

## Téma 9: Parametrické úlohy o jednom a dvou nezávislých náhodných výběrech z alternativních rozložení

### Úkol 1.: Asymptotický interval spolehlivosti pro parametr $\vartheta$ alternativního rozložení

Může politická strana, pro niž se v předvolebním průzkumu vyslovilo 60 z 1000 dotázaných osob, očekávat se spolehlivostí 0,95, že by v této době ve volbách překročila 5% hranici pro vstup do parlamentu?

**Návod:** Zavedeme náhodné veličiny  $X_1, \dots, X_{1000}$ , přičemž  $X_i = 1$ , když  $i$ -tá osoba se vysloví pro danou politickou stranu a  $X_i = 0$  jinak,  $i = 1, \dots, 1000$ . Tyto náhodné veličiny tvoří náhodný výběr z rozložení  $A(\vartheta)$ . V tomto případě  $n = 1000$ ,  $m = 60/1000 = 0,06$ ,  $\alpha = 0,05$ ,  $u_{1-\alpha} = u_{0,95} = 1,645$ .

Ověření podmínky  $n\vartheta(1-\vartheta) > 9$ : parametr  $\vartheta$  neznáme, musíme ho nahradit výběrovým průměrem. Pak  $1000 \cdot 0,06 \cdot 0,94 = 56,4 > 9$ .

95% levostranný interval spolehlivosti pro  $\vartheta$  je

$$\left( m - \sqrt{\frac{m(1-m)}{n}} u_{1-\alpha} ; \infty \right) = \left( 0,06 - \sqrt{\frac{0,06(1-0,06)}{1000}} u_{0,95} ; \infty \right).$$

V našem případě

$$d = 0,06 - \sqrt{\frac{0,06 \cdot 0,94}{1000}} \cdot 1,645 = 0,0476$$

S pravděpodobností přibližně 0,95 tedy  $\vartheta > 0,048$ . Protože tento interval zahrnuje i hodnoty nižší než 0,05, nelze vyloučit, že strana získá méně než 5% hlasů.

### Postup ve STATISTICE:

**Přesný způsob:** Vytvoříme datový soubor o jedné proměnné (nazveme ji  $d$ ) a o jednom případě. Do Dlouhého jména proměnné  $d$  napíšeme

$$=0,06 - \text{sqrt}(0,06 \cdot 0,94 / 1000) \cdot \text{VNormal}(0,95;0;1)$$

Vyjde 0,047647.

**Přibližný způsob:** Do nového datového souboru o jedné proměnné  $X$  a 1000 případech uložíme 60 jedniček (indikují volbu dané politické strany) a 940 nul (indikují volbu jiné politické strany).

Statistika – Základní statistiky a tabulky – Popisné statistiky – OK – Proměnné  $X$  – OK – Detailní výsledky – zaškrtneme Meze spolehl. prům. – Interval 90,00 – Výpočet.

Dostaneme tabulku:

| Proměnná | Popisné statistiky (Tabulka1) |          |                        |                      |
|----------|-------------------------------|----------|------------------------|----------------------|
|          | N platných                    | Průměr   | Int. spolehl. -90,000% | Int. spolehl. 90,000 |
| X        | 1000                          | 0,060000 | 0,047630               | 0,072370             |

Protože dolní mez oboustranného 90% intervalu spolehlivosti pro střední hodnotu je shodná s dolní mezí 95% levostranného intervalu spolehlivosti, můžeme konstatovat, že voliči budou volit danou politickou stranu s pravděpodobností aspoň 4,76%. Na základě uvedených dat strana tedy nemá zaručeno, že překročí 5% hranici pro vstup do parlamentu.

**Úkol k samostatnému řešení:** Přírůstky cen akcií na burze (v %) u 10 náhodně vybraných společností dosáhly těchto hodnot: 10, 16, 5, 10, 12, 8, 4, 6, 5, 4. Sestrojte 95% asymptotický empirický interval spolehlivosti pro pravděpodobnost, že přírůstek ceny akcie překročí 8,5%.

**Výsledek:**  $0,096 < \vartheta < 0,704$  s pravděpodobností aspoň 0,95.

Znamená to, že pravděpodobnost, že přírůstek ceny akcie překročí 8,5%, je aspoň 9,6% a nanejvýš 70,4% (při spolehlivosti 95%).

## Úkol 2.: Testování hypotézy o parametru $\vartheta$ alternativního rozložení

Určitá cestovní kancelář organizuje zahraniční zájezdy podle individuálních přání zákazníků. Z několika minulých let ví, že 30% všech takto organizovaných zájezdů má za cíl zemi X. Po zhoršení politických podmínek v této zemi se cestovní kancelář obává, že se zájem o tuto zemi mezi zákazníky sníží. Ze 150 náhodně vybraných zákazníků v tomto roce má 38 za cíl právě zemi X. Potvrzují nejnovější data pokles zájmu o tuto zemi? Volte hladinu významnosti 0,05.

**Návod:** Máme náhodný výběr  $X_1, \dots, X_{150}$  z rozložení  $A(0,3)$ . Testujeme  $H_0: \vartheta = 0,3$  proti levostranné alternativě  $H_1: \vartheta < 0,3$ . V tomto případě je testovým kritériem statistika

$$T_0 = \frac{M - c}{\sqrt{\frac{c(1-c)}{n}}}, \text{ která v případě platnosti nulové hypotézy má asymptoticky rozložení } N(0,1).$$

Musíme ověřit splnění podmínky  $n\vartheta(1-\vartheta) > 9$ :  $150 \cdot 0,3 \cdot 0,7 = 31,5 > 9$ . Vypočteme realizaci

$$\text{testového kritéria: } t_0 = \frac{m - c}{\sqrt{\frac{c(1-c)}{n}}} = \frac{\frac{38}{150} - 0,3}{\sqrt{\frac{0,3(1-0,3)}{150}}} = -1,24722. \text{ Kritický obor:}$$

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

### Postup ve STATISTICE:

Přesný způsob: Vytvoříme datový soubor o dvou proměnných (nazveme je  $t_0$  a kvantil) a jednom případě. Vypočteme realizaci testového kritéria tak, že do Dlouhého jména proměnné  $t_0$  napíšeme

$$=(38/150-0,3)/\text{sqrt}(0,3*0,7/150)$$

Do Dlouhého jména proměnné kvantil napíšeme

$$=V\text{Normal}(0,95;0;1)$$

Tím získáme kvantil  $u_{0,95}$  a testové kritérium porovnáme s opačnou hodnotou tohoto kvantilu.

|   | 1           | 2        |
|---|-------------|----------|
|   | $t_0$       | kvantil  |
| 1 | -1,24721913 | 1,644854 |

Jelikož testové kritérium je větší než -1,644854,  $H_0$  nezamítáme na asymptotické hladině významnosti 0,05.

Přibližný způsob: Do nového datového souboru o jedné proměnné X a 150 případech uložíme 38 jedniček (indikují zájem o danou zemi) a 112 nul (indikují nezáměr o danou zemi).

Statistika – Základní statistiky a tabulky – t-test, samost. vzorek – OK – Proměnné X – OK,

Test všech průměrů vůči 0,3 – Výpočet.

| Proměnná | Test průměrů vůči referenční konstantě (hodnotě) (Tabulka4) |          |     |          |                      |          |     |          |
|----------|---|----------|-----|----------|----------------------|----------|-----|----------|
|          | Průměr  | Sm.odch. | N   | Sm.chyba | Referenční konstanta | t        | SV  | p        |
| X        | 0,253333  | 0,436377 | 150 | 0,035630 | 0,300000             | -1,30976 | 149 | 0,192294 |

Hodnota testové statistiky je při tomto přibližném způsobu -1,30976. Odpovídající p-hodnota je 0,1923, ovšem to je p-hodnota pro oboustranný test. Tuto p-hodnotu tedy musíme dělit dvěma a dostaneme 0,0961. Na asymptotické hladině významnosti 0,05 nelze zamítnout hypotézu, že zájem o danou zemi se nezměnil.

### Úkol 3.: Asymptotický interval spolehlivosti pro parametrickou funkci $\vartheta_1 - \vartheta_2$

Při výstupní kontrole bylo náhodně vybráno 150 výrobků vyrobených na ranní směně a rovněž 150 výrobků vyrobených na odpolední směně. U ranní směny bylo zjištěno 16 zmetků a u odpolední 12 zmetků. Sestrojte 95% asymptotického interval spolehlivosti pro rozdíl pravděpodobností vyrobení zmetku v obou směnách.

**Návod:** Zavedeme náhodnou veličinu  $X_{1i}$ , která bude nabývat hodnoty 1, když  $i$ -tý výrobek z ranní směny je zmetek, 0 jinak,  $i = 1, \dots, 150$ . Náhodné veličiny  $X_{1,1}, \dots, X_{1,150}$  tvoří náhodný výběr z rozložení  $A(\vartheta_1)$ . Dále zavedeme náhodnou veličinu  $X_{2i}$ , která bude nabývat hodnoty 1, když  $i$ -tý výrobek z odpolední směny je zmetek, 0 jinak,  $i = 1, \dots, 150$ . Náhodné veličiny  $X_{2,1}, \dots, X_{2,150}$  tvoří náhodný výběr z rozložení  $A(\vartheta_2)$ .

$n_1 = 150, n_2 = 150, m_1 = 16/150 = 0,1067, m_2 = 12/150 = 0,08$ .

Ověření podmínek  $n_1 \vartheta_1 (1 - \vartheta_1) > 9$  a  $n_2 \vartheta_2 (1 - \vartheta_2) > 9$ : Parametry  $\vartheta_1$  a  $\vartheta_2$  neznáme, nahradíme je odhady  $m_1$  a  $m_2$ :  $16 \cdot (1 - 16/150) = 14,29 > 9, 12 \cdot (1 - 12/150) = 11,04 > 9$ .

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

$$d = m_1 - m_2 - \sqrt{\frac{m_1(1-m_1)}{n_1} + \frac{m_2(1-m_2)}{n_2}} u_{1-\alpha/2} = \\ = \frac{16}{150} - \frac{12}{150} - \sqrt{\frac{\frac{16}{150}(1-\frac{16}{150})}{150} + \frac{\frac{12}{150}(1-\frac{12}{150})}{150}} 1,96 = -0,039$$

$$h = m_1 - m_2 + \sqrt{\frac{m_1(1-m_1)}{n_1} + \frac{m_2(1-m_2)}{n_2}} u_{1-\alpha/2} = \\ = \frac{16}{150} - \frac{12}{150} + \sqrt{\frac{\frac{16}{150}(1-\frac{16}{150})}{150} + \frac{\frac{12}{150}(1-\frac{12}{150})}{150}} 1,96 = 0,092$$

Zjistili jsme tedy, že s pravděpodobností přibližně 0,95:  $-0,039 < \vartheta_1 - \vartheta_2 < 0,092$ .

#### Postup ve STATISTICE:

Otevřeme nový datový soubor se dvěma proměnnými  $d$  a  $h$  a o jednom případě. Do Dlouhého jména proměnné  $d$  napíšeme:

$$=16/150-12/150-\text{sqrt}((16/150)*(134/150)/150+(12/150)*(138/150)/150)*\text{VNormal}(0,975;0;1)$$

Do Dlouhého jména proměnné  $h$  napíšeme:

$$=16/150-$$

$$12/150+\text{sqrt}((16/150)*(134/150)/150+(12/150)*(138/150)/150)*\text{VNormal}(0,975;0;1)$$

Dostaneme tabulku

|   | 1       | 2        |
|---|---------|----------|
|   | d       | h        |
| 1 | -0,0391 | 0,092433 |

S pravděpodobností přibližně 0,95 se rozdíl pravděpodobností vyrobení zmetku na ranní a odpolední směně nachází v intervalu  $(-0,039; 0,092)$ .

### Úkol 4.: Testování hypotézy o parametrické funkci $\vartheta_1 - \vartheta_2$

Pro údaje z úkolu 3 testujte na asymptotické hladině významnosti 0,05 hypotézu, že pravděpodobnost vyrobení zmetků v obou směnách je táž.

#### Postup ve STATISTICE:

Statistiky – Základní statistiky a tabulky – Testy rozdílů: r, %, průměry – OK – vybereme Rozdíl mezi dvěma poměry – do políčka P 1 napíšeme 0,1067, do políčka N1 napíšeme 150, do políčka P 2 napíšeme 0,08, do políčka N2 napíšeme 150 –Výpočet. Dostaneme p-hodnotu 0,4274, tedy nezamítáme nulovou hypotézu na hladině významnosti 0,05.

**Testy rozdílů: r, %, průměry: Tabulka4**

Poslat/tisknout výsledky každ. výpočtu do okna protokolu Storno

**Rozdíl mezi dvěma korelačními koeficienty**

r1: 0,00 N1: 10 p: 1,0000  Jednostr. Výpočet  
r2: 0,00 N2: 10  Oboustr.

**Rozdíl mezi dvěma průměry (normální rozdělení)**

Pr1: ,25333 SmOd1: ,43637 N1: 150 p: ,0961 Výpočet  
Pr2: ,3 SmOd2: 1, N2: 10  Jednostr.  
 Výběrový průměr vs. střední hodnota  Oboustr.

**Rozdíl mezi dvěma poměry**

P 1: ,10670 N1: 150 p: ,4274 Výpočet  
P 2: ,08000 N2: 150  Oboustr.