

Úlohy pro 4ST414 – 3. Úloha

Termín – 31. říjen 2025, 20:00

Vysvětlivky: P – povinná úloha, T – teoretická úloha, * - bonusová, obtížnější úloha.

K řešení prosím přiložte všechny použité kódy v R, Excelu.

1. Poměrový/regresní odhad (P, 5 b + 1 b)

Soubor MU284 (Saerndal a kol. 1992, <http://lib.stat.cmu.edu/datasets/mu284>) obsahuje různé údaje o N=284 švédských municipalitách. Soubor je obsažen např. R package sampling (příkaz data (MU284)). Cílem šetření je odhadnout počet obyvatel v roce 1985, tj úhrn proměnné P85 s využitím znalosti o P75, což je počet obyvatel municipality v tisících v příslušných letech (1985 a 1975). Jako pomocnou proměnnou budeme brát P75, o níž předpokládáme, že její hodnoty jsou předem za všechny municipality známy. Cílem je odhadnout počet obyvatel v roce 1985 v menších a středně velkých municipalitách, tj. v těch, které v roce 1975 měly maximálně 150 tis. obyvatel. Podobná úloha se například řeší, pokud je cílem získat bleskový předběžný odhad sčítání lidu. Tyto výsledky za malé střední municipality se pak kombinují s odhady za zbylé (zejména) velké municipality.

- i) Uvažujte prostý náhodný bez vracení jedné desetiny municipalit. Pro spolehlivost 99 % spočítejte asymptotické intervaly spolehlivosti pro prostý náhodný výběr, poměrový a regresní odhad. Porovnejte šířky intervalu spolehlivosti odhadů. **(3 b)**
- ii) S využitím znalosti proměnné P75 pro požadovanou přesnost $\Delta = 1 \%$ populačního průměru obyvatel v roce 1975 u municipalit s 150 tis. obyv. a méně a spolehlivost 90 % vypočítejte rozsah výběru pro prostý náhodný výběr, poměrový a regresní odhad. Návod – populační průměry a rozptyly vypočítejte z pomocné proměnné P75. Využijte znalosti hodnoty korelačního koeficientu mezi P85 a P75 a skutečnosti, že rozptyl poměrového a regresního odhadu je o $100\rho_{P75,P85}^2 \%$ nižší. **(2 b)**
- iii) Pomocí simulace ověřte, zda se i v praxi je rozptyl u regresního odhadu je o o $100\rho_{P75,P85}^2 \%$ nižší než výběrový úhrn. **(1 b, bonus)**