

M9750 Robustní a neparametrické statistické metody

cvičení 7 - robustní odhady v lineárním regresním modelu

1. Uvažujte data **heptathlon** z knihovny **HSAUR** (proměnné **hurdles** a **shot**).
 - (a) Pomocí klasické lineární regrese modelujte závislost výsledku vrhu koulí na výsledku z běhu na 100 metrů překážek.
 - (b) Odhadnutou regresní přímku přidejte do grafu.
 - (c) Posuďte vhodnost zvoleného modelu. Obsahují data nějaká vlivná nebo odlehlá pozorování?
 - (d) Odhadněte regresní koeficienty pomocí 50% regresního kvantilu a opět odhadnutou regresní přímku přidejte do grafu.
 - (e) A konečně, uvažujte model bez odlehlého pozorování a proveďte odhad metodou nejmenších čtverců na těchto datech; odhadnutou regresní přímku přidejte do grafu.
 - (f) Všechny tři odhady porovnejte. Který z nich nejlépe popisuje sledovanou závislost?
2. Uvažujte data **CYGOB1** z knihovny **HSAUR** (logaritmus povrchové teploty **logst** a logaritmus světelné intenzity **logli** 47 hvězd).
 - (a) Data si nejprve vykreslete a poté pomocí lineární regrese modelujte závislost logaritmu světelné intenzity na logaritmu povrchové teploty.
 - (b) Odhadnutou regresní přímku přidejte do grafu.
 - (c) Posuďte vhodnost zvoleného modelu. Obsahují data nějaká vlivná nebo odlehlá pozorování?
 - (d) Odhadněte regresní koeficienty pomocí 50% regresního kvantilu a opět odhadnutou regresní přímku přidejte do grafu.
 - (e) A konečně, uvažujte model bez odlehlých pozorování a proveďte odhad metodou nejmenších čtverců na těchto datech; odhadnutou regresní přímku přidejte do grafu.
 - (f) Všechny tři odhady porovnejte. Který z nich nejlépe popisuje sledovanou závislost?
3. Do vytvořeného grafu z minulého bodu přidejte další odhady:
 - (a) M-odhad s Huberovou funkcí ψ .
 - (b) GM-odhad s Huberovou funkcí ψ .
 - (c) nejmenší useknuté čtverce (LTS).
 - (d) nejmenší medián čtverců (LMS).
 - (e) Theilův-Senův odhad.
 - (f) Siegelův opakovaný medián.
4. **Porovnání robustních odhadů v regresním modelu.** Zkoumejte chování předchozích odhadů v regresním modelu přímky ($\beta_0 = \beta_1 = 1$) na základě náhodného

výběru (Y_i, x_i) o rozsahu $n = 30$. Nejprve si nasimulujte realizace regrese x_1, \dots, x_{30} (z rovnoměrného spojitého rozdělení $Rs(0, 10)$) a dále je uvažujte pevné.

- (a) Generujte chyby modelu e_1, \dots, e_{30} ze standardizovaného normálního rozdělení a spočtěte hodnoty $Y_i = \beta_0 + \beta_1 x_i + e_i$.
- (b) Odhadněte směrnici regresní přímky β_1 pomocí LSE, LAD a všech odhadů z předchozího bodu.
- (c) Celý postup opakujte 10 000 krát. Na základě těchto simulací odhadněte střední hodnotu a střední čtvercovou chybu všech odhadů (průměr empirických středních čtvercových chyb).
- (d) Celý postup opakujte s chybami modelu e_i z Cauchyho rozdělení.
- (e) Celý postup opakujte s daty, kde první pozorování nahradíte hodnotou $(Y_1, x_1) = (10, 0)$.

Funkce, které by se mohly hodit: `points`, `abline`, funkce `rq` z knihovny `quantreg`, funkce `rlm`, `lqs` z knihovny `MASS`, funkce `mblm` z knihovny `mblm`.