Téma 11: Parametrické úlohy o jednom a dvou nezvislých náhodných výběrech z alternativních rozložení

Úkol 1.: Asymptotický interval spolehlivosti pro parametr ϑ alternativního rozložení
Může politická strana, pro niž se v předvolebním průzkumu vyslovilo 60 z 1000 dotázaných osob (tj. 6%), očekávat se spolehlivostí 0,95, že by v této době ve volbách překročila 5% hranici pro vstup do parlamentu?

Návod: Zavedeme náhodné veličiny $X_1, ..., X_{1000}$, přičemž $X_i = 1$, když i-tá osoba se vysloví pro danou politickou stranu a $X_i = 0$ jinak, i = 1, ..., 1000. Tyto náhodné veličiny tvoří náhodný výběr z rozložení A(ϑ). 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 ϑ neznáme, musíme ho nahradit výběrovým průměrem. Pak 1000.0,06.0,94 = 56,4 > 9.
95% levostranný interval spolehlivosti pro ϑ 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

$$d = 0{,}06 - \sqrt{\frac{0{,}06 \cdot 0{,}94}{1000}} \cdot 1{,}645 = 0{,}047$$

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 případu. Do Dlouhého jména proměnné d napíšeme
=0,06-sqrt(0,6*0,94/1000)*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:

	Popisné statistiky (Tabulka1)			
Promě	N platn	Prům	Int. spole -90,00(	Int. spole 90,00(
X	10(	0,060	0,047	0,072

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 ϑ 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 (tj. 25,3%). Potvrzují nejnovější data pokles zájmu o tuto zemi? Volte hladinu významnosti 0,05.

Návod: Máme náhodný výběr $X_1, ..., X_{150}$ 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 $T_0 = \frac{M - \vartheta}{\sqrt{\frac{\vartheta(1-\vartheta)}{n}}}$, která v případě platnosti nulové hypotézy má asymptoticky rozložení N(0,1).

Musíme ověřit splnění podmínky $n\vartheta(1-\vartheta) > 9$: 150.0,3.0,7 = 31,5 > 9. Vypočteme realizaci testového kritéria: $t_0 = \frac{m - \vartheta}{\sqrt{\frac{\vartheta(1-\vartheta)}{n}}} = \frac{\frac{38}{150} - 0,3}{\sqrt{\frac{0,3(1-0,3)}{150}}} = -1,247$. 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 t0 a kvantil) a jednom případu. Vypočteme realizaci testového kritéria tak, že do Douhého jména proměnné t0 napíšeme
=(38/150-0,3)/sqrt(0,3*0,7/150)
Do Dlouhého jména proměnné kvantil napíšeme
=VNormal(0,95;0;1)
Tím získáme kvantil $u_{0,95}$ a testové kritérium porovnáme s opačnou hodnotou tohoto kvantilu.

	1 t0	2 kvant
1	-1,2472	1,644

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 nezájem o danou zemi).
Statistika – Základní statistiky a tabulky – t-test, samost. vzorek – OK – Proměnné X – OK, Test všech průměrů vůči 0,3 – Výpočet.

Proměn	Test průměrů vůči referenční konstantě (hodnotě						
	Prům	Sm.od	N	Sm.chy	Referen konstan	t	SV	p
X	0,253	0,436	15	0,035	0,300	-1,30	14	0,192

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, …, 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, …, 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 ϑ_1 a ϑ_2 neznáme, nahradíme je odhady m_1 a m_2: 16.(1-16/150) = 14,29 > 9, 12.(1-12/150) = 11,04 > 9.

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

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

$$= \frac{16}{150} - \frac{12}{150} - \sqrt{\frac{\frac{16}{150}\left(1-\frac{16}{150}\right)}{150} + \frac{\frac{12}{150}\left(1-\frac{12}{150}\right)}{150}}\,1{,}96 = -0{,}039$$

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

$$= \frac{16}{150} - \frac{12}{150} + \sqrt{\frac{\frac{16}{150}\left(1-\frac{16}{150}\right)}{150} + \frac{\frac{12}{150}\left(1-\frac{12}{150}\right)}{150}}\,1{,}96 = 0{,}092$$

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 případu. Do Dlouhého jména proměnné d napíšeme:

=16/150-12/150-sqrt((16/150)*(134/150)/150+(12/150)*(138/150)/150)*VNormal(0,975;0;1)

Do Dlouhého jména proměnné h napíšeme:

=16/150-
12/150+sqrt((16/150)*(134/150)/150+(12/150)*(138/150)/150)*VNormal(0,975;0;1)

Dostaneme tabulku

	1 d	2 h
1	-0,0:	0,092

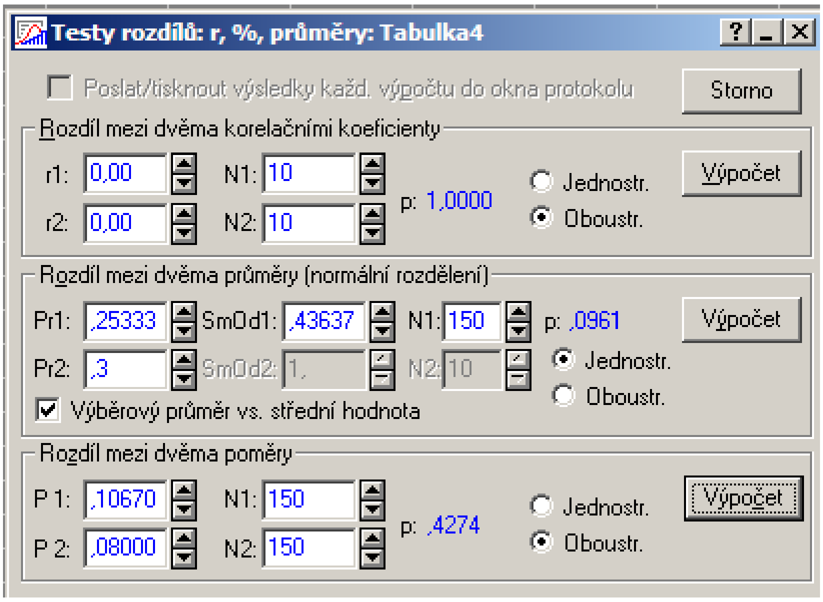
S pravděpodobností přibližně 0,95 se rozdíl pravděpodobností vyrobení zmetku na ranní a odpolední směně nachází v intervalu (-0,039; 0,092.

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.



Úkol k samostatnému řešení: V anketě, která se týkala očkování proti chřipce, odpovědělo z 200 náhodně vybraných mužů 97, že se podrobí očkování a z 300 náhodně vybraných žen chtělo podstoupit očkování 162. Na asymptotické hladině významnosti 0,05 testujte hypotézu, že pravděpobnost uskutečnění očkování proti chřipce je u mužů nižší než u žen.
Výsledek: Podmínky dobré aproximace jsou splněny. Relativní četnost mužů ochotných k očkování je 0,485, relativní četnost žen ochotných k očkování je 0,54. Na asymptotické hladině významnosti 0,05 se neprokázalo, že pravděpodobnost ochoty k očkování je u mužů nižší než u žen, protože p-hodnota pro levostrannou alternativu je 0,1142