

VÝSTAVBA REGRESNÍHO MODELU V ANALYTICKÉ LABORATOŘI

MILAN MELOUN

Katedra analytické chemie, Univerzita Pardubice, Čs. legií 565, 532 10 Pardubice,

e-mail: milan.meloun@upce.cz

Při výstavbě regresních modelů se užívá metody nejmenších čtverců. Metoda nejmenších čtverců poskytuje postačující odhady parametrů jenom při současném splnění všech předpokladů o datech a o regresním modelu. Pokud není následujících 7 předpokladů splněno, ztrácí metoda nejmenších čtverců své vlastnosti:

I. Regresní parametry β mohou nabývat libovolných hodnot.

II. Regresní model je lineární v parametrech a platí {xe "Model:aditivní"} aditivní model měření.

III. Matice nenáhodných, nastavovaných hodnot vysvětlujících proměnných \mathbf{X} má hodnotu rovnou právě m , žádné její dva sloupce \mathbf{x}_j , \mathbf{x}_k nejsou kolineární.

IV. {xe "Chyby:náhodné"} Náhodné chyby ε_i mají nulovou střední hodnotu $E(\varepsilon_i) = 0$.

V. {xe "Člen:absolutní"} Náhodné chyby ε_i mají konstantní a konečný rozptyl $E(\varepsilon_i^2) = \sigma^2$

VI. Náhodné chyby ε_i jsou vzájemně {xe "Chyby:nekorelované"} nekorelované a platí $\text{cov}(\varepsilon_i, \varepsilon_j) = E(\varepsilon_i \varepsilon_j) = 0$. VII. Chyby ε_i mají normální rozdělení $N(0, \sigma^2)$.

Postup hledání regresního modelu se skládá z kroků: 1. Návrh modelu začíná vždy od nejjednoduššího modelu, lineárního. 2. Předběžná analýza dat sleduje proměnlivost proměnných na rozptylových diagramech, indexových grafech. Vyšetřuje se multikolinearita, heteroskedasticita, autokorelace a vlivné body. 3. Odhadování parametrů se provádí klasickou metodou nejmenších čtverců, následuje testování významnosti parametrů Studentovým t-testem. Střední kvadratická chyba predikce MEP a Akaikovo informační kritérium AIC jsou rozhodčí kritéria při hledání modelu. 4. Regresní diagnostika provádí identifikaci vlivných bodů a ověření předpokladů metody nejmenších čtverců. V případě více vysvětlujících proměnných se posoudí vhodnost proměnných pomocí parciálních regresních grafů a parciálních reziduálních grafů. 5. Konstrukce zpřesněného modelu: parametry zpřesněného modelu jsou odhadovány s využitím (a) metody vážených nejmenších čtverců při nekonstantnosti rozptylu, (b) metody zobecněných nejmenších čtverců při autokorelaci, (c) metody podmínkových nejmenších čtverců při omezení kladených na parametry,

(d) metody racionálních hodnotí u multikolinearity, (e) metody rozšířených nejmenších čtverců pro případ, že všechny proměnné jsou zatíženy náhodnými chybami, a konečně (f) robustních metod pro jiná rozdělení než normální a data s vybočujícími hodnotami a extrémny. Regresní diagnostika: metoda MNČ nezajišťuje obecně nalezení přijatelného modelu. Musí být splněny podmínky, odpovídající složkám tzv. *regresního* {xe "Triplet:regresní"} *tripletu* [data, model, metoda odhadu]: a) {xe "Kvalita:dat"} vhodnosti dat pro navržený regresní model (*data*), b) {xe "Kvalita:modelu"} vhodnosti modelu pro daná data (*model*), c) splnění základních předpokladů MNČ (*metoda*).

1. Data: sleduje se výskyt *vlivných bodů*: *Vybočující pozorování*, které se liší v hodnotách vysvětlované (závisle) proměnné y od ostatních, a *extrémy*, které se liší v hodnotách nezávisle proměnných x . Jsou však i body jak vybočující, tak i extrémní.

2. Model: kvalitu regresního modelu lze posoudit v případě jedné vysvětlující proměnné x přímo z grafu závislosti y na x . Omezíme se na {xe "Grafy:parciální regresní"} parciální regresní grafy, jež umožňují posouzení kvality navrženého regresního modelu a indikují i vybočující body.

3. Metoda: v praxi bývají některé předpoklady MNČ porušeny: a) Na parametry jsou kladena omezení, následuje užití *podmínkových nejmenších čtverců*). b) Kovarianční matice chyb není diagonální (autokorelace) a data nemají stejný rozptyl (heteroskedasticita), pak je nutno užít *metody zobecněných nejmenších čtverců*, či *metody vážených nejmenších čtverců*. c) Rozdělení dat nelze považovat za normální nebo se v datech vyskytují vlivné body a užije se {xe "Regrese:robustní"} *robustního* kritéria, které je na porušení předpokladu o rozdělení chyb a na vlivné body málo citlivé. Pro odhad parametrů \mathbf{b} se užívá {xe "Metoda:opakovaných vážených NČ"} *iterační metody vážených nejmenších čtverců*. d) Také proměnné x mohou být zatíženy náhodnými chybami, což vede na užití *metody rozšířených nejmenších čtverců*. e) Pro špatně podmíněné matice $\mathbf{X}^T\mathbf{X}$ se používá *metoda racionálních hodnotí*, vedoucí k systému vychýlených odhadů, kde vychýlení je řízeno jedním parametrem.

LITERATURA

- (1) Milan Meloun a Jiří Militký: *Statistické zpracování experimentálních dat*, Plus Praha 1994, resp. East Publishing Praha 1998.
- (2) Milan Meloun a Jiří Militký: *Statistické zpracování experimentálních dat - Sbírka úloh*, Univerzita Pardubice 1996.
- (3) ADSTAT 1.25, 2.0 a verze 3.0, TriloByte Statistical Software Pardubice, 1999.

Název souboru: Milan Meloun
Adresář: E:\Konference\Konfer-prednasky\2000\CSspolecCB
Šablona: D:\Program Files\Microsoft Office\Sablony\Normal.dot
Název: VÝSTAVBA REGRESNÍHO MODELU V ANALYTICKÉ

LABORATOŘI

Předmět:
Autor: Milan Meloun
Klíčová slova:
Komentáře:
Datum vytvoření: 12.04.00 14:55
Číslo revize: 3
Poslední uložení: 13.04.00 13:06
Uložil: Milan Meloun
Celková doba úprav: 3 min.
Poslední tisk: 14.09.00 11:00
Jako poslední úplný tisk
Počet stránek: 2
Počet slov: 688 (přibližně)
Počet znaků: 3 925 (přibližně)