

Úlohy pro 4ST414 – 4. úloha

(termín odevzdání 20. 11. 2025, 20:00)

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

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

1. Alokace výběru pro oblastní (stratifikovaný) náhodný výběr (10 b + 2 b bonus v úloze iv))

Soubor Mzdy.txt obsahuje údaje o mzdě a vzdělání N=1134 zaměstnanců. Vzdělání je kódováno náslovně: 1 – ZŠ vzdělání, 2 – SŠ bez matury, 3 – SŠ s maturitou, 4 – VŠ. Má se vybrat $n = 50 + |MM - 6| * 10$ zaměstnanců (MM je měsíc data vašeho narození). Za strata zvolíme vzdělání zaměstnanců.

- i) Proveďte rozklad rozptylu mezd, posuďte podíl vnitroskupinové a meziskupinové variability mezd. **(2 b)**
- ii) Proveďte alokaci výběru:
 - a. rovnoměrnou
 - b. proporcionalní
 - c. optimální. **(2 b)**
- iii) Pro 95% spolehlivost určete intervalu spolehlivosti odhadu průměrné mzdy pro prostý náhodný výběr bez vracení z celé populace a pro stratifikované náhodné výběry z bodu 2. Povede využití stratifikovaného výběru k zvýšení přesnosti odhadu průměrné mzdy? Tratíme hodně na přesnosti, když použijeme proporcionalní alokaci místo optimální? Porovnejte přesnost výběrových plánů mezi sebou. **(3 b)**
- iv) Pro oblastní výběry dle ii) spočítejte pro 95% spolehlivost rovněž intervaly spolehlivosti pro jednotlivá strata. Dávají dostatečně přesné odhady průměrných příjmů jednotlivých skupin (zadavatel považuje odhad průměrné mzdy za dostatečně přesný, když šířka intervalu spolehlivosti 2Δ nepřesáhne 10 % populačního průměru). Která z alokací výběru je podle vás vhodná, pokud je cílem výběrového šetření získat při daném rozsahu výběru n co nejkvalitnější odhady průměrných mezd jednotlivých skupin zaměstnanců? (3b)
Použijte vhodné grafické metody pro srovnání. **(2 b bonus)**