

Regresní analýza - komentář

Před regresí

- Teorie -> výběr proměnných
 - Identifikace důležitých faktorů
 - teorie konfliktních linií, kontextuální faktory
 - Vyplývá z literatury nebo nějaké argumentované úvahy
 - Strana=kandidát / strana \neq kandidát
 - Mechanismus efektu?
 - Charakter efektu – kompozitní x kontextuální

- proměnné: věk, susedský efekt, vzdělání, nezaměstnanost, velikost obce a volební účast, kdy každá z těchto proměnných má svůj důvod.
- budou využita demografická data jako například věk voličských skupin, nezaměstnanost v obcích, či nejvyšší dosažené vzdělání obyvatel obcí
- voličmi SMERU-SD sú primárne občania v dôchodcovskom veku.
- Kdybych měl predikovat, tipoval bych, že kandidátka KDU-ČSL bude jasně nejvíce volena katolíky

Problematické proměnné

- Volební účast
- Kombinace proměnných měřících tutéž věc:
 - Podíl zš, podíl sš, podíl vš
 - Mladí, střední, starší
 - ...
- Velikost obce
 - Volba nějakých funkčních kategorií
 - 1) Velkoměsto (Obce nad 1000 obyvatel)
 - 2) Maloměsto (Obce od 251-1000 obyvatel)

Formulace hypotéz

- Obce s vysokým podílem osob nad 65 let budou volit KSČM
- ANO přilákalo spíše starší voliče
- Kandidáta ODS budou preferovat spíše podnikatelé/OSVČ než zaměstnanci
- Sousedský efekt pro KDU-ČSL bude silnější v druhém kole než v prvním.

Po výběru/vytvoření proměnných

- operacionalizace
- Seznámení se s vlastnostmi proměnných
 - Variance
 - Chybějící hodnoty
 - korelace

Regrese

- Využití váhy
- Nepřehlčení proměnnými
- Kontrola multikolinearity

multikolinearita

	KOEFICIENT		KOEFICIENT	KOLINEARITA
	B	Std. Error	Beta	VIF
Konstanta	-5,309	12,403		
volební účast_p	0,072	0,113	0,059	2,210
nezaměstnanost_říjen16p	1,842	0,372	0,443	2,067
mládež_p	0,400	0,345	0,137	3,596
dospělý_p	0,461	0,215	0,217	2,646
důchodce_p	0,125	0,277	0,049	2,992
základní_p	-0,158	0,313	-0,104	11,020
Soused_CK	1,020	3,088	0,021	1,031
střední_p	-0,112	0,251	-0,070	6,422
vysoké_p	-0,746	0,316	-0,403	7,577

Prezentace výsledků

- Postačuje R^2 , B a Beta (+ konstanta)
- Signifikance není nutná, nikam se nezobecňuje
- Pro prezentaci výsledku je vhodné přejmenovat proměnné
- Desetinná místa
- Přeložení do češtiny

Coefficients^a

Model		Unstandardized Coefficients		Standardized	Collinearity Statistics	
		B	Std. Error	Beta	Tolerance	VIF
1	(Constant)	-18,268	30,552			
	bez_zs	,037	,435	,019	,198	5,038
	stredoskolske	,037	,323	,155	,234	4,265
	vos_vs	-,037	,420	-,230	,113	8,843
	duchodci	,037	,317	,267	,351	2,847
	mladi	,037	,396	,198	,480	2,084
	ves	,058	1,240	,005	,838	1,194

Interpretace

- Vyjádření se k teoretickým předpokladům
- Založena na hodnotách B a Beta
- Srovnání mezi kandidáty

R2

- model nemá vysvětlovací sílu
- 44,1 % případů tento model vysvětluje
- Index determinace vysvětluje „jen“ 44 %.
- 40 % variability je vysvětleno proměnnými

Nestandardizované koeficienty

- Dummy
 - Obce v sousedství Králík volili o 15 % více tohoto kandidáta než ostatní obce
 - Nejvíce pozitivní vazbu tak můžeme vidět u výsledku kandidátky s obyvateli z vesnic pod 1000 obyvatel, kdy s každým obyvatelem takovéto vesnice vzrostl elektorát Bochníčkové o 0,07 pr. Bodu
 - Kandidáta v okolí domácí obce podpořilo o 3 p.b. více voličů

Efekty kardinálních proměnných

- nedá se tedy jednoznačně potvrdit, že obce s vyšším podílem seniorů výrazně více volí ANO
- je patrné, že u zaměstnanců je volba kandidáta ODS méně pravděpodobná
- S každým procentem podnikatelů v populaci klesala pravděpodobnost volby kandidáta
- Na zisky Staňka působí nejsilněji výskyt lidí v důchodovém věku a mladí lidé
- proměnné podíl nezaměstnaných osob má na kandidáta X pozitivní efekt
- Kladný vztah měly výsledky kandidáta také k zemědělcům
- vyšší podíl vysokoškolsky vzdělaných v obci kandidátovi škodil

Standardizovaný koeficient

- nejsilnější efekt má proměnná sousedství

	Nestandardizovaný koeficient		Standardizovaný koeficient
	B	Std. Error	Beta
Konstanta	-5,309	12,403	
volební účast_p	0,072	0,113	0,059
nezaměstnanost_říjen16p	1,842	0,372	0,443
mládež_p	0,400	0,345	0,137
dospělý_p	0,461	0,215	0,217
důchodce_p	0,125	0,277	0,049
základní_p	-0,158	0,313	-0,104
Soused_CK	1,020	3,088	0,021

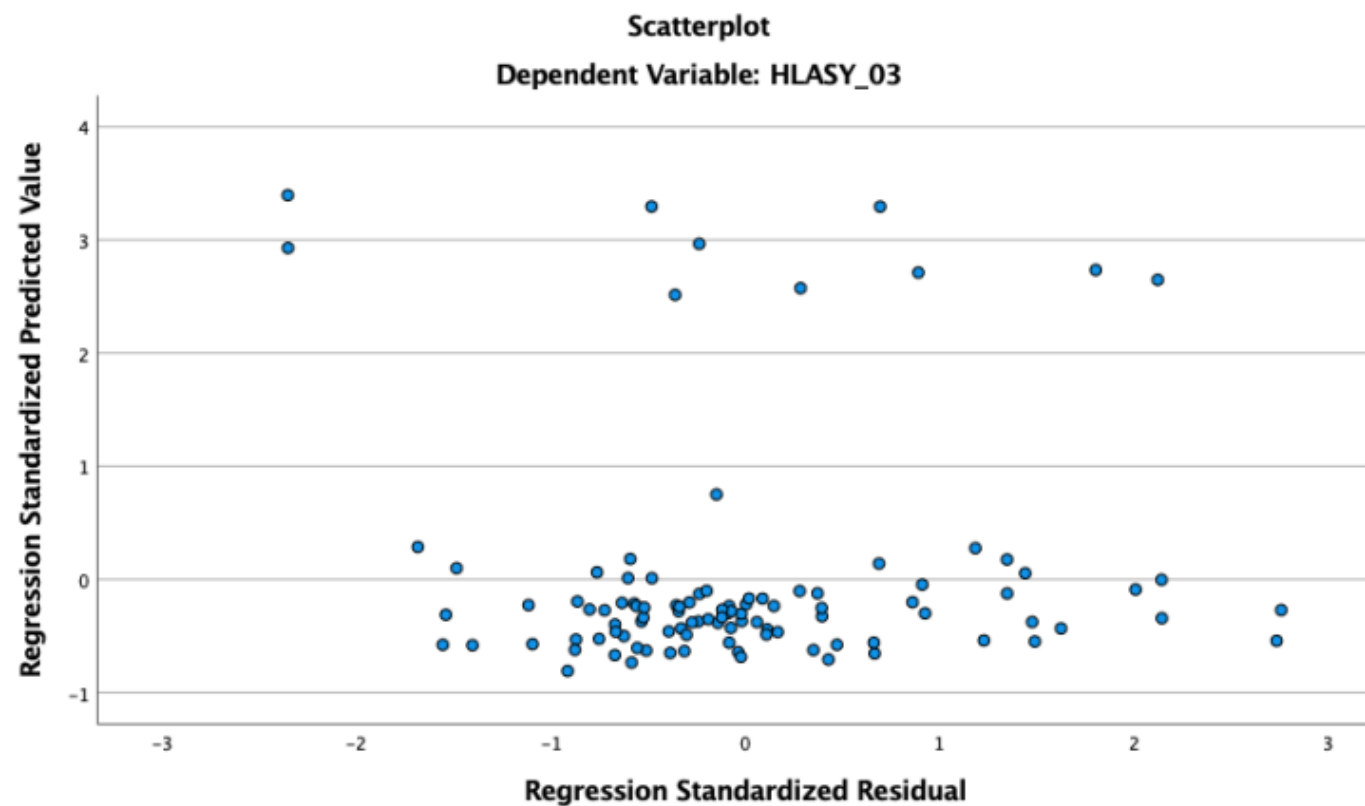
2. kolo

- Do úvah je vhodné zařadit konstelaci volební soutěže
- Např. v duelu KDU-ČSL x ČSSD se z KDU stane pravicová strana
- Ale v duelu KDU X ODS bude struktura podpory KDU spíše levicová
- Obvykle dochází k poklesu efektů z prvního kola
- Zejména u sousedského efektu

- obecně je v tomto spíše venkovském obvodu velmi málo VŠ vzdělaných
- Moc nechápu, jak je možné, že kandidát ČSSD byl více volen v místech s vyšším počtem OSVČ
- vidíme jasný negativní vliv „velkoměsta“ na výsledek KSČM

- pozor na rozdíl procentní bod a procento
- Ze sledovaných proměnných jsou jen dvě statisticky významné

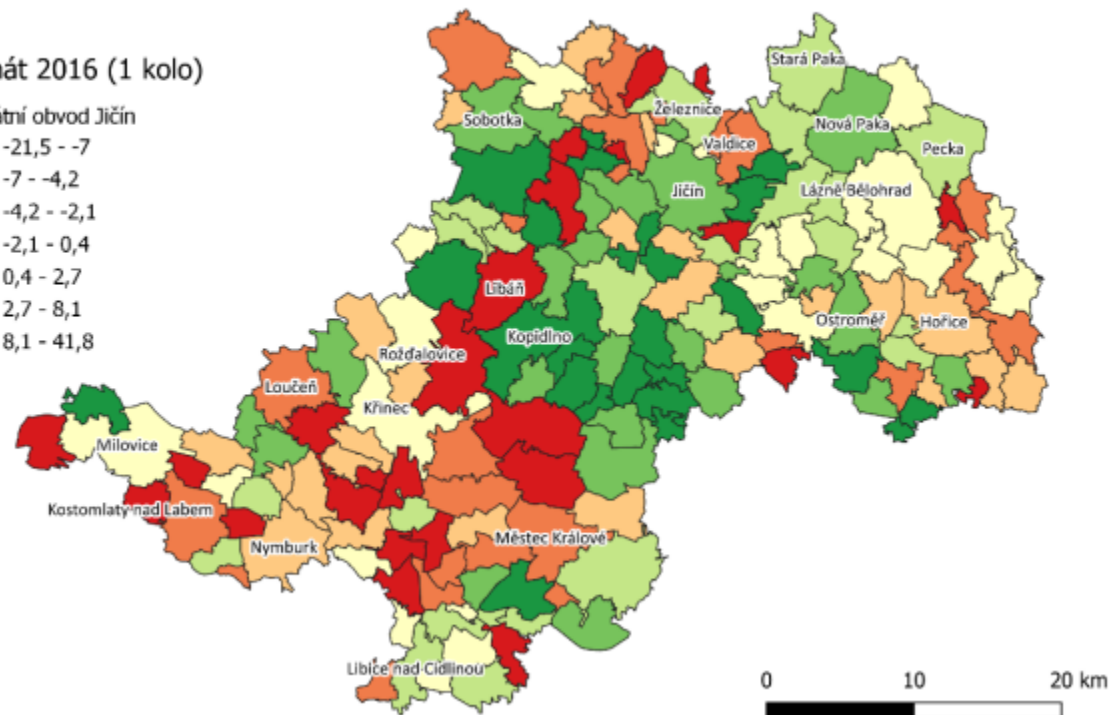
Mapa reziduí



Mapa reziduí - kandidát Josef Táborský (ČSSD)

Senát 2016 (1 kolo)

Senátní obvod Jičín



Data za městské části

- Jen pro členěná statutární města
 - Praha, Brno, Ostrava, Plzeň, Ústí nad Labem, Pardubice, Liberec, Opava
- <https://www.czso.cz/csu/czso/vysledky-scitani-2021-otevrena-data>
- [https://www.czso.cz/csu/czso/otevrena data pro vysledky scitani li du domu a bytu 2011 sl db 2011](https://www.czso.cz/csu/czso/otevrena_data_pro_vysledky_scitani_li du domu a bytu 2011 sl db 2011)
- Otevřená data volby.cz obsahují v položce obec rovnou data za městské části
 - V případě „běžných“ měst potřeba pracovat s volebními okrsky

SLDB 2011

- Seznam území
- Vyfiltrování městských částí
- Kód + sloupec městská část
- Připojení k datům

- Pro „běžná“ města lze vytvořit z dat na úrovni ZSJ
 - Zdá se že již/ještě nejsou veřejně dostupná
 - Možné získat na žádost

SLDB 2021

- Data v csv
- vložení dat do Excelu
 - Data -> Načíst externí data -> Z textu a nastavit správný oddělovač (čárka) a typ souboru (UTF-8)
- Městské části filtr -> cis=44