PŘEDPOKLADY MNČ

1. Regresní parametry (b_i) mohou nabývat libovolných hodnot. (V praxi obvykle existují omezení fyzikálního smyslu.)
2. Regresní model je lineární v parametrech. (Pomocí LR lze řešit např. i polynomické závislosti.)
3. Žádné dva sloupce matice **X** nejsou kolineární, tj. rovnoběžné vektory. (V datech matice **X** se nevyskytuje multikolinearita, matice $\mathbf{X}^T\mathbf{X}$ je dobře podmíněná.)
4. Chyby ε_i mají nulovou střední hodnotu. (Chyby "rovnoměrně" kolísají kolem 0, jejich hodnoty nejsou systematicky ovlivňovány.)
5. Náhodné chyby ε_i mají konstantní rozptyl (homoskedasticita).
6. Jednotlivé hodnoty závisle proměnné jsou získány nezávislým měřením. (Vzájemně nekorelované chyby ε_i.)
7. Nezávisle proměnné veličiny nejsou zatíženy systematickými či náhodnými chybami nebo jsou náhodné chyby zanedbatelné vůči náhodným chybám

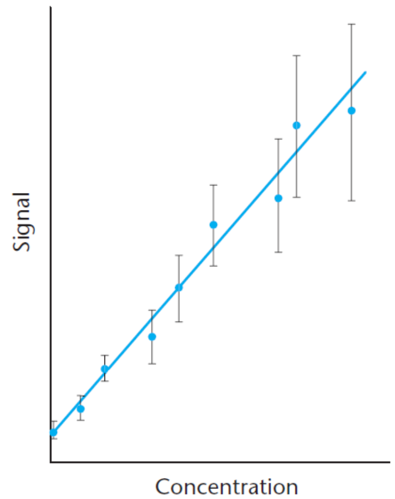
POSTUPY PŘI PORUŠENÍ PŘEDPOKLADŮ MNČ

METODA PODMÍNKOVÝCH NEJMENŠÍCH ČTVERCŮ

- Používá se, pokud jsou na parametry b_i kladena omezení. Patří sem úlohy, kdy:
  - Některé parametry musí nabývat zadaných hodnot.
  - Některé parametry musí zachovávat vzájemné poměry.
  - Regresní model musí procházet zadanými body.

METODA VÁŽENÝCH NEJMENŠÍCH ČTVERCŮ

- Aplikujeme ji v případě, kdy náhodné chyby nemají konstantní rozptyl (byla prokázána heteroskedasticita v datech).
- Do rovnice pro výpočet účelové funkce se vkládá váhový koeficient w_i.



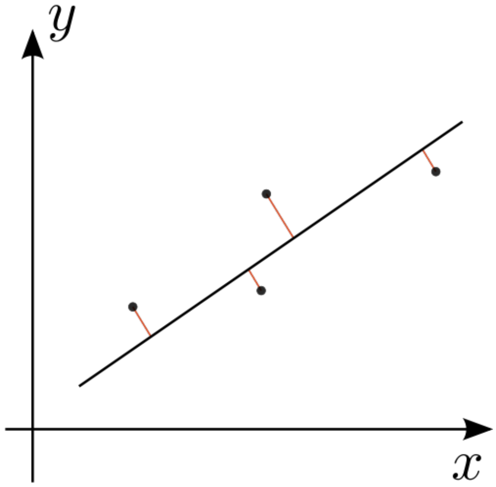
$$RSC = U(b) = \sum_i \left[y_i w_i - \sum_j w_i x_{ij} b_j \right]^2 \quad kde \quad w_i \ je \quad např. \frac{1}{y} \ či \ \frac{1}{\sigma}$$

METODA RACIONÁLNÍCH HODNOSTÍ

- Používá se v případě prokázané multikolinearity v datech.
- Obvykle se aplikuje při řešení polynomické regrese.

ORTOGONÁLNÍ REGRESE

- Jde o zvláštní případ metody rozšířených nejmenších čtverců pro případ, kdy chyby na ose x a y jsou srovnatelné ($s_x = s_y$).
- Je to obdoba MNČ, zde se minimalizuje suma kolmých vzdáleností experimentálních bodů od přímky.



ROBUSTNÍ REGRESE

- V případě porušení předpokladu normality chyb závisle proměnné nejsou odhady parametrů získané MNČ efektivní.
- Odhady parametrů získané MNČ jsou silně citlivé na přítomnost OB a EB.
- V takových případech místo kritéria MNČ použijeme robustního kritéria, které je jak na porušení předpokladu normality, tak na vlivné body málo citlivé. Nejznámější jsou M-odhady vycházející z minimalizace kritéria:

$$U_M = \sum_{i=1}^{n} \rho\left(\frac{e_i}{\sigma}\right) \quad kde \quad \rho \quad je\ vhodná\ \ funkce$$

Př.: Welschův M-odhad – minimalizuje se čtverec vážených normovaných reziduí:

$$\rho = \frac{w^2}{2} \cdot \left(1 - \exp(-\frac{e_i / \sigma}{w})^2 \right)$$

L_p – regrese

- Je příkladem robustních metod, kde se na rozdíl od MNČ minimalizuje součet:

$$U = |e_i|^p$$

- pro p = 1 – ***L_1 aproximace*** neboli ***robustní mediánová regrese***, je vhodná pro data s OB na obou stranách nebo chybami s Laplaceovým rozdělením (špičatost rozdělení chyb > 3)
- pro p =2 – MNČ
- pro 1<p<2 – robustní regrese vůči odlehlým hodnotám (standardně se p volí 1,5)
- pro p > 2 (obvykle p = 5-10) – ***metoda nejmenší maximální chyby***, což je metoda silně nerobustní k OB a je vhodná pro rovnoměrné rozdělení chyb (špičatost rozdělení chyb < 3)

Pro L_1-aproximaci je účelová funkce $$U = \sum_{i=1}^{n} |e_i|$$

a v případě přímky je účelová funkce $$U = \sum_{i=1}^{n} |y_i - b_0 - b_1 x_i|$$