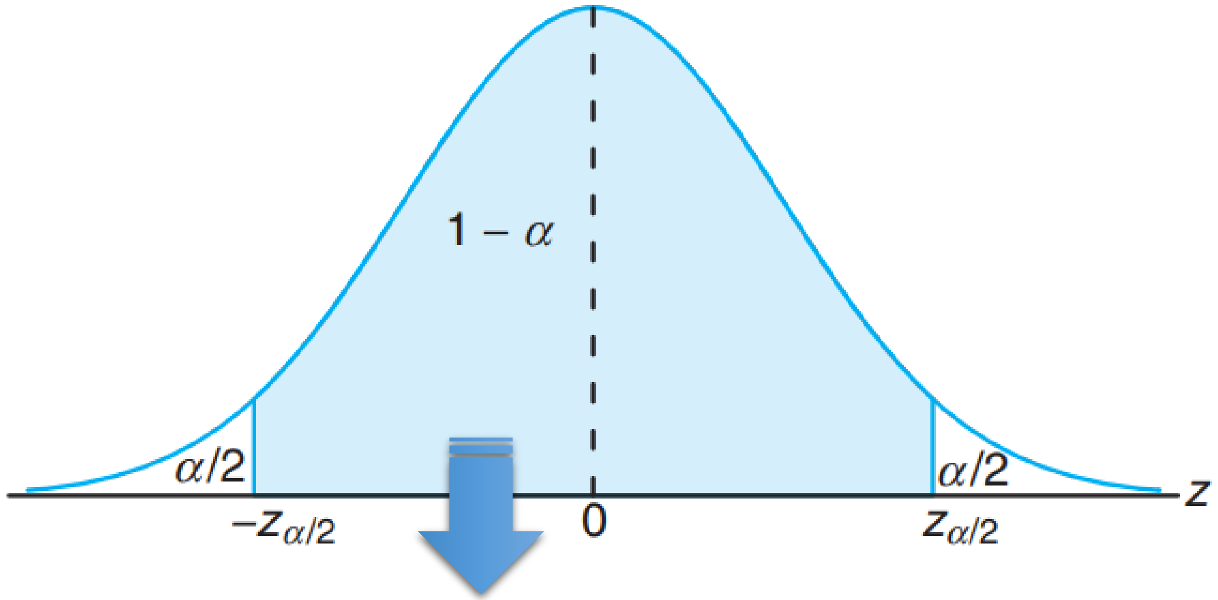
$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$$

dengan

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

sehingga

$$P\left(-z_{\alpha/2} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) = 1 - \alpha \quad \Rightarrow \quad P\left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$



$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$$

Estimasi Interval Rata-rata

1	Deviasi standar (σ) diketahui (sampel diambil dari populasi berdistribusi normal atau sampel besar)	
	$\bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$	Upper one-sided bound: $\bar{x} + z_{\alpha} \frac{\sigma}{\sqrt{n}}$ Lower one-sided bound: $\bar{x} - z_{\alpha} \frac{\sigma}{\sqrt{n}}$
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 ($n < 30$)	
	$\bar{x} - t_{\frac{\alpha}{2},(n-1)} \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\frac{\alpha}{2},(n-1)} \frac{s}{\sqrt{n}}$	Upper and lower one-sided bound: $\bar{x} + t_{\alpha,(n-1)} \frac{s}{\sqrt{n}}$ and $\bar{x} - t_{\alpha,(n-1)} \frac{s}{\sqrt{n}}$

Selang kepercayaan untuk μ

Sampling (pengambilan sampel) diasumsikan dari populasi yang berdistribusi normal

σ diketahui?

ya

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

tidak

***n* besar?**

ya

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

tidak

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

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

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

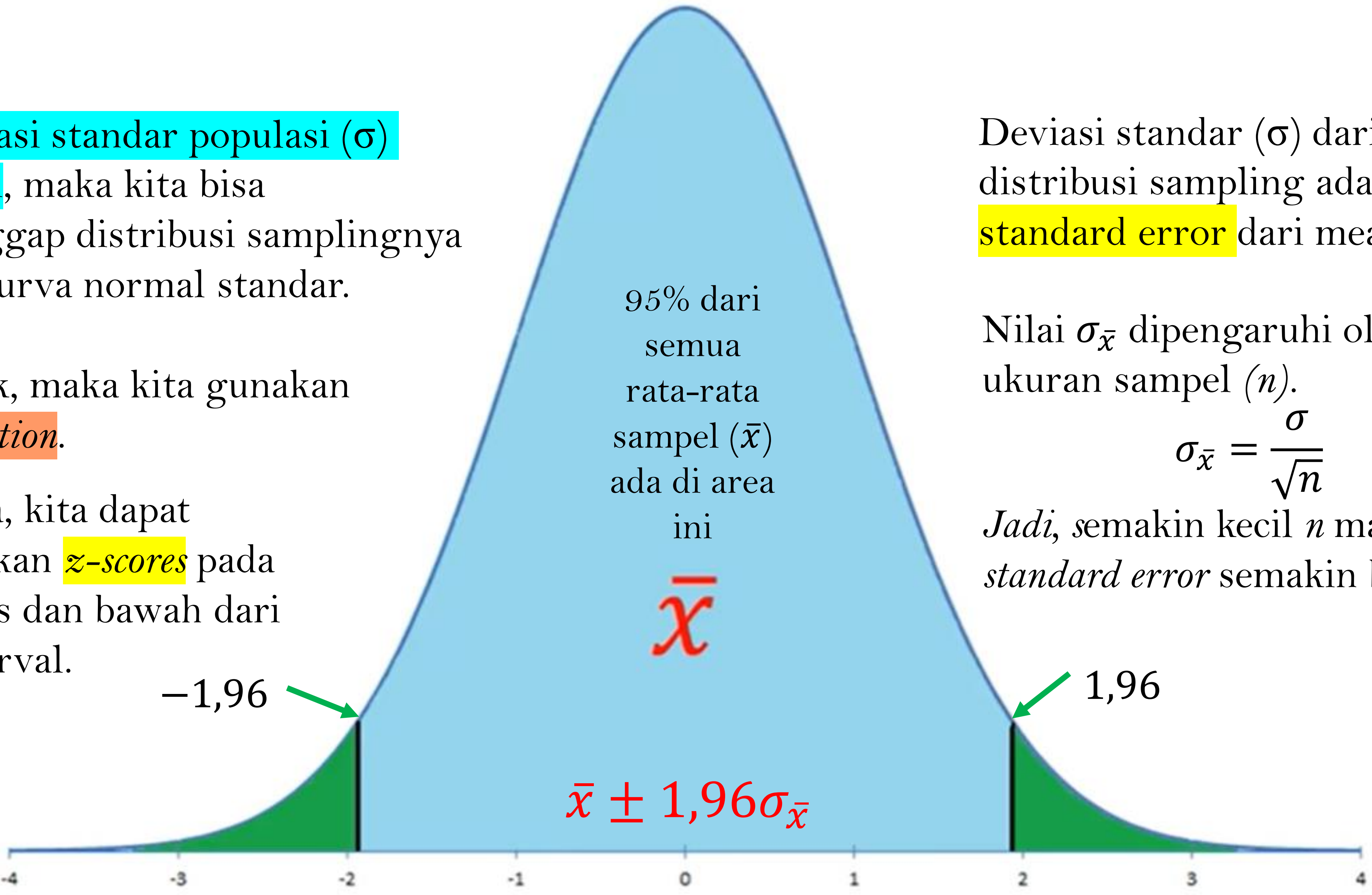
Sehingga, kita dapat menetapkan *z-scores* pada batas atas dan bawah dari 95% interval.

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

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

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

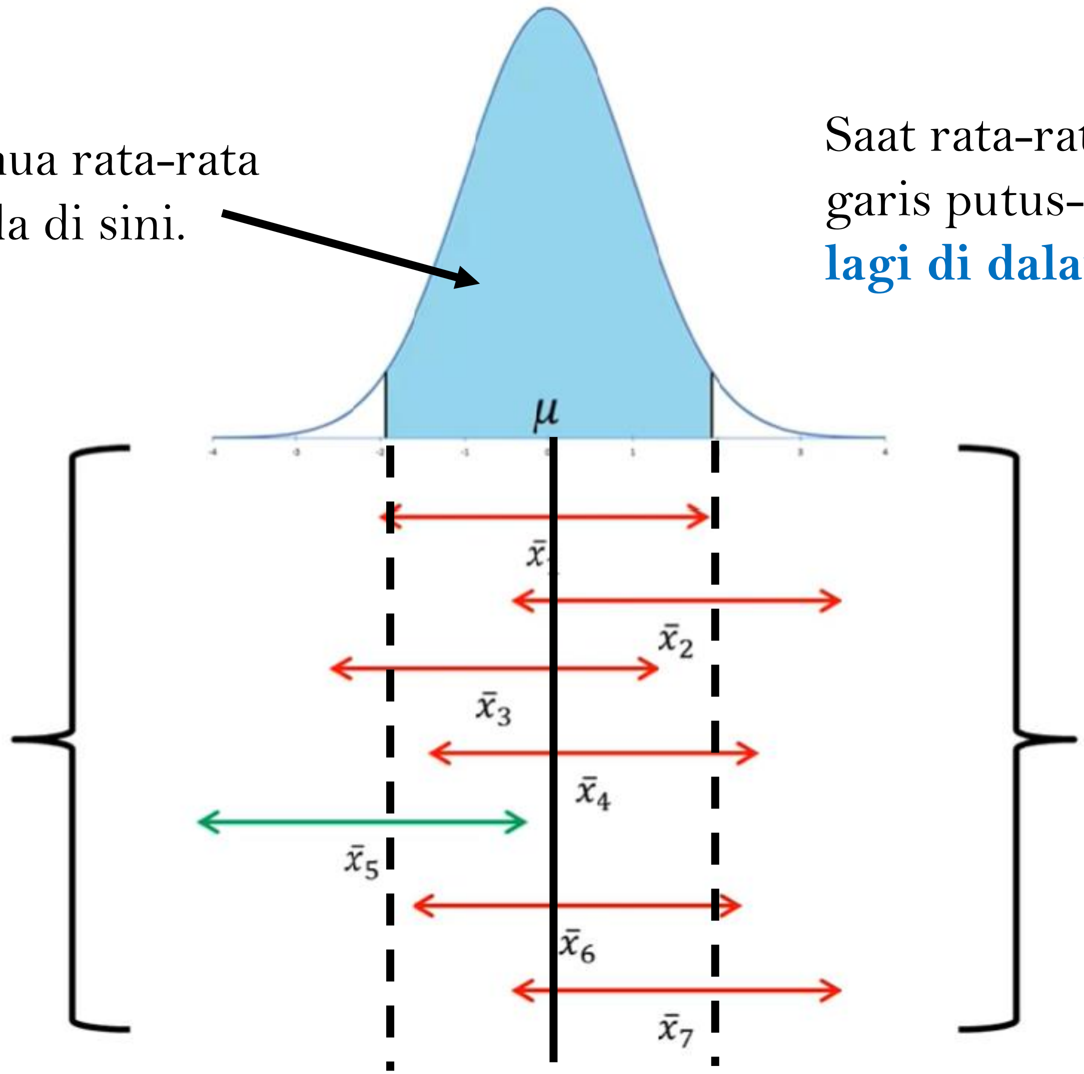
Jadi, semakin kecil *n* maka *standard error* semakin besar.



https://youtu.be/BQ88ni4bJNA

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.**



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.

https://www.youtube.com/watch?v=9GtaIHFuEZU

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

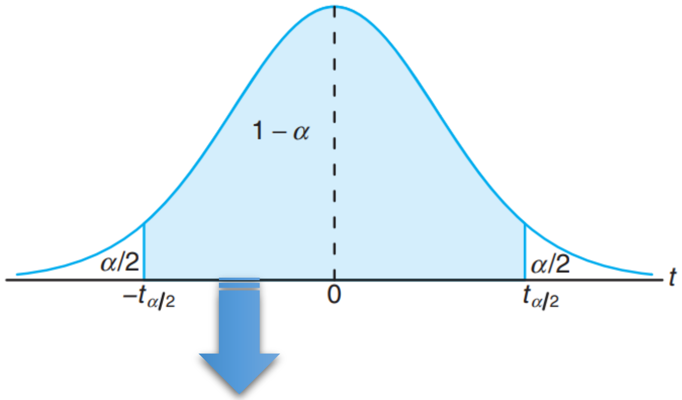
Seringkali, kita harus mencoba memperkirakan rata-rata suatu populasi ketika variansnya tidak diketahui. Jika kita memiliki sampel acak dari distribusi normal, maka variabel acak

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

memiliki distribusi-t Student dengan $\boldsymbol{n-1}$ derajat kebebasan dan S adalah deviasi standar sampel.

Dalam situasi ini, dengan σ tidak diketahui, T dapat digunakan untuk membangun interval kepercayaan pada μ. Prosedurnya sama dengan yang σ diketahui kecuali bahwa σ diganti dengan s dan distribusi normal standar diganti dengan distribusi-t.

$$P\left(\bar{X} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha/2}\frac{s}{\sqrt{n}}\right) = 1 - \alpha$$



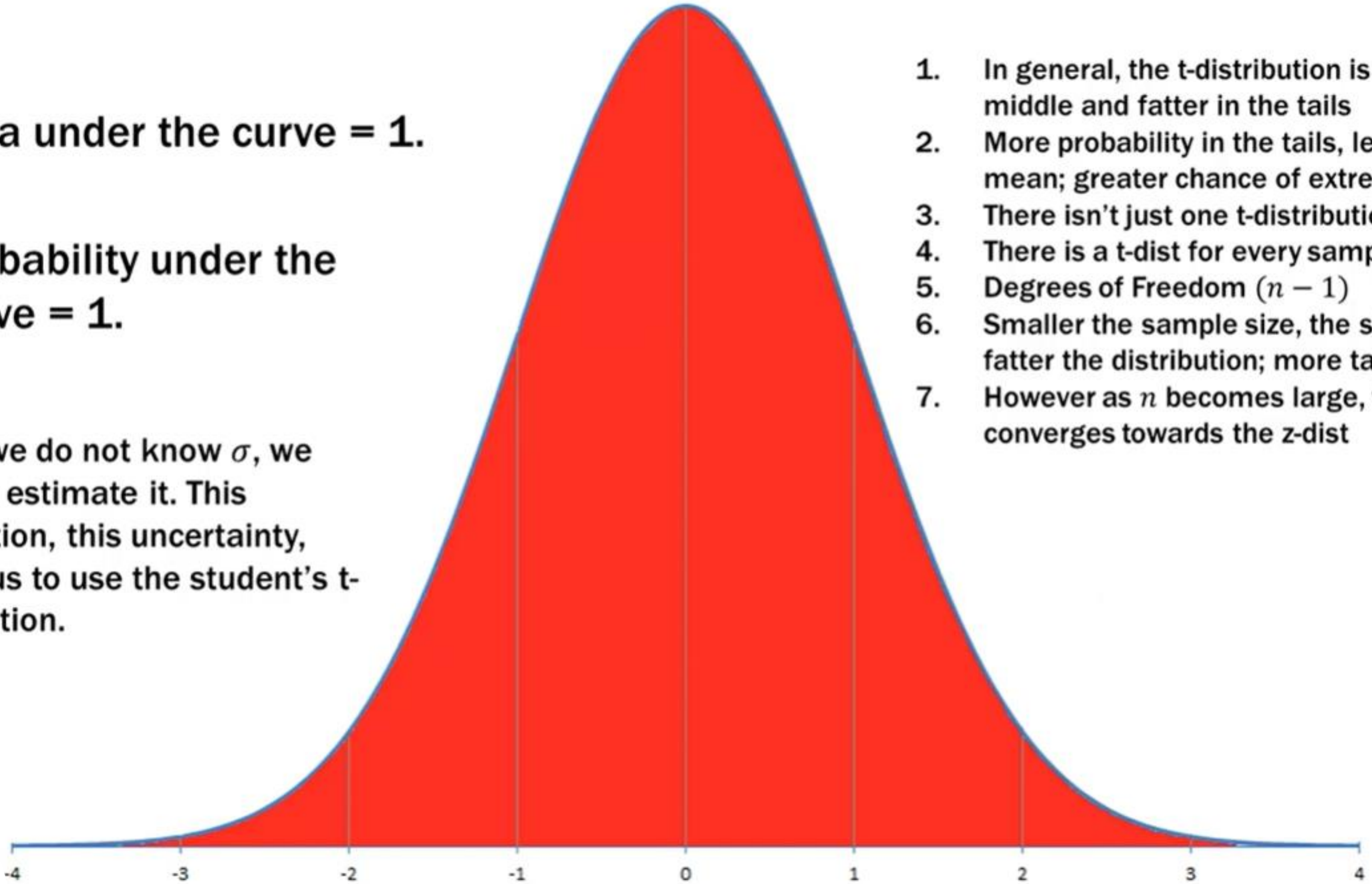
$$P\left(-t_{\alpha/2} < \mathrm{T} < t_{\alpha/2}\right) = 1 - \alpha$$

The Student's T-distribution

Area under the curve = 1.

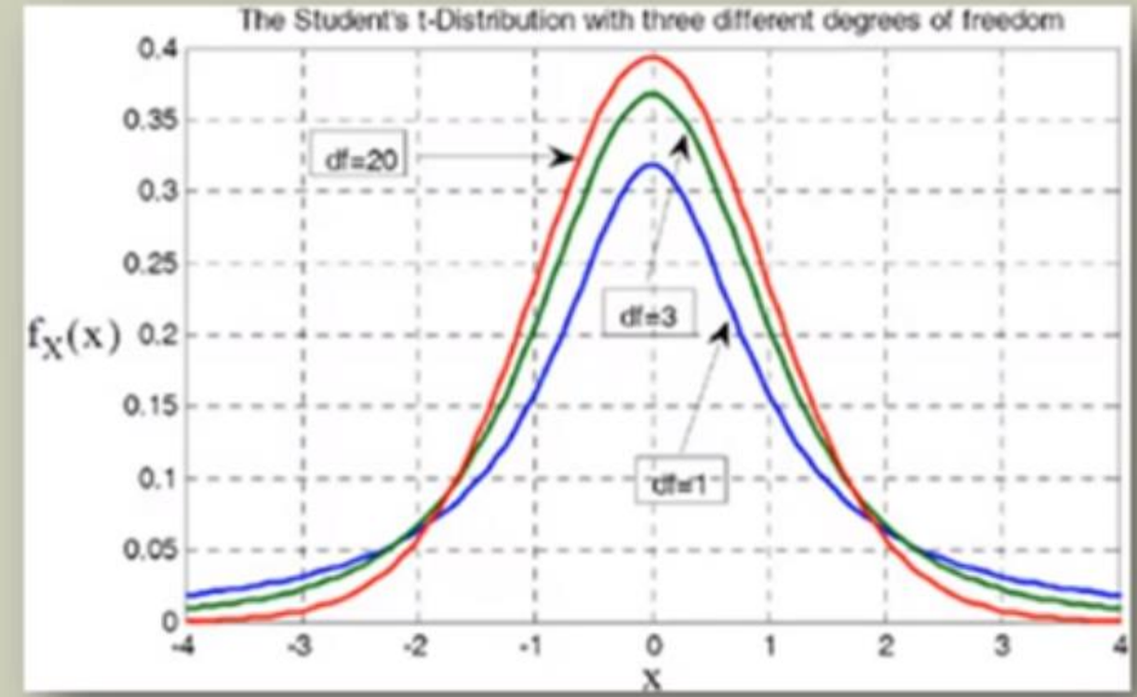
Probability under the curve = 1.

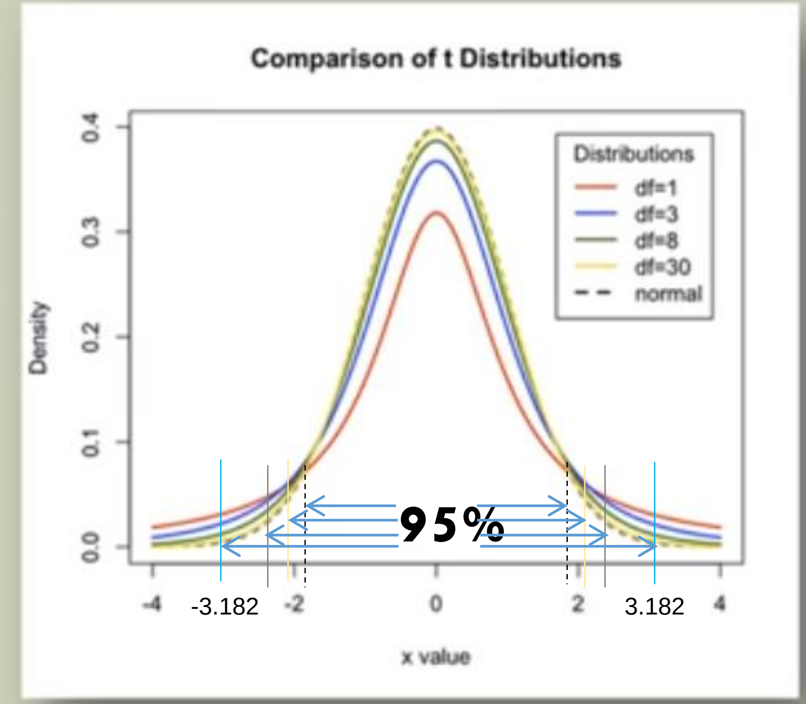
When we do not know σ, we have to estimate it. This estimation, this uncertainty, forces us to use the student's t-distribution.



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

https://youtu.be/BQ88ni4bJNA





Greater uncertainty in tails due to small sample sizes.

As n increases, the curves become tall and pointy.

Note the $df = 30$ and "normal" on the right; almost overlap

https://www.youtube.com/watch?v=Kzqm8F9Le_4

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

Contoh 1:

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…?

Estimasi Interval μ, σ tidak diketahui

Contoh 2 (sampel besar):

Nilai tes kemampuan Bahasa Inggris dari sampel acak 500 calon mahasiswa UNAIR dikumpulkan, dan mean sampel dan deviasi standar sampel masing-masing adalah 501 dan 112. Dapatkan selang kepercayaan 99% dari mean nilai tes kemampuan Bahasa Inggris untuk calan mahasiswa UNAIR.

$$\bar{x} \pm Z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$

$$= 501 \pm Z_{\frac{0.01}{2}} \frac{112}{\sqrt{500}}$$

$$= 501 \pm 2{,}576 \frac{112}{\sqrt{500}}$$

$$= [488{,}1; 513{,}9]$$

Interpretasi…?

Atau $488{,}1 < \mu < 513{,}9$

Jika n **ditambah**, *standard error* dari $\bar{X}$ akan berkurang dan ketidakpastian tentang paramater yang diestimasi akan **menyempit**

Pilih mana ??

Jika ukuran sampel **tidak dapat ditingkatkan**, tetapi masih **ingin** selang kepercayaan yang lebih **sempit**, maka **confidence level diturunkan**

Semakin besar *confidence level*, semakin lebar interval (yang lainnya tetap)

Semakin besar ukuran sampel, semakin sempit interval (yang lainnya tetap)

$$\bar{x} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = 122 \pm Z_{\frac{0.05}{2}} \frac{20}{\sqrt{25}} = 122 \pm 1.96 \frac{20}{\sqrt{25}} = 122 \pm 7.84 = [114.16;\ 129.84]$$

95% CI untuk μ

$$\bar{x} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = 122 \pm Z_{\frac{0.2}{2}} \frac{20}{\sqrt{25}} = 122 \pm 1.28 \frac{20}{\sqrt{25}} = 122 \pm 5.12 = [116.88; 127.12]$$

80% CI untuk μ

$$\bar{x} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = 122 \pm Z_{\frac{0.2}{2}} \frac{20}{\sqrt{25}} = 122 \pm 1.28 \frac{20}{\sqrt{25}} = 122 \pm 5.12 = [116.88; 127.12]$$

***n* = 25**

$$\bar{x} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = 122 \pm Z_{\frac{0.2}{2}} \frac{20}{\sqrt{2500}} = 122 \pm 1.28 \frac{20}{\sqrt{2500}} = 122 \pm 0.512 = [121.49; 122.51]$$

***n* = 2500**

Jika peneliti ingin keyakinan yang tinggi (misal 95%) **dan** selang kepercayaan yang sempit, peneliti harus mengambil **sampel lebih besar**

Latihan Soal

1. Untuk tujuan periklanan, Dewan Industri Daging Sapi perlu memperkirakan kandungan kalori rata-rata dari potongan steak *top loin* seberat 3 ons. Sampel acak dari 400 potong memberikan rata-rata sampel 212 kalori dan sampel deviasi standar 38 kalori. Berikan interval kepercayaan 95% dan 98% untuk kandungan kalori rata-rata potongan 3 ons steak *top loin*. Bandingkan hasilnya.
2. Seorang pedagang seni ingin memperkirakan nilai rata-rata suatu karya seni pada periode dan jenis tertentu. Sampel acak dari 20 karya seni dinilai. Rata-rata sampel adalah $ 5.139 dan deviasi standar sampel $ 640. Beri interval kepercayaan 95% untuk nilai rata-rata semua karya seni jenis ini.
3. Sebuah mesin menghasilkan potongan logam yang berbentuk silinder. Sampel dari potongan-potongan ini diambil dan diameternya ditemukan menjadi 1,01, 0,97, 1,03, 1,04, 0,99, 0,98, 0,99, 1,01, dan 1,03 cm. Asumsikan distribusi mendekati normal. Rata-rata sampel dan deviasi standar untuk data yang diberikan adalah $\bar{x} = 1{,}0056$ dan $s = 0{,}0246$. Carilah selang kepercayaan 99% pada diameter rata-rata.

Latihan Soal

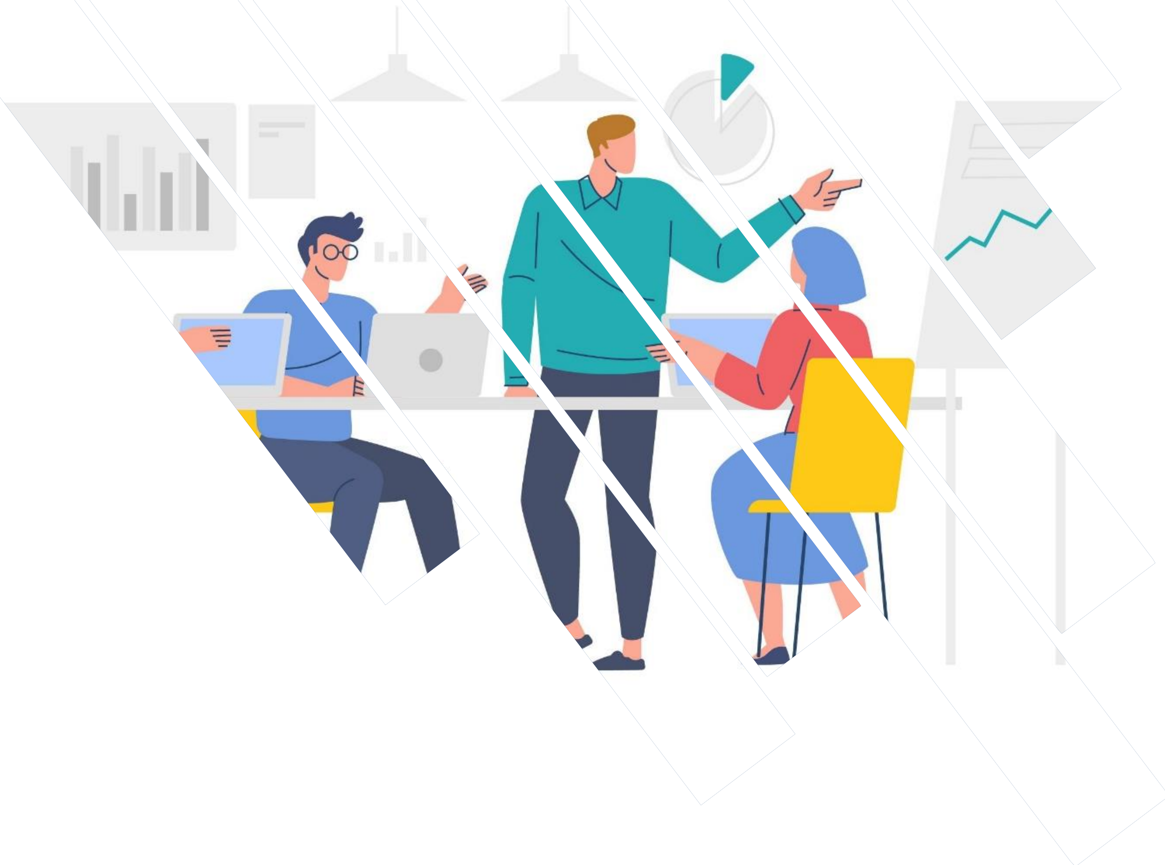
4. Harga berbagai beras tertentu, per kilogram, yang dikumpulkan dari 48 toko lokal di daerah pinggiran kota bervariasi dengan rata-rata Rp 10.500 dan standar deviasi Rp 3.000. Buatlah selang kepercayaan 95% untuk mean harga.
5. Pengukuran berikut dicatat untuk waktu pengeringan, dalam jam, dari merek cat lateks tertentu:

   3,4 2,5 4,8 2,9 3,6

   2,8 3,3 5,6 3,7 2,8

   4,4 4,0 5,2 3,0 4,8

   Dengan asumsi bahwa pengukuran mewakili sampel acak dari populasi normal, temukan interval kepercayaan 95% untuk waktu pengeringan untuk percobaan cat berikutnya.
6. Sebuah sampel acak dari 12 lulusan sekolah sekretaris tertentu mengetik rata-rata 79,3 kata per menit dengan standar deviasi 7,8 kata per menit. Dengan asumsi distribusi normal untuk jumlah kata yang diketik per menit, temukan interval kepercayaan 95% untuk jumlah rata-rata kata yang diketik oleh semua lulusan sekolah ini.



Estimasi Interval untuk 2-Sampel

2-Sampel: Estimasi Selisih Dua Mean

Jika $\bar{x}_1$ dan $\bar{x}_2$ adalah mean sampel random independen berukuran n_1 dan n_2 dari populasi dengan variansi yang diketahui masing-masing σ_1^2 dan σ_2^2, $100(1-\alpha)\%$ *confidence interval* untuk $\mu_1 - \mu_2$ didapatkan dari

$$(\bar{x}_1 - \bar{x}_2) - z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

dengan $Z_{\alpha/2}$ adalah nilai-*z* yang meninggalkan luas $\frac{\alpha}{2}$ ke kanan.

Interval kepercayaan untuk $\mu_1 - \mu_2$, σ_1^2 dan σ_2^2 diketahui

---

Jika $\bar{x}_1$ dan $\bar{x}_2$ adalah mean sampel random independen berukuran n_1 dan n_2 dari populasi mendekati normal dengan variansi tidak diketahui tetapi sama, $100(1-\alpha)\%$ *confidence interval* untuk $\mu_1 - \mu_2$ didapatkan dari

$$(\bar{x}_1 - \bar{x}_2) - t_{\frac{\alpha}{2}} s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\frac{\alpha}{2}} s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

dimana s_p adalah *pooled estimate* dari deviasi standar populasi dan $t_{\alpha/2}$ adalah nilai-*t* dengan derajat bebas $v = n_1 + n_2 - 2$, meninggalkan luas $\frac{\alpha}{2}$ ke kanan.

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Interval kepercayaan untuk $\mu_1 - \mu_2$, σ_1^2 dan σ_2^2 tidak diketahui, $\sigma_1^2 = \sigma_2^2$

CONTOH 3

Sebuah studi dilakukan di mana dua jenis mesin, A dan B, dibandingkan. Jarak tempuh gas, dalam mil per galon, diukur. Lima puluh percobaan dilakukan dengan menggunakan mesin tipe A dan 75 percobaan dilakukan dengan mesin tipe B. Bensin yang digunakan dan kondisi lainnya dipertahankan konstan. Jarak tempuh rata-rata gas adalah 36 mil per galon untuk mesin A dan 42 mil per galon untuk mesin B. Carilah selang kepercayaan 96% pada $\mu_B - \mu_A$, di mana μ_A dan μ_B adalah rata-rata populasi jarak tempuh gas untuk mesin A dan B, masing-masing. Asumsikan bahwa simpangan baku populasi masing-masing adalah 6 dan 8 untuk mesin A dan B.

Jawab:

$$(\bar{x}_B - \bar{x}_A) - z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_B^2}{n_B} + \frac{\sigma_A^2}{n_A}} < \mu_B - \mu_A < (\bar{x}_B - \bar{x}_A) + z_{\alpha/2}\sqrt{\frac{\sigma_B^2}{n_B} + \frac{\sigma_A^2}{n_A}}$$

$$(42 - 36) - z_{0,02}\sqrt{\frac{8^2}{75} + \frac{6^2}{50}} < \mu_B - \mu_A < (42 - 36) + z_{0,02}\sqrt{\frac{8^2}{75} + \frac{6^2}{50}}$$

$$6 - 2,05\sqrt{\frac{64}{75} + \frac{36}{50}} < \mu_B - \mu_A < 6 + 2,05\sqrt{\frac{64}{75} + \frac{36}{50}}$$

$$3,43 < \mu_B - \mu_A < 8,57$$

CONTOH 4

Dua stasiun pengambilan sampel independen dipilih untuk penelitian indeks keanekaragaman spesies makroinvertebrata untuk menunjukkan degradasi perairan karena drainase asam tambang, satu terletak di hilir dari titik pembuangan asam tambang dan yang lainnya terletak di hulu. Untuk 12 sampel bulanan yang dikumpulkan di stasiun hilir, indeks keanekaragaman spesies memiliki nilai rata-rata $\bar{x}_1 = 3{,}11$ dan standar deviasi $s_1 = 0{,}771$, sedangkan 10 sampel bulanan yang dikumpulkan di stasiun hulu memiliki nilai indeks rata-rata $\bar{x}_2 = 2{,}04$ dan simpangan baku $s_2 = 0{,}448$. Carilah selang kepercayaan 90% untuk perbedaan antara rata-rata populasi untuk dua lokasi, dengan asumsi bahwa populasi berdistribusi hampir normal dengan variansi yang sama.

Jawab: $\bar{x}_1 - \bar{x}_2 = 3{,}11 - 2{,}04 = 1{,}07$; $t_{0,1/2} = 1{,}725$ dengan derajat bebas $v = 12 + 10 - 2 = 20$

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{(11)(0{,}771)^2 + (9)(0{,}448)^2}{12 + 10 - 2} = 0{,}417 \;;\; s_p = 0{,}646$$

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$1{,}07 - 1{,}725(0{,}646)\sqrt{\frac{1}{12} + \frac{1}{10}} < \mu_1 - \mu_2 < 1{,}07 + 1{,}725(0{,}646)\sqrt{\frac{1}{12} + \frac{1}{10}}$$

$$0{,}593 < \mu_1 - \mu_2 < 1{,}547$$

2-Sampel: Estimasi Selisih Dua Mean

Interval kepercayaan untuk $\mu_1 - \mu_2$, σ_1^2 dan σ_2^2 tidak diketahui, $\sigma_1^2 \neq \sigma_2^2$

Jika $\bar{x}_1$ dan $\bar{x}_2$ adalah mean sampel random independen berukuran n_1 dan n_2 dari populasi mendekati normal dengan variansi tidak diketahui dan tidak sama, pendekatan $100(1-\alpha)\%$ *confidence interval* untuk $\mu_1 - \mu_2$ didapatkan dari

$$(\bar{x}_1 - \bar{x}_2) - t_{\frac{\alpha}{2}}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\frac{\alpha}{2}}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

dimana $t_{\alpha/2}$ adalah nilai-*t* dengan derajat bebas

$$v = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\left[\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1}\right] + \left[\frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}\right]}$$

meninggalkan luas $\frac{\alpha}{2}$ ke kanan.

Interval kepercayaan untuk $\mu_D = \mu_1 - \mu_2$, untuk observasi berpasangan

Jika $\bar{d}$ dan s_d masing-masing adalah mean dan deviasi standar berdistribusi secara normal selisih dari n pengukuran berpasangan acak, $100(1-\alpha)\%$ *confidence interval* untuk $\mu_D = \mu_1 - \mu_2$ didapatkan dari

$$\bar{d} - t_{\frac{\alpha}{2}}\frac{s_d}{\sqrt{n}} < \mu_D < \bar{d} + t_{\frac{\alpha}{2}}\frac{s_d}{\sqrt{n}}$$

dimana $t_{\alpha/2}$ adalah nilai-*t* dengan derajat bebas $v = n - 1$, meninggalkan luas $\frac{\alpha}{2}$ ke kanan.

CONTOH 5

Sebuah studi dilakukan oleh Departemen Ilmu Biologi di Virginia Tech untuk memperkirakan perbedaan jumlah ortofosfat kimia yang diukur di dua stasiun berbeda di Sungai James. Ortofosfat diukur dalam miligram per liter. Lima belas sampel diambil dari stasiun 1, dan 12 sampel diperoleh dari stasiun 2. Lima belas sampel dari stasiun 1 memiliki kandungan ortofosfat rata-rata 3,84 miligram per liter dan deviasi standar 3,07 miligram per liter, sedangkan 12 sampel dari stasiun 2 memiliki kandungan rata-rata 1,49 miligram per liter dan deviasi standar 0,80 miligram per liter. Carilah selang kepercayaan 95% untuk perbedaan kandungan ortofosfat rata-rata sebenarnya pada dua stasiun ini, dengan asumsi bahwa pengamatan berasal dari populasi normal dengan variansi yang berbeda.

Jawab:

$$v = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{[(s_1^2/n_1)^2/(n_1-1)] + [(s_2^2/n_2)^2/(n_2-1)]}$$
$$= \frac{(3{,}07^2/15 + 0{,}80^2/12)^2}{[(3{,}07^2/15)^2/(15-1)] + [(0{,}80^2/12)^2/(12-1)]}$$
$$= 16{,}3 \approx 16$$

$$\bar{x}_1 - \bar{x}_2 = 3{,}84 - 1{,}49 = 2{,}35$$
$$t_{\left(\frac{0{,}05}{2}\right),16} = 2{,}120$$

selang kepercayaan 95% untuk $\mu_1 - \mu_2$

$$2{,}35 - 2{,}120\sqrt{\frac{3{,}07^2}{15} + \frac{0{,}80^2}{12}} < \mu_1 - \mu_2 < 2{,}35 + 2{,}120\sqrt{\frac{3{,}07^2}{15} + \frac{0{,}80^2}{12}}$$
$$0{,}60 < \mu_1 - \mu_2 < 4{,}10$$

Oleh karena itu, kami yakin 95% bahwa interval dari 0,60 hingga 4,10 miligram per liter mengandung perbedaan kandungan ortofosfat rata-rata sebenarnya untuk kedua lokasi ini.

CONTOH 6

Sebuah studi yang diterbitkan di *Chemosphere* melaporkan tingkat dioksin TCDD dari 20 veteran Massachusetts Vietnam yang mungkin terpapar *Agent Orange*. Tingkat TCDD dalam plasma dan jaringan lemak tercantum dalam Tabel. Temukan interval kepercayaan 95% untuk $\mu_1 - \mu_2$, di mana μ_1 dan μ_2 masing-masing mewakili tingkat TCDD rata-rata yang sebenarnya dalam plasma dan dalam jaringan lemak. Asumsikan distribusi perbedaan mendekati normal.

Veteran	TCDD Levels in Plasma	TCDD Levels in Fat Tissue	d_i	Veteran	TCDD Levels in Plasma	TCDD Levels in Fat Tissue	d_i
1	2.5	4.9	−2.4	11	6.9	7.0	−0.1
2	3.1	5.9	−2.8	12	3.3	2.9	0.4
3	2.1	4.4	−2.3	13	4.6	4.6	0.0
4	3.5	6.9	−3.4	14	1.6	1.4	0.2
5	3.1	7.0	−3.9	15	7.2	7.7	−0.5
6	1.8	4.2	−2.4	16	1.8	1.1	0.7
7	6.0	10.0	−4.0	17	20.0	11.0	9.0
8	3.0	5.5	−2.5	18	2.0	2.5	−0.5
9	36.0	41.0	−5.0	19	2.5	2.3	0.2
10	4.7	4.4	0.3	20	4.1	2.5	1.6

Reprinted from *Chemosphere*, Vol. 20, Nos. 7-9 (Tables I and II), Schecter et al., "Partitioning 2,3,7,8-chlorinated dibenzo-p-dioxins and dibenzofurans between adipose tissue and plasma lipid of 20 Massachusetts Vietnam veterans," *pp.* 954–955, Copyright ©1990, with permission from Elsevier.

CONTOH 7

Jawab:

$\mu_1 - \mu_2 = \mu_D$, estimasi titik dari μ_D adalah $\bar{d} = -0,87$.

Deviasi standar sampel s_d adalah

$$s_d = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2} = \sqrt{\frac{168,4220}{19}} = 2,9773$$

Dengan tabel distribusi t untuk $\alpha = 0,05$ dan derajat bebas $n - 1 = 19$, $t_{0,025} = 2,093$

Selang kepercayaan 95% untuk $\mu_1 - \mu_2$ adalah

$$-0,87 - (2,093)\left(\frac{2,9773}{\sqrt{20}}\right) < \mu_D < -0,87 + (2,093)\left(\frac{2,9773}{\sqrt{20}}\right)$$

$$-2,2634 < \mu_D < 0,5234$$

dari sini kita dapat menyimpulkan bahwa **tidak ada perbedaan yang signifikan** antara tingkat TCDD rata-rata dalam plasma dan tingkat TCDD rata-rata dalam jaringan lemak (karena interval berisi nilai 0).

Sumber

- Aczel, A. and Sounderpandian, J., 2008, *Complete Business Statistics*, 7th *Edition*, McGraw-Hill/Irwin, USA.
- https://youtu.be/BQ88ni4bJNA
- https://www.youtube.com/watch?v=9GtaIHFuEZU
- https://www.youtube.com/watch?v=Kzqm8F9Le_4
- 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.