

Vliv socioekonomických faktorů na podporu Donalda Trumpa v Georgii

Úvod

Volební preference jsou ovlivňovány celou řadou socioekonomických a demografických faktorů. V amerických prezidentských volbách v roce 2020 byla zřetelná polarizace mezi venkovskými a městskými oblastmi, která se odrazila v podpoře kandidátů. Cílem tohoto projektu je zjistit, které faktory ovlivnily podporu Donalda Trumpa na úrovni okresů ve státě Georgia. Inspirací jsou dosavadní studie ukazující, že socioekonomické podmínky (např. vlastnictví domů, nezaměstnanost, chudoba) mohou predikovat volební preference.

Data a proměnné

Dataset obsahuje socioekonomické ukazatele pro jednotlivé okresy státu Georgia. Použité proměnné jsou:

- **Závislá proměnná:**
 - Podíl hlasů pro Donalda Trumpa v prezidentských volbách 2020 (Trump_percent).
- **Nezávislé proměnné:**
 - Podíl obyvatel ve vlastním domě (homeownership),
 - Podíl bytových domů (multi_unit),
 - Míra chudoby (poverty),
 - Populace okresu,
 - Míra nezaměstnanosti (unemployment_rate),
 - Změna v populaci,
 - Příjem na hlavu (per capita income),
 - Medián příjmu domácnosti (median household income).

Data byla čištěna odstraněním záznamů s chybějícími hodnotami.

Metodika

Nejprve byly vypočteny Pearsonovy korelace mezi podílem hlasů pro Donalda Trumpa a jednotlivými socioekonomickými ukazateli, aby byl zmapován základní vztah mezi proměnnými. Pro ověření vlivu více proměnných současně byl následně sestaven vícenásobný lineární regresní model s pěti prediktory: podíl obyvatel ve vlastním domě, podíl bytových domů, míra chudoby, velikost populace a míra nezaměstnanosti. Analýza byla provedena v programu R, model byl dále podroben diagnostice a rozšířen o robustní odhady pro ověření stability výsledků.

Korelace mezi proměnnými

Pro základní orientaci byly vypočteny Pearsonovy korelace mezi podílem hlasů pro Donalda Trumpa a jednotlivými socioekonomickými proměnnými. Výsledky shrnuje následující tabulka:

Proměnná	Korelace s podílem hlasů pro Trumpa
Podíl obyvatel ve vlastním domě (homeownership)	+0.446
Podíl bytových domů (multi_unit)	−0.588
Míra chudoby	−0.265
Míra nezaměstnanosti	−0.298
Populace (2017)	−0.436
Populace (2010)	−0.446
Populace (2000)	−0.465
Změna v populaci	+0.146
Příjem na hlavu	−0.069
Medián příjmu domácnosti	+0.034

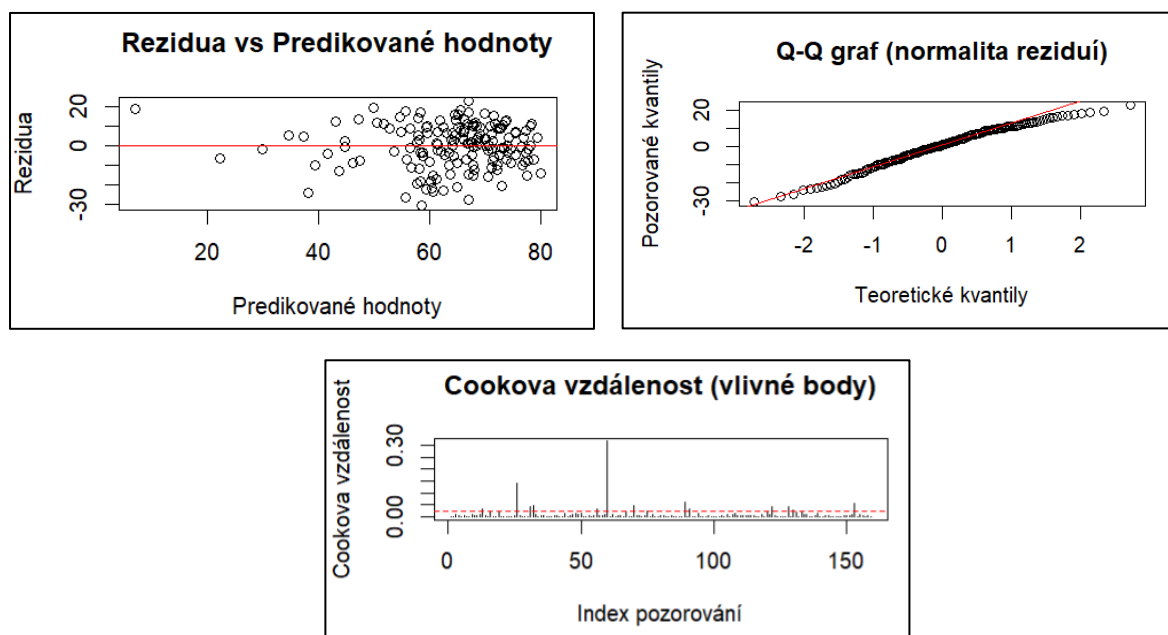
Tabulka č. 1: Korelace proměnných

Nejsilnější negativní vztah byl zjištěn u proměnné **podíl bytových domů (multi_unit)** (-0.59), což ukazuje, že okresy s vyšší koncentrací bytových domů měly nižší podporu Donalda Trumpa. Podobně negativní vztah vykazovaly i velikost populace a míra nezaměstnanosti. Naopak podíl obyvatel žijících ve vlastním domě měl pozitivní korelaci s podporou republikánského kandidáta.

Diagnostika modelu

Diagnostické testy a grafy ukázaly, že model obecně splňuje předpoklady lineární regrese. Byly zjištěny pouze mírné odchylky od normality reziduí, které nejsou kritické vzhledem k robustnosti lineární regrese vůči tomuto předpokladu. Test Breusch-Pagan prokázal přítomnost heteroskedasticity ($p = 0.024$), což znamená, že rezidua nevykazují konstantní rozptyl. Vlivné body byly identifikovány (okresy č. 26 a 60), ale jejich vliv není natolik silný, aby výrazně ovlivnil model. Test multikolinearity pomocí VIF ukázal, že žádná z proměnných nepřekračuje kritickou hranici ($VIF < 5$).

Z důvodu heteroskedasticity byly následně spočteny robustní standardní chyby, aby byly výsledky modelu spolehlivější.



Graf č. 1: Rezidua vs. predikované hodnoty

Graf č. 2: Q-Q graf pro kontrolu normality reziduí

Graf č. 3: Cookova vzdálenost (vlivné body)

Robustní diagnostika

Aby byla zajištěna správnost výsledků vzhledem k mírné heteroskedasticitě zjištěné diagnostikou modelu, byly spočteny robustní standardní chyby (HC1). Výsledky ukazují, že většina klíčových prediktorů zůstává statisticky významná i po této úpravě.

Model potvrdil, že podpora Donalda Trumpa výrazně klesala v okresech:

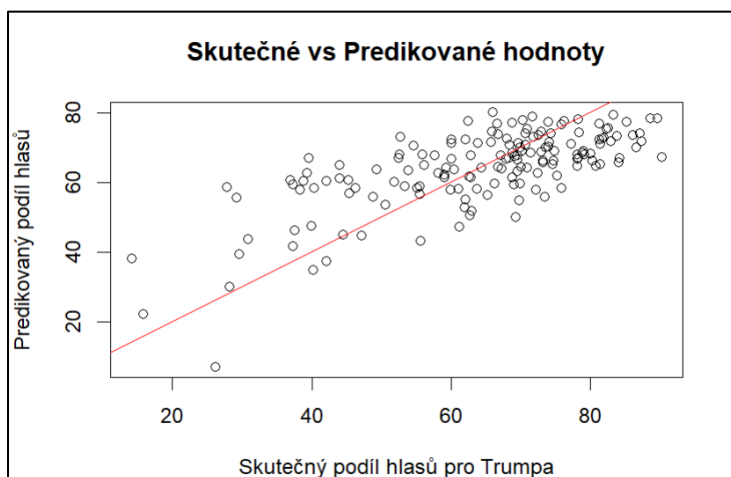
- s vyšším podílem **bytových domů** (multi_unit, $\beta = -1.02$, $p < 0.001$),
- s vyšší mírou **chudoby** (poverty, $\beta = -0.56$, $p = 0.009$),
- s větší **populací** (pop2017, $\beta = -0.0000319$, $p < 0.001$),
- a s vyšší **nezaměstnaností** (unemployment_rate, $\beta = -3.70$, $p < 0.001$).

Naopak podíl obyvatel žijících ve vlastním domě (**homeownership**) nevykázal statisticky významný vliv ($\beta = -0.29$, $p = 0.25$).

Výsledky tímto potvrzují robustnost modelu a ukazují, že urbanizační a socioekonomické faktory hrály zásadní roli v podpoře republikánského kandidáta v Georgii.

Proměnná	β koeficient	p-hodnota
Podíl obyvatel ve vlastním domě (homeownership)	-0.29	0.25
Podíl bytových domů (multi_unit)	-1.02	<0.001
Míra chudoby	-0.56	0.009
Populace (2017)	-0.0000319	<0.001
Míra nezaměstnanosti	-3.70	<0.001

Tabulka č. 2: Výsledky robustní lineární regrese (HC1 standardní chyby)



Graf č. 4: Skutečné vs. predikované hodnoty modelu

Graf ukazuje vztah mezi skutečným podílem hlasů pro Donalda Trumpa v jednotlivých okresech a hodnotami predikovanými lineárním regresním modelem. Čím blíže jsou body diagonální čáře, tím lépe model odpovídá skutečným výsledkům.

Robustní regrese (M-estimace)

Pro ověření robustnosti modelu byla provedena také robustní regrese metodou M-estimace (rlm). Výsledky potvrdily stabilitu hlavních efektů: podpora Donalda Trumpa klesala v okresech s vyšším podílem bytových domů ($\beta = -0.93$), vyšší mírou chudoby ($\beta = -0.60$), větší populací ($\beta \approx -0.00003$) a vyšší nezaměstnaností ($\beta = -3.36$). Podíl obyvatel ve vlastním domě (homeownership) se opět neprokázal jako významný faktor. To potvrzuje, že vztahy nalezené v modelu jsou robustní vůči odlehlým hodnotám a vlivným bodům.

Porovnání vysvětlené variance ukázalo, že klasická lineární regrese dosáhla $R^2 = 0.501$, zatímco robustní regrese vykázala $R^2 = 0.499$. Oba modely tedy vysvětlují přibližně 50 % rozptylu v podpoře Donalda Trumpa mezi okresy státu Georgia. Velmi podobné hodnoty R^2 potvrzují, že robustní regrese dosahuje srovnatelné míry vysvětlení rozptylu a zároveň poskytuje vyšší odolnost vůči případným odlehlým bodům a vlivným pozorováním.

Závěr

Výsledky ukazují, že podpora Donalda Trumpa byla v Georgii ovlivněna především mírou urbanizace a socioekonomickými faktory. Okresy s vyšším podílem bytových domů, vyšší mírou chudoby, větší populací a vyšší nezaměstnaností vykazovaly nižší podporu republikánského kandidáta. Vliv vlastnictví bydlení se neprokázal jako statisticky významný. Robustní analýzy potvrdily stabilitu těchto výsledků. Hlavním omezením této zprávy je geografické zaměření pouze na stát Georgia a použití agregátních dat na úrovni okresů. Budoucí výzkum by mohl rozšířit analýzu na celonárodní úroveň nebo zohlednit další kulturní a politické faktory.

Data a další informace o této zprávě jsou dostupné na adrese <https://dostal.vyzkum-psychologie.cz/stat4?i=604>.

Literatura:

- Použité balíčky v R: haven, car, lmtest, sandwich, MASS