

ODHADY PARAMETRŮ ZÁKLADNÍHO SOUBORU



9. cvičení

Statistická indukce

Metody statistické indukce jsou zaměřeny na řešení dvou základních úloh:

- ❖ **odhady populačních parametrů** - používáme, chceme-li určit velikost parametru NV, resp. velikost efektu (rozdílu, resp. poměru parametrů dvou NV).
- ❖ **testování statistických hypotéz o populačních parametrech a rozděleních populace** - používáme, chceme-li ověřit platnost předem definované hypotézy (s předem danou hladinou významnosti).

Typy odhadů parametrů populace

- ❖ **Bodový odhad** - číselná hodnota, která aproximuje hledaný parametr základního souboru.

Použití - pokud potřebujeme vyjádřit hledaný parametr jedinou hodnotou – např. pro další výpočty.

- ❖ **Intervalový odhad** - číselný interval, který s předem danou spolehlivostí vymezí prostor obsahující hledaný parametr.

Použití - pokud potřebujeme znát přesnost nalezeného odhadu.

Základní pojmy z teorie odhadu

- ❖ **odhad parametru θ** – ozn. T , je výběrová charakteristika (statistika), která nabývá hodnot „blízkých“ neznámému parametru θ

	konstanty			obecně značíme θ	
Základní soubor (populace)	stř. hodnota $E(X)$, resp. μ	medián $x_{0,5}$	rozptyl $D(X)$, resp. σ^2	směr. odchylka σ	pravděpodobnost π
Výběrový soubor (výběr)	(výběrový) průměr $\bar{X}$	výběrový medián $\tilde{X}_{0,5}$	výběrový rozptyl S^2	výběrová směr. odchylka S	rel. četnost p
	náhodné veličiny			obecně značíme $T(X)$	

- ❖ **parametrický prostor** – množina všech hodnot, kterých může parametr θ nabývat

Vlastnosti dobrého bodového odhadu

- ❖ **Nestrannost** (=nevychýlenost, nezkreslenost) - střední hodnota odhadu je rovna hledanému parametru.

$$E(T) = \theta$$

- ❖ **Vydatnost** (=eficience) – vydatný je takový nestranný odhad, jehož rozptyl je nejmenší mezi rozptyly všech nestranných odhadů. (=nejlepší nestranný odhad)

- ❖ **Konzistence** – s rostoucím rozsahem výběru n se odhad $T = T_n$ zpřesňuje

$$\lim_{n \rightarrow \infty} E(T_n) = \theta$$

$$\lim_{n \rightarrow \infty} D(T_n) = 0$$

- ❖ **Dostatečnost** – odhad obsahuje informaci z co nejvíce prvků výběrového souboru (např. medián x průměr)

Chyba bodového odhadu

- ❖ Bodový odhad – náhodná veličina
- ❖ **Výběrová chyba**
 - ❖ odchylka bodového odhadu od skutečné hodnoty parametru $(T - \theta)$
 - ❖ určuje velikost chyby, které se dopouštíme při odhadu na základě jednoho výběrového souboru.
- ❖ **Střední kvadratická chyba**
 - ❖ směrodatná odchylka **nezkresleného** bodového odhadu
$$\sigma_T = \sqrt{D(T)} = \sqrt{E(T - \theta)^2}$$
 - ❖ udává „průměrnou“ kvadratickou chybu odhadů určených z různých výběrových souborů daného rozsahu

Nejlepší bodové odhady

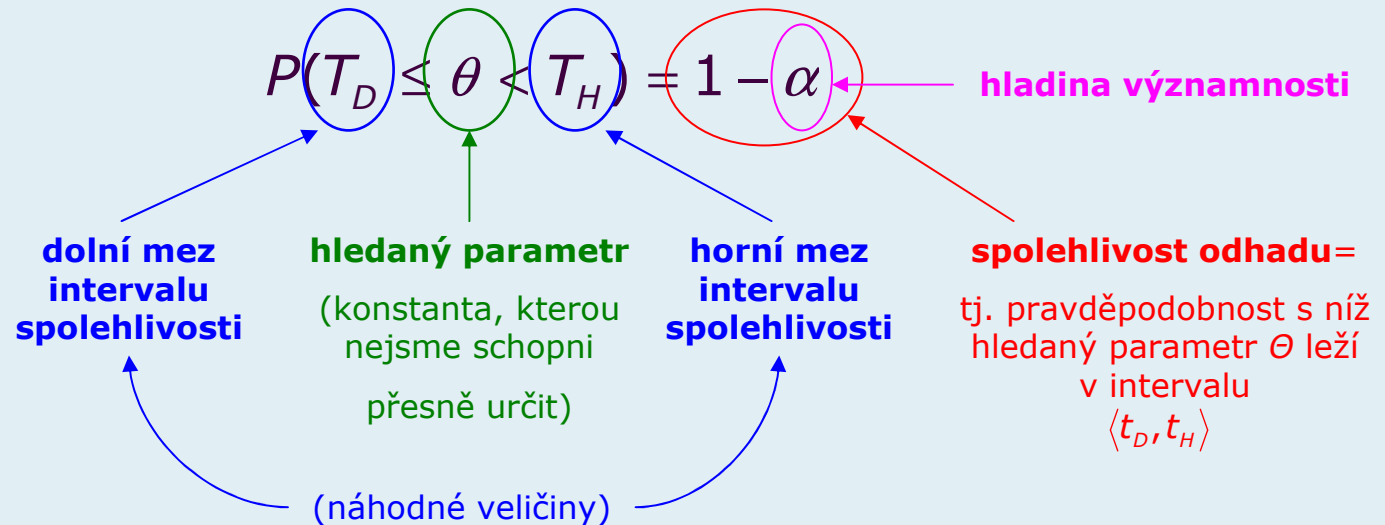
Bodovým odhadem:

- ❖ střední hodnoty μ je průměr $\bar{x}$,
- ❖ rozptylu σ^2 je výběrový rozptyl s^2 ,
- ❖ směrodatné odchylky σ je výběrová směrodatná odchylka s ,
- ❖ relativní četnosti π je výběrová relativní četnost p .

Intervalové odhady

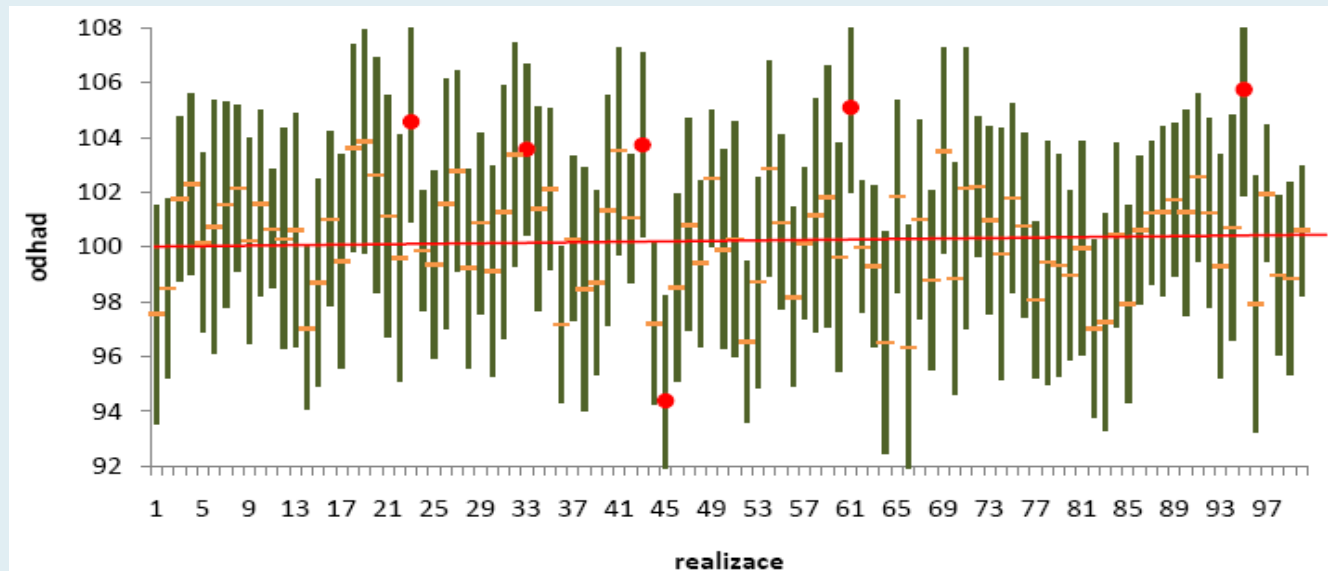
- ❖ **interval spolehlivosti** (konfidenční interval) – pro parametr Θ se spolehlivostí $1-\alpha$ (kde $\alpha \in \langle 0,1 \rangle$) je taková dvojice statistik $(T_D; T_H)$, že $P(T_D \leq \theta < T_H) = 1 - \alpha$
- ❖ **intervalový odhad parametru Θ** se spolehlivostí $1-\alpha$ je interval $\langle t_D, t_H \rangle$, kde t_D, t_H jsou hodnoty statistik T_D, T_H na daném statistickém souboru $(x_1, \dots, x_n)$. Intervalový odhad je tedy jednou z realizací intervalu spolehlivosti
- ❖ **spolehlivost odhadu** ($1-\alpha$) – předem určená pravděpodobnost (např. pro průmyslové aplikace většinou 95 %, pro biomedicínské 99 %)
- ❖ **hladina významnosti α**

Intervalové odhady



Spolehlivost odhadu $1 - \alpha$

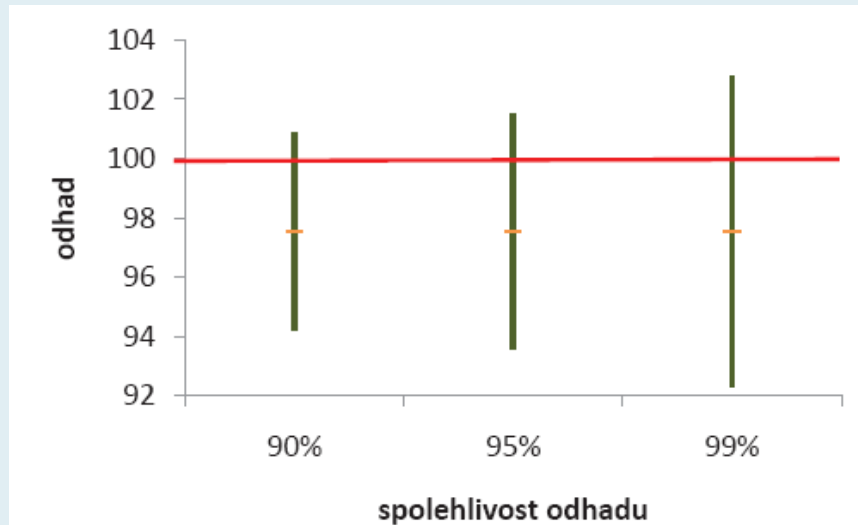
- ❖ Při opakovaných výběrech s konstantním rozsahem n z dané populace přibližně $100(1 - \alpha)\%$ intervalových odhadů obsahuje skutečnou hodnotu odhadovaného parametru θ a naopak $100 \cdot \alpha \%$ intervalových odhadů skutečnou hodnotu odhadovaného parametru θ neobsahuje.



Simulace intervalových odhadů střední hodnoty (spolehlivost 0,95) získaných na základě opakovaných výběrů o rozsahu 30 z populace se střední hodnotou 100. 6 intervalů ze 100 neobsahuje skutečnou střední hodnotu.

Spolehlivost odhadu $1 - \alpha$

- ❖ čím vyšší spolehlivost odhadu požadujeme, tím širší intervalový odhad získáme (hledaná hodnota se v něm musí nacházet s vyšší pravděpodobností) => kompromis mezi spolehlivostí a významností odhadu



intervaly jsou symetrické podle průměru (oranžová úsečka)

java applet: intervalové odhady
<http://mi21.vsb.cz/modul/uvod-do-statistiky>

- ❖ s rostoucím rozsahem výběru n se intervalový odhad populačních charakteristik zpřesňuje, tzn. šířka příslušných intervalových odhadů se zmenšuje a to úměrně $\sqrt{n}$

Typy intervalových odhadů

❖ Dvoustranné (T_D, T_H) $P(T_D \leq \theta < T_H) = 1 - \alpha$
 $P(\theta < T_D) = P(\theta \geq T_H) = \frac{\alpha}{2}$

❖ Jednostranné

❖ Levostranné $(T_D; \infty)$, $P(T_D < \theta) = 1 - \alpha$

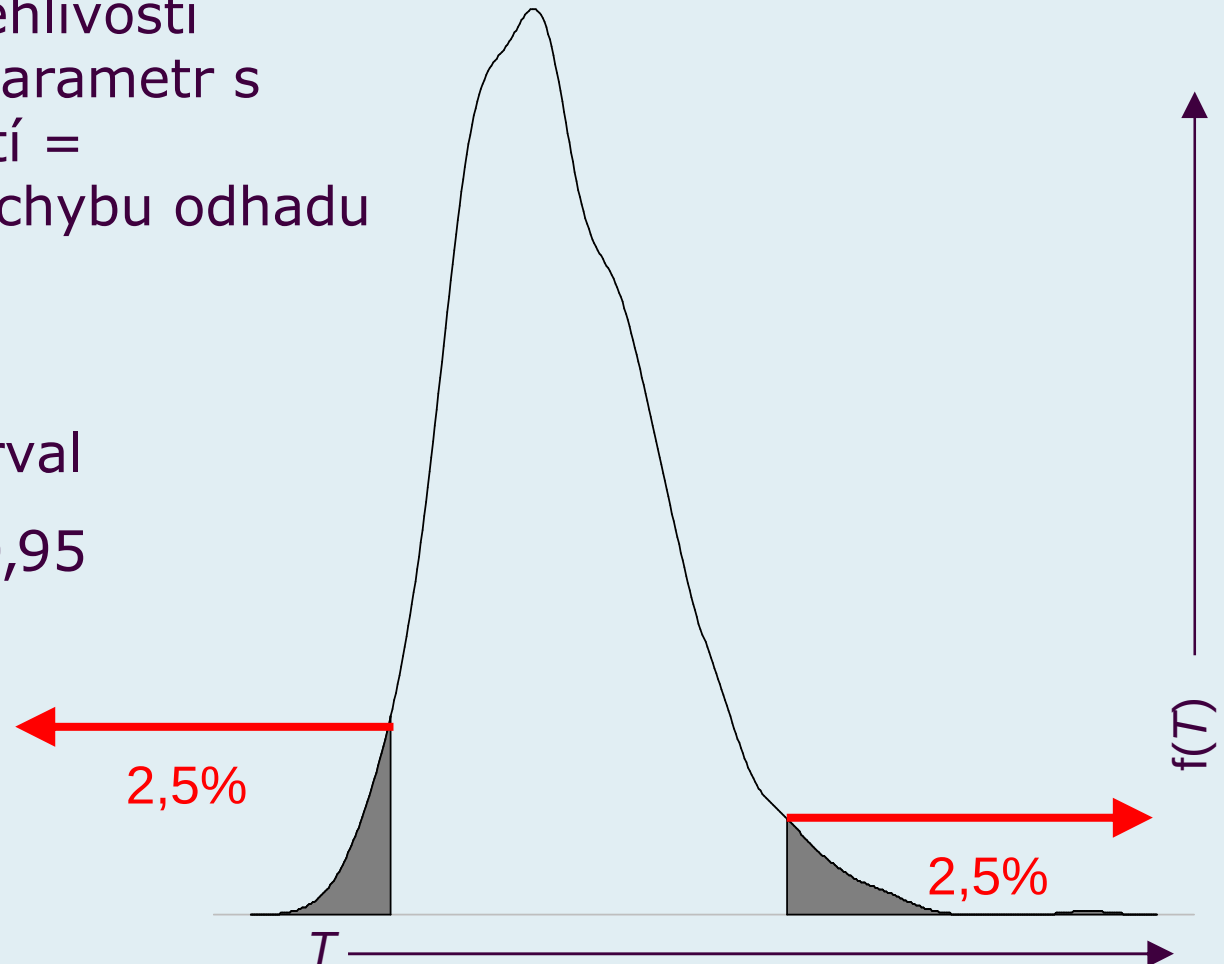
❖ Pravostranné $(-\infty; T_H)$ $P(\theta < T_H) = 1 - \alpha$

Grafová prezentace intervalového odhadu

Nechť interval spolehlivosti obsahuje hledaný parametr s 95%-ní spolehlivostí = připouštíme 5%-ní chybu odhadu ($\alpha = 0,05$)

Dvoustranný interval

$$P(T_D \leq \theta < T_H) = 0,95$$

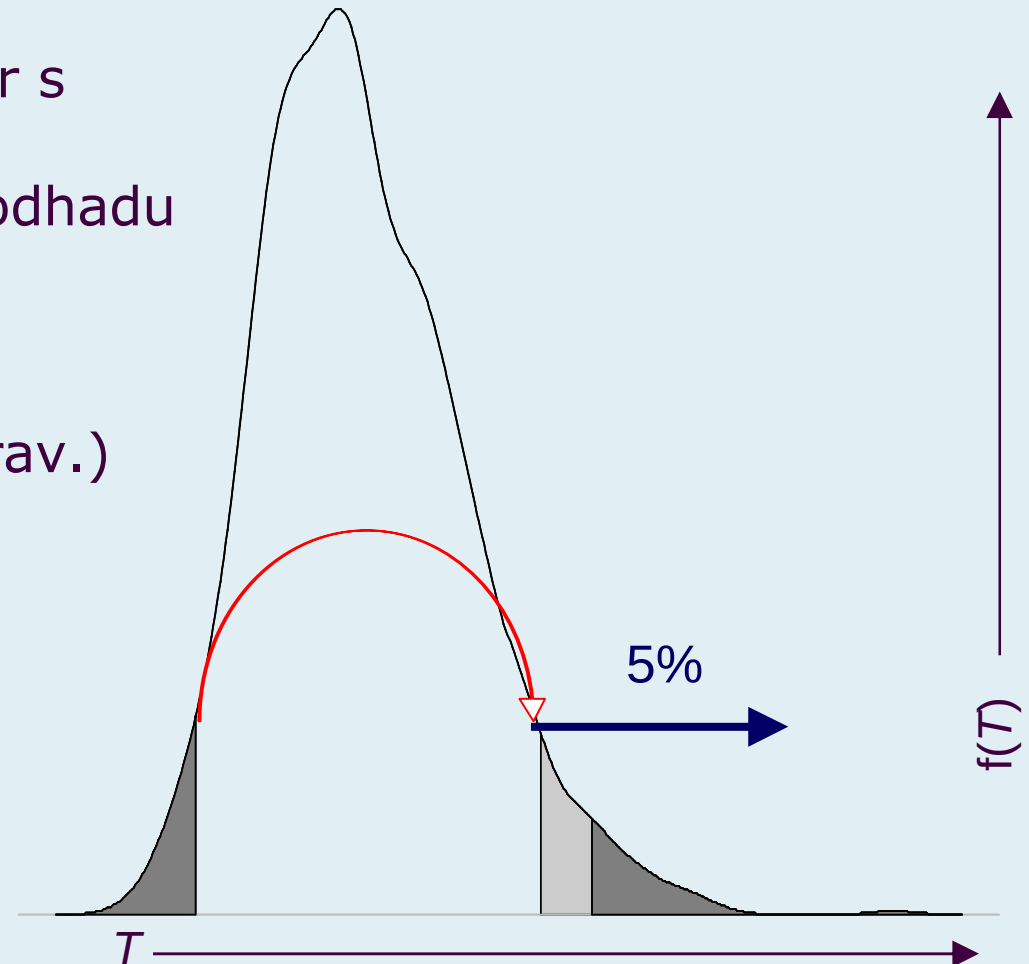


Grafová prezentace intervalového odhadu

Nechť interval spolehlivosti obsahuje hledaný parametr s 95%-ní spolehlivostí = připouštíme 5%-ní chybu odhadu ($\alpha = 0,05$)

Jednostranný interval (prav.)

$$P(\theta < T_H) = 0,95$$



Konstrukce intervalových odhadů

- ❖ Zvolíme vhodnou výběrovou charakteristiku, jejíž rozdělení známe (testová statistika) – $T(X)$

$$T(X) \approx F(x)$$

- ❖ PRAVOSTRANNÉ ODHADY:

Nechť: $P(T(X) < x_H) = 1 - \alpha$

Potom:

$$F(x_H) = 1 - \alpha$$

$$x_H = x_{1-\alpha}$$

Závěr: $P(T(X) < x_{1-\alpha}) = 1 - \alpha$

Konstrukce intervalových odhadů

- ❖ Zvolíme vhodnou výběrovou charakteristiku, jejíž rozdělení známe (testová statistika) – $T(X)$

$$T(X) \approx F(x)$$

- ❖ LEVOSTRANNÉ ODHADY:

Nechť: $P(x_D < T(X)) = 1 - \alpha$

Potom: $1 - F(x_D) = 1 - \alpha$

$$F(x_D) = \alpha$$

$$x_D = x_\alpha$$

Závěr: $P(x_\alpha < T(X)) = 1 - \alpha$

Konstrukce intervalových odhadů

- ❖ Zvolíme vhodnou výběrovou charakteristiku, jejíž rozdělení známe (testová statistika) – $T(X)$

$$T(X) \approx F(x)$$

- ❖ **DVOUSTRANNÉ ODHADY:**

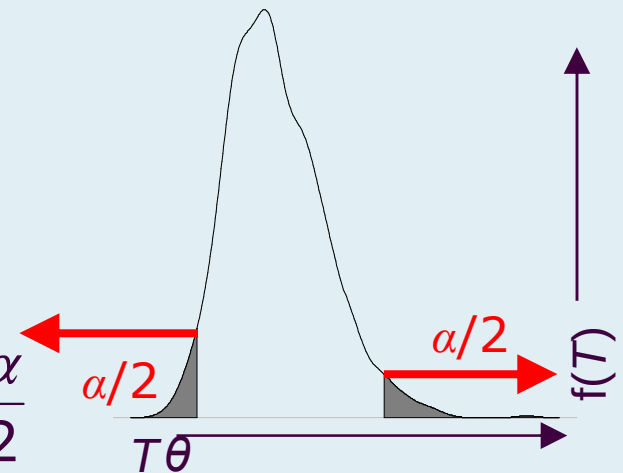
Nechť: $P(x_D < T(X) < x_H) = 1 - \alpha$

$$P(T(X) < x_D) = P(T(X) \geq x_H) = \frac{\alpha}{2}$$

Potom: $F(x_D) = 1 - F(x_H) = \frac{\alpha}{2}$

$$x_D = x_{\frac{\alpha}{2}} \quad x_H = x_{1-\frac{\alpha}{2}}$$

Závěr: $P\left(x_{\frac{\alpha}{2}} < T(X) < x_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$



Odhady parametrů normálního rozdělení

- ❖ Obecné metody jsou značně náročné => omezíme se na intervaly spolehlivosti pro parametry normálního rozdělení.
- ❖ Jestliže základní soubor nepochází z normálního rozdělení => používáme neparametrické (robustní) metody odhadu.

Odhad střední hodnoty - μ

- ❖ Nejlepším **bodovým odhadem** střední hodnoty μ je průměr $\bar{x}$.
- ❖ **Intervalový odhad** střední hodnoty μ počítáme jinak, jestliže známe rozptyl σ^2 základního souboru a jinak, když populační rozptyl σ^2 neznáme.

Odhad střední hodnoty - μ , známe σ^2

Známe rozptyl

$$X \rightarrow N(\mu; \sigma^2) \leftarrow \text{známe}$$

Volba vhodné testové statistiky: $Z = \frac{\bar{X} - \mu}{\sigma} \cdot \sqrt{n}; \quad Z \rightarrow N(0;1)$

Levostranný interval spolehlivosti:

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

$$P\left(\frac{\bar{X} - \mu}{\sigma} \cdot \sqrt{n} < z_{1-\alpha}\right) = 1 - \alpha$$

$$P\left(-\mu < z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}} - \bar{X}\right) = 1 - \alpha$$

$$P\left(\mu > \bar{X} - z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Odhad střední hodnoty - μ , známe σ^2

Známe rozptyl

$$X \rightarrow N(\mu; \sigma^2) \leftarrow \text{známe}$$

Volba vhodné testové statistiky: $Z = \frac{\bar{X} - \mu}{\sigma} \cdot \sqrt{n}; \quad Z \rightarrow N(0;1)$

Pravostranný interval spolehlivosti:

$$P(Z > z_\alpha) = 1 - \alpha$$

$$P\left(\frac{\bar{X} - \mu}{\sigma} \cdot \sqrt{n} > z_\alpha\right) = 1 - \alpha$$

$$P\left(-\mu > z_\alpha \cdot \frac{\sigma}{\sqrt{n}} - \bar{X}\right) = 1 - \alpha$$

$$P\left(\mu < \bar{X} - z_\alpha \cdot \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

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

Odhad střední hodnoty - μ , známe σ^2

Známe rozptyl $X \rightarrow N(\mu, \sigma^2)$ ← známe

Volba vhodné testové statistiky: $Z = \frac{\bar{X} - \mu}{\sigma} \cdot \sqrt{n}$; $Z \rightarrow N(0;1)$

Dvoustranný interval spolehlivosti:

$$P\left(z_{\frac{\alpha}{2}} < Z < z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

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

$$P\left(\bar{X} + \frac{\sigma}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}} > \mu > \bar{X} - \frac{\sigma}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

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

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

Odhad střední hodnoty - μ , neznáme σ^2

Neznáme rozptyl

$$X \rightarrow N(\mu, \sigma^2) \leftarrow \text{neznáme}$$

Uvedené intervalové odhady používáme nejen v případech, kdy známe směrodatnou odchylku σ , ale i v případech, kdy **máme dostatečně velký výběr ($n \geq 30$) a směrodatnou odchylku σ neznáme**. V těchto případech lze ve výše uvedených vzorcích nahradit směrodatnou odchylku σ výběrovou směrodatnou odchylkou s , aniž by tím vznikla významná chyba.

Pro vzorky s rozsahem menším než 30 prvků =>

Odhad střední hodnoty - μ , neznáme σ^2

Neznáme rozptyl

$$X \rightarrow N(\mu, \sigma^2) \leftarrow \text{neznáme}$$

Volba vhodné testové statistiky: $T_{n-1} = \frac{\bar{X} - \mu}{S} \cdot \sqrt{n}$; $T_{n-1} \rightarrow t_{n-1}$

Levostranný interval spolehlivosti:

$$P\left(\bar{X} - \frac{S}{\sqrt{n}} \cdot t_{1-\frac{\alpha}{2}, n-1} < \mu\right) = 1 - \alpha$$

Pravostranný interval spolehlivosti:

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

Dvoustranný interval spolehlivosti:

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

1. Při kontrole rychlosti 24 vozidel ze 100 překročilo rychlost 60 km/h, průměrná rychlost byla 65 km/h a příslušná výběrová směrodatná odchylka 7 km/h. Sestrojte 95% interval spolehlivosti pro střední hodnotu rychlosti vozidel.

Řešení:

Hledáme interval spolehlivosti pro střední hodnotu a neznáme rozptyl. Ale máme rozsah výběru větší než 30 ($100 > 30$) => položíme $\sigma \approx s$:

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

Hladina významnosti ze zadání: $\alpha = 1 - 0,95 = 0,05$

$$\frac{\alpha}{2} = 0,025, \quad 1 - \frac{\alpha}{2} = 0,975; \quad z_{0,975} = 1,96 \leftarrow \text{viz. Tabulka 1}$$

Výběrový soubor:

$$n = 100$$

$$\bar{x} = 65 \text{ km/h}$$

$$s = 7 \text{ km/h}$$

Řešení:

Dosadíme:

$$P\left(65 - \frac{7}{\sqrt{100}} \cdot 1,96 < \mu < 65 + \frac{7}{\sqrt{100}} \cdot 1,96\right) = 0,95$$

Po úpravě:

$$P(63,63 < \mu < 66,37) = 0,95$$

Tzn., že s 95% spolehlivostí můžeme tvrdit, že se střední rychlost vozidel pohybuje v rozmezí 64 km/h až 66 km/h.

2. Obchodní řetězec TETO si v dubnu 2006 zadal studii týkající se počtu zákazníků v prodejně TETO Poruba v pátek odpoledne (od 12:00 do 18:00) hodin. Po jednom měsíci sledování prodejny jsme získali tyto údaje:

Datum	Počet zákazníků
2.5.2006	3756
9.5.2006	2987
16.5.2006	3042
23.5.2006	4206
30.5.2006	3597

- a) Zamyslete se nad důvody, které výzkumníka vedly k analýze výběru o malém rozsahu (mnohem méně než 30 hodnot) a jaké jsou důsledky volby výběru o malém rozsahu.
- b) Určete pro management řetězce TETO intervalový odhad středního počtu zákazníků v prodejně TETO Poruba v pátek odpoledne (se spolehlivostí 95%).

Řešení:

- a) finanční a časová náročnost (pro výběr o rozsahu 30 hodnot by sledování trvalo déle než půl roku)
- b) hledáme odhad střední hodnoty a neznáme směrodatnou odchylku:

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

Hladina významnosti ze zadání: $\alpha = 1 - 0,95 = 0,05$

$$\frac{\alpha}{2} = 0,025, \quad 1 - \frac{\alpha}{2} = 0,975; \quad t_{0,975;4} = 2,78 \quad \leftarrow \text{viz. Tabulka 2}$$

Řešení:

b)

Výběrový soubor:

$$n = 5$$

$$\bar{x} = \frac{\sum_{i=1}^5 x_i}{5} = \frac{3756 + 2987 + 3042 + 4206 + 3597}{5} = 3517,6$$

$$s^2 = \frac{\sum_{i=1}^5 (x_i - \bar{x})^2}{5 - 1} = \frac{(3756 - 3517,6)^2 + \dots + (3597 - 3517,6)^2}{4} = 261191,3$$

$$s = 511,1$$

Dosadíme:

$$P\left(3517,6 - \frac{511,1}{\sqrt{5}} \cdot 2,78 < \mu < 3517,6 + \frac{511,1}{\sqrt{5}} \cdot 2,78\right) = 0,95$$

Řešení:

b)

Po úpravě:

$$P(2882,2 < \mu < 4153,0) = 0,95$$

Tzn., že s 95%-ní spolehlivostí můžeme tvrdit, že návštěvnost TETO Poruba se v libovolný pátek v odpoledních hodinách bude pohybovat v rozmezí 2882 až 4153 zákazníků.

Odhad rozptylu - σ^2

Volba vhodné testové statistiky: $\chi = \frac{(n-1)s^2}{\sigma^2}$; $\chi \rightarrow \chi_{n-1}^2$

Levostranný interval spolehlivosti:

$$P\left(\frac{(n-1)}{\chi_{1-\alpha, n-1}} \cdot s^2 < \sigma^2\right) = 1 - \alpha$$

Pravostranný interval spolehlivosti:

$$P\left(\sigma^2 < \frac{(n-1)}{\chi_{\alpha, n-1}} \cdot s^2\right) = 1 - \alpha$$

Dvoustranný interval spolehlivosti:

$$P\left(\frac{(n-1)}{\chi_{1-\frac{\alpha}{2}, n-1}} \cdot s^2 < \sigma^2 < \frac{(n-1)}{\chi_{\frac{\alpha}{2}, n-1}} \cdot s^2\right) = 1 - \alpha$$

Odhad směrodatné odchylky - σ

Volba vhodné testové statistiky: $\chi = \frac{(n-1)s^2}{\sigma^2}$; $\chi \rightarrow \chi_{n-1}^2$

Levostranný interval spolehlivosti:

$$P\left(\sqrt{\frac{(n-1)}{\chi_{1-a, n-1}}} \cdot s < \sigma\right) = 1 - a$$

Pravostranný interval spolehlivosti:

$$P\left(\sigma < \sqrt{\frac{(n-1)}{\chi_{a, n-1}}} \cdot s\right) = 1 - a$$

Dvoustranný interval spolehlivosti:

$$P\left(\sqrt{\frac{(n-1)}{\chi_{1-\frac{a}{2}, n-1}}} \cdot s < \sigma < \sqrt{\frac{(n-1)}{\chi_{\frac{a}{2}, n-1}}} \cdot s\right) = 1 - a$$

3. Automat vyrábí pístové kroužky o daném průměru. Při kontrole kvality bylo náhodně vybráno 80 kroužků a vypočtena směrodatná odchylka jejich průměru 0,04mm. Určete dolní hranici rozptylu a směrodatné odchylky průměru pístových kroužků. (Předpokládejte, že průměr pístových kroužků lze modelovat pomocí normálního rozdělení.)

Řešení:

Nejdříve hledáme levostranný intervalový odhad rozptylu:

$$P\left(\frac{(n-1)}{\chi_{1-a, n-1}} \cdot s^2 < \sigma^2\right) = 1 - a$$

Hladina významnosti ze zadání: $\alpha = 1 - 0,95 = 0,05$

$$\chi_{0,95;79} \cong 100,7 \leftarrow \text{viz. Tabulka 3}$$

Výběrový soubor:

$$n = 80, \quad s^2 = (0,04)^2 = 0,0016 \text{ mm}^2$$

Dosadíme:

$$P\left(\frac{79}{100,7} \cdot 0,0016 < \sigma^2\right) = 0,95$$

Řešení:

Po úpravě:

$$P(0,0013 < \sigma^2) = 0,95$$

S 95% spolehlivostí je rozptyl průměru pístových kroužků větší než 0,0013 mm².

Pro intervalový odhad směrodatné odchylky:

$$P(\sqrt{0,0013} < \sigma) = 0,95$$

$$P(0,036 < \sigma) = 0,95$$

S 95% spolehlivostí tedy můžeme tvrdit, že směrodatná odchylka průměru pístových kroužků je větší než 0,036 mm.

Odhad relativní četnosti - π

Relativní četnost π je z intervalu $\langle 0;1 \rangle$

Jestliže máme výběrový soubor, jehož rozsah :

- ❖ je dostatečně velký ($n > 30$),
- ❖ je menší než 5% rozsahu základního souboru ($n/N < 0,05$),
- ❖ splňuje podmínku $n \cdot p \cdot (1-p) > 9$,

pak lze relativní četnost π odhadnout následovně:

Odhad relativní četnosti - π

Volba vhodné testové statistiky: $P_1 = \frac{p - \pi}{\sqrt{\pi(1 - \pi)}} \cdot \sqrt{n}; \quad P_1 \rightarrow N(0;1)$

Levostranný interval spolehlivosti:

$$P\left(p - \sqrt{\frac{p \cdot (1 - p)}{n}} \cdot z_{1-\alpha} < \pi\right) = 1 - \alpha$$

Pravostranný interval spolehlivosti:

$$P\left(\pi < p + \sqrt{\frac{p \cdot (1 - p)}{n}} \cdot z_{1-\alpha}\right) = 1 - \alpha$$

Dvoustranný interval spolehlivosti:

$$P\left(p - \sqrt{\frac{p \cdot (1 - p)}{n}} \cdot z_{1-\frac{\alpha}{2}} < \pi < p + \sqrt{\frac{p \cdot (1 - p)}{n}} \cdot z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

4. Při kontrole jakosti bylo ze série 7000 televizorů vybráno náhodně 300 výrobků, z toho bylo 25 vadných. Odhadněte podíl vadných televizorů v sérii.

Řešení:

Rozsah výběru je dostatečně velký ($n > 30$, $n \cdot p \cdot (1-p) > 9$) a nepřevyšuje 5% rozsahu populace ($n/N < 0,05$). Intervalový odhad podílu (relativní četnosti) vadných televizorů v sérii lze tedy stanovit jako

$$P\left(p - \sqrt{\frac{p \cdot (1-p)}{n}} \cdot z_{1-\frac{\alpha}{2}} < \pi < p + \sqrt{\frac{p \cdot (1-p)}{n}} \cdot z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

Hladina významnosti ze zadání: $\alpha = 1 - 0,95 = 0,05$

$$\frac{\alpha}{2} = 0,025, \quad 1 - \frac{\alpha}{2} = 0,975; \quad z_{0,975} = 1,96 \quad \leftarrow \text{viz. Tabulka 1}$$

Výběrový soubor:

$$n = 300, \quad p = \frac{25}{300} = 0,0833$$

Řešení:

Dosadíme:

$$P\left(0,083 - \sqrt{\frac{0,083 \cdot (1 - 0,083)}{300}} \cdot 1,96 < \pi < 0,083 + \sqrt{\frac{0,083 \cdot (1 - 0,083)}{300}} \cdot 1,96\right) = 0,95$$

$$P(0,0517 < \pi < 0,1143) = 0,95$$

S 95% spolehlivostí můžeme tvrdit, že se v dané sérii nachází mezi 5,2% a 11,4% vadných televizorů.

Rozsah výběru

Přesnost odhadu x finanční a časová náročnost

=>

v praxi hledáme kompromis, který pro požadovanou přesnost
povede na co nejmenší rozsah výběru

Rozsah výběru

Přesnost odhadu $\times$ finanční a časová náročnost
 $=>$

v praxi hledáme kompromis, který pro požadovanou přesnost povede na co nejmenší rozsah výběru

- ❖ **Přípustná chyba odhadu Δ** - hodnota, o kterou jsme ochotni se zmýlit oproti skutečné hodnotě odhadovaného parametru při dané spolehlivosti odhadu (hladině významnosti)

Předvýběr

- ❖ Provádíme, pokud odhadujeme požadovaný rozsah výběru, který závisí na prozatím neznámých výběrových charakteristikách (nemáme proveden výběr).
- ❖ Pro předvýběr o rozsahu n_1 vypočteme požadované výběrové charakteristiky, pomocí nich zjistíme rozsah výběru n , který pak doplníme o dalších $n - n_1$ hodnot.
- ❖ Intervalové odhady potom provádíme na výběru rozsahu n .

↑ Iterační heuristická metoda.

Rozsah výběru při odhadu střední hod.

Známe rozptyl

Oboustranný intervalový odhad: $P\left(\bar{X} - \frac{\sigma}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}} < \mu < \bar{X} + \frac{\sigma}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$

Příslušný intervalový odhad tedy můžeme vyjádřit ve tvaru:

$$\left(\bar{X} - \frac{\sigma}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}}; \bar{X} + \frac{\sigma}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}}\right) = \left(\bar{X} \pm \frac{\sigma}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}}\right)$$

Přípustná chyba odhadu Δ : $\Delta = \frac{\sigma}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}}$

Požadujeme:

$$\Delta \geq \frac{\sigma}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}}$$

$$n \geq \left(\frac{\sigma}{\Delta} \cdot z_{1-\frac{\alpha}{2}}\right)^2$$

Rozsah výběru při odhadu střední hod.

Neznáme rozptyl

Přípustná chyba odhadu Δ :
$$\Delta = \frac{s}{\sqrt{n}} \cdot t_{1-\frac{\alpha}{2}, n-1}$$

PŘEDVÝBĚR:

$$\Delta \geq \frac{s_1}{\sqrt{n}} \cdot t_{1-\frac{\alpha}{2}, n-1}$$

$$n \geq \left(\frac{s_1}{\Delta} \cdot t_{1-\frac{\alpha}{2}, n-1} \right)^2$$

Rozsah výběru při odhadu relativní čet.

Oboustranný intervalový odhad : $\left(p - \sqrt{\frac{p \cdot (1-p)}{n}} \cdot z_{1-\frac{\alpha}{2}}; p + \sqrt{\frac{p \cdot (1-p)}{n}} \cdot z_{1-\frac{\alpha}{2}} \right)$

Přípustná chyba odhadu Δ : $\Delta = \sqrt{\frac{p \cdot (1-p)}{n}} \cdot z_{1-\frac{\alpha}{2}}$

PŘEDVÝBĚR

$$\Delta \geq \sqrt{\frac{p_1 \cdot (1-p_1)}{n}} \cdot z_{1-\frac{\alpha}{2}}$$

$$n \geq \frac{p_1 \cdot (1-p_1)}{\Delta^2} \cdot z_{1-\frac{\alpha}{2}}^2$$

Nejhorší možnost ($p=0,5$): $n \geq \frac{1}{4\Delta^2} \cdot z_{1-\frac{\alpha}{2}}^2$

5. Výběrovým šetřením bychom chtěli odhadnout průměrnou výši úspor novomanželů. Žádáme spolehlivost 95% a maximální chybu 200 Kč. Směrodatná odchylka byla předběžně odhadnuta na 2 500 Kč. Kolika párů se musíme zeptat, abychom zajistili požadovanou přesnost a spolehlivost?

Řešení:

Známe směrodatnou odchylku populace

$$n \geq \left(\frac{\sigma}{\Delta} \cdot z_{1-\frac{\alpha}{2}} \right)^2$$

Hladina významnosti ze zadání: $\alpha = 1 - 0,95 = 0,05$

$$\frac{\alpha}{2} = 0,025, \quad 1 - \frac{\alpha}{2} = 0,975; \quad z_{0,975} = 1,96 \quad \leftarrow \text{viz. Tabulka 1}$$

Výběrový soubor: $\sigma = 2500$ Kč, $\Delta \leq 200$ Kč

$$\text{Dosadíme: } n \geq \left(\frac{2500}{200} \cdot 1,96 \right)^2 \Rightarrow n \geq 600,25$$

Chceme-li dosáhnout přípustné chyby ve výši maximálně 200 Kč, musíme pro nalezení intervalového odhadu průměrné výše úspor se spolehlivostí 95% provést výběrové šetření na výběrovém souboru o rozsahu minimálně 601 novomanželských párů.

6. Odhadujeme podíl prvotřídních výrobků. Kolik jich je třeba přezkoušet, aby se s **spolehlivostí 95%** chyba nepřekročila $\pm 3\%$? Co když víme, že hledaný podíl prvotřídních výrobků bude přes 80%?

Řešení:

- a) Výběrová relativní četnost není odhadnuta pomocí předvýběru, musíme počítat s nejhorší variantou $\Rightarrow p=0,5$

$$n \geq \frac{1}{4\Delta^2} \cdot z_{1-\frac{\alpha}{2}}^2$$

Hladina významnosti ze zadání: $\alpha = 1 - 0,95 = 0,05$

$$\frac{\alpha}{2} = 0,025, \quad 1 - \frac{\alpha}{2} = 0,975; \quad z_{0,975} = 1,96 \quad \leftarrow \text{viz. Tabulka 1}$$

$$\Delta \leq 3 \% \qquad n \geq \frac{1}{4 \cdot 0,03^2} \cdot 1,96^2 \qquad n \geq 1067,11$$

Chceme-li dosáhnout přípustné chyby ve výši maximálně 3 %, musíme pro nalezení intervalového odhadu relativní četnosti prvotřídních výrobků se spolehlivostí 95% provést výběrové šetření na výběrovém souboru o rozsahu minimálně 1068 výrobků.

Řešení:

b) Výběrová relativní četnost je odhadnuta na 80 %, $\Rightarrow p=0,8$

$$\Delta = \sqrt{\frac{p \cdot (1-p)}{n}} \cdot z_{1-\frac{\alpha}{2}}$$

Hladina významnosti ze zadání: $\alpha = 1 - 0,95 = 0,05$

$$\frac{\alpha}{2} = 0,025, \quad 1 - \frac{\alpha}{2} = 0,975; \quad z_{0,975} = 1,96 \quad \leftarrow \text{viz. Tabulka 1}$$

$$\Delta \leq 3 \% \quad n \geq \frac{0,8 \cdot (1 - 0,8)}{0,03^2} \cdot 1,96^2 \quad n \geq 682,95$$

Chceme-li dosáhnout přípustné chyby ve výši maximálně 3 %, musíme pro nalezení intervalového odhadu relativní četnosti prvotřídních výrobků se spolehlivostí 95% provést výběrové šetření na výběrovém souboru o rozsahu minimálně 683 výrobků.

Dvouvýběrové intervalové odhady

- ❖ Nástroj pro porovnání populačních charakteristik **dvou výběrů** pocházejících z nezávislých základních souborů s normálním rozdělením.
- ❖ Intervalový odhad poměru rozptylů normálních rozdělení
 - ❖ Pokud interval zahrnuje i hodnotu **jedna**, pak mohou být oba populační rozptyly shodné!
- ❖ Intervalový odhad rozdílu středních hodnot
 - ❖ Pokud interval zahrnuje i hodnotu **nula**, pak mohou být oba populační průměry shodné!
- ❖ Intervalový odhad rozdílu relativních četností
 - ❖ Pokud interval zahrnuje i hodnotu **nula**, pak mohou být obě populační relativní četnosti shodné!

Intervaly spolehlivosti pro poměr dvou rozptylů normálních rozdělení

$$X_1 \rightarrow N(\mu_1, \sigma_1^2) \quad X_2 \rightarrow N(\mu_2, \sigma_2^2)$$

Volba vhodné testové statistiky:

$$F = \frac{\frac{s_1^2}{\sigma_1^2}}{\frac{s_2^2}{\sigma_2^2}}; \quad F \rightarrow F_{n1-1; n2-1}$$

Levostranný interval spolehlivosti:

$$P\left(\frac{1}{f_{1-\alpha}} \frac{s_1^2}{s_2^2} < \frac{\sigma_1^2}{\sigma_2^2}\right) = 1 - \alpha$$

Pravostranný interval spolehlivosti:

$$P\left(\frac{\sigma_1^2}{\sigma_2^2} < \frac{1}{f_\alpha} \frac{s_1^2}{s_2^2}\right) = 1 - \alpha$$

Dvoustranný interval spolehlivosti:

$$P\left(\frac{1}{f_{1-\frac{\alpha}{2}}} \frac{s_1^2}{s_2^2} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{1}{f_{\frac{\alpha}{2}}} \frac{s_1^2}{s_2^2}\right) = 1 - \alpha$$

Intervaly spolehlivosti pro rozdíl dvou parametrů normálních rozdělení

$$X_1 \rightarrow N(\mu_1, \sigma_1^2) \quad X_2 \rightarrow N(\mu_2, \sigma_2^2)$$

Intervalový odhad rozdílu středních hodnot dvou populací s normálním rozdělením, z nichž byly pořízeny náhodné výběry, lze provádět pokud:

- ❖ Známe rozptyly σ_1^2 a σ_2^2 obou populací.
- ❖ Neznáme rozptyly obou populací, ale lze předpokládat, že jsou shodné.
- ❖ Neznáme rozptyly obou populací a nelze předpokládat, že jsou shodné.

Odhad rozdílu středních hodnot

Známe rozptyly

Volba vhodné testové statistiky:
$$Z_2 = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}; \quad Z_2 \rightarrow N(0;1)$$

Levostranný interval spolehlivosti:

$$P\left((\overline{X}_1 - \overline{X}_2) - \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \cdot z_{1-\alpha} < (\mu_1 - \mu_2)\right) = 1 - \alpha$$

Pravostranný interval spolehlivosti:

$$P\left((\mu_1 - \mu_2) < (\overline{X}_1 - \overline{X}_2) + \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \cdot z_{1-\alpha}\right) = 1 - \alpha$$

Dvoustranný interval spolehlivosti:

$$P\left((\overline{X}_1 - \overline{X}_2) - \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \cdot z_{1-\frac{\alpha}{2}} < (\mu_1 - \mu_2) < (\overline{X}_1 - \overline{X}_2) + \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \cdot z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

Odhad rozdílu středních hodnot

Neznáme rozptyly a lze předpokládat, že jsou shodné

Volba vhodné testové statistiky: $T_2 = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}; \quad T_2 \rightarrow t_{n_1+n_2-1}$

Levostranný interval spolehlivosti:

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

$$P\left((\overline{X}_1 - \overline{X}_2) - s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \cdot t_{1-\alpha, n_1+n_2-2} < (\mu_1 - \mu_2)\right) = 1 - \alpha$$

Pravostranný interval spolehlivosti:

$$P\left((\mu_1 - \mu_2) < (\overline{X}_1 - \overline{X}_2) + s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \cdot t_{1-\frac{\alpha}{2}, n_1+n_2-2}\right) = 1 - \alpha$$

Dvoustranný interval spolehlivosti:

$$P\left((\overline{X}_1 - \overline{X}_2) - s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \cdot t_{1-\frac{\alpha}{2}, n_1+n_2-2} < (\mu_1 - \mu_2) < (\overline{X}_1 - \overline{X}_2) + s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \cdot t_{1-\frac{\alpha}{2}, n_1+n_2-2}\right) = 1 - \alpha$$

Odhad rozdílu středních hodnot

Neznáme rozptyly a nelze předpokládat, že jsou shodné

Volba vhodné testové statistiky:

$$T_2 = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}; T_2 \rightarrow t_\nu$$

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left(\frac{s_1^2}{n_1}\right)^2 \frac{1}{n_1+1} + \left(\frac{s_2^2}{n_2}\right)^2 \frac{1}{n_2+1}}$$

Levostranný interval spolehlivosti:

$$P\left((\bar{X}_1 - \bar{X}_2) - \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \cdot t_\nu < (\mu_1 - \mu_2)\right) = 1 - \alpha$$

Pravostranný interval spolehlivosti:

$$P\left((\mu_1 - \mu_2) < (\bar{X}_1 - \bar{X}_2) + \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \cdot t_{1-\frac{\alpha}{2}, \nu}\right) = 1 - \alpha$$

Dvoustranný interval spolehlivosti:

$$P\left((\bar{X}_1 - \bar{X}_2) - \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \cdot t_{1-\frac{\alpha}{2}, \nu} < (\mu_1 - \mu_2) < (\bar{X}_1 - \bar{X}_2) + \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \cdot t_{1-\frac{\alpha}{2}, \nu}\right) = 1 - \alpha$$

Odhad rozdílu relativních četností

Jestliže máme výběrové soubory s rozsahy, které:

- ❖ jsou dostatečně velký ($n_1 > 30$; $n_2 > 30$),
- ❖ jsou menší než 5% rozsahu základního souboru ($n_1/N_1 < 0,05$; $n_2/N_2 < 0,05$),
- ❖ splňuje podmínku $n_1 \cdot p_1 \cdot (1-p_1) > 9$; $n_2 \cdot p_2 \cdot (1-p_2) > 9$,

pak lze rozdíl relativních četností $\pi_1 - \pi_2$ odhadnout následovně:

Odhad rozdílu relativních četností

Volba vhodné testové statistiky:

$$P_2 = \frac{(p_1 - p_2) - (\pi_1 - \pi_2)}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}; \quad P_2 \rightarrow N(0;1)$$

Levostranný interval spolehlivosti:

$$p = \frac{X_1 + X_2}{n_1 + n_2}$$

$$P\left((p_1 - p_2) - \sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \cdot z_{1-\frac{\alpha}{2}} < (\pi_1 - \pi_2)\right) = 1 - \alpha$$

Pravostranný interval spolehlivosti:

$$P\left((\pi_1 - \pi_2) < (p_1 - p_2) + \sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \cdot z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

Dvoustranný interval spolehlivosti:

$$P\left((p_1 - p_2) - \sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \cdot z_{1-\frac{\alpha}{2}} < (\pi_1 - \pi_2) < (p_1 - p_2) + \sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \cdot z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

7. V Ostravě byla zjištěna hmotnost 17 novorozenců s průměrnou hmotností 2870 g a výběrovou směrodatnou odchylkou 840 g. V Opavě bylo sledováno 20 novorozenců, s průměrnou hmotností 3105 g a směrodatnou odchylkou 875 g. Jsou za předpokladu normálního rozdělení obou základních souborů průměrné hmotnosti v obou městech stejné? Zjistěte s použitím hladiny spolehlivosti 0,05.

Řešení:

Známe výběrové charakteristiky:

$$x_1 = 2870 \text{ g}$$

$$s_1 = 840 \text{ g}$$

$$n_1 = 17$$

$$x_2 = 3105 \text{ g}$$

$$s_2 = 875 \text{ g}$$

$$n_2 = 20$$

To, že se oba výběrové průměry liší neznamena, že se liší i střední hodnoty ($\mu_1 = \mu_2$?).

Pro sestrojení intervalu spolehlivosti pro rozdíl průměrů je potřeba vědět, jestli se populační rozptyly rovnají nebo ne.

95% interval spolehlivosti pro podíl rozptylů:

$$P\left(\frac{1}{f_{1-\frac{\alpha}{2}, n_1-1, n_2-1}} \frac{s_1^2}{s_2^2} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{1}{f_{\frac{\alpha}{2}, n_1-1, n_2-1}} \frac{s_1^2}{s_2^2}\right) = 1 - \alpha$$

Řešení:

Dosadíme:
$$P\left(\frac{1}{2,59} \frac{840^2}{875^2} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{1}{0,37} \frac{840^2}{875^2}\right) = 0,95$$

$$P\left(0,36 < \frac{\sigma_1^2}{\sigma_2^2} < 2,49\right) = 0,95$$

Protože je interval v rozmezí (0,36; 2,49), obsahuje i hodnotu jedna => nelze usoudit, že populační rozptyly nebyly shodné => pro výpočet intervalu spolehlivosti rozdílu středních hodnot použijeme:

$$P\left((\bar{X}_1 - \bar{X}_2) - s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \cdot t_{1-\frac{\alpha}{2}, n_1+n_2-2} < (\mu_1 - \mu_2) < (\bar{X}_1 - \bar{X}_2) + s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \cdot t_{1-\frac{\alpha}{2}, n_1+n_2-2}\right) = 1 - \alpha$$

Dopočítáme charakteristiku s_p :

$$s_p = \sqrt{\frac{(17-1)840^2 + (20-1)875^2}{17+20-2}} = 859$$

Řešení:

Dosadíme:

$$P \left(\begin{array}{c} (2870 - 3105) - 859 \cdot \sqrt{\frac{1}{17} + \frac{1}{20}} \cdot 2,03 < \\ (\mu_1 - \mu_2) < \\ (2870 - 3105) + 859 \cdot \sqrt{\frac{1}{17} + \frac{1}{20}} \cdot 2,03 \end{array} \right) = 0,95$$

$$P(-810 < (\mu_1 - \mu_2) < 340) = 0,95$$

S pravděpodobností 95 % se bude rozdíl středních hodnot hmotností pohybovat v intervalu $(-810; 340)$. Interval obsahuje i hodnotu **nula**. Rozdíl mezi středními hodnotami tedy není významný. Nelze říci, že by průměrná porodní hmotnost v Ostravě a Opavě byla rozdílná.

Řešení:

V případě, že by rozdíl mezi středními hodnotami hmotnosti v Ostravě a Opavě byl záporný ($\mu_1 - \mu_2 < 0$), znamenalo by to, že novorozenci v Ostravě mají nižší hmotnost než novorozenci v Opavě ($\mu_1 < \mu_2$).

Analogicky v případě, když by byl rozdíl mezi středními hodnotami hmotnosti v Ostravě a Opavě kladný ($\mu_1 - \mu_2 > 0$), znamenalo by to, že novorozenci v Ostravě mají vyšší hmotnost než novorozenci v Opavě ($\mu_1 > \mu_2$).

8. Diskety dvou velkých výrobců - Sonik a 5M byly podrobeny zkoušce kvality. Diskety obou výrobců jsou baleny po 20-ti kusech. Ve 40-ti balíčcích fy Sonik bylo nalezeno 24 vadných disket, ve 30-ti balíčcích 5M bylo nalezeno 14 vadných disket. Určete 95%-ní interval spolehlivosti pro rozdíl v procentu vadných disket v celkové produkci firem Sonik a 5M.

Řešení:

$$P\left((p_1 - p_2) - \sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \cdot z_{1-\frac{\alpha}{2}} < (\pi_1 - \pi_2) < (p_1 - p_2) + \sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \cdot z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

Hladina významnosti ze zadání: $\alpha = 1 - 0,95 = 0,05$

$$\frac{\alpha}{2} = 0,025, \quad 1 - \frac{\alpha}{2} = 0,975; \quad z_{0,975} = 1,96$$

Výběrové soubory:

Sonik:

$$x_1 = 24$$

$$n_1 = 40 \cdot 20 = 800$$

$$p_1 = \frac{24}{800} = 0,030$$

výběrový podíl
vadných disket fy
Sonik

5M:

$$x_2 = 14$$

$$n_2 = 30 \cdot 20 = 600$$

$$p_2 = \frac{14}{600} = 0,023$$

výběrový podíl
vadných disket fy
5M

$$p = \frac{24 + 14}{800 + 600} = 0,027$$

Řešení:

Dosadíme:

$$P((0,007) - 0,017 < (\pi_1 - \pi_2) < (0,007) + 0,017) = 0,95$$

$$P(-0,010 < (\pi_1 - \pi_2) < 0,024) = 0,95$$

$$P(-1,0 \% < (\pi_1 - \pi_2) < 2,4 \%) = 0,95$$

S 95%-ní spolehlivostí můžeme tvrdit, že rozdíl mezi podílem vadných disket firmy Sonik a podílem vadných disket firmy 5M je v rozmezí $-1,0 \%$ a $2,4\%$. Tzn., že nemůžeme říci, které diskety jsou kvalitnější.

Řešení:

V případě, že by rozdíl mezi podílem vadných disket firmy Sonik a podílem vadných disket firmy 5M byl záporný ($\pi_1 - \pi_2 < 0$), znamenalo by to, že diskety firmy Sonik jsou kvalitnější (obsahují menší podíl vadných) než diskety firmy 5M ($\pi_1 < \pi_2$).

Obdobně v případě, že by rozdíl mezi podílem vadných disket firmy Sonik a podílem vadných disket firmy 5M byl kladný ($\pi_1 - \pi_2 > 0$), znamenalo by to, že diskety firmy Sonik mají horší kvalitu (obsahují větší podíl vadných) než diskety firmy 5M ($\pi_1 > \pi_2$).

Test

1. Chceme-li najít nejlepší možný odhad směrodatné odchylky vybrané vlastnosti nekonečné populace, měli bychom

- a) použít co možná největší výběrový soubor,,
- b) použít co možná nejmenší výběrový soubor,
- c) zjistit hodnotu sledované vlastnosti u všech prvků populace,
- d) použít výběrový soubor o rozsahu nejvýše 10 000 prvků populace.

1. Chceme-li najít nejlepší možný odhad směrodatné odchylky vybrané vlastnosti nekonečné populace, měli bychom

- ☒ a) použít co možná největší výběrový soubor,,
- ☐ b) použít co možná nejmenší výběrový soubor,
- ☐ c) zjistit hodnotu sledované vlastnosti u všech prvků populace,
- ☐ d) použít výběrový soubor o rozsahu nejvýše 10 000 prvků populace.

2. Chceme-li najít nejlepší možný odhad směrodatné odchylky vybrané vlastnosti populace o rozsahu 50 000 jednotek (prvků), pak by rozsah výběru neměl překročit

- a) 49 999 jednotek,
- b) 10 000 jednotek,
- c) 5 000 jednotek,
- d) 2 500 jednotek,
- e) 1 000 jednotek.

2. Chceme-li najít nejlepší možný odhad směrodatné odchylky vybrané vlastnosti populace o rozsahu 50 000 jednotek (prvků), pak by rozsah výběru neměl překročit

~~a)~~ 49 999 jednotek,

~~b)~~ 10 000 jednotek,

~~c)~~ 5 000 jednotek,

 d) 2 500 jednotek,

~~e)~~ 1 000 jednotek.

3. Doplňte:

- a) Průměr je (náhodná veličina, konstanta).
- b) Střední hodnota je (výběrová, populační) charakteristika.
- c) Odhadujeme-li populační charakteristiku jedním číslem, hovoříme o (bodovém, intervalovém) odhadu.
- d) Řekneme, že odhad je (nestranný, vydatný, konzistentní), jestliže se jeho střední hodnota rovná hledanému parametru.

3. Doplňte:

- a) Průměr je náhodná veličina (náhodná veličina, konstanta).
- b) Střední hodnota je (výběrová, populační) charakteristika.
- c) Odhadujeme-li populační charakteristiku jedním číslem, hovoříme o (bodovém, intervalovém) odhadu.
- d) Řekneme, že odhad je (nestranný, vydatný, konzistentní), jestliže se jeho střední hodnota rovná hledanému parametru.

3. Doplňte:

- a) Průměr je **náhodná veličina** (náhodná veličina, konstanta).
- b) Střední hodnota je **populační** (výběrová, populační) charakteristika.
- c) Odhadujeme-li populační charakteristiku jedním číslem, hovoříme o (bodovém, intervalovém) odhadu.
- d) Řekneme, že odhad je (nestranný, vydatný, konzistentní), jestliže se jeho střední hodnota rovná hledanému parametru.

3. Doplňte:

- a) Průměr je **náhodná veličina** (náhodná veličina, konstanta).
- b) Střední hodnota je **populační** (výběrová, populační) charakteristika.
- c) Odhadujeme-li populační charakteristiku jedním číslem, hovoříme o **bodovém** (bodovém, intervalovém) odhadu.
- d) Řekneme, že odhad je (nestranný, vydatný, konzistentní), jestliže se jeho střední hodnota rovná hledanému parametru.

3. Doplňte:

- a) Průměr je**náhodná veličina**..... (náhodná veličina, konstanta).
- b) Střední hodnota je**populační**..... (výběrová, populační) charakteristika.
- c) Odhadujeme-li populační charakteristiku jedním číslem, hovoříme o**bodovém**..... (bodovém, intervalovém) odhadu.
- d) Řekneme, že odhad je**nestranný**..... (nestranný, vydatný, konzistentní), jestliže se jeho střední hodnota rovná hledanému parametru.

3. Doplňte:

- e) Nestranný odhad, jehož rozptyl je
(nejmenší, největší) mezi rozptyly všech
nestranných odhadů příslušného parametru, se
nazývá nejlepší nestranný (vydatný, eficientní)
odhad.
- f) Mějme náhodný výběr. S rostoucí spolehlivostí
odhadu se obvykle intervalové odhady populačních
parametrů (zužují, rozšiřují).
- g) S rostoucí spolehlivostí odhadu $1 - \alpha$
(roste, klesá) hladina významnosti α .

3. Doplňte:

- e) Nestranný odhad, jehož rozptyl je **nejmenší**
(nejmenší, největší) mezi rozptyly všech
nestranných odhadů příslušného parametru, se
nazývá nejlepší nestranný (vydatný, eficientní)
odhad.
- f) Mějme náhodný výběr. S rostoucí spolehlivostí
odhadu se obvykle intervalové odhady populačních
parametrů (zužují, rozšiřují).
- g) S rostoucí spolehlivostí odhadu $1 - \alpha$
(roste, klesá) hladina významnosti α .

3. Doplňte:

- e) Nestranný odhad, jehož rozptyl je **nejmenší**
(nejmenší, největší) mezi rozptyly všech
nestranných odhadů příslušného parametru, se
nazývá nejlepší nestranný (vydatný, eficientní)
odhad.
- f) Mějme náhodný výběr. S rostoucí spolehlivostí
odhadu se obvykle intervalové odhady populačních
parametrů **rozšiřují** (zužují, rozšiřují).
- g) S rostoucí spolehlivostí odhadu $1 - \alpha$
(roste, klesá) hladina významnosti α .

3. Doplňte:

- e) Nestranný odhad, jehož rozptyl je **nejmenší**
(nejmenší, největší) mezi rozptyly všech
nestranných odhadů příslušného parametru, se
nazývá nejlepší nestranný (vydatný, eficientní)
odhad.
- f) Mějme náhodný výběr. S rostoucí spolehlivostí
odhadu se obvykle intervalové odhady populačních
parametrů **rozšiřují** (zužují, rozšiřují).
- g) S rostoucí spolehlivostí odhadu $1 - \alpha$ **klesá**
(roste, klesá) hladina významnosti α .

3. Doplňte:

- h) Při dané spolehlivosti odhadu se obvykle intervalové odhady populačních parametrů s rostoucím rozsahem výběru (zužují, rozšiřují).
- i) V technické praxi se obvykle volí spolehlivost odhadu rovna (0,80; 0,90; 0,95; 0,99; 0,20; 0,10; 0,05; 0,01).
- j) V technické praxi se obvykle volí hladina významnosti rovna (0,80; 0,90; 0,95; 0,99; 0,20; 0,10; 0,05; 0,01).
- k) Horní mez pravostranného intervalového odhadu je (stejná jako, menší než, větší než) horní mez příslušného oboustranného odhadu.

3. Doplňte:

- h) Při dané spolehlivosti odhadu se obvykle intervalové odhady populačních parametrů s rostoucím rozsahem výběru **zužují** (zužují, rozšiřují).
- i) V technické praxi se obvykle volí spolehlivost odhadu rovna (0,80; 0,90; 0,95; 0,99; 0,20; 0,10; 0,05; 0,01).
- j) V technické praxi se obvykle volí hladina významnosti rovna (0,80; 0,90; 0,95; 0,99; 0,20; 0,10; 0,05; 0,01).
- k) Horní mez pravostranného intervalového odhadu je (stejná jako, menší než, větší než) horní mez příslušného oboustranného odhadu.

3. Doplňte:

- h) Při dané spolehlivosti odhadu se obvykle intervalové odhady populačních parametrů s rostoucím rozsahem výběru **zužují** (zužují, rozšiřují).
- i) V technické praxi se obvykle volí spolehlivost odhadu rovna **0,95** (0,80; 0,90; 0,95; 0,99; 0,20; 0,10; 0,05; 0,01).
- j) V technické praxi se obvykle volí hladina významnosti rovna (0,80; 0,90; 0,95; 0,99; 0,20; 0,10; 0,05; 0,01).
- k) Horní mez pravostranného intervalového odhadu je (stejná jako, menší než, větší než) horní mez příslušného oboustranného odhadu.

3. Doplňte:

- h) Při dané spolehlivosti odhadu se obvykle intervalové odhady populačních parametrů s rostoucím rozsahem výběru **zužují** (zužují, rozšiřují).
- i) V technické praxi se obvykle volí spolehlivost odhadu rovna **0,95** (0,80; 0,90; 0,95; 0,99; 0,20; 0,10; 0,05; 0,01).
- j) V technické praxi se obvykle volí hladina významnosti rovna **0,05** (0,80; 0,90; 0,95; 0,99; 0,20; 0,10; 0,05; 0,01).
- k) Horní mez pravostranného intervalového odhadu je (stejná jako, menší než, větší než) horní mez příslušného oboustranného odhadu.

3. Doplňte:

- h) Při dané spolehlivosti odhadu se obvykle intervalové odhady populačních parametrů s rostoucím rozsahem výběru **zužují** (zužují, rozšiřují).
- i) V technické praxi se obvykle volí spolehlivost odhadu rovna **0,95** (0,80; 0,90; 0,95; 0,99; 0,20; 0,10; 0,05; 0,01).
- j) V technické praxi se obvykle volí hladina významnosti rovna **0,05** (0,80; 0,90; 0,95; 0,99; 0,20; 0,10; 0,05; 0,01).
- k) Horní mez pravostranného intervalového odhadu je **menší než** (stejná jako, menší než, větší než) horní mez příslušného oboustranného odhadu.