

# Aplikace shlukování časových řad na analýzu trendů úhrnné plodnosti v ČR

Ing. Karolína Bakuncová, Ing. Tereza Frömmelová

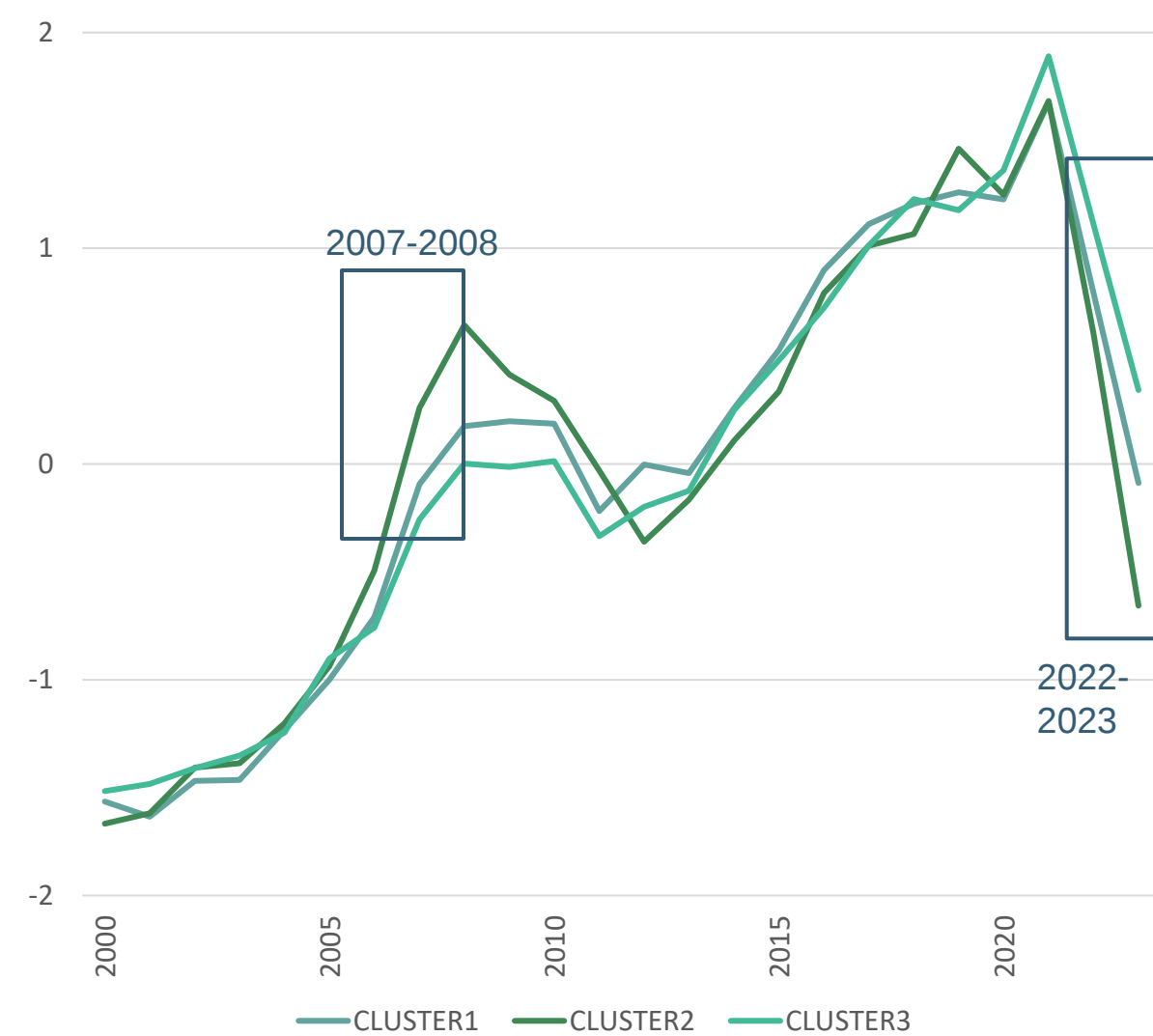
Vysoká škola ekonomická v Praze, Fakulta informatiky a statistiky

Příspěvek se zabývá analýzou regionálních vzorců úhrnné plodnosti v krajích České republiky pomocí shlukování časových řad prostřednictvím algoritmu K-Shape a jeho příslušející SBD vzdálenosti. Cílem výzkumu je identifikovat podobnosti a rozdíly v časovém vývoji plodnosti mezi regiony. První výsledky ukazují, že metoda dokázala oddělit region na severozápadě České republiky a vytvořit samostatný cluster z východních krajů republiky. Poster představuje tyto úvodní výsledky, přičemž další analýza se zaměří na podrobnější charakterizaci clusterů a identifikaci faktorů, které ovlivňují jejich odlišnosti.

## Metodologie shlukování

Klasifikovaná data představují soubor 13 časových řad zaznamenávající hodnoty úhrnné plodnosti pro jednotlivé kraje České (celkem 24 hodnot od roku 2000 do roku 2023). Pro samotné shlukování bylo vynecháno Hlavní město Praha.

Z hlediska shlukování, algoritmy používané u časových řad často předpokládají, že shlukování je prováděno na bázi porovnání shody tvaru vývoje časové řady nežli konkrétní úrovně hodnot. Vybraný algoritmus shlukování časových řad K-shape pak vychází ze škálovatelného iterativního postupu, jehož výsledkem jsou vnitřně homogenní dobře oddělené shluky. Základ pro K-shape algoritmus tvoří vzdálenost SBD, což je v podstatě normalizovaná verze míry vyjadřující vzájemnou korelaci dvou časových řad, používanou pro zobrazení a porovnání tvaru dvou časových řad a výpočet skupinových centroidů, umožňující další třídění vstupních časových řad do výsledných shluků (Paparrizos & Gravano, 2016).



Graf 2: Porovnání skupinových centroidů dle K-shape s vyznačením období největších rozdílů

## Rozdíl centroidů clusterů

Vzhledem k tomu, že i v původních časových řadách úhrnné plodnosti odpovídaly vrcholy a lokální minima napříč clusterovanými kraji stejným okamžikům v čase, bylo možné vrátit se při interpretaci výsledků analýzy opět na časovou osu.

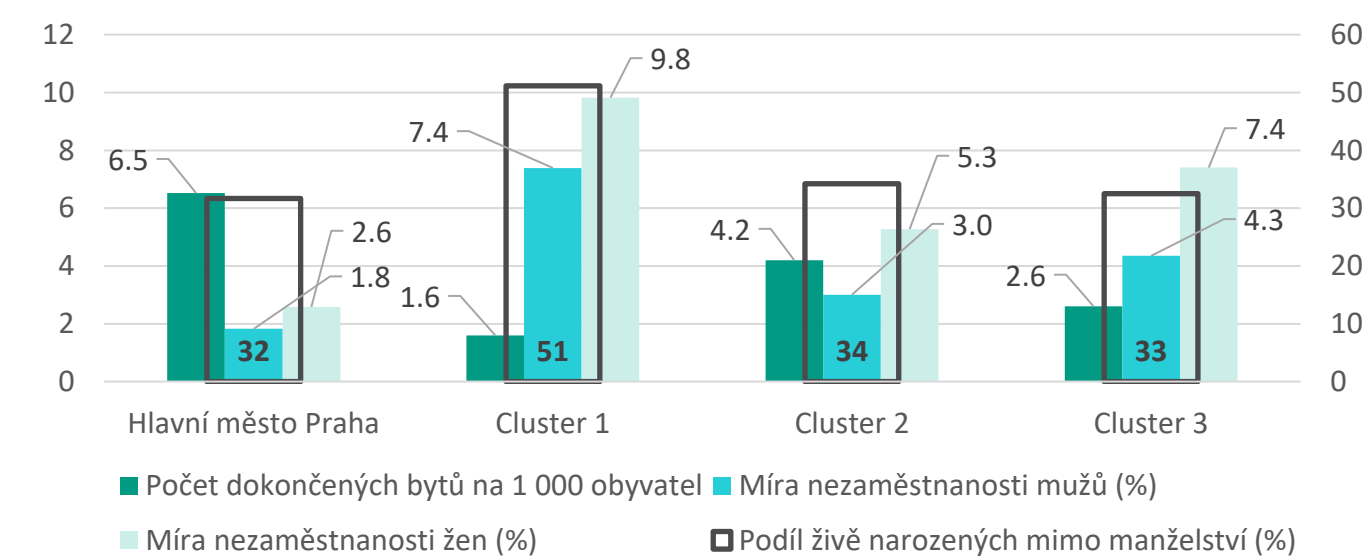
Na základě vzdáleností centroidů jednotlivých clusterů bylo identifikováno, že okamžik, kdy pozorujeme největší rozdíly v úhrnné plodnosti odpovídá období let 2007 až 2008 a následně letům 2022 až 2023.

## Struktura rozdílů

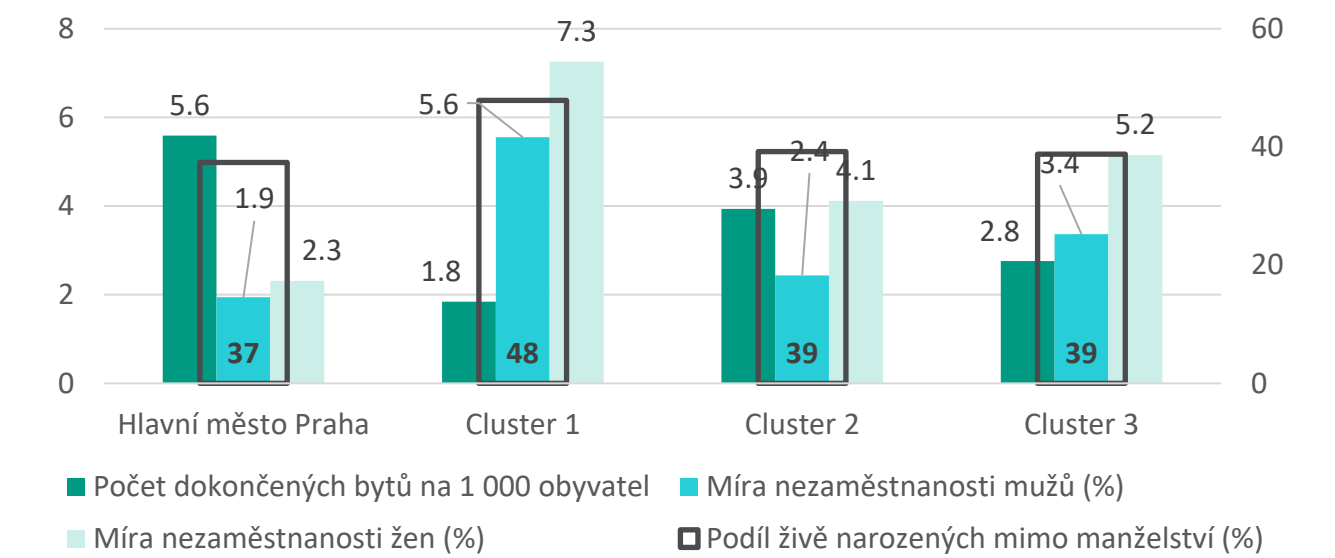
Za účelem charakteristiky struktury odlišností mezi clustery byly zvoleny proměnné, které mají na základě rešerše literatury potenciál ovlivňovat reprodukční chování populace v jednotlivých krajích.:

- Počet dokončených bytů na obyvatele (Hašková, 2012)
- Míra nezaměstnanosti mužů a žen (Paloncyová et al., 2023)
- Podíl dětí narozených mimo manželství (Rychtaříková, 1996, Cherlin, 2021)
- Průměrná a mediánová mzda mužů a žen (pouze pro období 2022-2023) (Hašková, 2012)

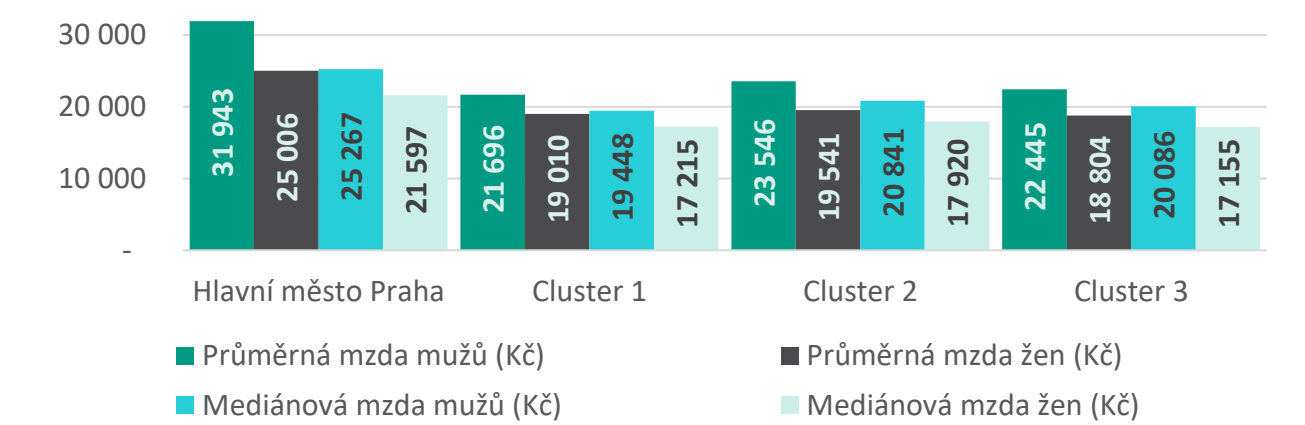
Pro kompletnost dat je v přehledu hodnot v dílčích proměnných v grafech 3, 4 a 5 doplněna i Praha.



Graf 3: Průměrné hodnoty dílčích proměnných v clusterch a hlavním městě v období 2007-2008



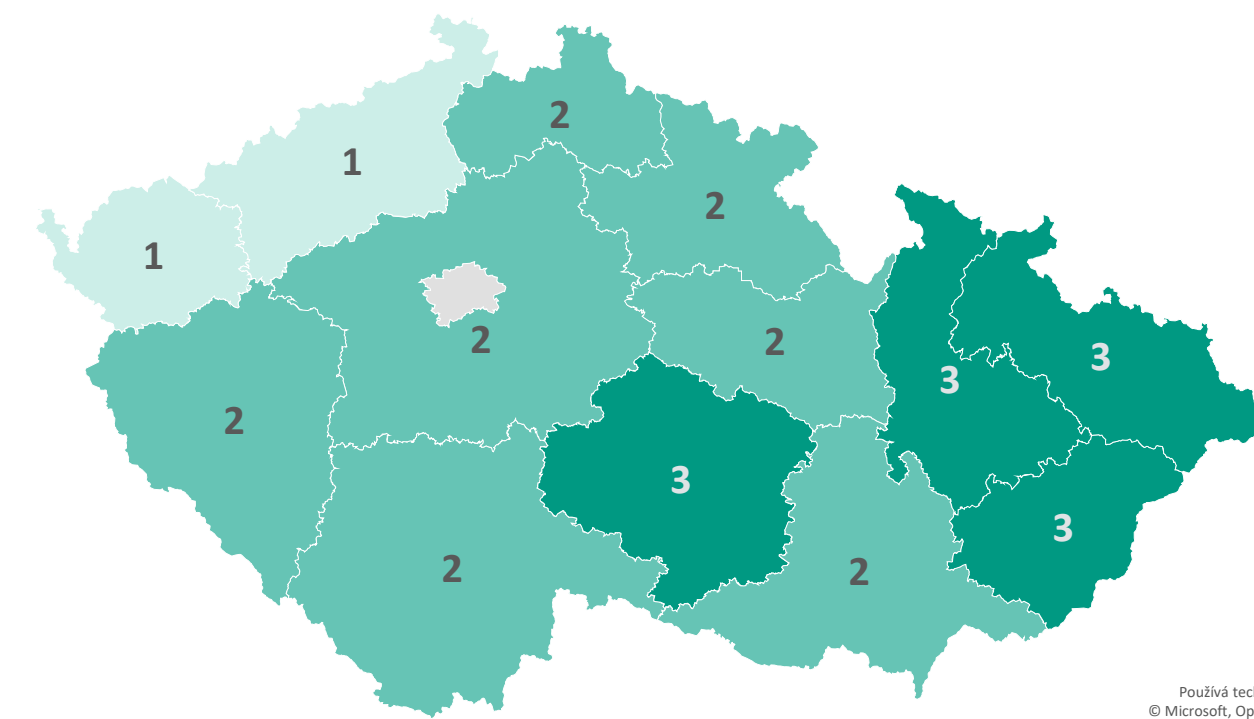
Graf 4: Průměrné hodnoty dílčích proměnných v clusterch a hlavním městě v období 2022-2023



Graf 5: Průměrná a mediánová mzda v clusterch a hlavním městě v období 2022-2023

## Literatura

- Cherlin, A. J. (2021). Rising nonmarital first childbearing among college-educated women: Evidence from three national studies. *Proceedings of the National Academy of Sciences*, 118(37), e2109016118. <https://doi.org/10.1073/pnas.2109016118>
- Hašková, H., Dudová, R., & Pospíšilová, K. (2019). Kde se berou jedináčci? Faktory související s jednodětností v ČR. *Demografie*, 61(2), 93–110.
- Paloncyová, J., Höhne, S., Kreckova Tumova, N., Svobodova, K., & Vidovičová, L. (2023). Zpráva o rodině 2023. Praha: RILSA. Dostupné z: [https://katalog.vupsv.cz/fulltext/ui\\_2580.pdf](https://katalog.vupsv.cz/fulltext/ui_2580.pdf)
- Paparrizos, J., & Gravano, L. (2016). K-shape. *ACM SIGMOD Record*, 45(1), 69–76. <https://doi.org/10.1145/2949741.2949758>
- Rychtaříková, J. (1996). Současné změny charakteru reprodukce v České republice a mezinárodní situace. *Demografie*, 38(2), 77–89.



Graf 1: Clustery krajů ČR (při vynechání hlavního města), označení Cluster 1, Cluster 2, Cluster 3