8 ODHADY PARAMETRŮ ZÁKLADNÍHO SOUBORU

Čas ke studiu kapitoly: 90 minut

Cíl: Po prostudování tohoto odstavce budete:

- rozumět pojmům: bodový odhad, intervalový odhad
- znát vlastnosti bodového odhadu
- umět zkonstruovat intervalové odhady pro vybrané parametry normálního rozdělení: střední hodnotu, rozptyl, směrodatnou odchylku, relativní četnost (podíl), rozdíl dvou středních hodnot a rozdíl relativních četností (podílů)

Výklad:

8.1 Základní soubor, výběrový soubor

Náhodnou veličinu *X*, jejíž hodnoty při realizaci náhodného pokusu pozorujeme, můžeme popsat pomocí různých číselných charakteristik (v souvislosti s náhodnou veličinou hovoříme častěji o **parametrech základního souboru (populace),** popř. o **parametrech rozdělení** náhodné veličiny). K parametrům základního souboru patří: střední hodnota μ, rozptyl σ^2, směrodatná odchylka σ, relativní četnost π, atd… Parametry populace jsou **konstantní hodnoty** (pro určitou náhodnou veličinu).

Ve výběrovém souboru (výběru ze základního souboru (populace)) lze najít příslušné protějšky parametru populace. Říká se jim **výběrové charakteristiky** a jejich **hodnoty se mění** podle aktuálního výběru.

Přehled nejpoužívanějších parametrů populace a příslušných výběrových charakteristik, včetně jejich značení je uveden v následující tabulce:

Základní soubor (populace)	střední hodnota μ (EX)	rozptyl σ^2	směrodatná odchylka σ	podíl (relativní četnost) π
Výběrový soubor (výběr)	průměr $\bar{x}$	výběrový rozptyl s^2	výběrová směrodatná odchylka s	výběrová relativní četnost p

Z pravděpodobnostního hlediska mají výběrové charakteristiky charakter náhodných veličin (na základě různosti jednotlivých výběrů, nelze hodnoty výběrových charakteristik určit předem). Každá výběrová charakteristika má tedy svoje rozdělení pravděpodobnosti, které se nazývá **výběrové rozdělení**. Známe-li výběrové rozdělení, dokážeme **odhadnout** příslušný parametr základního souboru.

Průvodce studiem

Nyní se pokusíme výše uvedenou terminologii propojit s praxí. Na následujícím konkrétním příkladu se pokusíme ukázat rozdíl mezi výběrem (parametry výběru) a populací (parametry populace). Dále bychom si na tomto příkladu měli ujasnit, proč potřebujeme parametry populace odhadovat:

Mějme např. denní produkci tyčí (o daném průměru) ocelářské firmy - 600 ocelových tyčí. Naším cílem je určit střední hodnotu tažnosti těchto tyčí.

Populace je v tomto případě tvořena všemi tyčemi z denní produkce a střední hodnota tažnosti je jeden z parametrů této populace. Je zřejmé, že požadovaný úkol je neřešitelný – k jeho splnění bychom museli určit tažnost všech tyčí (destruktivní zkouška) a z naměřených hodnot určit průměr. To je v praxi neproveditelné. Jediné možné řešení je – pokusit se o **odhad** tohoto parametru.

Jestliže vybereme náhodně například 10 tyčí (10 tyčí můžeme „obětovat") a určíme jejich průměrnou tažnost, určujeme průměr a je zřejmé, že jeho hodnota závisí na konkrétním výběru (vybereme-li jiných 10 tyčí, jejich průměrná tažnost bude jiná než v předcházejícím případě). Průměr je **výběrovou charakteristiku** denní produkce tyčí a je tedy **náhodnou veličinou.** Proto mu můžeme **přiřadit** nějaké **rozdělení** (viz. Limitní věty). Známe-li rozdělení průměru, můžeme vytvářet různé úsudky o střední hodnotě. Např. dokážeme určit jaká je pravděpodobnost, že střední hodnota leží v námi zvoleném intervalu.

Výklad:

8.2 Bodový a intervalový odhad

V této podkapitole se dozvíte, jak na základě znalosti výběrového souboru (a jeho charakteristik) najít co nejlepší odhad parametrů základního souboru.

Nejdříve si musíme ujasnit, co si pod pojmem „nejlepší odhad" představujeme.

Z metodického hlediska používáme dva typy odhadů parametrů:

- **bodový odhad**, kdy parametr základního souboru aproximujeme jediným číslem

a

- **intervalový odhad**, kdy tento parametr aproximujeme intervalem, v němž s velkou pravděpodobností daný parametr leží

O tom, který z výše uvedených odhadů použijeme, rozhoduje konkrétní situace, v níž se nacházíme. Pokud potřebujeme hledaný parametr vyjádřit jedinou hodnotou (většinou v případech, kdy jej budeme používat v dalších výpočtech), použijeme bodový odhad. Potřebujeme-li přesnější odhad, použijeme intervalový odhad, tzn., že najdeme tzv. interval spolehlivosti.

Interval spolehlivosti (**konfidenční interval**) je interval, v němž hledaný parametr leží s danou pravděpodobností. Této pravděpodobnosti se říká **spolehlivost odhadu**.

Příklad:
90%-ní interval spolehlivosti pro střední hodnotu je interval, v němž střední hodnota leží s pravděpodobností 90%.

Je zřejmé, že čím vyšší spolehlivost odhadu požadujeme, tím širší interval spolehlivosti bude (hledaná hodnota se v něm musí nacházet s vyšší pravděpodobností). Bohužel to však ubírá na jeho vypovídací schopnosti, jeho významnost klesá. (Uvědomte si jaká je vypovídací schopnost informace, že průměrný věk všech lidí na zemi leží se 100%-ní spolehlivostí v intervalu (0; 142) let.) Proto v praxi vždy hledáme kompromis mezi spolehlivostí a **významností**.

Označíme-li **spolehlivost odhadu (1-α)**, pak **α** se nazývá **hladinou významnosti**. S rostoucí spolehlivosti odhadu klesá hladina významnosti. V technické praxi se spolehlivost odhadu se volí nejčastěji 95% nebo 99% (hladina významnosti tedy bývá 5% nebo 1%).
Při konstrukci bodových a intervalových odhadů budeme používat následující pojmy:

Nechť máme náhodný výběr $(X_1, \ldots, X_n)$ z rozdělení s distribuční funkcí $F(x,\theta)$ s **neznámým parametrem θ** . Množinu všech uvažovaných hodnot parametru θ nazýváme **parametrický prostor**. Statistiku $\theta = T(X_1, \ldots, X_n)$, která bude sloužit pro účely odhadu neznámého parametru θ, budeme nazývat **odhadem** parametru θ, její pozorovanou hodnotu pak **bodovým odhadem** θ.

8.3 Vlastnosti „dobrého“ bodového odhadu

„Dobrý“ (věrohodný) odhad musí splňovat určité vlastnosti. Mezi základní vlastnosti věrohodných odhadů patří:

- nestrannost (nevychýlenost, nezkreslenost)
- vydatnost (eficience)
- konzistence
- dostatečnost

8.3.1 Nestranný odhad

Řekneme, že odhad je **nestranný**, jestliže se jeho střední hodnota rovná hledanému parametru ($E\hat{\theta} = \theta$). Znamená to, že tento odhad systematicky nenadhodnocuje ani nepodhodnocuje odhadovaný parametr.

Slabší formou nestrannosti je **asymptotická nestrannost**. Říkáme, že odhad je asymptoticky nestranný pokud: $\lim_{n\to\infty} E\hat{\theta} = \theta$

Příklady nestranných odhadů:

- $\overline{X}$ je nestranným odhadem střední hodnoty (limitní věty)
- Výběrová relativní četnost p je nestranným odhadem relativní četnosti (podílu) π
- V případě náhodného výběru z normálního rozdělení je výběrový rozptyl s^2 nestranným odhadem rozptylu σ^2

Je třeba říci, že existuje mnoho dobrých odhadů, které nejsou nestranné.

8.3.2 Vydatný (eficientní) odhad

Nestrannost sama o sobě nezaručuje, že je odhad „dobrý“. Rádi bychom dosáhli také toho, aby bodové odhady byly rozloženy co nejtěsněji kolem odhadovaného parametru. Pokud budeme mít dva nestranné odhady $\hat{\theta}_1$ a $\hat{\theta}_2$, vybereme si ten, který bude mít menší rozptyl. Tato vlastnost se nazývá **vydatnost** (eficience).

Jestliže pro dva nestranné odhady $\hat{\theta}_1$ a $\hat{\theta}_2$ platí $D\hat{\theta}_1 > D\hat{\theta}_2$, potom je **relativní eficience** odhadu $\hat{\theta}_1$ vzhledem k odhadu $\hat{\theta}_2$ dána podílem $D\hat{\theta}_1 / D\hat{\theta}_2$, což je číslo mezi 0 a 1.

Nestranný odhad, jehož rozptyl je nejmenší mezi všemi nestrannými odhady příslušného parametru, se nazývá **nejlepší nestranný (eficientní) odhad**.

Příklady nejlepších nestranných odhadů:

- $\overline{X}$ je nejlepším nestranným odhadem střední hodnoty (limitní věty)
- Výběrová relativní četnost p je nejlepším nestranným odhadem rel. četnosti (podílu) π
- V případě náhodného výběru z normálního rozdělení je výběrový rozptyl s^2 nejlepším nestranným odhadem rozptylu σ^2

8.3.3 Konzistentní odhad

Další žádoucí vlastností dobrého odhadu je konzistence. Odhad je konzistentní pokud se s rostoucím rozsahem výběru (*n*) zpřesňuje, k čemuž dochází pokud:

a) $\hat{\theta}$ je asymptoticky nestranný, tj. $E\hat{\theta} \to \theta$

b) $\lim\limits_{n\to\infty} D\hat{\theta} = \theta$

Vlastnost b) říká, že se s rostoucím *n* (rozsahem výběru) rozdělení $\hat{\theta}$ zužuje kolem hledaného parametru.

Příklady konzistentních odhadů:

- $\overline{X}$ je konzistentním odhadem střední hodnoty, protože $D\overline{X} = \dfrac{\sigma^2}{n} \to 0 \quad \text{pro} \quad n \to \infty$

- Výběrová relativní četnost p je konzistentním odhadem rel. četnosti (podílu) π, protože $Dp = \dfrac{\pi(1-\pi)}{n} \to 0 \ pro \ n \to \infty$

8.3.4 Dostatečný (postačující) odhad

Odhad parametru je **dostatečný**, jestliže obsahuje veškerou informaci o sledovaném parametru, kterou může výběrový soubor poskytnout. Znamená to, že žádný jiný parametr neobsahuje větší množství informace o výběrovém souboru.

Příklady dostatečných odhadů:

- $\overline{X}$ je dostatečným odhadem střední hodnoty, protože pro jeho výpočet jsou použity všechny hodnoty výběrového souboru (nese největší informaci, srovnejte například s mediánem)

- Výběrová relativní četnost p je konzistentním odhadem rel. četnosti (podílu) π, protože pro její výpočet jsou použity všechny hodnoty výběrového souboru

8.3.5 Chyba bodového odhadu

Bodový odhad je náhodná veličina. I v případě, kdy bude bodový odhad splňovat všechny výše uvedené požadavky je zřejmé, že jeho hodnota, vypočtena na základě jednoho výběru, bude odlišná od skutečné hodnoty parametru populace. Důsledkem této odlišnosti je tzv. **výběrová chyba** $(\theta - \theta)$, která určuje velikost chyby při odhadu na základě jednoho výběrového souboru. Je-li bodový odhad θ nezkresleným odhadem parametru θ, pak měřítkem přesnosti odhadu je jeho směrodatná odchylka $\sqrt{D\theta}$, pro níž se často používá název **střední chyba**. Střední chyba udává „průměrnou" chybu odhadů určených z různých výběrových souboru daného rozsahu.

8.4 Konstrukce intervalových odhadů

V praktických aplikacích častěji určujeme odhad příslušného parametru pomocí intervalového odhadu. Tento odhad je reprezentován intervalem (T_D; T_H), v němž hledaný parametr leží s předem určenou pravděpodobností (spolehlivostí), kterou označujeme (1-α).

Intervaly spolehlivosti konstruujeme jako **jednostranné (**důležitá je pouze jedna mez, odhadujeme-li například délku života nějakého zařízení, je pro nás důležitá pouze dolní mez.) nebo **dvoustranné**.

8.4.1 Jednostranné intervaly spolehlivosti

U jednostranných intervalů se udává pouze dolní mez (T_D) nebo pouze horní mez (T_H) odhadu.

Je-li dána pouze dolní mez odhadu T_D (T_H = ∞), mluvíme o **levostranném intervalu spolehlivosti** a platí pro něj:

$$P(\theta > T_D) = 1 - \alpha$$

Interval $(T_D; \infty)$ se pak nazývá **100.(1-α)%-ní levostranný interval spolehlivosti pro parametr θ**.

Je-li dána pouze horní mez odhadu T_H (T_D = -∞), mluvíme o **pravostranném intervalu spolehlivosti** a platí pro něj:

$$P(\theta < T_H) = 1 - \alpha$$

Interval $(-\infty; T_H)$ se pak nazývá **100.(1-α)%-ní pravostranný interval spolehlivosti pro parametr θ**.

8.4.2 Oboustranný interval spolehlivosti

Zajímají-li nás obě meze odhadu (dolní i horní), konstruujeme oboustranný interval spolehlivosti. Většinou tyto meze určujeme tak, aby platilo, že pravděpodobnost, že parametr populace leží pod dolní mezí byla stejná jako pravděpodobnost, že leží nad horní mezí a byla rovna α/2:

$$P(\theta < T_D) = P(\theta \geq T_H) = \frac{\alpha}{2}$$

Tyto dvě podmínky zaručují, že:

$$P(T_D \le \theta < T_H) = 1 - \alpha$$

Interval (T_D, T_H) se pak nazývá **100.(1-α) %-ní interval spolehlivosti pro parametr θ**.

Obecné metody konstrukce intervalů spolehlivosti jsou značně náročné. Pro naše účely se omezíme na **intervaly spolehlivosti pro parametry normálního rozdělení**, které jsou dobře prozkoumané (i proto se tak často setkáme s požadavkem na normalitu zpracovávaných dat). V případě, že základní soubor nemá normální rozdělení, musíme přistoupit k tzv. **neparametrickým metodám odhadu** (ty však nejsou obsahem těchto materiálů).

8.5 Interval spolehlivosti pro střední hodnotu

Nejlepším (nestranným, vydatným, konzistentním a dostatečným) bodovým odhadem střední hodnoty μ je průměr $\bar{x}$. Nyní si ukážeme jak najít intervalový odhad střední hodnoty.

8.5.1 Odhad střední hodnoty μ, známe-li směrodatnou odchylku σ

Předpokládejme, že sledovaná náhodná veličina X má normální rozdělení, jehož rozptyl σ^2 známe. Zvolme výběrový soubor z dané populace. Nechť má tento výběrový soubor rozsah n a průměr $\bar{x}$.

Využijeme poznatku o asymptotickém rozdělení průměru (viz. Lindebergova-Lévyho věta (kap. 7.4.1)). Víme, že pro dostatečně velký rozsah výběru $(n \to \infty)$ je rozdělení průměru asymptoticky normální se střední hodnotou μ a rozptylem σ^2/n:

$$\bar{X} \to N\left(\mu; \frac{\sigma^2}{n}\right)$$

Definujeme-li náhodnou veličinu Z jako:

$$Z = \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} = \frac{\bar{X} - \mu}{\sigma} \cdot \sqrt{n},$$

víme, že Z má normované normální rozdělení: $Z \to N(0;1)$

Nechť $z_{\frac{\alpha}{2}}$ a $z_{1-\frac{\alpha}{2}}$ jsou $100.\left(\frac{\alpha}{2}\right)$%-ní a $100.1-\left(\frac{\alpha}{2}\right)$%-ní kvantily normovaného normálního rozdělení. Pak můžeme tvrdit, že:

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

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

Úpravou tohoto vztahu, při využití vlastnosti symetrie normovaného normálního rozdělení $\left(z_{\frac{\alpha}{2}} = -z_{1-\frac{\alpha}{2}} \right)$ pak dostaneme požadovaný **oboustranný interval**:

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

Obdobně bychom mohli ukázat, že **levostranný interval** spolehlivosti je vymezen vztahem:

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

a **pravostranný interval** najdeme podle vztahu:

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

Všimněte si, že **s rostoucím rozsahem náhodného výběru** (n) šířka intervalu klesá, takže **se odhad zpřesňuje** (při konstantní spolehlivosti). Naopak, při konstantním rozsahu výběru se s rostoucí spolehlivostí šířka intervalu zvětšuje.

Výše 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. (viz. 8.5.2)

Průvodce studiem:

V tomto průvodci studiem najdete **podrobné odvození oboustranného intervalu spolehlivosti pro střední hodnotu (známe-li σ)**:

$$Z = \frac{\overline{X} - \mu}{\sigma} \cdot \sqrt{n}; \quad Z \to N(0;1)$$

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

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

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

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

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

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

Výklad:

8.5.2 Odhad střední hodnoty μ, neznáme-li směrodatnou odchylku σ

V praxi se většinou setkáváme s tím, že směrodatnou odchylku σ neznáme. Pokud nemáme ani dostatečný rozsah výběru $(n \geq 30)$, nemůžeme použít výše odvozené intervaly spolehlivosti pro střední hodnotu. Je i v takovém případě možné najít intervalový odhad střední hodnoty?

S ohledem na zadání vezmeme opět vhodné výběrové rozdělení – teď to bude takové, které neobsahuje σ a přitom z něj můžeme získat interval spolehlivosti pro μ: z kapitoly 6.10. víme, že náhodná veličina definovaná jako.

$$T_{n-1} = \frac{\overline{X} - \mu}{s} \cdot \sqrt{n}$$

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

$$T_{n-1} \rightarrow t_{n-1}$$

Z toho plyne, že můžeme zapsat následující pravděpodobnost:

$$P(t_{\frac{\alpha}{2},\, n-1} < T_{n-1} < t_{1-\frac{\alpha}{2},\, n-1} = 1-\alpha$$

$$P(t_{\frac{\alpha}{2},\, n-1} < \frac{\overline{X} - \mu}{s} \cdot \sqrt{n} < t_{1-\frac{\alpha}{2},\, n-1}) = 1-\alpha\,,$$

kde $t_{\frac{\alpha}{2},\, n-1}; t_{1-\frac{\alpha}{2},\, n-1}$ jsou příslušné kvantily Studentova rozdělení s n-1 stupni volnosti.

Úpravou tohoto vztahu, při využití vlastnosti symetrie Studentova rozdělení $\left(t_{\frac{\alpha}{2},\, n-1} = -t_{1-\frac{\alpha}{2},\, n-1} \right)$ pak dostaneme požadovaný **oboustranný interval**:

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

Obdobně bychom mohli ukázat, že **levostranný interval** spolehlivosti je vymezen vztahem:

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

a **pravostranný interval** najdeme podle vztahu:

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

Víme, že pro $n\to\infty$ (vysoký počet stupňů volnosti, v praxi pro $n\geq 30$) se Studentovo t rozdělení blíží normovanému normálnímu rozdělení. Pro $n\geq 30$ tedy můžeme kvantily Studentova rozdělení nahradit kvantily normovaného normálního rozdělení a pak vztahy pro určení intervalů spolehlivosti střední hodnoty v případě neznámé směrodatné odchylky přecházejí ve vztahy pro určení intervalů spolehlivosti střední hodnoty v případě známé směrodatné odchylky, v nichž směrodatnou odchylku aproximujeme výběrovou směrodatnou odchylkou.

Řešený příklad:

Útvar kontroly podniku Edison testoval životnost žárovek. Kontroloři vybrali z produkce podniku náhodně 50 žárovek a došli k závěru, že průměrná doba života těchto 50-ti žárovek je 950 hodin a příslušná výběrová směrodatná odchylka doby života je 100 hodin. Určete 95%-ní interval spolehlivosti životnosti žárovek firmy Edison.

Řešení:

Chceme najít 95%-ní interval spolehlivosti pro střední hodnotu životnosti žárovek firmy Edison, přičemž neznáme směrodatnou odchylku životnosti těchto žárovek. Máme k dispozici informace pocházející z výběru o rozsahu 50 žárovek, tj. rozsah výběru je vyšší než 30 a proto k nalezení příslušného intervalového odhadu můžeme použít následující vztah (jde o intervalový odhad střední hodnoty pro známé σ, kde jsme položili σ=s) :

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

Spolehlivost intervalového odhadu: $1-\alpha=0{,}95$

$\Rightarrow$ Hladina významnosti: $\alpha=1-0{,}95=0{,}05$

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

$\Rightarrow \qquad z_{0,975} = 1{,}96 \quad$ (viz. Tabulka 1)

Výběrový soubor: $\bar{x} = 950 \; hodin$
$s = 100 \; hodin$
$n = 50$

Dosadíme: $$P\left(950 - \frac{100}{\sqrt{50}} \cdot 1{,}96 < \mu < 950 + \frac{100}{\sqrt{50}} \cdot 1{,}96\right) \qquad = 0{,}95$$

Po úpravě dostáváme: $$P(922{,}3 < \mu < 977{,}7) \qquad = 0{,}95$$

Tzn., že s 95%-ní spolehlivostí můžeme tvrdit, že životnost žárovek firmy Edison se pohybuje v rozmezí 922 hodin 18 minut až 977 hodin 42 minut.

Řešený příklad:

Obchodní řetězec TETO si v dubnu 2006 zadal studii týkající se počtu zákazníku 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 v TETO Poruba (12:00-18:00) hodin**
2.5.2006	3756
9.5.2006	2987
16.5.2006	3042
23.5.2006	4206
30.5.2006	3597

a) Objasněte, proč jsme nezískali výběrový soubor o rozsahu alespoň 30 hodnot a jaké jsou důsledky volby výběru o malém rozsahu.
b) Určete pro managment řetězce TETO 95%-ní interval spolehlivosti počtu zákazníku v prodejně TETO Poruba v pátek odpoledne.

Řešení:

ada) Pro získání výběru o rozsahu minimálně 30 hodnot bychom museli danou prodejnu sledovat minimálně 30 pátku (tj. déle než půl roku), což by vedlo jednak k zvýšení finanční náročnosti studie, jednak bychom museli dlouho čekat na výsledky. Z těchto důvodu jsme zvolili menší rozsah výběru (n=5) odpovídající měsíčnímu sledování prodejny. Nevýhodou malého rozsahu výběru je nízká přesnost odhadu (poměrně široký interval).

adb) Určujeme intervalový odhad střední hodnoty s neznámou směrodatnou odchylkou a malým rozsahem výběru, proto pro jeho výpočet použijeme následující vztah:

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

Spolehlivost intervalového odhadu: $1-\alpha=0{,}95$

$\Rightarrow$ Hladina významnosti: $\alpha=1-0{,}95=0{,}05$

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

$\Rightarrow$ $t_{0{,}975,\,4}=2{,}78$ (viz. Tabulka 2)

Výběrový soubor:

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

$$s^2=\frac{\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2}{n-1}=\frac{(3756-3517{,}6)^2+\ldots+(3597-3517{,}6)^2}{4}=261191{,}3\quad\Rightarrow\quad s=511{,}1$$

$n=5$

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

Po úpravě dostáváme: $P(2882{,}2<\mu<4153{,}0)=0{,}95$

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.

Výklad:

8.6 Interval spolehlivosti pro rozptyl

Nejlepším (nestranným, vydatným, konzistentním a dostatečným) bodovým odhadem rozptylu σ^2 je výběrový rozptyl s^2.

Intervalový odhad rozptylu σ^2 se hledá jinak v případě že známe střední hodnotu populace (základního souboru) a jinak, když tuto střední hodnotu neznáme. Protože znalost střední hodnoty μ při neznalosti rozptylu σ^2 není příliš reálná, omezíme se pouze na druhý případ.

Předpokládejme, že sledovaná náhodná veličina X má normální rozdělení. Zvolme výběrový soubor z dané populace. Nechť má tento výběrový soubor rozsah n a výběrový rozptyl s^2.

Z vlastností rozdělení Chí-kvadrát (kap. 6.9) víme, že definujeme-li si náhodnou veličinu χ jako:

$$\chi = \frac{(n-1)s^2}{\sigma^2},$$

pak má tato náhodná veličina rozdělení Chí-kvadrát s (n-1) stupni volnosti: $\chi \to \chi^2_{n-1}$

Z toho plyne, že můžeme zapsat následující pravděpodobnost:

$$P(x_{\frac{\alpha}{2},\, n-1} < \chi < x_{1-\frac{\alpha}{2},\, n-1}) = 1-\alpha$$

$$P(x_{\frac{\alpha}{2},\, n-1} < \frac{(n-1)s^2}{\sigma^2} < x_{1-\frac{\alpha}{2},\, n-1}) = 1-\alpha,$$

kde $x_{\frac{\alpha}{2},\, n-1}$; $x_{1-\frac{\alpha}{2},\, n-1}$ jsou příslušné kvantily χ^2 rozdělení s *n*-1 stupni volnosti.

Úpravou tohoto vztahu (pozor, rozdělení χ^2 není symetrické) pak dostaneme požadovaný **oboustranný interval**:

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

Obdobně bychom mohli ukázat, že **levostranný interval** spolehlivosti je vymezen vztahem:

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

a **pravostranný interval** najdeme podle vztahu:

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

8.7 Interval spolehlivosti pro směrodatnou odchylku

Nejlepším (nestranným, vydatným, konzistentním a dostatečným) bodovým odhadem směrodatné odchylky σ je výběrová směrodatná odchylka s.

Intervalový odhad směrodatné odchylky σ najdeme snadno uvědomíme-li si, že směrodatná odchylka je odmocninou z rozptylu. Stačí tedy upravit intervalové odhady pro rozptyl.

Opět předpokládejme, že sledovaná náhodná veličina X má normální rozdělení. Zvolme výběrový soubor z dané populace. Nechť má tento výběrový soubor rozsah n a výběrovou směrodatnou odchylku s.

Oboustranný interval spolehlivosti určíme jako:

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

Obdobně je **levostranný interval** spolehlivosti vymezen vztahem:

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

a **pravostranný interval** najdeme podle vztahu:

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

Řešený příklad:

Automat vyrábí pístové kroužky o daném průměru. Při kontrole kvality bylo náhodně vybráno 80 kroužků a vypočtena směrodatná odchylka jejich průměru 0,04mm. Odhadněte 95%-ní levostranný interval spolehlivosti pro rozptyl a směrodatnou odchylku průměru pístových kroužků.

Řešení:

Nejdříve najdeme 95%-ní levostranný interval spolehlivosti pro rozptyl. Pro jeho nalezení použije následující vztah:

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

Spolehlivost intervalového odhadu: $1-\alpha=0{,}95$

$\Rightarrow$ $x_{0{,}95;\,79}\cong 60{,}4$ (viz. Tabulka 3)

Výběrový soubor: $s^2=(0{,}04)^2=0{,}0016\; mm^2$

n = 80

Po dosazení: $P\left(\frac{79}{60{,}4}\cdot 0{,}0016<\sigma^2\right)=0{,}95$

$P\left(0{,}0021<\sigma^2\right)=0{,}95$

Jednoduchou úpravou pak získáme 95%-ní levostranný interval spolehlivosti pro směrodatnou odchylku:

$$P\left(\sqrt{0{,}0021}<\sigma\right)=0{,}95$$

$$P\left(0{,}046<\sigma\right)=0{,}95$$

S 95%-ní spolehlivostí tedy můžeme tvrdit, že rozptyl průměru pístových kroužků je větší než $2{,}2.10^{-3}$ mm^2 (resp., že s 95%-ní spolehlivostí je směrodatná odchylka průměru pístových kroužků větší než $4{,}6.10^{-2}$ mm).

Výklad:

8.8 Interval spolehlivosti pro relativní četnost (podíl)

Nejlepším (nestranným, vydatným, konzistentním a dostatečným) bodovým odhadem relativní četnosti π je výběrová relativní četnost p.

Jsou-li splněny podmínky Moivreovy-Laplaceovy věty ($n\geq 30$, popř. $n\cdot p\cdot(1-p)>9$), pak známe rozdělení relativní četnosti (podílu) (viz. kap. 7.5.1):

Je-li náhodná veličina X definována jako:

$$P_1=\frac{p-\pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}=\frac{p-\pi}{\sqrt{\pi(1-\pi)}}\cdot\sqrt{n}\,,$$

pak má náhodná veličina X normované normální rozdělení: $P_1\to N(0;1)$

Nechť $z_{\frac{\alpha}{2}}$ a $z_{1-\frac{\alpha}{2}}$ jsou $100.\left(\frac{\alpha}{2}\right)$%-ní a $100.1-\left(\frac{\alpha}{2}\right)$%-ní kvantily normovaného normálního rozdělení. Pak můžeme tvrdit, že:

$$P\left(z_{\frac{\alpha}{2}}\leq P_1<z_{1-\frac{\alpha}{2}}\right)=1-\alpha$$

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

Úpravou tohoto vztahu, při využití vlastnosti symetrie normovaného normálního rozdělení $\left(z_{\frac{\alpha}{2}} = -z_{1-\frac{\alpha}{2}} \right)$ pak dostaneme požadovaný **oboustranný interval**:

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

Uvážíme-li, že pro dostatečně velké výběry můžeme relativní četnost aproximovat výběrovou relativní četností (viz. Bernoulliho věta), můžeme tvrdit, že:

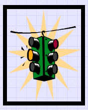
$$P\left(p - \sqrt{\frac{p \cdot (1-p)}{n}} \cdot z_{1-\frac{\alpha}{2}} < \pi < p + \sqrt{\frac{p \cdot (1-p)}{n}} \cdot z_{1-\frac{\alpha}{2}} \right) = 1-\alpha$$

Obdobně bychom mohli ukázat, že **levostranný interval** spolehlivosti je vymezen vztahem:

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

a **pravostranný interval** najdeme podle vztahu:

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



Řešený příklad:

Při kontrole data spotřeby určitého druhu masové konzervy ve skladech produktů masného průmyslu bylo náhodně vybráno 320 konzerv a zjištěno, že 59 z nich má prošlou záruční lhůtu. Stanovte 95% interval spolehlivosti pro odhad procenta konzerv s prošlou záruční lhůtou.

Řešení:

Pro nalezení 95%-ního intervalu spolehlivosti pro relativní četnost použijeme následující vztah:

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

Spolehlivost intervalového odhadu: $1-\alpha = 0{,}95$

$\Rightarrow$ Hladina významnosti: $\alpha = 1 - 0{,}95 = 0{,}05$

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

$\Rightarrow$ $z_{0{,}975} = 1{,}96$ (viz. Tabulka 1)

Výběrový soubor: $p = \dfrac{59}{320} \cong 0{,}18$

$n = 320$

Po dosazení:

$$P\left(0{,}18 - \sqrt{\frac{0{,}18\cdot(1-0{,}18)}{320}}\cdot 1{,}96 < \pi < 0{,}18 + \sqrt{\frac{0{,}18\cdot(1-0{,}18)}{320}}\cdot 1{,}96\right) = 0{,}95$$

$$P(0{,}138 < \pi < 0{,}222) = 0{,}95$$

S 95%-ní spolehlivostí můžeme tvrdit, že mezi masovými konzervami se v daném skladu nachází mezi 13,8% a 22,2% konzerv s prošlou záruční lhůtou.

Výklad:

8.9 Rozsah výběru

Ještě před zahájením výběrového šetření musíme stanovit velikost výběrového souboru. Ukázali jsme si, že velikost výběru má přímý vliv na přesnost odhadu parametrů základního souboru: čím větší rozsah výběru, tím přesnější je intervalový odhad. V řešeném příkladu věnovaném studii pro obchodní řetězec TETO jsme si však také ukázali, že ekonomické a časové důvody nás mnohdy nutí volit rozsah výběru co nejmenší. V praxi proto hledáme kompromis, který pro požadovanou přesnost výpočtu povede k co nejmenšímu rozsahu výběru.

Požadovanou přesnost výpočtu vyjadřujeme pomocí tzv. maximální **přípustné chyby odhadu Δ.** Jde o hodnotu, o kterou jsme ochotni se zmýlit oproti skutečné hodnotě odhadovaného parametru při dané spolehlivosti odhadu (hladině významnosti). Přípustná chyba odhadu je rovna polovině šířky oboustranného intervalu spolehlivosti.

8.9.1 Rozsah výběru při odhadu střední hodnoty

Obdobně jako při hledání intervalu spolehlivosti pro střední hodnotu, musíme i zde rozlišit dva případy: situaci kdy známe směrodatnou odchylku populace a situaci, kdy tuto směrodatnou odchylku neznáme.

a) Známe σ

Oboustranný intervalový odhad je dán vztahem:

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

Příslušný intervalový odhad tedy můžeme vyjádřit ve tvaru:

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

Polovina šířky oboustranného intervalu spolehlivosti a tedy přípustná chyba odhadu Δ je:

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

Požadujeme-li, aby přípustná chyba odhadu Δ dosahovala při dané spolehlivosti odhadu maximálně určité přípustné hodnoty, pak rozsah výběru určíme jako funkci této chyby:

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

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

b) Neznáme σ

Obdobně jako v předcházejícím případě bychom mohli ukázat, že přípustná chyba odhadu je:

$$\Delta = \frac{s}{\sqrt{n}} \cdot t_{1-\frac{\alpha}{2}, n-1}$$

Přípustná chyba odhadu je v tomto případě nejen funkcí hladiny významnosti a rozsahu výběru, ale závisí také na výběrové směrodatné odchylce, kterou v případě, že ještě nemáme stanovený výběr, neznáme. Její hodnotu tedy musíme odhadnout. Obvykle se za tímto účelem provádí tzv. **předvýběr**, tj. výběr o malém rozsahu n_1, z něhož vypočteme výběrovou odchylku s_1, kterou považujeme za odhad výběrové směrodatné odchylky s. Pak určíme minimální rozsah výběru úpravou příslušného vztahu:

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

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

Po zjištění požadovaného rozsahu n pak stačí doplnit předvýběr o chybějících ($n-n_1$) prvků a intervalový odhad pak provést z výběru o rozsahu n.

8.9.2 Rozsah výběru při odhadu relativní četnosti (podílu)

Oboustranný interval spolehlivosti je dán jako:

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

Polovina šířky oboustranného intervalu spolehlivosti a tedy přípustná chyba odhadu Δ je:

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

Vidíme, že přípustná chyba odhadu závisí tentokrát na výběrové relativní četnosti, kterou neznáme. Nemáme-li žádné informace o výběrové relativní četnosti, můžeme dále postupovat dvěma způsoby:

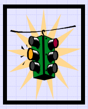
a) Provedeme **předvýběr**, z něhož vypočteme výběrovou relativní četnost p_1, kterou považujeme za odhad výběrové relativní četnosti p. Pak určíme minimální rozsah výběru úpravou příslušného vztahu:

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

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

Po zjištění požadovaného rozsahu n pak stačí doplnit předvýběr o chybějících ($n-n_1$) prvků a intervalový odhad pak provést z výběru o rozsahu n.

b) Druhou možností je odhadnout výběrovou relativní četnost nejhorší možnou variantou, tj. maximální hodnotou rozptylu p.(1-p), které je dosaženo pro p = 0,5.



Řešený příklad:

Výběrovým šetřením bychom chtěli odhadnout průměrnou mzdu pracovníků určitého výrobního odvětví. Z vyčerpávajícího šetření, které probíhalo před několika měsíci, víme, že směrodatná odchylka mezd byla 750,-Kč. Odhad chceme provést s 95% spolehlivosti a jsme ochotni připustit maximální chybu ve výši 50,-Kč. Jak velký musíme provést výběr, abychom zajistili požadovanou přesnost a spolehlivost?

Řešení:

Chceme odhadnou rozsah výběru pro intervalový odhad střední hodnoty známe-li směrodatnou odchylku σ (vyčerpávající šetření = zkoumání celého základního souboru (populace)).

$$1-\alpha = 0{,}95 \Rightarrow \alpha = 0{,}05 \Rightarrow 1-\frac{\alpha}{2} = 0{,}975$$

$$z_{0,975} = 1{,}96 \quad (\mathit{Tabulka}\ 1)$$

$$\sigma = 750\ K\check{c}$$

$$\Delta \leq 50\ K\check{c}$$

Rozsah výběru odhadneme v tomto případě podle vztahu:

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

Po dosazení:

$$n \geq \left(\frac{750}{50} \cdot 1{,}96\right)^2$$

$$n \geq 864{,}4$$

Chceme-li dosáhnout přípustné chyby ve výši maximálně 50,- Kč, musíme pro nalezení 95%-ního intervalového odhadu provést výběrové šetření na souboru o rozsahu minimálně 865 pracovníků.

Výklad:

Na závěr této kapitoly si ještě ukážeme jak najít intervalové odhady pro rozdíl středních hodnot dvou populací a pro rozdíl relativních četností dvou populací.

V praktických případech většinou nedokážeme přesně určit **parametry základního souboru** (populace). K jejich odhadu používáme charakteristiky příslušných výběrových souboru – **výběrové charakteristiky.**

Z metodického hlediska používáme dva typy odhadů parametrů:

- **bodový odhad**, kdy parametr základního souboru aproximujeme jediným číslem

a

- **intervalový odhad** (konfidenční interval), kdy tento parametr aproximujeme intervalem, v němž parametr leží s danou pravděpodobností. Této pravděpodobnosti říkáme **spolehlivost odhadu** a označujeme ji (1-α), α nazýváme **hladinou významnosti**.

„Dobrý“ (věrohodný) odhad musí splňovat určité vlastnosti. Mezi základní vlastnosti věrohodných odhadů patří:

- **nestrannost** (nevychýlenost, nezkreslenost)
- **vydatnost** (eficience)

- **konzistence**
- **dostatečnost**

V praktických aplikacích, častěji než bodový odhad, určujeme intervalový odhad příslušného parametru. Tento odhad je reprezentován intervalem (T_D; T_H), v němž hledaný parametr leží s předem určenou pravděpodobností (spolehlivostí), kterou označujeme (1-α).

Intervaly spolehlivosti konstruujeme jako **jednostranné** nebo **dvoustranné**. V následující tabulce najdete přehled intervalových odhadů pro parametry normálního rozdělení včetně použitých výběrových charakteristik.

Intervaly spolehlivosti pro parametry normálního rozdělení

Odhadovaný parametr	Vhodná výběrová charakteristika	Rozdělení výběrové char.	Meze oboustranného intervalu spolehlivosti		Dolní mez levostranného intervalu spolehlivosti	Horní mez pravostranného intervalu spolehlivosti
			T_D	**T_H**	**T_D**	**T_H**
μ, známe σ	$Z=\frac{\overline{X}-\mu}{\sigma}\cdot\sqrt{n}$	N(0;1)	$\overline{X}-\frac{\sigma}{\sqrt{n}}\cdot z_{1-\frac{\alpha}{2}}$	$\overline{X}+\frac{\sigma}{\sqrt{n}}\cdot z_{1-\frac{\alpha}{2}}$	$\overline{X}-\frac{\sigma}{\sqrt{n}}\cdot z_{1-\alpha}$	$\overline{X}+\frac{\sigma}{\sqrt{n}}\cdot z_{1-\alpha}$
μ, neznáme σ	$T_{n-1}=\frac{\overline{X}-\mu}{s}\cdot\sqrt{n}$	t_{n-1}	$\overline{X}-\frac{s}{\sqrt{n}}\cdot t_{1-\frac{\alpha}{2},\,n-1}$	$\overline{X}+\frac{s}{\sqrt{n}}\cdot t_{1-\frac{\alpha}{2},\,n-1}$	$\overline{X}-\frac{s}{\sqrt{n}}\cdot t_{1-\alpha,\,n-1}$	$\overline{X}+\frac{s}{\sqrt{n}}\cdot t_{1-\alpha,\,n-1}$
σ^2	$\chi=\frac{(n-1)s^2}{\sigma^2}$	χ^2_{n-1}	$\frac{(n-1)}{x_{1-\frac{\alpha}{2},\,n-1}}\cdot s^2$	$\frac{(n-1)}{x_{\frac{\alpha}{2},\,n-1}}\cdot s^2$	$\frac{(n-1)}{x_{1-\alpha,\,n-1}}\cdot s^2$	$\frac{(n-1)}{x_{\alpha,\,n-1}}\cdot s^2$
σ	intervalový odhad je odvozen z intervalového odhadu σ^2		$\sqrt{\frac{(n-1)}{x_{1-\frac{\alpha}{2},\,n-1}}}\cdot s$	$\sqrt{\frac{(n-1)}{x_{\frac{\alpha}{2},\,n-1}}}\cdot s$	$\sqrt{\frac{(n-1)}{x_{1-\alpha,\,n-1}}}\cdot s$	$\sqrt{\frac{(n-1)}{x_{\alpha,\,n-1}}}\cdot s$
π	$P_1=\frac{p-\pi}{\sqrt{\pi(1-\pi)}}\cdot\sqrt{n}$	N(0;1)	$p-\sqrt{\frac{p\cdot(1-p)}{n}}\cdot z_{1-\frac{\alpha}{2}}$	$p+\sqrt{\frac{p\cdot(1-p)}{n}}\cdot z_{1-\frac{\alpha}{2}}$	$p-\sqrt{\frac{p\cdot(1-p)}{n}}\cdot$	$p+\sqrt{\frac{p\cdot(1-p)}{n}}\cdot z_{1-\alpha}$

Velikost výběru má přímý vliv na přesnost odhadu parametrů základního souboru: čím větší rozsah výběru, tím přesnější je intervalový odhad. Ekonomické a časové důvody nás však mnohdy nutí volit rozsah výběru co nejmenší. V praxi proto hledáme kompromis, který pro požadovanou přesnost výpočtu (**přípustnou chybu odhadu Δ)** povede k co nejmenšímu rozsahu výběru.

Odhadovaný parametr	Rozsah výběru
μ, známe σ	$n\geq\left(\frac{\sigma}{\Delta}\cdot z_{1-\frac{\alpha}{2}}\right)^2$
μ, neznáme σ	$n\geq\left(\frac{s_1}{\Delta}\cdot t_{1-\frac{\alpha}{2},\,n-1}\right)^2$
π	$n\geq\frac{p_1\cdot(1-p_1)}{\Delta^2}\cdot z^2_{1-\frac{\alpha}{2}}$

Otázky

1. Objasněte rozdíl mezi základním souborem (populací) a výběrovým souborem.

2. Jaké znáte způsoby odhadu parametrů základního souboru ?

3. Vysvětlete co je to „dobrý“ odhad (vysvětlete pojmy: nestrannost, konzistence, vydatnost, dostatečnost).

4. Popište obecně oboustranný (levostranný, pravostranný) 100.(1-α)%-ní interval spolehlivosti pro nějaký parametr θ.

5. Najděte oboustranný interval spolehlivosti pro střední hodnotu μ při zvolené hladině významnosti α, pro zadaný náhodný výběr z normálního rozdělení, jehož rozptyl σ^2 známe (resp. neznáme).

6. Najděte oboustranný interval spolehlivosti pro rozptyl σ^2 při zvolené hladině významnosti α, pro zadaný náhodný výběr z normálního rozdělení.

7. Najděte oboustranný interval spolehlivosti pro směrodatnou odchylku σ při zvolené hladině významnosti α, pro zadaný náhodný výběr z normálního rozdělení.

8. Najděte oboustranný interval spolehlivosti pro relativní četnost (podíl) π při zvolené hladině významnosti α, pro zadaný náhodný výběr z normálního rozdělení.

Úlohy k řešení

1. Náhodný výběr pěti států USA má následující rozlohy (v 1000 čtverečních mil):

   147, 84, 24, 85, 159

   Vypočtěte 95% interval spolehlivosti pro střední rozlohu každého z 50-ti států USA.

2. V náhodném výběru čipů vyráběných velkou světovou společnosti 10% čipů nevyhovuje novým požadavkům na kvalitu. Sestrojte 95% interval spolehlivosti pro podíl čipů (v celé produkci společnosti), které nevyhovují dané normě, jestliže rozsah výběru je:
   a) n = 10
   b) n= 25
   c) n = 50
   d) n = 200

3. Pro realizaci rozsáhlého šetření o diferenciaci mezd ve velkém průmyslovém podniku musíme velmi rychle získat určitou představu o průměrné odchylce mezd. Z celkového počtu 5.000 zaměstnanců jsme jich náhodně vybrali 30 a určili průměrnou mzdu 9.450,-Kč a směrodatnou odchylku ve výši 1.200,-Kč. V jakém intervalu lze s 95% pravděpodobnosti očekávat směrodatnou odchylku mezd v celém podniku? Předpokládáme, že rozdělení mezd v základním souboru všech pracovníků podniku je normální.

4. Jaký minimální rozsah výběru pro odhad podílu chybně zúčtovaných položek musíme navrhnout, chceme-li při 90% spolehlivosti zajistit přípustnou chybu ± 3 %. O možném podílu chybných položek nemáme při prováděném auditu žádnou informaci

5. Hypermarket Hyper chce pro zkvalitnění služeb poskytovaných zákazníkům zkrátit dobu jejich čekání u pokladen. Náhodně bylo vybráno 10 zákazníků a byla změřena doba jejich čekání u pokladny (předpokládáme normalitu rozdělení dob čekání). Výsledky šetření (v sekundách): 310, 225, 390, 265, 358, 255, 170, 265, 150, 240.
   a) V jakých mezích lze s pravděpodobnosti 0,95 očekávat průměrnou dobu čekání zákazníka na obsluhu (v minutách)?
   b) Jaká je horní hranice doby čekání, která nebude s pravděpodobností 0,95 překročena?

6. Agentura provádějící průzkum veřejného mínění plánuje šetření, na základě kterého chce odhadnout, kolik procent voličů podporuje současnou vládní koalici. Předpokládejme (v praxi tomu tak ovšem není), že jsou dotazování vybírání zcela náhodně. Kolik dotazovaných by mělo být do výběru zařazeno, jestliže si vedení agentury přeje, aby se odhad z výběru nelišil od skutečného podílu příznivců koalice o více než 3%? (Volte hladinu významnosti 0,05.)

7. Z 90 zkoušek meze kluzu konstrukční oceli z produkce určité ocelárny byl vypočten výběrový průměr 251,34 MPa a výběrový rozptyl 319,48 MPa^2. Najděte 80% intervaly

spolehlivosti pro střední hodnotu a směrodatnou odchylku meze kluzu. (za předpokladu normality dat)

8. Agentura STAT udává, že v lednu 1999 byla v populaci České republiky 30%-ní podpora ČSSD (1000 respondentů) a při průzkumu v květnu 1999 (1600 respondentů) zjistili pouze 25%-ní podporu této strany. Na základě květnového průzkumu učiňte 90% intervalový odhad ohledně procentuálního zastoupení voličů ČSSD v celé populaci.

Řešení:

1. $P(31{,}9 < \mu < 167{,}7) = 0{,}95$

2. a) $P(-0{,}086 < \pi < 0{,}286) = 0{,}95$
   b) $P(-0{,}018 < \pi < 0{,}218) = 0{,}95$
   c) $P(0{,}017 < \pi < 0{,}183) = 0{,}95$
   d) $P(0{,}058 < \pi < 0{,}142) = 0{,}95$

   Všimněte si, že rostoucí rozsah výběru vede k zpřesňování konfidenčního intervalu

3. $P(955{,}70 Kč < \sigma < 1613{,}2 Kč) = 0{,}95$

4. $n \geq 751{,}7 \Rightarrow n = 752$

5. a) $P(209{,}1 < \mu < 316{,}5) = 0{,}95$
   b) $P(301{,}9 < \mu) = 0{,}95$

6. $n \geq 1067{,}1 \Rightarrow n = 1068$

7. $P(248{,}91 < \mu < 253{,}77) = 0{,}80$, $P(16{,}34 < \sigma < 19{,}81) = 0{,}80$

8. $P(0{,}23 < \pi < 0{,}25) = 0{,}95$