

MA012 Statistika II

cvičení 9

Ondřej Pokora (pokora@math.muni.cz)

Ústav matematiky a statistiky, Přírodovědecká fakulta, Masarykova univerzita, Brno

(podzim 2015)



Příklad 1 (z přednášky)

V letech 1953 – 1983 byly měřeny ztráty vody při distribuci do domácností. Výsledky měření jsou uloženy v souboru `voda.csv`. Proměnná x označuje množství vyrobené vody, proměnná Y ztrátu. Ověřte, zda se v datech vyskytuje autokorelace 1. řádu a případně ji odstraňte.

`reseni-09-1.R` [Vhodný model: polynom 7. stupně, autokorelace 1. řádu se nezamítá, normalita reziduí u nového modelu se nezamítá.]

Příklad 2

V proměnné `longley` v R jsou uloženy makroekonomické údaje USA z let 1947–1962. Modelem lineární regrese zkoumejte závislost zaměstnanosti na HDP a velikosti populace starší 14 let: `Employed ~ GNP + Population`. Ověřte, zda se v datech vyskytuje autokorelace 1. řádu a případně se ji pokuste odstranit. Uvažte i autokorelaci ve tvaru $AR(1)$ v závislosti na roce sběru dat (`Year`).

`reseni-09-2.R`

Příklad 3

V proměnné `LakeHuron` v R jsou uloženy roční údaje o hloubce jezera Huron (ve stopách) v letech 1875 – 1972. Nalezněte vhodný regresní model (nevýše 7. stupně) a ověřte, zda se v datech vyskytuje autokorelace 1. řádu. Případně se ji pokuste odstranit. Zkoumejte také normalitu reziduí.

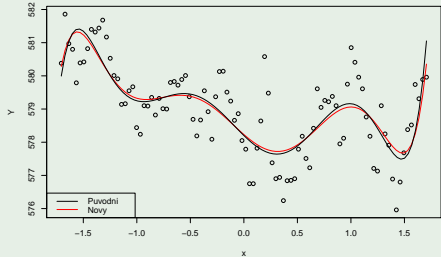
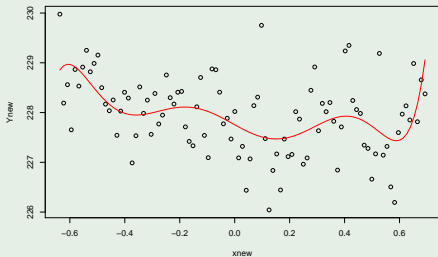
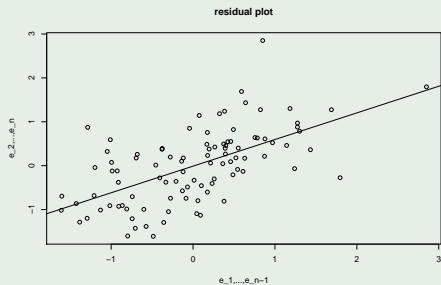
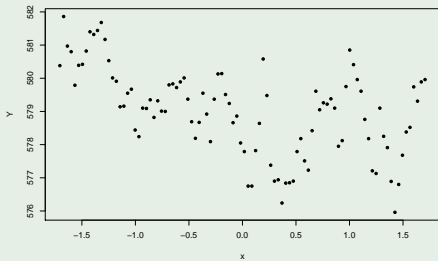
[reseni-09-3.R](#)

Příklad 4 (z přednášky)

V souboru `vydaje.csv` jsou uložena data o 20 náhodně vybraných domácnostech. Sloupce proměnné „domácnosti“ obsahují postupně tyto údaje: výdaje za potraviny a nápoje (Y), počet členů domácnosti (X_1), počet dětí (X_2), průměrný věk výdělečně činných (X_3) a příjem domácnosti (X_4). Metodou postupné regrese zkonstruujte model s nejlepší podmíněností regresorů.

[reseni-09-4.R](#)

Příklad 3



Příklad 5

V souboru `cement.csv` jsou uloženy údaje, které se týkají chemického složení portlandského cementu:

- y množství tepla v kaloriích na gram cementu
- x_1 Tricalcium aluminate $3\text{CaO} \cdot \text{Al}_2\text{O}_3$ v %
- x_2 Tricalcium silicate $3\text{CaO} \cdot \text{SiO}_2$ v %
- x_3 Tetracalcium alumino ferrite $4\text{CaO} \cdot \text{Al}_2\text{O}_3 \cdot \text{Fe}_2\text{O}_3$ v %
- x_4 Dicalcium silicate $2\text{CaO} \cdot \text{SiO}_2$ v %

Testujte multikolinearitu v lineárním modelu závislosti množství tepla na obsahu 4 hlavních složek cementu. Metodou postupné regrese nalezněte vhodný model. Poté ověřte normalitu reziduí.

`reseni-09-5.R` [Multikolinearita se nezamítá, vhodný model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_4 X_4, \text{ normalita reziduí se nezamítá.}]$$

Příklad 6

V proměnné `mtcars` v R jsou uložena data pro modelování závislosti spotřeby paliva osobních automobilů (proměnná `mpg`, počet mil/galon) na vlastnostech motoru, které jsou popsány následujícími proměnnými:

| | |
|-------------------|--|
| <code>cyl</code> | počet válců |
| <code>disp</code> | objem válců (kubické palce) |
| <code>hp</code> | výkon (počet koní) |
| <code>drat</code> | převodový poměr zadní nápravy |
| <code>wt</code> | hmotnost vozidla (kilolibry) |
| <code>qsec</code> | zrychlení (počet sekund z 0 na 1/4 míle) |
| <code>vs</code> | uspořádání válců (1 – „V“, 0 – za sebou) |
| <code>am</code> | převodovka (0 – automat, 1 – manuál) |
| <code>gear</code> | počet převodových stupňů |
| <code>carb</code> | počet karburátorů |

Testujte multikolinearitu v daném modelu. Metodou postupné regrese nalezněte vhodný model. Ověřte také normalitu reziduí.

`reseni-09-6.R` [Multikolinearita se nezamítá, vhodný model:

$$\text{mpg} = \beta_0 + \beta_1 \text{wt} + \beta_2 \text{cyl}, \text{ normalita reziduí se nezamítá.}]$$