

Odpovědi (téma 8)

1.1 ne

1.2 ano

1.3 ano

2.1 otázka se jinými slovy ptá na směrodatnou chybu, její hodnota se rovná $\sigma / \sqrt{n} = 15 / 3 = 5$

2.2 $z = (105-100) / 5 = 1 \dots 16\%$
pro 110 je to 2,3%

2.3 cca 68%; 95%

2.4 přibližně ano

2.5 25

2.6 1

2.7 cca 68%

2.8 ano

3.1 ano

3.2 ne, $\sigma^2/n = \sigma_m^2$ a ne σ_m

3.3 ano

3.4 ano

3.5 ne

3.6 ne, to druhé je m

3.7 ne, suma odchylek od průměru je 0

3.8 ano

4.1 95%

4.2 a)

4.3 intervalový

4.4 d)

4.5 $n = 1$

4.6 vyšší hladina spolehlivosti znamená širší interval

4.7 centrální limitní teorém

4.8 ano, ale hodnota σ_m se pro různá n liší

4.9 ne, protože v obou případech budeme interval spolehlivosti konstruovat okolo jiného výběrového průměru

4.10 $n = 100$

5.1 $m_{IQ}=98$; $s_{IQ}11$; $n=30$

- pokud bychom chtěli využít receptář Oseckých, pak budeme hledat sekci li (I výběr, intervalová proměnná) a v ní recept na proceduru **I μ S** (μ -průměr, interval Spolehlivosti)
- hledáme interval spolehlivosti se středem v m a takovou šířkou, aby s 95% pravděpodobností zahrnoval μ .
- $\alpha = 0,05$... 95% interval spolehlivosti
- průměr má výběrové rozložení t s průměrem m a výběrovou chybou $s_0 = s/\sqrt{n} = 11/5,5 = 2$
- interval tedy bude mít podobu $[m-X.s_0 ; m+X.s_0]$, kde X je hodnota t -rozložení odpovídající 2,5. a 97,5. percentilu (kvantil 0,025 a 0,975; tj. $\alpha/2$ a $1-\alpha/2$), mezi nimiž se nalézá 95% rozložení výběrových průměrů
- naše t -rozložení má $v = n - 1 = 29$ stupňů volnosti
- kvantily nalezneme nejnázem pomocí Excelu nebo v tabulkách t -rozložení – protože t -rozložení je symetrické $_{\alpha/2}t(v) = -_{1-\alpha/2}t(v)$; Excel umí hledat jen $_{\alpha/2}t(v)$
- $(X) = _{\alpha/2}t(v) = \text{TINV}(\alpha;v) = \text{TINV}(0,05;29) = 2,05$ (ta α místo $\alpha/2$ je ve vzorci proto, že Excel si z nějakého důvodu dodanou hodnotu v tomto vzorci sám vydělí dvěma)
- zkonstruujeme interval spolehlivosti: $[m-X.s_0 ; m+X.s_0] = [98-2.05*2; 98+2.05*2] = [93,9; 102,1]$

5.2 $r = -0,1$; $n = 30$

- pro tohle recept u Oseckých nenaleznete, ale zkuste to porovnat s receptem I p H na str. 19
- postup je stejný jako v předchozím případě, pouze s jedním krokem navíc – z-transformací
- $\alpha = 0,05$... 95% interval spolehlivosti
- výběrové rozložení korelace neznáme (Hendl 252); když se ale korelační koef. urč. způsobem přetransformuje, pak výběrové rozložení této transformované statistiky známe – jde o normální rozložení s $s_0=1/\sqrt{(n-3)}$
- jde o Fisherovu z-transformaci: $z = 0,5 \ln((1+r)/(1-r))$ – to je totéž, co funkce hyperbolický arkustangens (arctgh), není nutné to počítat, v Excelu to počítá funkce FISHER(r)
- takže v našem případě: $z = \text{FISHER}(-0,1) = -0,10034$ (čím dále od nuly, tím více se bude z a r lišit, maximem z je nekonečno)
- interval tedy bude mít podobu $[z-X*s_0 ; z+X*s_0]$, kde X je hodnota normálního rozložení odpovídající 2,5. a 97,5. percentilu (kvantil 0,025 a 0,975; tj. $\alpha/2$ a $1-\alpha/2$), mezi nimiž se nalézá 95% rozložení výběrových z-transformovaných korelací
- $s_0 = 1 / \sqrt{(30-3)} = 0,19$
- kvantily nalezneme nejnázem pomocí Excelu nebo v tabulkách normálního rozložení (nebo si vzpomeneme na 1,96
- protože normální rozložení je symetrické $_{\alpha/2}u = -_{1-\alpha/2}u$
- $(X) = _{\alpha/2}u = \text{NORMSINV}(\alpha/2) = \text{NORMSINV}(0,025) = -1,96$ (nás zajímá jen abs. hodnota)
- zkonstruujeme interval spolehlivosti: $[z-X*s_0 ; z+X*s_0] = [-0,10037-1,96*0,19 ; -0,10037+1,96*0,19] = [-0,47 ; 0,27]$
- tohle je ale interval v z-transformovaných hodnotách, musíme tedy ještě jeho meze transformovat zpět na koeficient r ; k tomu slouží v Excelu FISHERINV (neboli TGH
- $[\text{fisherinv}(-0,47) ; \text{fisherinv}(0,27)] = [-0,51 ; 0,28]$

6.1 Průměr je $M = 366,67$ a směrodatná odchylka $s = 149,35$.

6.2 Použije se t -rozdělení, neboť neznáme populační rozptyl rychlosti čtení. Kdybychom ho znali, pak můžeme použít normální rozdělení.

6.3 Směrodatná chyba odhadu průměru je $s_m = s / \sqrt{n} = 60,97$. Příslušné t -rozdělení bude mít 5 stupňů volnosti ($n-1$), kvantil $_{0,975}t(5) = 2,571$ (v excelu lze využít funkce $\text{TINV}(0,05; 5)$). Příslušný konfidenční interval bude mít meze $M - _{0,975}t(5) * s_m = 209,93$ a $M + _{0,975}t(5) * s_m = 523,40$. 95% interval spolehlivosti je (209,93; 523,40).

6.4 99% interval spolehlivosti bude širší. Kdybychom daný pokus a výpočet zopakovali stokrát, pak v případě 95% intervalu by skutečná hodnota průměru ležela v 95 případech v nalezeném intervalu. V 99% intervalu by skutečná hodnota měla ležet v 99 případech, je tedy logické, že tyto intervaly musí být širší. Analogickým postupem k minulému dospějeme k intervalu (120,81; 612,52).

6.5 Intervaly spolehlivosti budou výrazně užší, odhad průměru tedy bude přesnější. Změní se směrodatná chyba odhadu průměru ($s_m = s / \sqrt{n} = 35,20$) a t -rozdělení – bude mít 17 stupňů volnosti, vyjdou tedy nižší kvantily – $_{0,975}t(17) = 2,110$. 95% interval spolehlivosti je (292,39; 440,94).

7.1 Časy z jednotlivých měření není třeba brát v úvahu, stačí počítat pouze s rozdíly časů; průměr rozdílů je

M = 20,7, směrodatná odchylka s = 10,53, směrodatná chyba odhadu průměru $s_m = 3,33$ a příslušný kvantil $_{0,975}t(9) = 2,26$ (v excelu TINV(0,05; 9)).
 95% interval spolehlivosti je (13,17; 28,23).
 7.2 Směrodatná chyba odhadu průměru je $s_m = 1,090$, kvantil $_{0,975}t(46) = 2,013$.
 95% interval spolehlivosti je (14,17; 18,56).

8. Průměr je M = 0,4845, směrodatná odchylka s = 0,0191, směrodatná chyba odhadu průměru $s_m = 0,0068$ a kvantil $_{0,975}t(7) = 2,365$.
 95% interval spolehlivosti je (0,4685; 0,5005); neboli, na 95% hladině spolehlivosti nemohou prohlásit, že je výčepní okrádá (hodnota 0,5 spadá do intervalu).

9.1 Znalost populační směrodatné odchylky ovlivní rozdělení, které použijeme – místo t-rozdělení vezmeme normální rozdělení. Průměr je M = 125, směrodatnou chybu odhadu průměru je 4,74 (vypočítáme ji pomocí populační směrodatné odchylky). Příslušný kvantil normálního rozdělení je $u_{0,975} = 1,96$.
 95% interval spolehlivosti vychází (115,70; 134,30).

9.2 Nyní tedy použijeme t-rozdělení. Průměr je stejný M = 125, směrodatná odchylka vychází 14,99, směrodatná chyba odhadu průměru 4,74, kvantil $_{0,975}t(9) = 2,26$.
 95% interval spolehlivosti odhadu průměru je (114,27; 135,72). Při použití t-rozdělení je interval širší, ačkoli výběrová směrodatná odchylka se prakticky shoduje s populační. Příslušné kvantily t-rozdělení jsou vyšší než normálního, nejvýraznější jsou tyto rozdíly u nízkých stupňů volnosti.

9.3 Výběrový rozptyl je 224,67, rozptyl po transformaci má rozdělení chí-kvadrát, příslušné kvantily pro 9 stupňů volnosti a pro pravděpodobnosti 2,5% a 97,5% jsou 2,70 a 19,02 (lze je vypočítat v excelu pomocí CHIINV(0,025; 9) a CHIINV(0,975; 9)). Meze intervalu spolehlivosti vypočítáme jako $(n - 1) * s^2 / 19,02$ a $(n - 1) * s^2 / 2,70$.
 95% interval spolehlivosti pro rozptyl je (106,29; 748,78). Pokud by nás zajímal interval spolehlivosti pro směrodatnou odchylku, získáme jej odmocněním: (10,31; 27,36).

9.4 Rozdíl je jenom v kvantilech, pro 0,5% a 99,5% jsou 1,73 a 23,59.
 99% interval spolehlivosti pro rozptyl je (85,72; 1165,48), pro směrodatnou odchylku (9,26; 34,14).

10. Příslušné kvantily chí-kvadrát rozdělení jsou 140,169 a 67,328. Tedy:

$$\sqrt{\frac{(n-1)s^2}{\chi^2_R}} < \sigma < \sqrt{\frac{(n-1)s^2}{\chi^2_L}} = \sqrt{\frac{(101-1)(1,68)^2}{140,169}} < \sigma < \sqrt{\frac{(101-1)(1,68)^2}{67,328}}$$

$$1,419 < \sigma < 2,047$$

99% interval spolehlivosti pro směrodatnou odchylku tedy je (1,419; 2,047).

11. Směrodatná chyba odhadu průměru 0,748, kvantil t-rozdělení je 2,064.

95% interval spolehlivosti pro odhad průměru je (5,466; 8,554).

Příslušné kvantily chí-kvadrát rozdělení jsou 39,364 a 12,401, tedy:

$$\sqrt{\frac{(n-1)s^2}{\chi^2_R}} < \sigma < \sqrt{\frac{(n-1)s^2}{\chi^2_L}} = \sqrt{\frac{(25-1)(3,74)^2}{39,364}} < \sigma < \sqrt{\frac{(25-1)(3,74)^2}{12,401}}$$

$$2,92 < \sigma < 5,203$$

95% interval spolehlivosti pro odhad směrodatné odchylky tedy je (2,92; 5,203).

12. Známe populační směrodatnou odchylku, použijeme normální rozdělení. Směrodatná chyba odhadu průměru je $s_m = 1,652$, kvantil $u_{0,975} = 1,96$.

95% interval spolehlivosti je $135,89 < m < 142,37$.

13. Nejprve se použije Fisherova transformace (tzv. z-transformace):

$z = 0,5 * \ln((1+r)/(1-r))$ (ekvivalentní je funkce **arctanh**(r), hyperbolický arkustangens, kterou umí kalkulačky)

Rozdělení z je pro $n > 10$ přibližně normální s průměrem a rozptylem:

$$N\left(\frac{1}{2} \ln \frac{1+p}{1-p}; \frac{1}{n-3}\right)$$

Po této transformaci dostaneme $z = 0,4847$ a $s_z = 0,2425$. Příslušný kvantil normálního rozdělení je 1,96, netransformovaný interval spolehlivosti je (0,0093; 0,9601) – ten se ale týká z, nikoli samotné korelace.

Musíme provést inverzní transformaci:

$r = (e^{2z} - 1) / (e^{2z} + 1)$ (ekvivalentní je funkce **tanh**(z), hyperbolický tangens, kterou umí kalkulačky)

Po transformaci mezi intervalu vychází 95% interval spolehlivosti pro Pearsonův korelační koeficient (0,0093; 0,7443).

14. Nie je možné nulovú hypotézu zamietnuť, pretože interval spoľahlivosti ukazuje, že hodnota 20 je pravdepodobná hodnota parametra populácie.

15. a, d – štúdia testuje rozdiel medzi priemermi a významný rozdiel by mal byť väčší alebo menší ako 0. Potom intervaly spoľahlivosti, ktoré neobsahujú nulu reprezentujú štatisticky významné zistenia.

16. interval spoľahlivosti 99% obsahuje všetky hodnoty ktoré má interval spoľahlivosti 95% a zároveň ho rozširuje. Ak nie je niečo významné na hladine 0,05, nebude to významné ani na hladine 0,01.