

8. Parametrické úlohy o jednom náhodném výběru a dvou nezávislých náhodných výběrech z alternativních rozložení

8.1. Věta: Asymptotické rozložení statistiky odvozené z výběrového průměru.

8.2. Věta: Vzorec pro meze $100(1-\alpha)\%$ asymptotického empirického intervalu spolehlivosti pro parametr ϑ .

8.3. Příklad: Náhodně bylo vybráno 100 osob a zjištěno, že 34 z nich používá zubní kartáček zahraniční výroby. Najděte 95% asymptotický interval spolehlivosti pro pravděpodobnost, že náhodně vybraná osoba používá zubní kartáček zahraniční výroby.

Řešení:

Zavedeme náhodné veličiny X_1, \dots, X_{100} , přičemž $X_i = 1$, když i -tá osoba používá zahraniční zubní kartáček a $X_i = 0$ jinak, $i = 1, \dots, 100$. Tyto náhodné veličiny tvoří náhodný výběr z rozložení $A(\vartheta)$.

$n = 100$, $m = 34/100$, $\alpha = 0,05$, $u_{1-\alpha/2} = u_{0,975} = 1,96$.

Ověření podmínky $n\vartheta(1-\vartheta) > 9$: parametr ϑ neznáme, musíme ho nahradit výběrovým průměrem. Pak $100 \cdot 0,34 \cdot 0,66 = 22,44 > 9$.

$$d = 0,34 - \sqrt{\frac{0,34(1-0,34)}{100}} \cdot 1,96 = 0,2472, \quad h = 0,34 + \sqrt{\frac{0,34(1-0,34)}{100}} \cdot 1,96 = 0,4328.$$

S pravděpodobností přibližně 0,95 tedy $0,2472 < \vartheta < 0,4328$.

8.4. Příklad: Kolik osob musíme vybrat, abychom podíl modrookých osob v populaci odhadli se spolehlivostí 90% a šířka intervalu spolehlivosti byla nanejvýš a) 0,06, b) 0,01?

Řešení:

Šířka $100(1-\alpha)\%$ asymptotického empirického intervalu spolehlivosti pro parametr ϑ :

$$h - d = m + \sqrt{\frac{m(1-m)}{n}} u_{1-\alpha/2} - \left(m - \sqrt{\frac{m(1-m)}{n}} u_{1-\alpha/2} \right) = 2\sqrt{\frac{m(1-m)}{n}} u_{1-\alpha/2}$$

Požadujeme, aby $h - d \leq \Delta$, tedy $2\sqrt{\frac{m(1-m)}{n}} u_{1-\alpha/2} \leq \Delta$. Odtud vyjádříme

$$n \geq \frac{4m(1-m)u_{1-\alpha/2}^2}{\Delta^2}.$$

Předpokládejme, že nemáme žádné předběžné informace o podílu modrookých osob v populaci. Musíme tedy zvolit takové m , aby šířka intervalu spolehlivosti byla maximální. Maximalizujeme výraz $m(1-m) = m - m^2$. Derivujeme podle m a položíme rovno 0:

$$1 - 2m = 0 \Rightarrow m = \frac{1}{2}. \text{ V tomto případě volíme relativní četnost } m = 0,5.$$

$$\text{ad a) } n \geq \frac{4m(1-m)u_{1-\alpha/2}^2}{\Delta^2} = \frac{4 \cdot 0,5 \cdot 0,5 \cdot u_{0,95}^2}{0,06^2} = \frac{4 \cdot 0,5 \cdot 0,5 \cdot 1,645^2}{0,06^2} = 751,67$$

Uvedenou podmínku tedy splníme, když vybereme aspoň 752 osob.

$$\text{ad b) } n \geq \frac{4m(1-m)u_{1-\alpha/2}^2}{\Delta^2} = \frac{4 \cdot 0,5 \cdot 0,5 \cdot u_{0,95}^2}{0,01^2} = \frac{4 \cdot 0,5 \cdot 0,5 \cdot 1,645^2}{0,01^2} = 27060,25$$

Chceme-li dosáhnout podstatně užšího intervalu spolehlivosti, musíme vybrat aspoň 27 061 osob.

Modifikace: Předpokládejme, že v populaci je nanejvýš 30% modrookých osob. Pak relativní četnost $m = 0,3$.

$$\text{ad a) } n \geq \frac{4m(1-m)u_{1-\alpha/2}^2}{\Delta^2} = \frac{4 \cdot 0,3 \cdot 0,7 \cdot u_{0,95}^2}{0,06^2} = \frac{4 \cdot 0,3 \cdot 0,7 \cdot 1,645^2}{0,06^2} = 631,41$$

V tomto případě stačí vybrat 632 osob.

Ve srovnání s předešlým případem vidíme, že rozsah výběru skutečně klesl.

ad b)

$$n \geq \frac{4m(1-m)u_{1-\alpha/2}^2}{\Delta^2} = \frac{4 \cdot 0,3 \cdot 0,7 \cdot u_{0,95}^2}{0,01^2} = \frac{4 \cdot 0,3 \cdot 0,7 \cdot 1,645^2}{0,01^2} = 22730,61$$

V tomto případě musíme vybrat aspoň 22 731 osob.

8.5. Věta: Testování hypotézy o parametru ϑ .

8.6. Příklad: Pravděpodobnost vyrobení zmetku při výrobě určité součástky činí $\vartheta = 0,01$.

Bylo náhodně vybráno 1000 výrobků a zjistilo se, že mezi nimi je 16 zmetků. Na asymptotické hladině významnosti 0,05 testujte hypotézu $H_0: \vartheta = 0,01$ proti oboustranné alternativě $H_1: \vartheta \neq 0,01$.

Řešení:

Ověření podmínky $n\vartheta(1-\vartheta) > 9$: $1000 \cdot 0,01 \cdot 0,99 = 9,9 > 9$. Realizace testového kritéria:

$$t_0 = \frac{0,016 - 0,01}{\sqrt{\frac{0,01 \cdot 0,99}{1000}}} = 1,907. \text{ Kritický obor: } W = (-\infty, -u_{0,975}) \cup (u_{0,975}, \infty) =$$

$= (-\infty, -1,96) \cup (1,96, \infty)$. Protože $1,907 \notin W$, H_0 nezamítáme na asymptotické hladině významnosti 0,05.

8.7. Příklad: Nový léčebný postup považujeme za úspěšný, pokud po jeho ukončení bude dosaženo zlepšení zdravotního stavu u alespoň 50% zúčastněných pacientů. Nová terapie byla vyzkoušena u 40 pacientů a ke zlepšení došlo u 24 osob. Je možné na asymptotické hladině významnosti zamítnout 0,05 hypotézu, že tato terapie nedosahuje úspěšnosti aspoň 50%?

Řešení:

Zavedeme náhodné veličiny X_1, \dots, X_{40} , přičemž $X_i = 1$, když terapie u i -tého pacienta byl úspěšná a $X_i = 0$ jinak, $i = 1, \dots, 40$. Tyto náhodné veličiny tvoří náhodný výběr z rozložení $A(\vartheta)$.

Testujeme hypotézu $H_0: \vartheta \leq 0,5$ proti pravostranné alternativě $H_1: \vartheta > 0,5$.

Známe: $n = 40$, $m = \frac{24}{40} = 0,6$, $c = 0,5$, $\alpha = 0,05$, $u_{1-\alpha/2} = u_{0,95} = 1,645$

Ověření podmínky $n\vartheta(1-\vartheta) > 9$: $40 \cdot 0,6 \cdot 0,4 = 9,6 > 9$.

Realizace testového kritéria: $t_0 = \frac{m-c}{\sqrt{\frac{c \cdot (1-c)}{n}}} = \frac{0,6-0,5}{\sqrt{\frac{0,5 \cdot 0,5}{40}}} = 1,2649$.

Kritický obor: $W = \langle u_{1-\alpha}, \infty \rangle = \langle u_{0,95}, \infty \rangle = \langle 1,645, \infty \rangle$. Protože $1,2649 \notin W$, H_0 nezamítáme na asymptotické hladině významnosti 0,05.

8.8. Věta: Asymptotické rozložení statistiky odvozené ze dvou výběrových průměrů.

8.9. Věta: Vzorec pro meze 100(1- α)% asymptotického empirického intervalu spolehlivosti pro parametrickou funkci $\vartheta_1 - \vartheta_2$.

8.10. Příklad: Management supermarketu vyhlásil týden slev a sledoval, zda toto vyhlášení má vliv na podíl větších nákupů (nad 500 Kč). Na základě náhodného výběru 200 zákazníků v týdnu bez slev bylo zjištěno 97 velkých nákupů, zatímco v týdnu se slevou z 300 náhodně vybraných zákazníků učinilo velký nákup 162 zákazníků. Sestrojte 95% asymptotický interval spolehlivosti pro rozdíl pravděpodobností uskutečnění většího nákupu v týdnu se slevou a v týdnu bez slevy.

Řešení:

Zavedeme náhodnou veličinu X_{1i} , která bude nabývat hodnoty 1, když v týdnu bez slevy i -tý náhodně vybraný zákazník uskuteční větší nákup a hodnoty 0 jinak, $i = 1, \dots, 200$. Náhodné veličiny $X_{1,1}, \dots, X_{1,200}$ tvoří náhodný výběr z rozložení $A(\vartheta_1)$. Dále zavedeme náhodnou veličinu X_{2i} , která bude nabývat hodnoty 1, když v týdnu se slevou i -tý náhodně vybraný zákazník uskuteční větší nákup a hodnoty 0 jinak, $i = 1, \dots, 300$. Náhodné veličiny $X_{2,1}, \dots, X_{2,300}$ tvoří náhodný výběr z rozložení $A(\vartheta_2)$.

$n_1 = 200$, $n_2 = 300$, $m_1 = 97/200$, $m_2 = 162/300$.

Ověření podmínek $n_1 \vartheta_1 (1 - \vartheta_1) > 9$ a $n_2 \vartheta_2 (1 - \vartheta_2) > 9$: Parametry ϑ_1 a ϑ_2 neznáme, nahradíme je odhady m_1 a m_2 . $97 \cdot (1 - 97/200) = 49,955 > 9$, $162 \cdot (1 - 162/300) = 74,52 > 9$.

Meze 100(1- α)% asymptotického empirického intervalu spolehlivosti pro parametrickou funkci $\vartheta_1 - \vartheta_2$ jsou:

$$d = m_1 - m_2 - \sqrt{\frac{m_1(1-m_1)}{n_1} + \frac{m_2(1-m_2)}{n_2}} u_{1-\alpha/2} =$$

$$= \frac{97}{200} - \frac{162}{300} - \sqrt{\frac{\frac{97}{200}(1-\frac{97}{200})}{200} + \frac{\frac{162}{300}(1-\frac{162}{300})}{300}} 1,96 = -0,1443$$

$$h = m_1 - m_2 + \sqrt{\frac{m_1(1-m_1)}{n_1} + \frac{m_2(1-m_2)}{n_2}} u_{1-\alpha/2} =$$

$$= \frac{97}{200} - \frac{162}{300} + \sqrt{\frac{\frac{97}{200}(1-\frac{97}{200})}{200} + \frac{\frac{162}{300}(1-\frac{162}{300})}{300}} 1,96 = 0,0343$$

Zjistili jsme tedy, že s pravděpodobností přibližně 0,95 $-0,1443 < \vartheta_1 - \vartheta_2 < 0,0343$.

8.11. Věta: Testování hypotézy o parametrické funkci $\vartheta_1 - \vartheta_2$

8.12. Poznámka: Postup při testování hypotézy $\vartheta_1 - \vartheta_2 = 0$

8.13. Příklad: Pro údaje z příkladu 8.10. testujte na asymptotické hladině významnosti 0,05 hypotézu, že týden se slevami nezvýší pravděpodobnost uskutečnění většího nákupu.

Řešení:

Testujeme hypotézu $\vartheta_1 - \vartheta_2 = 0$ proti levostranné alternativě $H_1: \vartheta_1 - \vartheta_2 < 0$ na asymptotické hladině významnosti 0,05.

$n_1 = 200, n_2 = 300, m_1 = 97/200, m_2 = 162/300, m_* = (97 + 162)/500 = 0,518$.

Podmínky dobré aproximace byly ověřeny v příkladu 8.8.

Testování pomocí intervalu spolehlivosti:

Pro levostrannou alternativu používáme pravostranný interval spolehlivosti:

$$h = m_1 - m_2 + \sqrt{\frac{m_1(1-m_1)}{n_1} + \frac{m_2(1-m_2)}{n_2}} u_{1-\alpha} = \\ = \frac{97}{200} - \frac{162}{300} + \sqrt{\frac{\frac{97}{200}(1-\frac{97}{200})}{200} + \frac{\frac{162}{300}(1-\frac{162}{300})}{300}} 1,645 = 0,02$$

Protože číslo $c = 0$ je obsaženo v intervalu $(-\infty; 0,02)$, H_0 nezamítáme na asymptotické hladině významnosti 0,05.

Testování pomocí kritického oboru:

Realizace testového kritéria:

$$t_0 = \frac{m_1 - m_2}{\sqrt{m_*(1-m_*)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{\frac{97}{200} - \frac{162}{300}}{\sqrt{0,518(1-0,518)\left(\frac{1}{200} + \frac{1}{300}\right)}} = -1,2058.$$

Kritický obor je $W = (-\infty, -u_{1-\alpha}) = (-\infty, -u_{0,95}) = (-\infty, -1,645)$. Protože testové kritérium nepatří do kritického oboru, H_0 nezamítáme na asymptotické hladině významnosti 0,05.

Testování pomocí p-hodnoty:

Pro levostrannou alternativu se p-hodnota počítá podle vzorce $p = P(T_0 \leq t_0)$:

$$p = P(T_0 \leq -1,2058) = \Phi(-1,2058) = 1 - \Phi(1,2058) = 1 - 0,8861 = 0,1139$$

Protože p-hodnota je větší než 0,05, H_0 nezamítáme na asymptotické hladině významnosti 0,05.