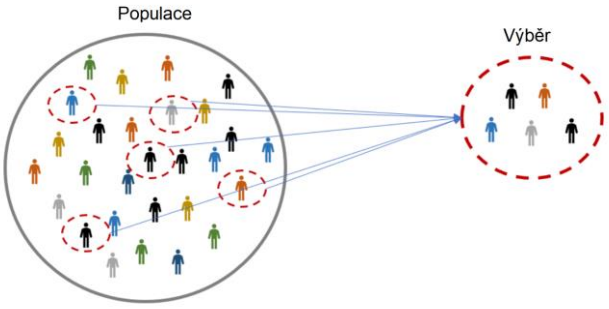
ANALYTICKÁ CHEMOMETRIE

odhady parametrů základního souboru



Patrik Kania
Ústav analytické chemie, VŠCHT Praha

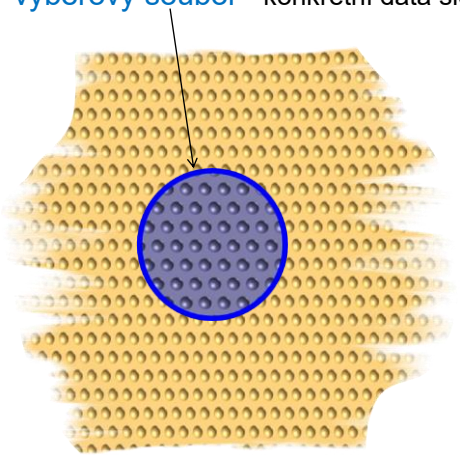
14. 10. 2025

Základní charakteristiky základního souboru:

- množina všech možných hodnot náhodné veličiny ξ
- distribuční funkce F náhodné veličiny ξ
- hustota pravděpodobnosti f nebo pravděpodobnostní funkce

- střední (očekávaná) hodnota $E(\xi) = \mu$
- rozptyl $D(\xi) = \sigma^2$
- kvantil x_p

výběrový soubor - konkrétní data sloužící k poznání základního souboru



z výběrového souboru určíme
výběrové statistiky
platící pouze pro výběr

z výběrových statistik odhadujeme
vlastnosti základního souboru (ZS)

POZOR!
neznáme celý ZS → parametry ZS mají nejistotu

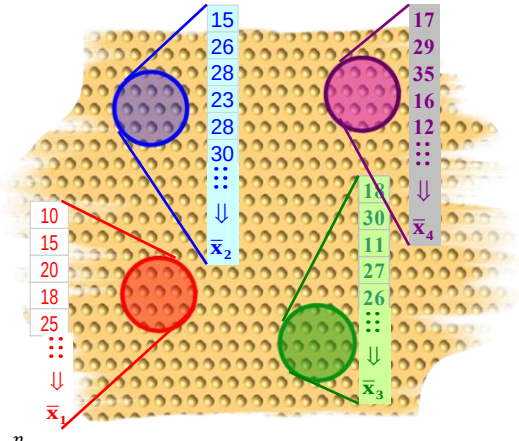
základní soubor - může být příliš velký či mít až nekonečné množství prvků a vše nelze změřit

POZOR, výběrové charakteristiky jsou náhodné veličiny!

např. výběrový průměr

náhodná veličina

$$\mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$$



důkaz:

$$E(\bar{x}) = E\left(\frac{1}{n}\sum_i^n x_i\right) = \frac{1}{n}\sum_i^n E(x_i) = \frac{1}{n}\sum_i^n \mu = \frac{1}{n}n\mu = \mu$$

$$D(\bar{x}) = D\left(\frac{1}{n}\sum_i^n x_i\right) = \frac{1}{n^2}D\left(\sum_i^n x_i\right) = \frac{1}{n^2}\sum_i^n D(x_i) = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}$$

Základní pojmy spojené s výběrovým šetřením:

Parametr - je statistická charakteristika základního souboru
(značí se řeckými písmeny, např. střední hodnota μ)

Statistika - je statistická charakteristika výběrového souboru
(značí se latinkou, např. výběrový průměr $\bar{x}$)

Odhad parametru základního souboru je určitá hodnota vhodné statistiky získaná postupem zvaným bodový nebo intervalový odhad.

Bodové odhady základních parametrů

základní vlastnosti bodových odhadů:

nespornost (konzistence)

- se vzrůstající velikostí výběru vzrůstá i pravděpodobnost, že odhad bude blízký skutečnému parametru

nevychýlenost (nestrannost)

- střední hodnota výběrové statistiky *X* se rovná odhadovanému parametru *P* základního souboru. Například pro $n \to \infty$ se $\bar{x} = \mu$.

$$E(X) = P$$

vydatnost (efektivnost)

- odhad je vydatný, pokud se jeho rozptyl okolo skutečné hodnoty parametru rychle zmenšuje s rostoucím rozsahem výběru

robustnost

- necitlivost odhadu na odchylky od předpokládaného rozdělení a na extrémy

Nejpoužívanější bodové odhady parametrů rozdělení:

VÝBĚROVÝ ARITMETICKÝ PRŮMĚR $\bar{x}$ (angl. *arithmetic mean of sample*)
=PRŮMĚR(A1:A10)

- nejlepší odhad μ (konzistentní, nestranný a vydatný)

$$\bar{x} = \frac{1}{n}\sum_i x_i$$

VÝBĚROVÝ ROZPTYL s^2 (angl. *sample variance*)
=VAR.VÝBĚR(A1:A10)

- nejlepší odhad σ^2 (konzistentní, nestranný a vydatný)

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

VÝBĚROVÁ SMĚRODATNÁ ODCHYLKA s (ang. *sample standard deviation*)
=SMODCH.VÝBĚR(A1:A10)

$$s = \sqrt{s^2}$$

SMĚRODATNÁ ODCHYLKA VÝBĚROVÉHO PRŮMĚRU či STŘEDNÍ CHYBA PRŮMĚRU $s_{\bar{x}}$ (angl. *standard error of the mean*)

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

RELATIVNÍ VÝBĚROVÁ SMĚRODATNÁ ODCHYLKA, *RSD*
(angl. *relative sample standard deviation*)

$$RSD = \frac{s}{\bar{x}}$$

VÝBĚROVÝ VARIAČNÍ KOEFICIENT (%), *CV* (angl. *coefficient of variation*)

$$CV = RSD \cdot 100$$

ZÁKLADNÍ SOUBOR

konečný

6,8	10,5	10,5	10,5
7,5	10,5	10,5	10,5
7,5	10,5	10,5	10,5
7,5	10,5	10,5	10,5
7,5	10,5	10,5	10,5
7,5	10,5	10,5	10,5
8,2	10,5	10,5	12
9,0	10,5	10,5	12
9,0	10,5	10,5	12
9,0	10,5	10,5	15
9,0	10,5	10,5	15
9,0	10,5	10,5	15
9,8	10,5	10,5	15
9,8	10,5	10,5	

$$\mu = 10{,}37$$

VÝBĚROVÝ SOUBOR

lidé s příjmením na „K“

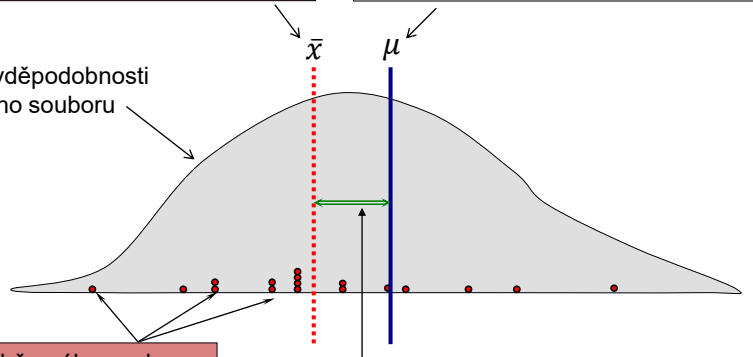
7,5
8,2
10,5
10,5
10,5
10,5
10,5
15

$$\bar{x} = 10{,}40$$

Slabá stránka bodového odhadu:

výběrový průměr - vypočítaný z červených bodů - tuto hodnotu známe

střední hodnota ZS - v praxi ji neznáme, znali bychom ji pouze, pokud bychom znali celý ZS



hustota pravděpodobnosti základního souboru

hodnoty výběrového souboru

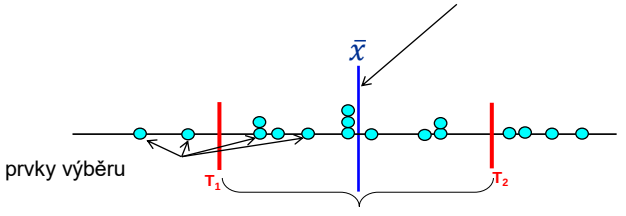
tato vzdálenost je pro jeden konkrétní výběr neznámá (protože neznáme polohu μ) a není možné tedy určit spolehlivost konkrétního odhadu (tj. míru vychýlení výběrového průměru) - proto se pro odhad parametrů ZS používají raději intervalové odhady

Intervalové odhady základních parametrů

Intervalový odhad (angl. *confidence interval*) pro parametr τ a pro hladinu významnosti α je taková dvojice statistik T_1 a T_2, že

$$P(T_1 \leq \tau \leq T_2) = 1 - \alpha$$

bodový odhad neznámé střední hodnoty μ určený z výběru
(nevíme nic o jeho vztahu ke skutečné střední hodnotě)



intervalový odhad neznámé střední hodnoty μ
předpokládáme, že s pravděpodobností $P = 1 - \alpha$
leží nám neznámé μ kdekoliv v tomto intervalu

Výhoda: intervalový odhad lze určit s pravděpodobností (spolehlivostí)

Pravděpodobnost odhadu určujeme pomocí hladiny významnosti α či hladiny spolehlivosti $1 - \alpha$.

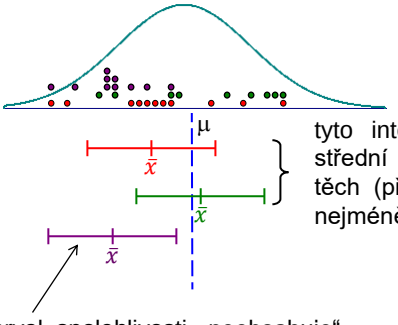
Hladina významnosti α (angl. *level of significance*)

- je pravděpodobnost („riziko“) toho, že intervalový odhad nebude skutečnou hodnotu parametru obsahovat.

Hladina spolehlivosti $(1 - \alpha)$ (angl. *level of confidence*)

- je pravděpodobnost, že intervalový odhad obsahuje skutečnou hodnotu parametru.

Nejčastější hodnotou α je 0,05 (tj. 5 %) nebo 0,01 (tj. 1 %).



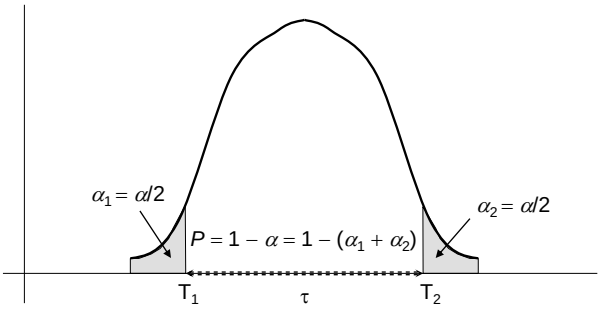
tyto intervaly „obsahují“ odhadovanou střední hodnotu (jsou tedy „správné“), těch (při opakovaných výběrech) bude nejméně $(1 - \alpha) \cdot 100$ %

tento interval spolehlivosti „neobsahuje“ střední hodnotu (je tedy „chybný“), těchto intervalů se objeví nejvýše $(100 \cdot \alpha)$ %

Oboustranný intervalový odhad parametru

(angl. *two-sided confidence interval*)

- pro odhad parametru ZS, který neporovnáváme s hraniční hodnotou
- při konstrukci intervalu musíme „rozložit“ pravděpodobnost α na obě strany, protože nevíme, kde je skutečná hodnota parametru, zda vlevo/vpravo od IS
- proto se obě hranice oboustranného IS konstruují na základě pravděpodobnosti $\alpha/2$, nikoliv α



Jednostranný intervalový odhad parametru

(ang. *one-sided confidence interval*)

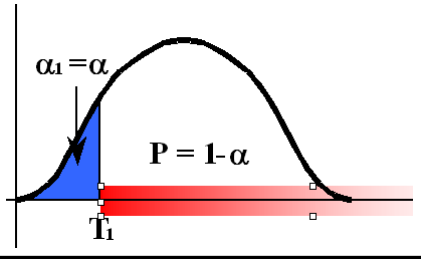
- pro odhad parametru ZS a následné porovnání s hraniční hodnotou
- na druhé straně intervalu nezáleží!
- konstrukce hranice je na základě pravděpodobnosti α

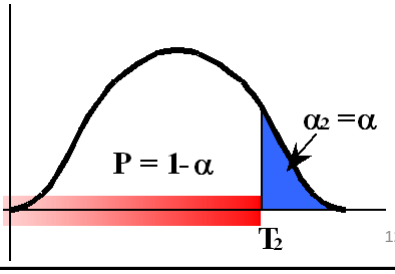
levostranný odhad

$$P(\tau > T_1) = 1 - \alpha$$

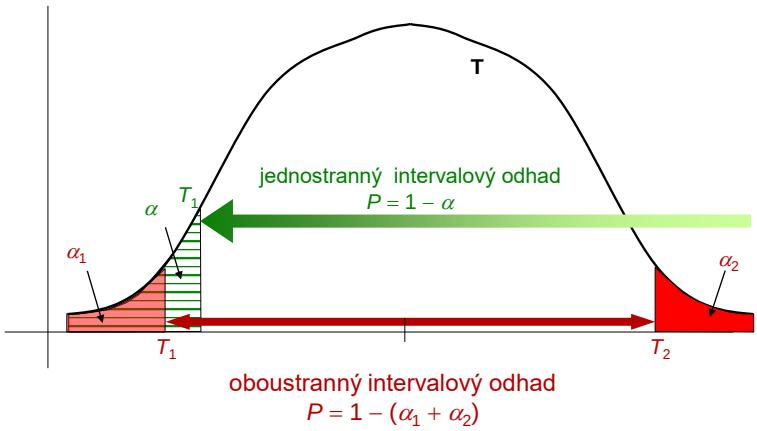
pravostranný odhad

$$P(\tau < T_2) = 1 - \alpha$$





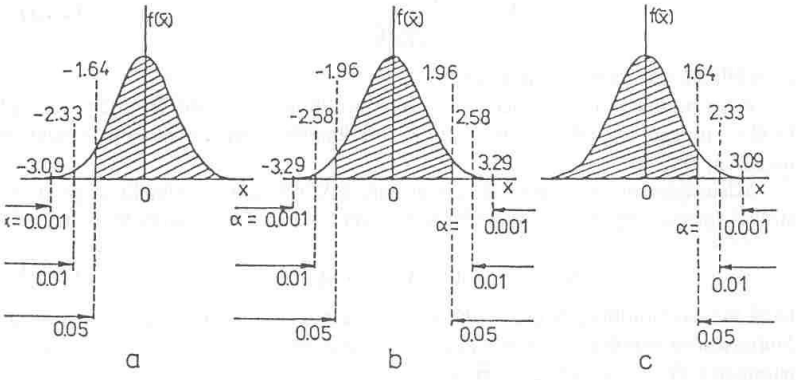
Porovnání jednostranného a oboustranného intervalového odhadu



Porovnání jednostranného a oboustranného intervalového odhadu

Kritické hodnoty normovaného normálního rozdělení pro:
$\alpha = 0{,}05$ (interval s 95% spolehlivostí)
$\alpha = 0{,}01$ (interval s 99% spolehlivostí)
$\alpha = 0{,}001$ (interval s 99,9% spolehlivostí)

pro: (a) - levostranný, (b) - oboustranný, (c) - pravostranný interval

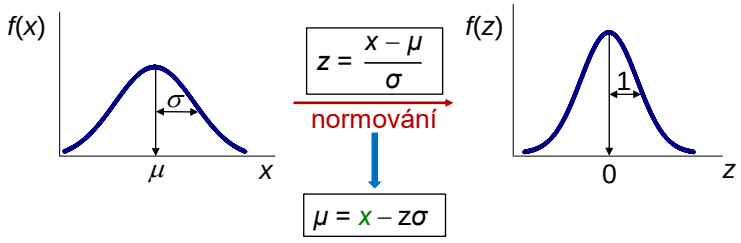


Odvození výrazu pro interval spolehlivosti střední hodnoty μ

známá směrodatná odchylka σ základního souboru nebo je používán velký výběr (> 30 prvků)

náhodná veličina ξ náhodná veličina ζ

normální rozdělení N(μ, σ^2) normované normální rozdělení N(0,1)



pro náhodnou proměnnou x v podobě průměru $\bar{x}$ mající rozdělení $\mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$

$$\mu = \bar{x} - z \cdot \frac{\sigma}{\sqrt{n}}$$

Oboustranný interval spolehlivosti střední hodnoty μ

známá směrodatná odchylka σ základního souboru nebo je používán velký výběr (> 30 prvků)

při testu hypotézy $\mu = \mu_0$ je hypotéza platná, pokud je μ_0 uvnitř intervalu

v případě velkého výběru lze použít místo σ výběrovou směrodatnou odchylku s

$$\boxed{\underbrace{\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}}_{\text{dolní hranice}} \leq \mu \leq \underbrace{\bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}}_{\text{horní hranice}}}$$

$z_{\alpha/2}$ je oboustranná kritická hodnota normovaného normálního rozdělení pro hladinu významnosti α

Oboustranný interval spolehlivosti střední hodnoty μ

neznámá směrodatná odchylka σ základního souboru a máme malý výběr (< 30 prvků)

Na následujících snímcích budou převážně probírány varianty pro malý výběr.

při testu hypotézy $\mu = \mu_0$ je hypotéza platná, pokud je μ_0 uvnitř intervalu

$$\boxed{\underbrace{\bar{x} - t_{(\frac{\alpha}{2},n-1)} \cdot \frac{s}{\sqrt{n}}}_{\text{dolní hranice}} \leq \mu \leq \underbrace{\bar{x} + t_{(\frac{\alpha}{2},n-1)} \cdot \frac{s}{\sqrt{n}}}_{\text{horní hranice}}}$$

$t_{(\alpha/2,\, n-1)}$ je oboustranná kritická hodnota Studentova rozdělení pro hladinu významnosti α a n-1 stupňů volnosti

VÝBĚROVÝ SOUBOR

lidé s příjmením na „K“

7,5
8,2
10,5
10,5
10,5
10,5
10,5
15

$\bar{x} = 10{,}40$ $s = 2{,}22$

$t_{(\frac{0{,}05}{2},8-1)} = 2{,}365$

$\mu = 10{,}37$

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

$$10{,}40 - 2{,}365 \cdot \frac{2{,}22}{\sqrt{8}} \leq \mu \leq 10{,}40 + 2{,}365 \cdot \frac{2{,}22}{\sqrt{8}}$$

$$8{,}51 \leq \mu \leq 12{,}23$$

Jednostranný interval spolehlivosti střední hodnoty μ

případ pro neznámou odchylku σ základního souboru a malý výběr (< 30 prvků)

levostranný odhad

$\langle L_1, \infty)$

$$L_1 = \bar{x} - t_{(\alpha,n-1)} \cdot \frac{s}{\sqrt{n}}$$

pokud je číslo μ_0 uvnitř intervalu, platí hypotéza $\mu \leq \mu_0$

pravostranný odhad

$(-\infty, L_2\rangle$

$$L_2 = \bar{x} + t_{(\alpha,n-1)} \cdot \frac{s}{\sqrt{n}}$$

pokud je číslo μ_0 uvnitř intervalu, platí hypotéza $\mu \geq \mu_0$

$t_{(\alpha, n-1)}$ je jednostranná kritická hodnota Studentova rozdělení pro hladinu významnosti α a n-1 stupňů volnosti

Příklad: Obsah nitroglycerinu ve farmaceutickém roztoku má podle normy obsahovat maximálně 2,052 %. Z právě vyrobené šarže tohoto roztoku byly náhodným výběrem a následnou analýzou získány tyto výsledky [%]:

2,07 2,04 2,03 2,05 2,06 2,04 2,12 2,04 2,05 2,07

Otestujte na hladině významnosti 5 %, zda vyrobená šarže splňuje normu.

$\bar{x}$ = 2,057 s = 0,0258 n = 10 $t_{0,05;9}$ = 1,833

=TINV(0,10;9)

levostranný odhad

$\langle L_1, \infty)$

$$L_1 = \bar{x} - t_{(\alpha,n-1)} \cdot \frac{s}{\sqrt{n}} = 2{,}057 - 1{,}833 \cdot \frac{0{,}0258}{\sqrt{10}} = 2{,}042$$

norma (číslo 2,052) je uvnitř intervalu $\Rightarrow$ šarže splňuje normu

MÁSLO (min. 82 % živočišného tuku)

Jak se to ověřuje?

- dle vyjádření Ústavu mléka, tuků a kosmetiky na VŠCHT Praha se většinou porovnává pouze aritmetický průměr jednotlivých stanovení s limitem
- ale započítání nejistoty je legislativně <u>správný</u> postup

Kunín Máslo	Tesco Value Máslo	Moravia Máslo	Meggle Alpenbutter	Lidl/Pilos Máslo	Molkerei Ammerland Německé máslo	Norma/ Landfein Deutsche Markenbutter	Arla Foods Deutsche Markenbutter
59,90	56,90	50,50 [1)]	64,70	56,90	56,90	56,90	51,70
250	250	250	250	250	250	250	250
24	22,80	20,20	25,90	22,80	22,80	22,80	20,70
uspokojivě 58 %	**uspokojivě 55 %**	**uspokojivě 52 %**	**uspokojivě 51 %**	**uspokojivě 47 %**	**nedostatečně 17 %**	**nedostatečně 17 %**	**nedostatečně 11 %**
neuvedeno	neuvedeno	neuvedeno	×	neuvedeno	×	✓	✓
82 / 82,9	82 / 82,6	82 / 82,2	82 / 82,2	82 / 82	82 / 80,8	82 / 80,7	82 / 80,1
15,2	15,4	16	15,6	15,6	15,9	16	15,9

Můžete jim věřit

Deklarovanou tučnost dodržely všechny vzorky. Tři másla původem z německých mlékáren – Norma/Landfein Deutsche Markenbutter, Molkerei Ammerland Německé máslo a Arla Foods Deutsche Markenbutter – <u>však slibovaným hodnotám dostála jen při započtení nejistoty měření.</u> Tento fakt jsme zohlednili v hodnocení deklarace a za dodržení slibované tučnosti jsme jim udělili dostatečnou známku.

dTest 11 - 2017

Jaký tedy použít interval?

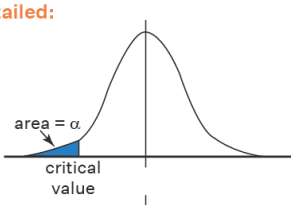
chci pouze odhadnout střední hodnotu μ nebo testovat její rovnost určité hodnotě $L_1 \leftarrow \underset{\mu}{\vert} \quad \bar{x} \quad \rightarrow L_2$	oboustranný interval spolehlivosti $\langle L_1, L_2 \rangle$ střední hodnoty μ $\bar{x} - t_{(\frac{\alpha}{2}, n-1)} \cdot \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{(\frac{\alpha}{2}, n-1)} \cdot \frac{s}{\sqrt{n}}$
chci odhadnout střední hodnotu μ a porovnat s limitem	
pokud je MAX uvnitř intervalu, platí hypotéza $\mu \leq$ MAX $L_1 \leftarrow \underset{\text{MAX}}{\vert} \quad \bar{x} \quad \rightarrow \infty$	jednostranný interval spolehlivosti $\langle L_1, \infty)$ střední hodnoty μ $L_1 = \bar{x} - t_{(\alpha, n-1)} \cdot \frac{s}{\sqrt{n}}$
pokud je MIN uvnitř intervalu, platí hypotéza $\mu \geq$ MIN $-\infty \leftarrow \quad \bar{x} \quad \rightarrow \underset{\text{MIN}}{\vert} L_2$	jednostranný interval spolehlivosti $(-\infty, L_2\rangle$ střední hodnoty μ $L_2 = \bar{x} + t_{(\alpha, n-1)} \cdot \frac{s}{\sqrt{n}}$

Jaký kvantil/kritickou hodnotu použít?

Výpočet intervalu spolehlivosti střední hodnoty μ	
znám rozptyl základního souboru σ nebo ho odhaduji pomocí s z velkého množství měření ($n > 30$)	použiji kvantil normálního rozdělení z =*NORMSINV*(hodnota kvantilu) např. 95% jednostranný interval =*NORMSINV*(0,95) výsledek 1,64 např. 95% oboustranný interval =*NORMSINV*(0,975) výsledek 1,96
rozptyl základního souboru odhaduji pomocí s z malého množství měření ($n < 30$)	použiji kvantil/kritickou hodnotu Studentova rozdělení t =*TINV*(α; n–1) pro 95% oboustranný interval =*TINV*(0,05; n–1) pro 95% jednostranný interval =*TINV*(2*0,05; n–1)

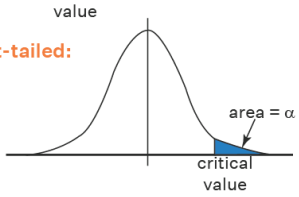
Jak vypočítat kritickou hodnotu Studentova rozdělení v Excelu?

left-tailed:



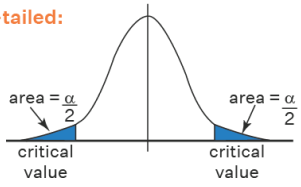
95% kritická hodnota (5% kvantil)
= –*TINV*(2*0,05; *n*–1)
nově =*T.INV*(0,05; *n*–1)

right-tailed:



5% kritická hodnota (95% kvantil)
=*TINV*(2*0,05; *n*–1)
nově =*T.INV*(0,95; *n*–1)

two-tailed:



97,5% kritická hodnota (2,5% kvantil)
= –*TINV*(0,05; *n*–1)
nově = –*T.INV.2T*(0,05; *n*–1)

2,5% kritická hodnota (97,5% kvantil)
=*TINV*(0,05; *n*–1)
nově =*T.INV.2T*(0,05; *n*–1)

Šířka intervalu spolehlivosti je ovlivněna:

- velikostí výběru n (čím větší výběr, tím užší interval)
- hladinou významnosti α (čím vyšší α, tím užší interval) → menší spolehlivost
- variabilitou dat (čím vyšší hodnota σ nebo s, tím širší interval)
- použitým vzorcem (pokud používáme *t*-rozdělení, je interval širší než při použití N(0, 1) a rozdíl je markantnější u malých výběrů)

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

$$\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Odvození výrazu pro interval spolehlivosti rozptylu σ^2:

Mějme normální náhodnou veličinu ξ s rozdělením N(μ, σ^2) a proveďme výběr hodnot této veličiny v rozsahu k. Pro každý získaný výběr vypočítejme hodnotu Y_j

$$\boxed{Y_j = \sum_{i=1}^{k} \left(\frac{x_i - \mu}{\sigma}\right)^2 = \sum_{i=1}^{k} z_i^2 \sim \chi^2(k)}$$

- získané hodnoty jsou hodnotami veličiny s $\chi^2(k)$ rozdělením
- je to tedy rozdělení sumy čtverců normované náhodné veličiny

Co se stane, pokud nahradíme neznámé μ výběrovým průměrem?

$$\boxed{Y_j = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\sigma^2} = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)}$$

Oboustranný interval spolehlivosti rozptylu σ^2

$$\boxed{\underbrace{\frac{(n-1)\cdot s^2}{\chi^2_{(\frac{\alpha}{2},n-1)}}}_{\text{dolní hranice}} \leq \sigma^2 \leq \underbrace{\frac{(n-1)\cdot s^2}{\chi^2_{(1-\frac{\alpha}{2},n-1)}}}_{\text{horní hranice}}}$$

Intervalový odhad směrodatné odchylky je nesymetrický!

$\chi^2_{(1-\frac{\alpha}{2},n-1)}$ a $\chi^2_{(\frac{\alpha}{2},n-1)}$ je dolní a horní kritická hodnota chí-kvadrát rozdělení pro hladinu významnosti α a n–1 stupňů volnosti

Jednostranný interval spolehlivosti rozptylu σ^2

levostranný odhad

$\langle L_1, \infty)$

$$L_1 = \frac{(n-1)\cdot s^2}{\chi^2_{(\alpha,n-1)}}$$

pokud je číslo σ_0^2 uvnitř intervalu, platí hypotéza $\sigma^2 < \sigma_0^2$

pravostranný odhad

$\langle 0, L_2\rangle$

$$L_2 = \frac{(n-1)\cdot s^2}{\chi^2_{(1-\alpha,n-1)}}$$

pokud je číslo σ_0^2 uvnitř intervalu, platí hypotéza $\sigma^2 > \sigma_0^2$

$\chi^2_{(1-\alpha,n-1)}$ a $\chi^2_{(\alpha,n-1)}$ je dolní a horní kritická hodnota chí-kvadrát rozdělení pro hladinu významnosti α a n-1 stupňů volnosti

<u>Příklad:</u> Bylo provedeno 10 náhodných analýz obsahu pseudoefedrinu v léčivu Sudafed [mg]:

120,1 118,1 114,7 121,5 122,3 123,2 120,7 121,8 119,9 118,7

Rozhodněte s chybou I. druhu 5 %, zda je směrodatná odchylka obsahu pseudoefedrinu v léčivu Sudafed rovna 1 mg.

$s^2 = 6{,}091$ $n = 10$ $\chi^2_{0{,}025;\,9} = 19{,}023$ $\chi^2_{0{,}975;\,9} = 2{,}700$

$$\sqrt{\frac{(n-1)\cdot s^2}{\chi^2_{(0{,}025;\,10-1)}}} \le \sigma \le \sqrt{\frac{(n-1)\cdot s^2}{\chi^2_{(0{,}975;10-1)}}}$$

$$\sqrt{\frac{(10-1)\cdot 6{,}091}{19{,}023}} \le \sigma \le \sqrt{\frac{(10-1)\cdot 6{,}091}{2{,}700}}$$

interval je $\langle 1{,}70;\ 4{,}51\rangle$

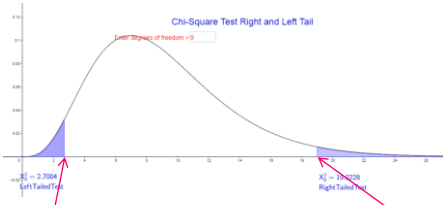
jelikož požadovaná hodnota 1 mg neleží uvnitř intervalu

$\Downarrow$

směrodatná odchylka není rovna 1 mg

Výpočet intervalu spolehlivosti rozptylu σ^2

rozdělení pravděpodobnosti odpovídá chí-kvadrát rozdělení



použiji kvantil chí-kvadrát rozdělení
=*CHINV*(α; n–1)

dolní kvantil $\chi^2_{(\frac{\alpha}{2},n-1)}$ =*CHINV*(1-α/2; n–1)

horní kvantil $\chi^2_{(1-\frac{\alpha}{2},n-1)}$ =*CHINV*(α/2; n–1)

např. pro α = 0,05 a n = 10 dostáváme:

2,5% kvantil (dolní) =*CHINV*(0,975; 10–1)
výsledek 2,700

97,5% kvantil (horní) =*CHINV*(0,025; 10–1)
výsledek 19,023

97,5% kvantil (97,5 % hodnot leží pod touto hranicí)
neboli
2,5% kritická hodnota (2,5% hodnot leží za hranicí)

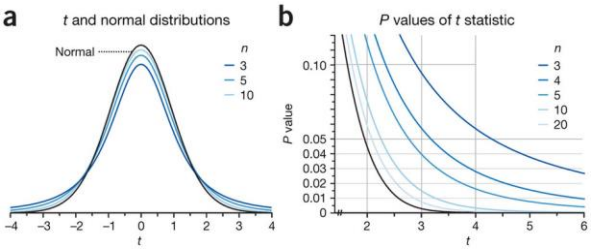
2,5% kvantil (2,5 % hodnot leží pod touto hranicí)
neboli
97,5% kritická hodnota (97,5% hodnot leží za hranicí)

t-rozdělení (Studentovo) (angl. *Student's t-distribution*)

Jsou-li X a Y nezávislé náhodné veličiny, přičemž X má rozdělení N(0, 1) a Y má rozdělení $\chi^2(k)$, pak náhodná veličina T

$$T = \frac{X}{\sqrt{\frac{Y}{k}}}$$

má Studentovo rozdělení $t(k)$ s k stupni volnosti



t-rozdělení (Studentovo) (angl. *Student's t-distribution*)

Pokud náhodné veličiny x_1, x_2, …, x_n mají normální rozdělení N(μ, σ^2) a jsou navzájem nezávislé, pak náhodná veličina definovaná jako

$$\boxed{\frac{\bar{x}-\mu}{s}\sqrt{n} \sim t(n-1)}$$

má Studentovo *t* rozdělení s (n–1) stupni volnosti.

důkaz:

$\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$ po standardizaci získáme $\frac{\bar{x}-\mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1)$

dále víme, že náhodná veličina $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$

Nyní obě náhodné veličiny dosadíme do definice Studentova rozdělení:

Jsou-li X a Y nezávislé náhodné veličiny, přičemž X má rozdělení N(0, 1) a Y má rozdělení $\chi^2(k)$, pak náhodná veličina T

$$\boxed{T = \frac{X}{\sqrt{\frac{Y}{k}}}}$$

má Studentovo rozdělení $t(k)$ s k stupni volnosti

$$t = \frac{X}{\sqrt{\frac{Y}{k}}} = \frac{\frac{\bar{x}-\mu}{\sqrt{\frac{\sigma^2}{n}}}}{\sqrt{\frac{\frac{(n-1)s^2}{\sigma^2}}{(n-1)}}} = \frac{\frac{\bar{x}-\mu}{\sqrt{\frac{\sigma^2}{n}}}}{\sqrt{\frac{s^2}{\sigma^2}}} = \frac{\bar{x}-\mu}{s}\sqrt{n} \sim t(n-1)$$

Konec