

APLIKACE STATISTICKÝCH MODELŮ NA ODHADOVÁNÍ V MALÝCH OBLASTECH

DOC. ING. TOMÁŠ HOBZA, PH.D.

Popis tématu

Odhadování v malých oblastech se zabývá problémem poskytnutí spolehlivých odhadů nějaké charakteristiky v oblastech, kde dostupné informace o této charakteristice jsou samy o sobě nedostatečné pro poskytnutí přesného odhadu.

Představme si například výběrové statistické šetření pro zjištění průměrného příjmu v České republice. Takováto výběrová šetření jsou většinou navržena pro získání dostatečně přesných odhadů na určité geografické úrovni, např. celostátní nebo krajské. Pokud je ale třeba zúžit poznatky z šetření na nějaké podoblasti, např. okresy, zřídkakdy máme v dané podoblasti k dispozici dostatečný vzorek dat, aby bylo možné získat odhady s požadovanou přesností. Takovým podoblastem pak říkáme malé oblasti. Mezi důvody, proč nemáme k dispozici dostatečný výběr na nižších úrovních, může patřit například vysoká ekonomická a časová náročnost nebo to, že požadavek po odhadech v těchto podoblastech přišel až po provedení šetření.

Pro získání odhadů v těchto malých oblastech se používají statistické modely, které si takzvaně "půjčují sílu" ze sousedních nebo jinak souvisejících oblastí. Tyto modely tedy využívají data ze sousedních oblastí, data sbíraná v jiných časových úsecích stejně jako dodatečné informace dostupné např. z posledního sčítání lidu (jako třeba procento obyvatel s VŠ vzděláním v jednotlivých oblastech) či aktuálních administrativních záznamů.

V současné době se zmíněné modely hojně využívají např. pro rozdělování fondů (EU, Světová banka) do chudších oblastí, pro modelování výskytu nemocí apod.

Příklad matematického modelu

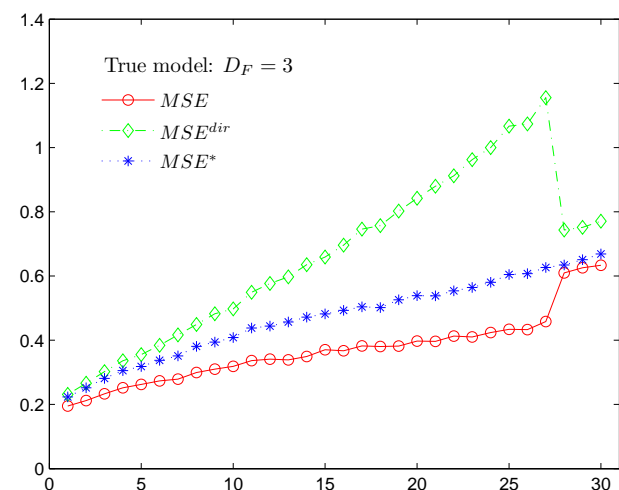
Smíšený lineární model s náhodnými regresními koeficienty:

$$y_{dj} = \sum_{k=1}^p \beta_k x_{kdj} + \sum_{k=1}^p u_{kd} x_{kdj} + u_{0d} + \epsilon_{dj}$$

- y_{dj} ... j -tá hodnota ($j = 1, \dots, n_d$) vysvětlované proměnné v oblasti d ($d = 1, \dots, D$), např. průměrný příjem nebo příznak zaměstnaný-nezaměstnaný atd.
- x_{kdj} ... vektory pomocných proměnných získaných např. z posledního sčítání lidu nebo administrativních záznamů, např. věk, pohlaví, dosažené vzdělání atd.
- β_k ... neznámé pevné regresní koeficienty společné pro všechny oblasti
- $u_{kd} \stackrel{iid}{\sim} N(0, \sigma_k^2)$... náhodné regresní koeficienty pro jednotlivé oblasti
- $u_{0d} \stackrel{iid}{\sim} N(0, \sigma_0^2)$... náhodné efekty pro jednotlivé oblasti
- $\epsilon_{dj} \stackrel{iid}{\sim} N(0, w_{dj}^{-1} \sigma_\epsilon^2)$... náhodné fluktuace

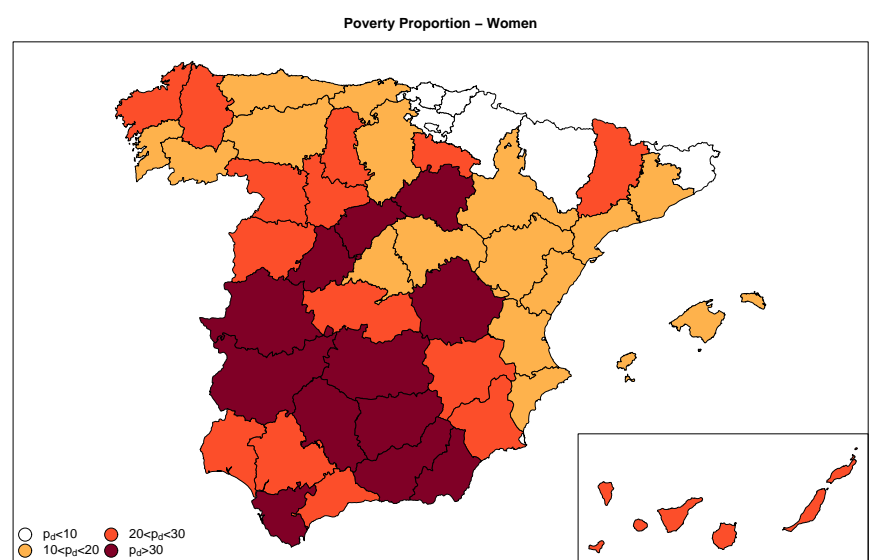
Testování modelu pomocí simulací

Porovnání středních kvadratických chyb pro různé modely



Aplikace modelu na reálná data

Grafické znázornění odhadnutých podílů žen pod oficiální hranici chudoby v jednotlivých provinciích Španělska



Čím byste se mohli zabývat

- Návrh vhodného statistického modelu
- Odvození teoretických formulí a algoritmů pro odhadování parametrů malých oblastí pomocí navrženého modelu
- Implementace a provedení simulační studie pro ověření vlastností zkoumaných metod
- Aplikace metod na reálná data a jejich porovnání s již používanými postupy