

METODY LOKÁLNÍHO ODHADU DEMOGRAFICKÝCH UKAZATELŮ

LOCAL ESTIMATORS OF THE DEMOGRAPHIC CHARACTERISTICS

Helena Nešetřilová

Adresa autorky

Katedra statistiky PEF, Česká zemědělská univerzita v Praze, 165 21 Praha 6 - Suchbát, tel. +4202/24382253, e-mail: nesetrilova@pef.czu.cz

Úvod

V uplynulých padesáti letech se censy v Československu, resp. v České republice konaly zhruba v desetiletých intervalech¹. To je poměrně dlouhá doba, během které podrobnější informace zastarávají a používat je k ekonomickým, demografickým nebo sociologickým analýzám není již po několika letech vhodné. V mezidobí mezi dvěma censy jsou ČSÚ a orgány státní zprávy zjišťovány pouze některé informace a často jsou zveřejňovány v globálnější podobě, než je v některých geograficky omezených analýzách zapotřebí. (Může se jednat např. o vztah mezi charakteristikami v regionech a mikroregionech.) Existují však některé statistické metody, které umožňují aktualizaci informací i na lokální úrovni. V tomto příspěvku bude o několika takových metodách stručně pojednáno.

Charakteristika metod

Jedna skupina těchto statistických metod je založena na poměrových nebo regresních odhadech.

Poměrový odhad lze použít, chceme-li pomocí výběrového šetření aktualizovat hodnotu populačního průměru nebo úhrnu znaku Y v běžném roce t v určitém regionu. Označme $\bar{Y}_t$ populační průměr, resp. Y_t populační úhrn tohoto znaku v příslušném regionu v roce t . Nechť $\bar{Y}_0$, resp. Y_0 jsou populační průměr, resp. populační úhrn tohoto znaku v uvažovaném regionu v roce s censem. Označme dále $\bar{y}_t$, resp. $\bar{y}_0$ výběrové průměry v běžném roce t , resp. v roce s censem (pro odhad úhrnu předpokládejme konstantní velikost populace). Pak můžeme odhadnout poměr

¹ Termíny posledních censů v Československu, resp. České republice: 1.3.1950, 1.3.1961, 1.12.1970, 1.11.1980, 3.3.1991, 1.3.2001.

$R = \frac{\bar{Y}_t}{\bar{Y}_0} = \frac{Y_t}{Y_0}$ pomocí poměru výběrových průměrů, tj. $\bar{R} = \frac{\bar{y}_t}{\bar{y}_0}$. Odtud dostáváme odhady

$$\bar{Y} = \bar{R} \bar{Y}_0, \quad (1)$$

resp. $Y = R Y$. (2)

Do skupiny regresních metod patří mj. použití vícenásobných regresních modelů s tzv. symptomatickými proměnnými, které v těchto modelech vystupují jako nezávisle proměnné. Předpokládejme, že chceme odhadnout úhrn znaku Y_{it} v běžném roce t v oblastech $i = 1, \dots, m$ (např. počet ekonomicky aktivního obyvatelstva v mikroregionech). Metodu, založenou na použití p symptomatických proměnných uvádí např. [2]. Označme pořadí $k = 0, 1, t$ roky posledních dvou censů, resp. běžný rok. Označme dále X_{ijk} hodnotu j -té symptomatické proměnné $j = 1, \dots, p$ v oblasti i v roce k a Y_{ik} úhrn znaku Y v roce k . Pro

označení poměrů budou použity následující symboly $y_{ik} = \frac{Y_{ik}}{\sum_i Y_{ik}}$, $x_{ijk} = \frac{X_{ijk}}{\sum_i X_{ijk}}$ a na

mikroregionální úrovni $R'_i = \frac{y_{i1}}{y_{i0}}$, $R_i = \frac{y_{it}}{y_{i1}}$, $r'_{ij} = \frac{x_{ij1}}{x_{ij0}}$, $r_{ij} = \frac{x_{ijt}}{x_{ij1}}$. Známe-li hodnoty

poměrů $R'_i, r'_{i1}, \dots, r'_{ip}$, $i = 1, \dots, m$, odhadneme nejprve regresní koeficienty β

vícenásobného regresního modelu, který charakterizuje vztah mezi poměry úhrnu znaku Y ve dvou letech s censem a odpovídajícími změnami poměrů symptomatických proměnných r'_{ij}

$$R'_i = \beta'_0 + \beta'_1 r'_{i1} + \dots + \beta'_p r'_{ip}. \quad (3)$$

Pak pro období po posledním censu odhadneme R_i pomocí vztahu (4), ve kterém jsou použity odhady regresních koeficientů β'_j a známé změny poměrů symptomatických proměnných r_{ij}

$$R_i = \beta'_0 + \beta'_1 r_{i1} + \dots + \beta'_p r_{ip}. \quad (4)$$

Je-li možné získat informaci o součtu $\sum_i Y_{it}$ v běžném roce, pak lze odhadnout

$$\bar{Y}_{it} = \bar{R}_i y_{i1} \sum_i Y_{it}. \quad (5)$$

Jinou možností jak získat odhady úhrnů Y_i v podoblastech $i, i = 1, \dots, m$, je použití tzv. syntetických odhadů. Tato metoda je založena na předpokladu, že je k dispozici nestranný odhad Y pro velkou oblast, který byl získán na základě výběrového šetření. Cílem je provést odhady pro jednotlivé podoblasti této oblasti.

Předpokládejme tedy ve shodě s [2], že populace je rozdělena na disjunktní oblasti a že pro každou oblast k je k dispozici přímý odhad $\bar{Y}_{.k}$ úhrnu $Y_{.k}$. Předpokládejme dále, že oblasti jsou rozděleny na podoblasti (pro podoblasti bude použit index i). Označme Y_{ik} úhrn pro i -tou podoblast v oblasti k . Necht' platí $Y_{.k} = \sum_i Y_{ik}$ a necht' jsou známy rozsahy podoblastí N_{ik} , $N_{.k} = \sum_i N_{ik}$. Syntetický odhad úhrnu Y_i je dán vztahem (6)

$$\bar{Y}_i^{\text{syn}} = \sum_k \frac{N_{ik}}{N_{.k}} \bar{Y}_{.k} \quad . \quad (6)$$

Přímý odhad $\bar{Y}_{.k}$, který byl v (6) použit pro konstrukci syntetického odhadu bývá často poměrový odhad

$$\bar{Y}_{.k} = \frac{\sum_{l \in S_k} w_l y_l}{\sum_{l \in S_k} w_l n_l} N_{.k} = \frac{\bar{Y}_{.k}}{\bar{N}_{.k}} N_{.k} \quad . \quad (7)$$

Dosadíme-li ze (7) do (6), dostáváme syntetický odhad v následujícím tvaru

$$\bar{Y}_i^{\text{syn}} = \sum_k \frac{\bar{Y}_{.k}}{\bar{N}_{.k}} N_{ik} \quad . \quad (8)$$

Poznamenejme závěrem, že syntetické odhady jsou konzistentní, ale nejsou nestranné. Pro jejich vychýlení B přibližně platí

$$B = \sum_k \left(\frac{Y_{.k}}{N_{.k}} - \frac{Y_{ik}}{N_{ik}} \right) N_{ik} \quad . \quad (9)$$

Z (9) je patrné, že vychýlení odhadu vymizí v tom případě, že průměr podoblastí v každé oblasti bude roven celkovému průměru oblasti. Tato podmínka však v mnoha případech splněna evidentně nebude, je však možné, že místní rozdíly v průměrné úrovni Y budou jenom malé.

Závěr

V tomto příspěvku byly uvedeny tři metody pro odhad demografických i jiných charakteristik v lokálním měřítku. Poznamenejme, že odhady získané pomocí těchto metod budou obecně vychýlené, zřejmě nejproblematictější může být v některých případech použití syntetických odhadů. Hrozí-li takové nebezpečí, je možné pokusit se vychýlení odstranit složitější konstrukcí odhadové statistiky, jak je tomu např. u kompozitních odhadů.

Literatura

- [1] Cochran, W.G.: Sampling Techniques, Wiley, New York 1977.
- [2] Govindarajulu, Zakkula: Elements of Sampling Theory and Methods. Prentice-Hall, Upper Saddle River, 1999.
- [3] Levy, Paul S., Lemeshow, Stanley: Sampling of Populations. Wiley, New York 1999.