## *Statistická indukce a intervalové odhady*

Je k výběru náhodného vzorku třeba, aby byly splněny následující podmínky (ano-ne)?
1. Aby byly skóry v populaci normálně rozložené.
2. Každý jedinec v populaci (jednotka zkoumání) musí mít stejnou pravděpodobnost, že bude vybrán do vzorku.
3. Výběr kteréhokoli jedince (jednotky zkoumání) musí být zcela nezávislý na výběru kohokoli jiného.

*U Wechslerových inteligenčních škál jsou skóry normálně rozložené s $\mu$=100 a $\sigma$=15. Představte si, že jsme otestovali náhodný vzorek 9 lidí, spočítali jejich průměrné skóre a celou proceduru zopakovali 1000krát.*
15. Odhadněte směrodatnou odchylku těch 1000 průměrů.
16. Zhruba jaká část (v %) těchto výběrových průměrů s $n$ = 9 by byla vyšší než 105? A než 110?
17. Zhruba jaká část (v %) těchto výběrových průměrů s $n$ = 9 by byla mezi 95 a 105? A mezi 90 a 110?
18. Budou tyto výběrové průměry normálně rozložené?
19. Jaký je rozptyl tohoto rozložení výběrových průměrů?
20. Kdyby $n$ = 225 (místo 9), jaká by byla hodnota směrodatné chyby?
21. Kdyby $n$ = 225, jaká část (v %) výběrových průměrů by se pohybovala nejvýše jeden bod od 100, tj. mezi 99 a 101?
23. Bude rozložení výběrových průměrů přibližně normální i tehdy, kdyby rozložení skórů v populaci nebylo normální?
24. Který matematický teorém tvrdí, že výběrové rozložení průměrů se s rostoucím $n$ blíží normálnímu, bez odhledu na tvar rozložení proměnné v populaci?
26. Známe-li $\sigma$, platí, že $m \pm 1{,}96\sigma_m$ tvoří 95% interval spolehlivosti pro jakékoli $n$?
27. Předpokládejme proměnnou s normálním rozložením a známou $\sigma$. Pokud vybereme náhodně 2 vzorky o $n$ = 100 a na obou spočítáme 68% interval spolehlivosti, budou tyto dva intervaly shodné?

Jsou následující páry ekvivalentními výrazy?
28. (1) směrodatná chyba $\bar{X}$ a (2) směrodatná odchylka výběrového rozložení $\bar{X}$.
29. (1) $\sigma^2/n$ a (2) směrodatná chyba $\bar{X}$.
30. (1) $\sigma_{\bar{X}}^2$ a (2) rozptyl výběrového rozložení $\bar{X}$.
31. (1) populační rozptyl $\sigma^2$ a (2) $n$ krát $\sigma_{\bar{X}}^2$
32. (1) průměr výběrového rozložení $\bar{X}$ a (2) $\sigma_{\bar{X}}^2$
34. (1) $\mu$ a (2) $(\Sigma X)/n$
35. (1) $\bar{X}$ a (2) $(\Sigma x)/n$
36. (1) $s^2$ a (2) $\Sigma x^2/(n-1)$

37. Pokud budeme provádět mnoho různých studií na různá témata a vždy v nich budeme k odhadovaným statistikám tvořit 95% intervaly spolehlivosti, kolik z těchto intervalů asi obsahuje odhadovaný parametr?
38. V kterém z následujících případů by nárůst velikosti vzorku způsobil největší zúžení intervalu spolehlivosti?

a) z 5 na 25
b) z 10 na 30
c) ze 40 na 60

39. Který typ odhadu lépe sděluje přesnost odhadu, bodový, nebo intervalový?

11.1 z se má k $\sigma_m$ jako $t$ se má k
a) $\sigma$ b) $\sigma^2$ c) $s$ d) $s_m$
11.4 Za jakých podmínek platí $\sigma = \sigma_m$ ?

A. Provádíme výzkum toho, zda či jak dlouhodobá hospitalizace škodí dětem ve vývoji. Jednou z výzkumných otázek je, zda nedochází k zabrždění vývoje intelektu. Pro tento účel jsme 30 dlouhodobě hospitalizovaným dětem v mladším školním věku rozdali inteligenční test s následujícími výsledky: $m_{IQ}$=98, $s_{IQ}$11.

i) Stanovte 95% interval spolehlivosti pro průměrnou hodnotu inteligence v populaci dlouhodobě hospitalizovaných dětí v mladším školním věku ($\mu_{IQ}$).

ii) Dále jsme zjistili, že délka hospitalizace (ve dnech) koreluje s IQ, r=-0,1. Stanovte 95% interval spolehlivosti pro korelaci mezi délkou hospitalizace a IQ ($\rho$) (viz Hendl 252, pozor na chybu, výběrové rozložení Fischerova Z je normální, nikoli $t$)

*řešení na následující straně*

**Odpovědi**

1. ne
2. ano
3. ano

15. 5
16. z=(105-100)/5=1 … 16%; pro 110 je to 2,3%
17. cca 68%; 95%
18. ano, přibližně ano
19. 25
20. 1
21. cca 68%

23. ano
24. centrální limitní teorém

26. ano, ale hodnota $\sigma_m$ se pro různá $n$ liší
27. ne, protože v obou případech budeme interval spolehlivosti konstruovat okolo jiného výběrového průměru.
28. ano
29. ne, $\sigma^2/n = \sigma_{\bar{X}}^2$ a ne $\sigma_{\bar{X}}$
30. ano
31. ano
32. ne

34. ne, (2) je $m$, $\bar{X}$
35. ne, suma odchylek od průměru je 0
36. ano
37. 95%
38. a)
39. intervalový

11.1 d)
11.4 $n$ = 1

A.

I.

$m_{IQ}$=98, $s_{IQ}$11, n=30

Pokud bychom chtěli využít receptář Oseckých, pak budeme hledat sekci Ii (**I** výběr, **i**ntervalová proměnná) a v ní recept na proceduru **I μ S** (μ-průměr, interval **S**polehlivosti)

Hledáme interval spolehlivosti se středem v $m$ a takovou šířkou, aby s 95% pravděpodobností zahrnoval μ.

1. α=0,05 ... 95% interval spolehlivosti
2. průměr má výběrové rozložení $t$ s průměrem $m$ a výběrovou chybou $s_0$=s/odm(n)

    $s_0$=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. α/2 a 1-α/2), mezi nimiž se nalézá 95% rozložení výběrových průměrů. Naše $t$-rozložení má $\nu = n - 1$ = 29 stupňů volnosti.

    Kvantily nalezneme nejsnáze pomocí Excelu nebo v tabulkách $t$-rozložení

    Protože $t$-rozložení je symetrické ${}_{\alpha/2}t(\nu) = -{}_{1-\alpha/2}t(\nu)$. Excel umí hledat jen ${}_{\alpha/2}t(\nu)$.

    (X) = ${}_{\alpha/2}t(\nu)$ = TINV(α; ν) = TINV(0,05;29)= 2,05 (ta α místo α/2 je ve vzorci proto, že excel si z nějakého důvodu dodanou hodnotu v tomto vzorci sám vydělí dvěma)

3. Zkonstruujeme interval spolehlivosti

    [$m$-X.$s_0$ ; $m$+X.$s_0$] = [98-2.05*2; 98+2.05*2] = **[93,9 ; 102,1]**

II.

r = -0,1 ; n = 30

Pro tohle recept u Oseckých nenaleznete, ale zkuste to porovnat s receptem I ρ H na str. 19.

Postup je stejný jako v předchozím případě, pouze s jedním krokem navíc – z-transformací.

1. a=0,05 – 95% interval spolehlivosti
2. 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/odm(n-3)**. Jde o Fisherovu z-transformaci:

    z = 0,5 ln((1+r)/(1-r)) (to je totéž, co funkce hyperbolický arkustangtens (arctgh), neni nutné to počítat - v excelu to počítá funkce FISHER(r))

    Takže v našem případě:

z= 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/odm(30-3)=0,19

Kvantily nalezneme nejsnáze 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)= $_{\alpha/2}u$ = NORMSINV($\alpha$/2) = NORMSINV(0,025) = -1,96 (nás zajímá jen abs. hodnota)

3. 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)

[fisherinv(-0,47) ; fisherinv(0,27)] = **[-0,51 ; 0,28]**