

Prostorová analýza – Seminář 2
popisné statistiky,
korelace

14. 4. 2015

Úkol č. 1

- Průměr, medián, směrodatná odchylka
- Variační koeficient
- Giniho koeficient

- Vizually přívětivé tabulky !!!
- Česky!!!

Možnosti interpretace

- Interpretace neznamená jen přepis hodnot z tabulky do textu
- Možnosti
 - Srovnání hodnot mezi stranami v rámci voleb
 - Srovnání se stejnou stranou v jiných volbách
 - Srovnání s jinými ukazateli (srovnání průměru a mediánu, srovnání giniho koeficientu a variačního koeficientu)
 - Jít zpět do dat (bude vysvětleno na příkladu)

Příklady dobré praxe

KRAJSKÉ VOLBY 2012						
	TOP 09	ČSSD	ODS	KSČM	KDU-ČSL	Jihočeši
Průměr	5,44	26,82	13,28	18,12	8,75	12
Medián	4,76	26,07	12,76	17,52	6,42	11,51
Směrodatná odchylka	4,05	9,57	7,22	8,81	8,18	7
Variační koeficient	0,74	0,36	0,54	0,49	0,93	0,58
<u>Giniho koeficient</u>	0,57	0,42	0,55	0,43	0,46	0,56

Průměr

- Srovnat s mediánem
- Vhodné srovnat také s reálnou podporou v rámci kraje (např. Jihočeši 14,5 x 12, ODS 12,5 x 13,5) – vliv velikosti obce
- KDUČSL má výrazně vyšší průměr, než medián, což je zapříčiněno odlehlými hodnotami.
- Při porovnání průměru s mediánem tak můžeme zjistit mezi kterými hodnotami se vyskytuje více extrémních hodnot. Z tabulky tak vidíme, že například KSČM v roce 2013 má prakticky stejný průměr jako medián, což znamená, že se pravděpodobně vyskytovalo málo extrémních hodnot. Naopak u ČSSD v roce 2012 vidíme rozdíl až 2,5 procent.

Variační koeficient

- Podpora strany se mezi jednotlivými obcemi příliš nelišila, jelikož variační koeficient, jehož hodnoty ukazují variaci podpory dané strany v jednotlivých obcích, je pouze 0,36 a blíží se více 0 než 1 (ta by ukazovala na velké rozdíly mezi obcemi).
- Volby do Poslanecké sněmovny 2013 vyhrála ČSSD. Její hodnota variačního koeficientu je oproti volbám do PS velmi vysoká (0,96). Podpora v jednotlivých obcích se tedy výrazně liší, což naznačuje, že se může lišit elektorát mezi jednotlivými úrovněmi voleb.

Giniho koeficient

- Nejvíce koncentrovanou podporu má v těchto volbách dle Giniho koeficientu strana TOP 09 (0,57). Tato situace je zapříčiněna tím, že strana získává vysokou podporu spíše ve velkých městech v rámci kraje a v menších obcích dosahuje obvykle nízkých zisků.
- Tato strana měla velmi koncentrovanou volební podporu, tedy byla volena někde hodně a někde vůbec.

Srovnání variační koeficient x Gini

- KDU-ČSL má podle hodnoty variačního koeficientu nejvíce odlišnou podporu mezi obcemi. Giniho koeficient má však v tomto případě spíše nižší hodnotu. Ačkoli jsou tedy výsledky značně různorodé, tak nedochází k významnější koncentraci podpory. To je dáno především tím, že obce s velmi nízkou i velmi vysokou podporou strany jsou spíše menší a nemají tedy na koncentraci volební podpory velký vliv.

Ne úplně dobrá praxe

- Směrodatná odchylka nám ukazuje o kolik procent od průměru se může v jednotlivých obcích lišit konkrétní hodnota od průměrného zisku hlasů strany. Zde vidíme, že největší odchylky až 11% procent dosáhla ČSSD .
- Díváme-li se na hodnoty variačního koeficientu, které jsou v procentuálním zobrazení ...
- Variační koeficient, který nám ukazuje podíl variability vzhledem k průměru, je u všech stran v pořádku; je v rozmezí hodnot 0-1.
- Giniho koeficient u ČSSD dosahuje středních hodnot

	Průměr	Medián	Variační koeficient	Směrodatná odchylka	Giniho koeficient
ANO	119,80	44	12,667	1517,481	0,453
ČSSD	109,71	38	12,663	1389,324	0,543
KDU-ČSL	40,14	13	12,663	508,308	0,676
KSČM	83,17	33	12,631	1050,562	0,390
ODS	42,99	13	12,697	545,858	0,416
TOP 09	76,26	25	12,815	977,323	0,537
Úsvit	47,73	20	12,587	600,782	0,386

Statistics

	KOALLIB_REL_2008	KSČM_REL_2008	ODS_REL_2008	SLK_REL_2008	SOS_REL_2008	SZ_REL_2008	USZ_REL_2008	ČSSD_REL_2008	TOPSTAN_REL_2012	ČS
N Valid	216	216	216	216	216	216	216	216	216	
Missing	1	1	1	1	1	1	1	1	1	
Mean	3,395327911	15,97938832	22,37284810	16,27015017	2,532834509	3,595496920	3,166716657	25,69271348	3,287423115	1
Median	2,166093212	14,88354739	21,88988095	12,41438356	1,291748936	3,252247488	2,784357638	25,36772658	2,525564443	1
Mode	,0000000000	,0000000000	16,66666667	,0000000000	,0000000000	,0000000000	,0000000000	26,08695652	,0000000000	8
Variance	18,199	64,761	81,189	148,632	15,691	9,273	8,534	63,953	12,035	

a. Multiple modes exist. The smallest value is shown

- silně koncentrovaný voličský elektorát
- nejbližše absolutně koncentrovanému rozložení oproti jiným stranám byla SOS a USZ, nicméně ani tyto strany nedosahují ideálu

Popis vývoje volební podpory

Popis vývoje volební podpory

- Bazické a řetězové indexy
- Korelační koeficient
- Území stabilní volební podpory

Území volební podpory

- Přehlednost, zohlednění populační velikosti
- Nevhodné pro lokální strany (např. SMK na Slovensku)
- Doplnující indikátory
 - Míra úspěšnosti: kolikrát je podpora strany vyšší v (jádru) území volební podpory oproti zbytku území

Bazické a řetězové indexy

- Řetězový index
 - Volby 2013/volby 2010; volby2010/volby2006 ...
 - Ukazuje postupný vývoj
- Bazický index
 - Volby 2013/volby2002, volby2010/volby 2002
 - Ukazuje změnu stav oproti stanovenému základu
 - Index volební úspěšnosti
 - =krajské volby 2012/poslanecké volby 2010 * 100
 - Index volební stability
 - Krajské volby 2012/ krajské 2008 * 100

Korelační analýza

- Míra souvislosti mezi dvěma proměnnými
- Lineární vztah
- Nárůst hodnot jedné proměnné je spojený s nárůstem / poklesem hodnot druhé proměnné
- Korelace neimplikuje kauzalitu
- Možnost vytvořit (libovolně) velkou matici
 - Souvislost následujících voleb i souvislost vzdálených voleb

Pearsonův korelační koeficient

- Předpoklady:
 - Kardinální data (možná výjimka)
 - Pokud zjišťujeme i statistickou signifikanci, tak i normální rozložení (nebo dostatečná velikost vzorku)
 - V prostorové analýze obvykle není důvod zabývat se signifikancí
- Hodnoty koeficientu: Rozsah od -1 po 1
- Čím více je hodnota vzdálena od nuly, tím je souvislost silnější
- Existují různé návody jak označovat různě silné korelace
 - NEDĚLEJTE TO !!!

Interpretace koeficientů

- Míra souvislosti
- Neukazuje kauzalitu
- Přímou lze hodnoty srovnávat, jen pokud jsou spočítány na stejném počtu případů (- to se týká především signifikance)

Území stabilní volební podpory

- Strany s koncentrovanou stabilní podporou X strany s nekoncentrovanou stabilní podporou
- Výpočet procenta hlasů přítomných v ÚSVP v každých volbách

Analýza normálního rozdělení

- Průměr, medián
- Směrodatná odchylka
- Kvartily
- Šikmost, strmost

- testy

Proč?

- Můžeme použít proměnné v další analýze?