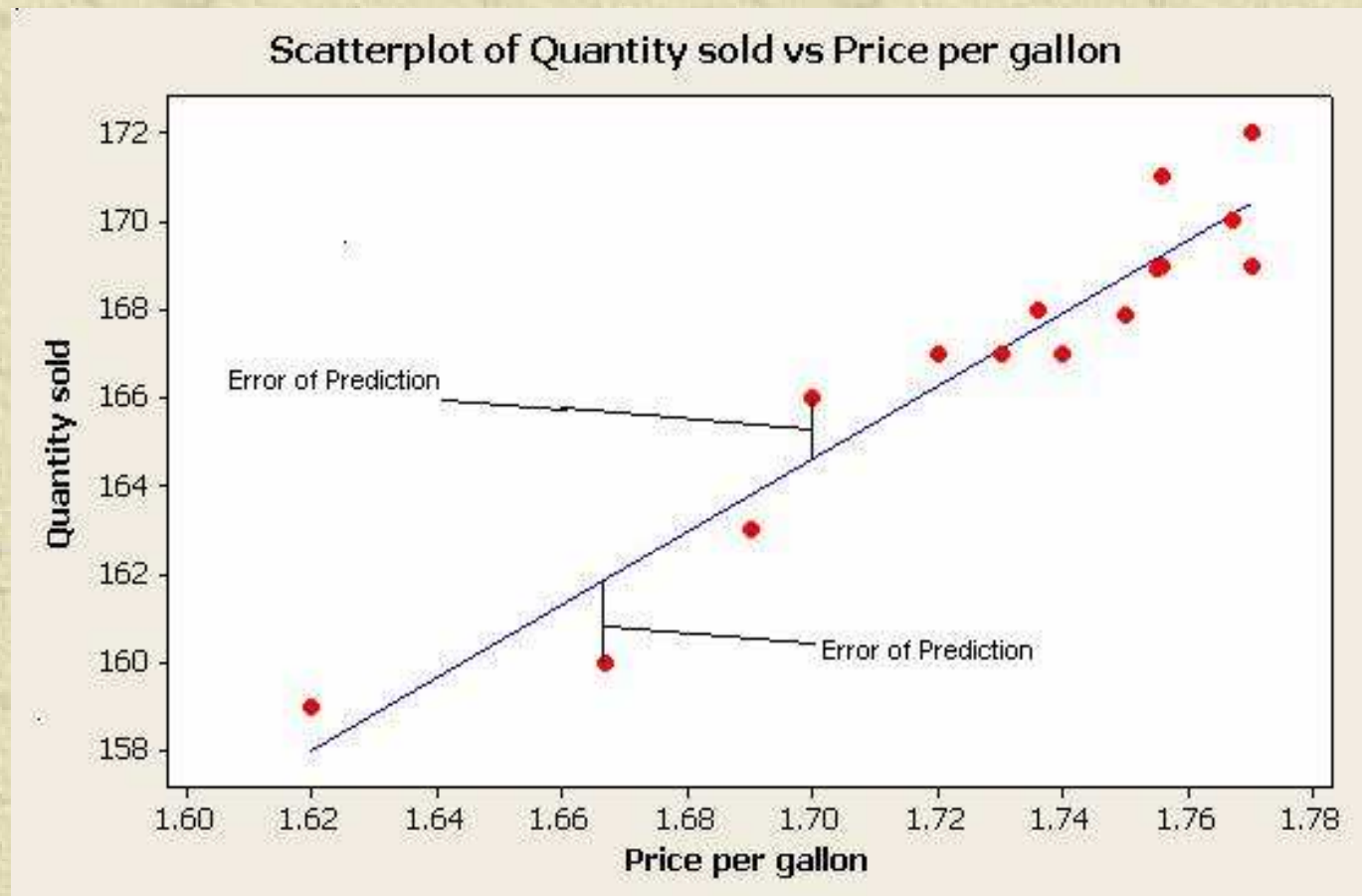


# Aplikace regresní analýzy v humánní geografii



# Všeobecná souvislost jevů

## ✦ Souvislost jevů

- ◆ náhodná
- ◆ příčinná souvislost
  - výskyt jedné proměnné podmiňuje výskyt jiné
  - nebo existuje další proměnná, která je společnou příčinou současného výskytu obou (všech) sledovaných proměnných

## ✦ Náhodnost a příčinnost souvislostí je relativní

- ◆ odpovídá stupni našeho poznání

## ✦ Příklad:

- ◆ srážky a nadmořská výška
  - příčinná souvislost
- ◆ věk nevěst a věk ženichů
  - není příčinná souvislost, podmíněno komplexem faktorů životního cyklu

## ✦ Zřetězení příčin a následků

- ◆ proměnné mohou být v jednom vztahu příčinami a v jiném následky

# Funkční a stochastická závislost

## ✦ Funkční (pevná) závislost

- vždy dostaneme stejný nebo velmi podobný výsledek
- typická v neživé přírodě (fyzika, chemie)
  - zákonitost
  - příklad: vztah teploty a objemu tělesa
- mezi společenskými (geografickými) jevy se nevyskytuje

## ✦ Stochastická (volná, statistická) závislost

- při změnách nezávisle proměnné se *zpravidla* (v průměru) mění hodnoty závisle proměnné
  - ale ne v každém jednotlivém případě

## ✦ Ve společenských procesech

- vždy komplex příčin a následků
- jednotlivé proměnné je obtížné identifikovat
  - mění se v prostoru a čase

# Postup zkoumání vztahu mezi proměnnými

## ✦ Začátek analýzy: formulace hypotéz

- ◆ o možných vnitřních příčinných souvislostech
  - logický rozbor vzájemného vztahu proměnných

## ✦ Ověření existence závislosti

- ◆ zda vůbec existuje (je významná)
- ◆ jaký je její základní charakter
  - pozitivní
  - negativní
- ◆ jaký je její specifický charakter
  - tvar (= regrese)
- ◆ zjištění těsnosti závislosti (síly, intenzity závislosti)
  - do jaké míry zjištěné schéma platí?
  - kolik případů vztahů mezi proměnnými mu vyhovuje?

# Korelační vs. regresní analýza

## ✦ Korelační analýza

- ◆ první informace o existenci vztahu a jeho základním charakteru
- ◆ měří intenzitu lineární závislosti mezi dvěma proměnnými

## ✦ Regresní analýza

- ◆ více ucelený proces studia vztahu mezi závisle proměnnou a souborem nezávisle proměnných

# Lineární regrese

- ✦ Předpokládá lineární závislost
  - závisle proměnná ( $y$ )
  - nezávisle proměnné (vysvětlující, *explanatory*) ( $x$ )
- ✦ Prokládá (vyrovnává) hodnoty přímkou
- ✦ Zkoumá účinek (*effect*) proměnné  $x$  na  $y$
- ✦ Významným výsledkem analýzy je regresní rovnice
  - umožňuje předpovídat hodnoty  $y$  z hodnot  $x$

$$\hat{y} = a + bx$$

- $\hat{y}$  ... předpokládaná hodnota  $y$  (*predicted value*)
- $x$  ... pozorovaná hodnota  $x$
- $a$  ... intercept (průsečík, konstanta) (místo kde regresní přímka protíná vertikální osu)
- $b$  ... sklon přímky
- $a, b$  ... parametry regresní rovnice (přímky)

# Lineární regrese

✦ Intercept (konstanta) nemusí mít vždy reálnou interpretaci

- ◆ příklad: vztah ceny domu a jeho velikosti

✦ Rezidua

- ◆ rozdíl mezi očekávanou a skutečnou (naměřenou) hodnotou

✦ Cíl regresní analýzy:

- ◆ najít parametry (a,b) nejlépe vyhovující přímky/křivky

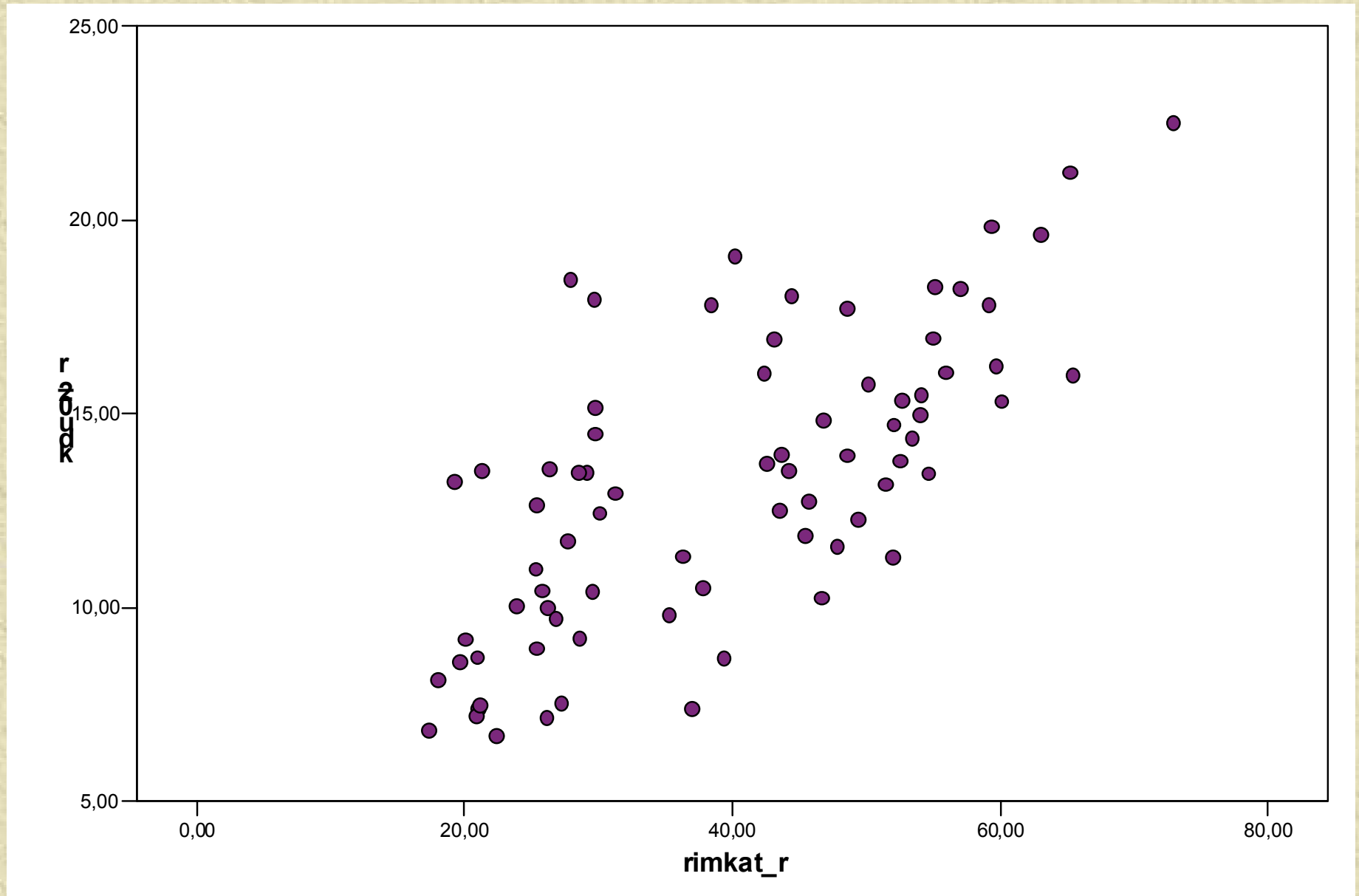
- metoda nejmenších čtverců

- ◆ minimalizace součtu čtverců vertikálních vzdáleností (= minimalizace součtu čtverců reziduí)

	d	krok	nazev	m	metro	morava	pohr	kod1	rdb201s1	Volici90odhad	hlasyc90
1		3101	Praha	1	1	3	9	9	1214174	961834	860341
2		3201	Benešov	1	9	1	1	1	88362,0	65812	63795
3		3202	Beroun	1	9	1	1	1	75859,0	57872	55161
4		3203	Kladno	1	9	1	1	1	149407,0	113368	106154
5		3204	Kolín	1	9	1	1	1	91319,0	68758	65833
6		3205	Kutná Hora	1	9	1	1	1	78938,0	60001	57079
7		3206	Mladá Boleslav	1	9	1	1	1	94402,0	70985	66392
8		3207	Mladá Boleslav	1	9	1	1	1	111694,0	83408	77613
9		3208	Nymburk	1	9	1	1	1	89154,0	67514	64685
10		3209	Praha východ	1	1	3	9	9	93626,0	71551	68049
11		3210	Praha západ	1	1	3	9	9	76164,0	59023	56134
12		3211	Píseň	1	9	1	1	1	108805,0	81302	78548
13		3212	Rakovník	1	9	1	1	1	55152,0	41415	39527
14		3301	Česká, Budějovice	1	2	1	1	2	173386,0	133064	125285
15		3302	Český Krumlov	1	9	1	5	5	57388,0	42385	41432
16		3303	Jindřichův Hradec	1	9	1	1	1	93048,0	69214	66824
17		3304	Pelhřimov	1	9	1	1	1	74614,0	55286	53475
18		3305	Písek	1	9	1	1	1	72074,0	55266	54180
19		3306	Prachatice	1	9	1	3	3	50985,0	37966	36886
20		3307	Strakonice	1	9	1	1	1	71978,0	53469	51639
21		3308	Tábor	1	9	1	1	1	104030,0	79158	75988
22		3401	Domažlice	1	9	1	3	3	58729,0	44069	43861
23		3402	Čech	1	9	1	5	5	86932,0	68655	66897
24		3403	Karlovy Vary	1	2	1	5	5	122430,0	92183	89367
25		3404	Klatovy	1	9	1	1	1	89767,0	69315	69269
26		3405	Plzeň	1	2	1	1	2	173008,0	134490	124954
27		3406	Plzeň-jih	1	2	1	1	2	68149,0	51870	50487
28		3407	Plzeň-sever	1	2	1	1	2	72500,0	54602	52409
29		3408	Rokycany	1	9	1	1	1	46118,0	36362	35343
30		3409	Sokolov	1	9	1	5	5	92623,0	67533	60809
31		3410	Tachov	1	9	1	5	5	50036,0	38089	37766
32		3501	Česká Lípa	1	9	1	5	5	102259,0	72644	67071
33		3502	Dobruška	1	9	1	5	5	133448,0	97005	89788
34		3503	Chomutov	1	3	1	8	8	124081,0	89426	80137

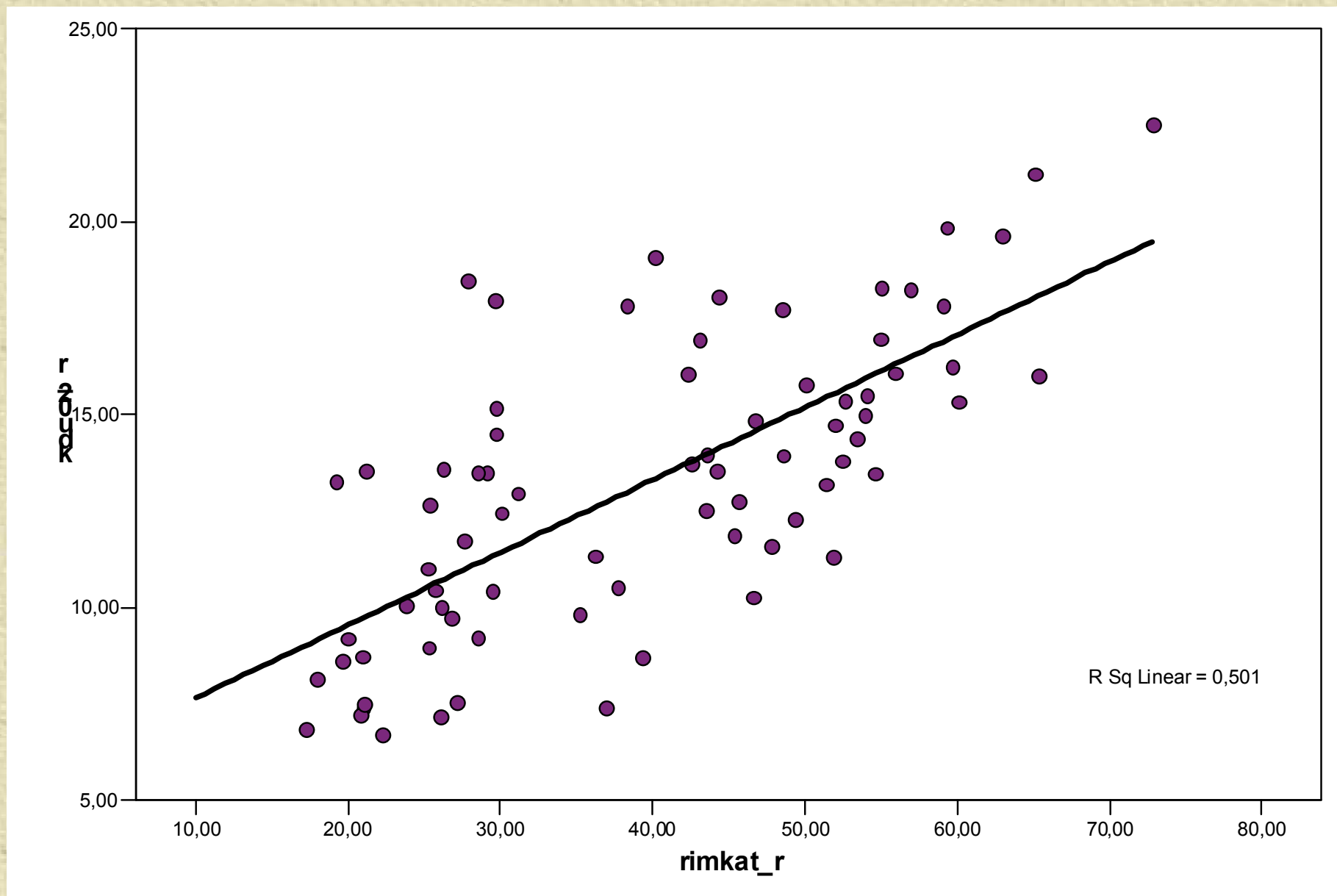


# Závislost volební podpory KDU-ČSL (y) na podílu katolíků (x)



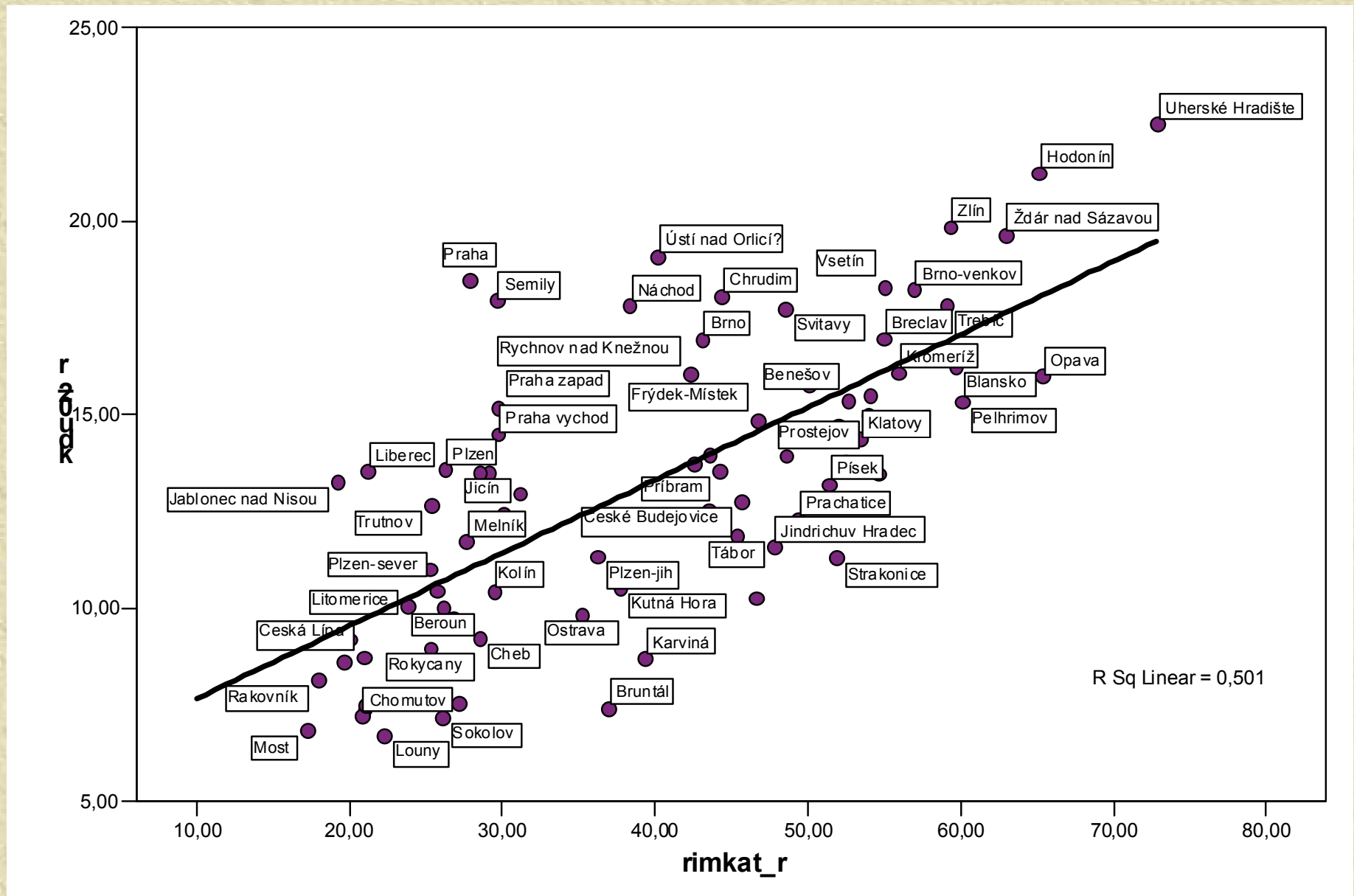
# Regresní přímka

Závislost volební podpory KDU-ČSL (y) na podílu katolíků (x)



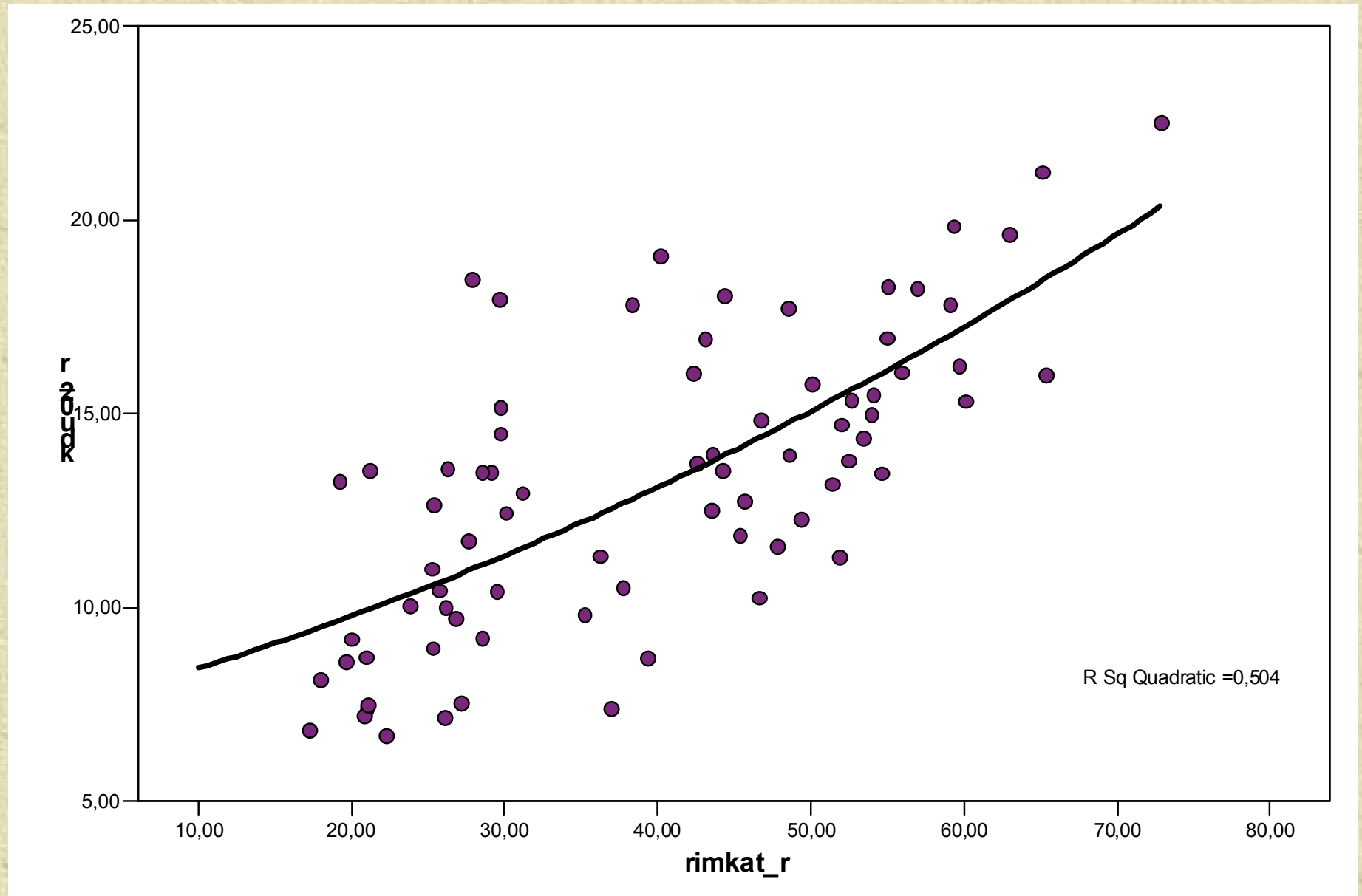
# Regresní přímka

Závislost volební podpory KDU-ČSL (y) na podílu katolíků (x)



# Regresní křivka (kvadratická rovnice)

Závislost volební podpory KDU-ČSL (y) na podílu katolíků (x)



101	Praha	1	1	3	9	9	1214174	961834	860341	89,45	3866
201	Benešov	1	9	1	1	1	88362,0	65812	63795	96,94	188
202	Beroun							57872	55161	95,32	211
203	Kladno							113368	106154	93,64	329
204	Kolín							68758	65833	95,75	230
205	Kutná Hora							60001	57079	95,13	211
206	Mělník							70985	66392	93,53	217
207	Mladá Boleslav							83408	77613	93,05	254
208	Nymburk							67514	64685	95,81	212
209	Praha východ							71551	68049	95,11	266
210	Praha západ							59023	56134	95,11	213
211	Příbram							81302	78548	96,61	231
212	Rakovník							41415	39527	95,44	95
301	České Budějovice							133064	125285	94,15	291
302	Český Krumlov							42385	41432	97,75	78
303	Jindřichův Hradec							69214	66824	96,55	124
304	Pelhřimov							55286	53475	96,72	114
305	Písek							55266	54180	98,03	124
306	Prachatice							37966	36886	97,16	59
307	Strakonice							53469	51639	96,58	98
308	Tábor							79158	75988	96,00	212
401	Domažlice							44069	43861	99,53	76
402	Cheb							68655	66897	97,44	164
403	Karlovy Vary	1	2	1	5	5	122430,0	92183	89367	96,95	261
404	Klatovy	1	9	1	1	1	89767,0	69315	69269	99,93	118
405	Plzeň	1	2	1	1	2	173008,0	134490	124954	92,91	505
406	Plzeň-jih	1	2	1	1	2	68149,0	51870	50487	97,33	115
407	Plzeň-sever	1	2	1	1	2	72500,0	54602	52409	95,98	140
408	Rokycany	1	9	1	1	1	46118,0	36362	35343	97,20	102
409	Sokolov	1	9	1	5	5	92623,0	67533	60809	90,04	155
410	Tachov	1	9	1	5	5	50036,0	38089	37766	99,15	100
501	Česká Lípa	1	9	1	5	5	102259,0	72644	67071	92,33	182
502	Děčín	1	9	1	5	5	133448,0	97005	89788	92,56	295
503	Chomutov	1	3	1	8	8	124081,0	89426	80137	89,61	324
504	Jablonec nad Nisou	1	2	1	3	3	88118,0	65323	60993	93,37	557
505	Liberec	1	2	1	5	5	159196,0	118344	110954	93,76	384
506	Litoměřice	1	9	1	3	3	113993,0	85473	80644	94,35	213

**Linear Regression** ✕

Dependent:

Block 1 of 1

Previous Next

Independent(s):

Method: Enter

Selection Variable:

Case Labels:

WLS Weight:

- d\_r
- kodok
- nazev
- m
- metro
- morava
- kod polohy - pohra
- kod [kod1]
- rdb201s1
- odhad90 [Volici90o
- hlasyc90
- ucast90e
- css\_3
- sb\_6
- of\_7
- vdspr\_8
- vszs\_9
- ksc\_10
- szv\_11

101	Praha	1	1	3	9	9	1214174	961834	860341	89,45	3866
201	Benešov	1	9	1	1	1	88362,0	65812	63795	96,94	188
202	Beroun							57872	55161	95,32	211
203	Kladno							113368	106154	93,64	329
204	Kolín							68758	65833	95,75	230
205	Kutná Hora							60001	57079	95,13	211
206	Mělník							70985	66392	93,53	217
207	Mladá Boleslav							83408	77613	93,05	254
208	Nymburk							67514	64685	95,81	212
209	Praha východ							71551	68049	95,11	266
210	Praha západ							59023	56134	95,11	213
211	Příbram							81302	78548	96,61	231
212	Rakovník							41415	39527	95,44	95
301	České Budějovice							133064	125285	94,15	291
302	Český Krumlov							42385	41432	97,75	78
303	Jindřichův Hradec							69214	66824	96,55	124
304	Pelhřimov							55286	53475	96,72	114
305	Písek							55266	54180	98,03	124
306	Prachatice							37966	36886	97,16	59
307	Strakonice							53469	51639	96,58	98
308	Tábor							79158	75988	96,00	212
401	Domažlice							44069	43861	99,53	76
402	Cheb							68655	66897	97,44	164
403	Karlovy Vary	1	2	1	5	5	122430,0	92183	89367	96,95	261
404	Klatovy	1	9	1	1	1	89767,0	69315	69269	99,93	118
405	Plzeň	1	2	1	1	2	173008,0	134490	124954	92,91	505
406	Plzeň-jih	1	2	1	1	2	68149,0	51870	50487	97,33	115
407	Plzeň-sever	1	2	1	1	2	72500,0	54602	52409	95,98	140
408	Rokycany	1	9	1	1	1	46118,0	36362	35343	97,20	102
409	Sokolov	1	9	1	5	5	92623,0	67533	60809	90,04	155
410	Tachov	1	9	1	5	5	50036,0	38089	37766	99,15	100
501	Česká Lípa	1	9	1	5	5	102259,0	72644	67071	92,33	182
502	Děčín	1	9	1	5	5	133448,0	97005	89788	92,56	295
503	Chomutov	1	3	1	8	8	124081,0	89426	80137	89,61	324
504	Jablonec nad Nisou	1	2	1	3	3	88118,0	65323	60993	93,37	557
505	Liberec	1	2	1	5	5	159196,0	118344	110954	93,76	384
506	Litoměřice	1	9	1	3	3	113993,0	85473	80644	94,35	213

**Linear Regression** [X]

Dependent:

Block 1 of 1

Independent(s):

Method: 

- Enter
- Stepwise
- Remove
- Backward

Selection Variable:

Case Labels:

WLS Weight:

101	Praha	1	1	3	9	9	1214174	961834	860341	89,45	3866
					1	1	88362,0	65812	63795	96,94	188
					1	1	75859,0	57872	55161	95,32	211
					1	1	149407,0	113368	106154	93,64	329
					1	1	91319,0	68758	65833	95,75	230
					1	1	78938,0	60001	57079	95,13	211
					1	1	94402,0	70985	66392	93,53	217
					1	1	111694,0	83408	77613	93,05	254
					1	1	89154,0	67514	64685	95,81	212
					9	9	93626,0	71551	68049	95,11	266
					9	9	76164,0	59023	56134	95,11	213
					1	1	108805,0	81302	78548	96,61	231
					1	1	55152,0	41415	39527	95,44	95
					1	2	173386,0	133064	125285	94,15	291
					5	5	57388,0	42385	41432	97,75	78
					1	1	93048,0	69214	66924	96,55	124
											114
											124
											59
											98
											212
											76
											164
											261
											118
											505
											115
											140
											102
											155
402	Cheb	1	9	1	5	5	52025,0	67533	60009	96,04	100
403	Karlovy Vary	1	2	1	5	5	50036,0	38089	37766	99,15	100
404	Klatovy	1	9	1	5	5	102259,0	72644	67071	92,33	182
405	Plzeň	1	2	1	5	5	133448,0	97005	89788	92,56	295
406	Plzen-jih	1	2	1	8	8	124081,0	89426	80137	89,61	324
407	Plzen-sever	1	2	1	3	3	88118,0	65323	60993	93,37	557
408	Rokycany	1	9	1	5	5	159196,0	118344	110954	93,76	384
409	Sokolov	1	9	1	3	3	113993,0	95473	89644	94,35	213
410	Tachov	1	9	1							
501	Česká Lípa	1	9	1							
502	Děčín	1	9	1							
503	Chomutov	1	3	1							
504	Jablonec nad Nisou	1	2	1							
505	Liberec	1	2	1							
506	Litoměřice	1	9	1							

### Linear Regression

Dependent:

Independent(s):

Method:

Selection Variable:

Case Labels:

WLS Weight:

Block 1 of 1

Previous Next

OK Paste Reset Cancel Help

Statistics... Plots... Save... Options...

### Linear Regression: Statistics

Regression Coefficients

- Estimates
- Confidence intervals
- Covariance matrix
- Model fit
- R squared change
- Descriptives
- Part and partial correlations
- Collinearity diagnostics

Residuals

- Durbin-Watson
- Casewise diagnostics
  - Outliers outside:  standard deviations
  - All cases

Continue Cancel Help

### Descriptive Statistics

	Mean	Std. Deviation	N
kdu02_r	13,2015	3,79089	76
rimkat_r	39,4720	14,25927	76

### Correlations

		kdu02_r	rimkat_r
Pearson Correlation	kdu02_r	1,000	,708
	rimkat_r	,708	1,000
Sig. (1-tailed)	kdu02_r	.	,000
	rimkat_r	,000	.
N	kdu02_r	76	76
	rimkat_r	76	76

### Variables Entered/Removed<sup>b</sup>

Model	Variables Entered	Variables Removed	Method
1	rimkat_r <sup>a</sup>	.	Enter

a. All requested variables entered.

b. Dependent Variable: kdu02\_r

### Model Summary<sup>b</sup>

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	,708 <sup>a</sup>	,501	,495	2,69503	,501	74,394	1	74	,000	1,053

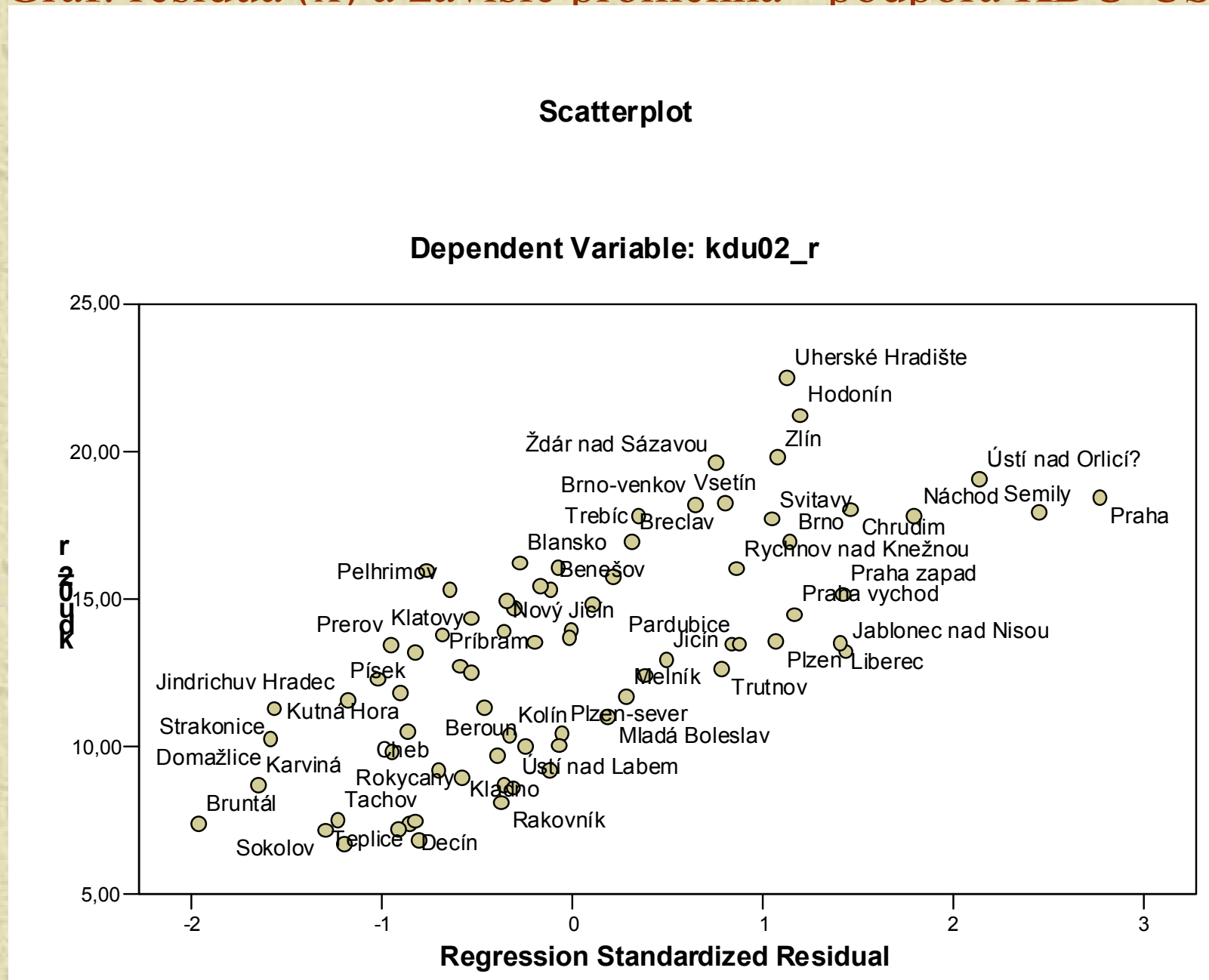
a. Predictors: (Constant), rimkat\_r

b. Dependent Variable: kdu02\_r



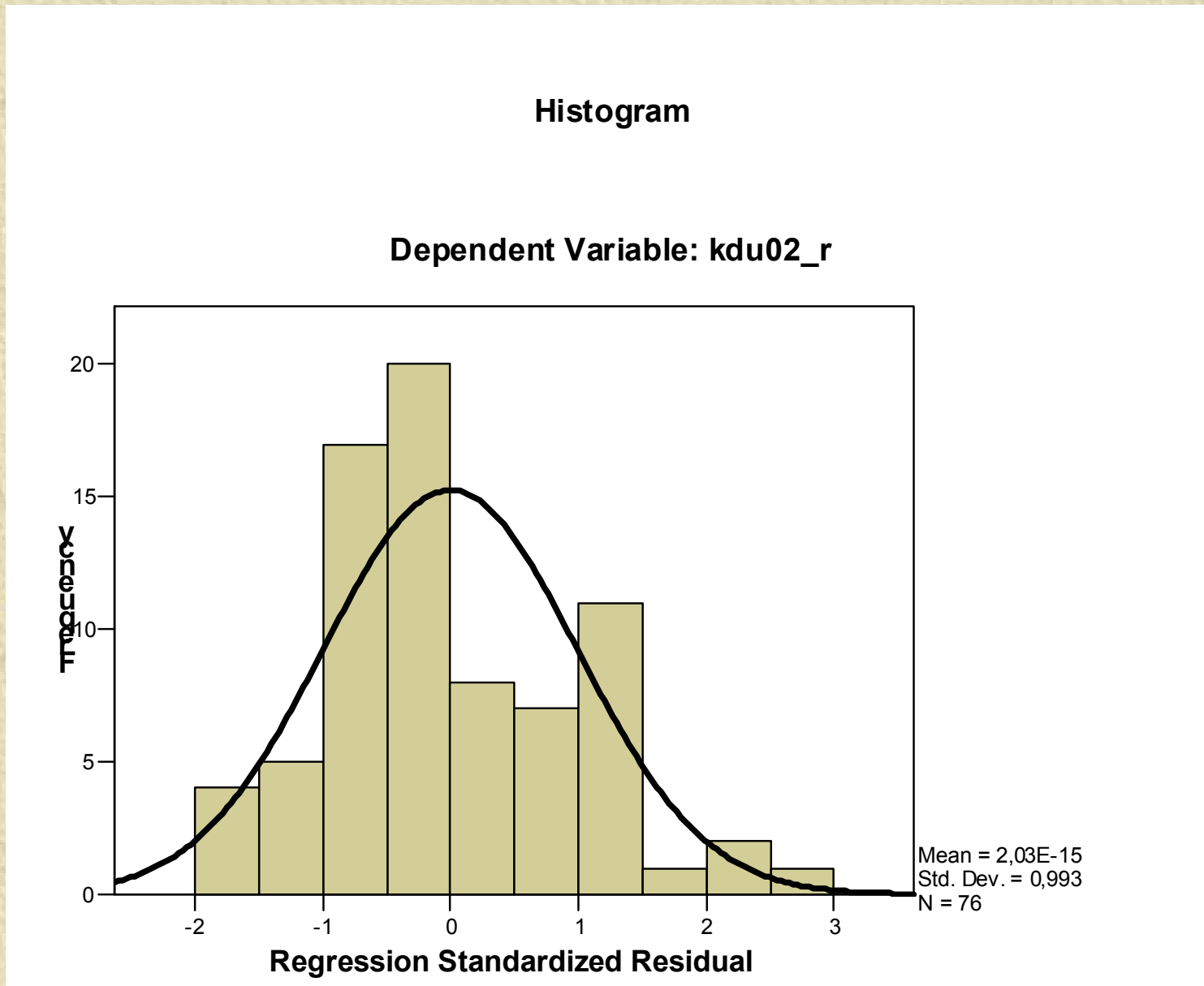
# Jednoduchá regresní analýza

Graf: residua (x) a závisle proměnná - podpora KDU-ČSL (y)



# Histogram reziduí

Histogram: residua (x) a četnost (*frequency*) (y)

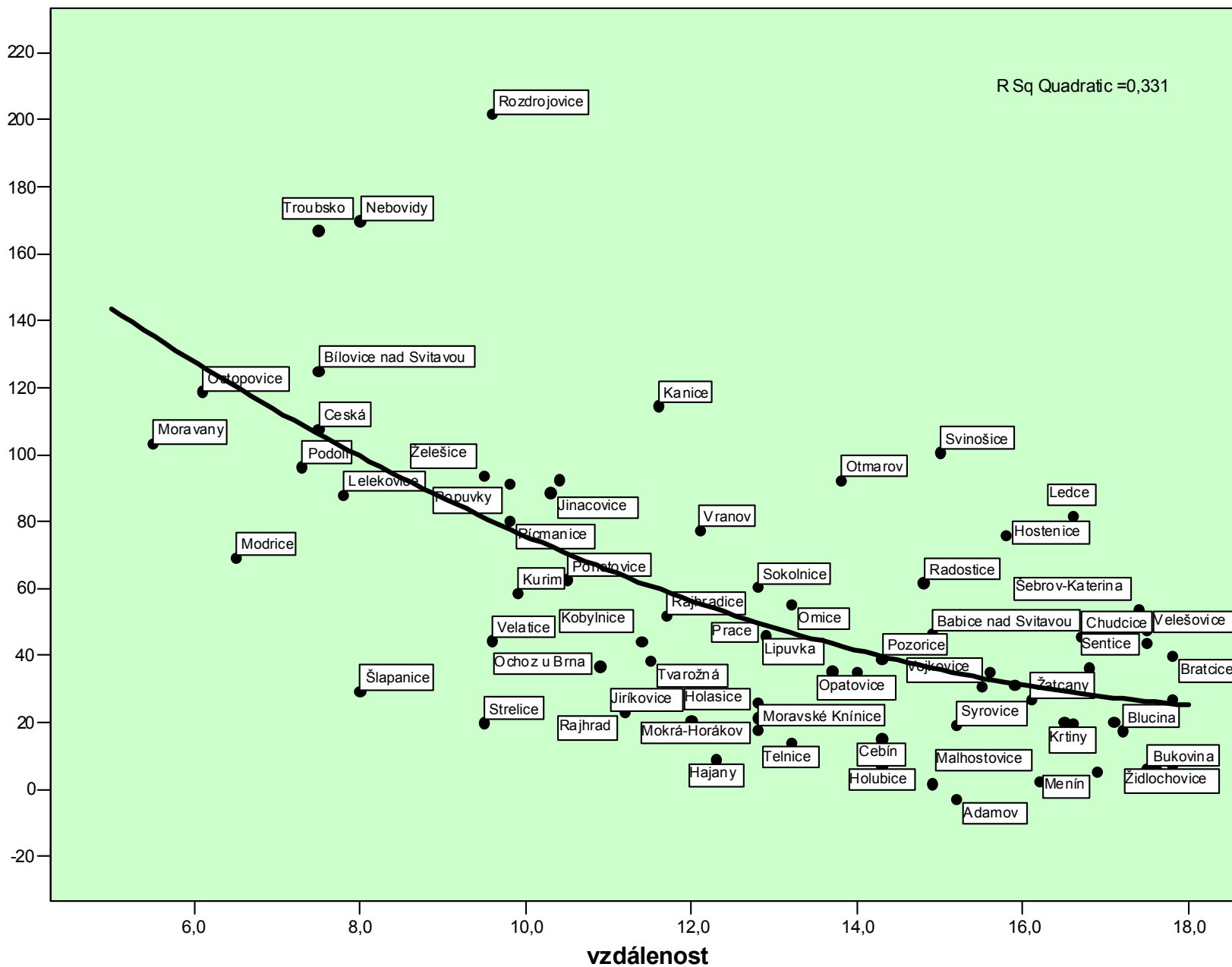


# Příklad: podmíněnost suburbanizace vzdáleností od centrálního města

Závisle proměnná: migrační saldo s Brnem na 1000 obyvatel

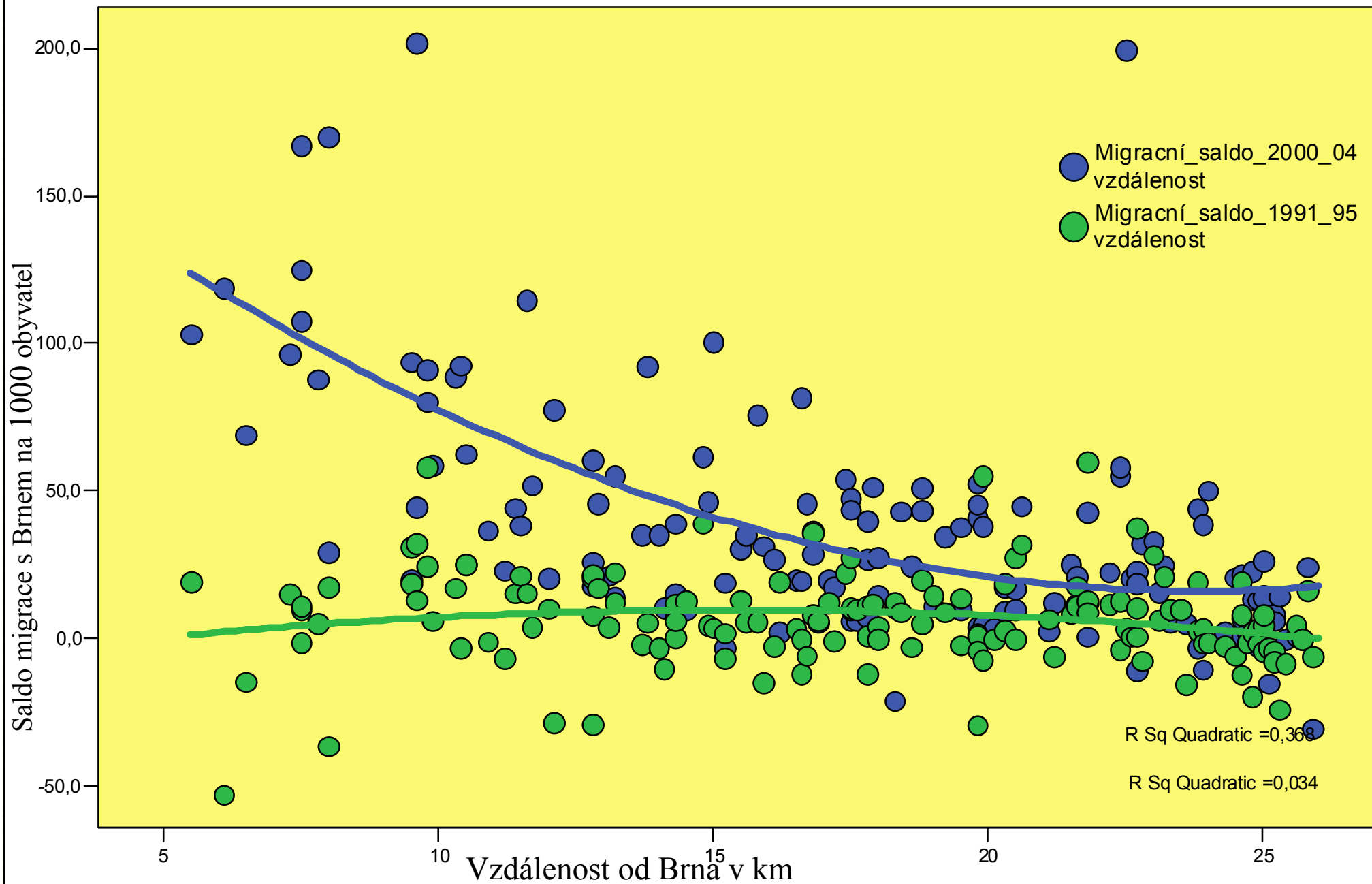
*Zázemí Brna, období 2000-04, obce do 18 km od Brna*

Migrační saldo s Brnem na 1000 obyvatel



# Vztah mezi migračním saldem s Brnem (na 1000 obyvatel) a vzdáleností od Brna

*Srovnání období 1991-1995 (zelená) a 2000-04 (modrá)*



# Vícenásobná regrese

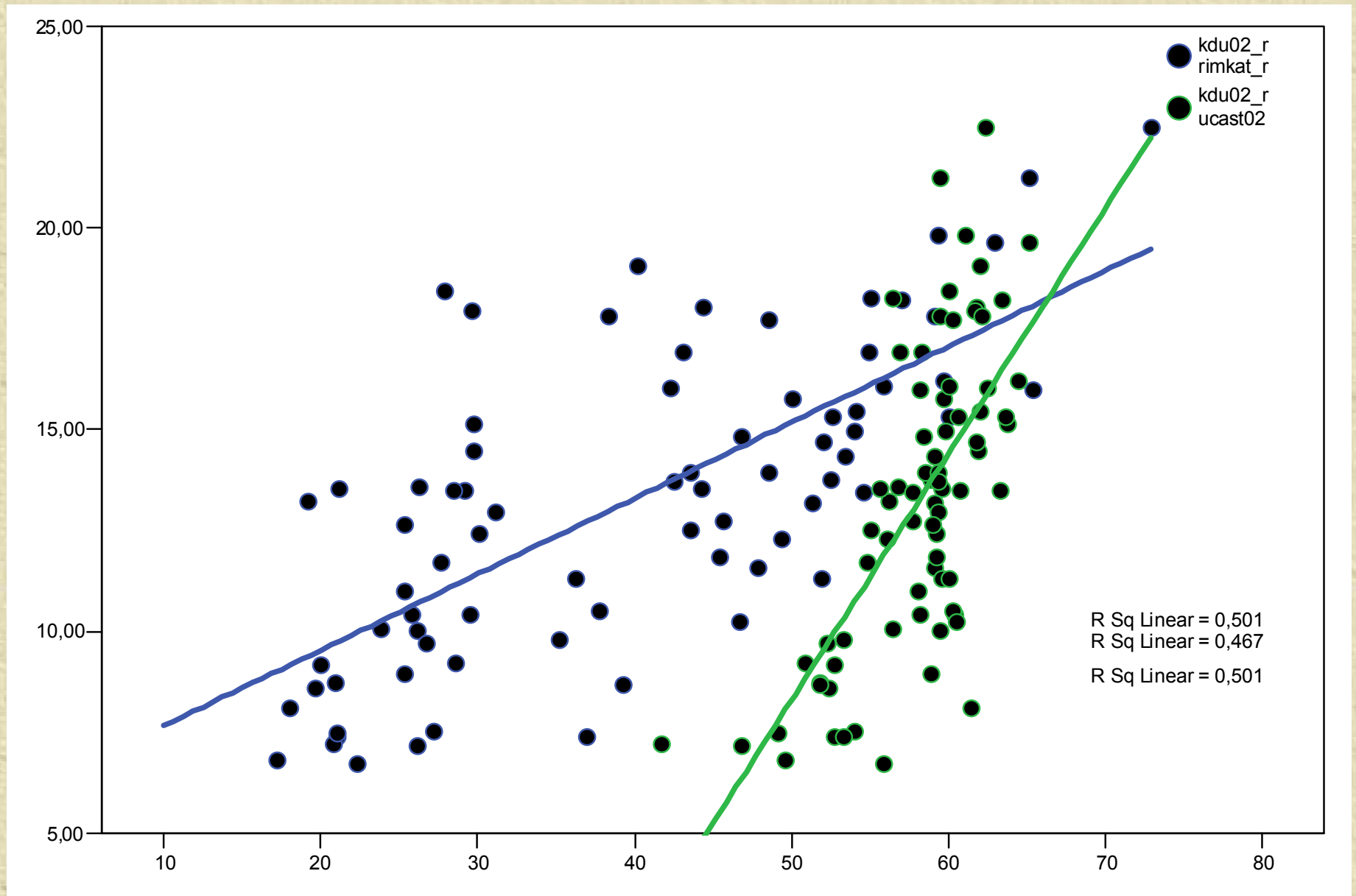
✦ Závisle proměnná:

- ◆ volební podpora KDU-ČSL

✦ Nezávisle proměnné

- ◆ podíl katolíků
- ◆ volební účast

# Graf (scatterplot)



## → Regression

### Descriptive Statistics

	Mean	Std. Deviation	N
kdu02_r	13,2015	3,79089	76
rimkat_r	39,4720	14,25927	76
ucast02	58,0206	4,24655	76

### Correlations

		kdu02_r	rimkat_r	ucast02
Pearson Correlation	kdu02_r	1,000	,708	,683
	rimkat_r	,708	1,000	,525
	ucast02	,683	,525	1,000
Sig. (1-tailed)	kdu02_r	.	,000	,000
	rimkat_r	,000	.	,000
	ucast02	,000	,000	.
N	kdu02_r	76	76	76
	rimkat_r	76	76	76
	ucast02	76	76	76

### Variables Entered/Removed<sup>b</sup>

Model	Variables Entered	Variables Removed	Method
1	ucast02 <sup>a</sup> rimkat_r	.	Enter

a. All requested variables entered.

b. Dependent Variable: kdu02\_r





# Příklad: Využití regresní analýzy v politické geografii

- ✦ Pramen: Tatu Vanhanen (1987) *The level of democratization related to socioeconomic variables in 147 countries*
- ✦ Výchozí otázka: co podmiňuje nerovnoměrné rozložení stupně demokracie ve světě?
- ✦ Východiska:
  - ◆ Demokracie je ve světě nerovnoměrně rozšířena
    - některé státy jsou více demokratické než jiné, některé nejsou demokratické vůbec
  - ◆ Existuje řada teorií, které vysvětlují stupeň demokracie charakterem společnosti
  - ◆ Jednotlivé teorie zdůrazňují význam různých faktorů
    - S. M. Lipset: úroveň hospodářského a sociálního rozvoje
    - R. Gastil: nezáleží tolik na hospodářství, ale na přijetí demokratických myšlenek
    - J. Linz: klíčové jsou politické proměnné („politické dovednosti“ charakter politického systému)
    - S. Huntington: důležité jsou kulturní faktory (vedle socioekonomických)
    - I. Wallerstein: záleží více na vnějších než na vnitřních podmínkách

# Vstupní hypotéza

## ✦ Formulace vstupní hypotézy:

- ◆ politika je zápasem o kontrolu zdrojů
- ◆ proto rozložení moci ve společnosti odpovídá rozložení hlavních zdrojů moci ve společnosti
  - různé formy moci: politická, ekonomická, ideologická, vojenská
- ◆ pokud jsou zdroje moci rozloženy mezi různé společenské skupiny – existují dobré předpoklady pro demokracii
- ◆ pokud jsou koncentrovány v rukou malé skupiny – špatné předpoklady (dobré předpoklady pro autokracii)
- ◆ tedy: relativní rozložení moci mezi jednotlivé části populace je předpokládanou podmínkou pro vysvětlené stupně demokracie

## ✦ Testování hypotézy

- ◆ hodnocení vztahu mezi stupněm demokracie a rozložením (některých) zdrojů moci (charakteristik společnosti)
  - existuje závislost?
  - je dostatečně silná?
- ◆ ale: empirická analýza poskytne vždy jen částečné vysvětlení
  - teoretické koncepce jsou neúplné
  - část rozdílů mezi politickými systémy je zřejmě náhodná
  - nemáme dostatek dat...

# Operacionalizace

## ✦ Metoda:

- ◆ závisle proměnná: stupeň demokracie
- ◆ nezávisle proměnné: vystihují rozložení zdrojů moci ve společnosti
- ◆ jednotky analýzy: 147 států (s více než 200 000 obyvateli, nezávislé, existují data)

## ✦ Operacionalizace: jak měřit stupeň demokracie?

- ◆ demokracie má dvě hlavní dimenze: konkurenci a participaci
- ◆ **konkurence**: stupeň konkurence mezi politickými subjekty (stranami)
  - podíl hlasů ve volbách pro všechny strany s výjimkou nejsilnější
    - ◆ byla jen jediná strana? → konkurence = 0
- ◆ **participace**: rozsah populace, která se účastní politiky (např. ve volbách)
  - volební účast
- ◆ kombinace obou: **index demokracie** (násobek)
  - stanovení prahové hodnoty

# Nezávisle proměnné

- ✦ snaha vybrat proměnné vycházející z teorie
- ✦ ale nutnost dělat kompromisy
  - ◆ některé faktory nejsou (přímo) měřitelné
- ✦ Proměnné, které použil Vanhanen:
  - ◆ IOD - index socioekonomické diverzifikace (socioekonomický rozvoj)
    - UP - urbanizace (podíl městské populace)
    - NAP – podíl nezemědělské populace
  - ◆ IKD - Index distribuce vzdělání
    - Students – počet studentů univerzit na 100 000 obyvatel
    - Literates – podíl gramotných obyvatel
  - ◆ ERD - Index distribuce ekonomických zdrojů
    - Family farms – podíl malých (rodinných) farem na celkové výměře zemědělské půdy
    - NARD – stupeň decentralizace nezemědělských zdrojů (Giniho koeficient – podíl veřejného, soukromého a zahraničního sektoru)
  - ◆ IPR - index zdrojů moci (kombinace tří předchozích) (násobek)
- ✦ alternativní proměnné
  - ◆ HDP na obyvatele
    - hypotéza: v bohatších společnostech (více zdrojů) jsou příznivější podmínky pro demokracii
  - ◆ index etnické homogenity
    - hypotéza: v etnicky/jazykově homogenních společnostech jsou příznivější podmínky pro demokracii

# Analýza

## ✦ Korelační analýza

- ◆ první informace o existenci vztahů mezi proměnnými

## ✦ Testování vztahu mezi nezávisle proměnnými

- ◆ závislost by měla být nízká
- ◆ každá proměnná vyjadřuje jinou vlastnost
- ◆ omezení problému multikolinearity

## ✦ Regresní analýza

- ◆ viz obr.