

# Regresní analýza v „prostorové“ analýze

Petr Voda

# Regrese - připomenutí

---

- ▶ Nástroj k analýze vlivu více nezávisle proměnných na jednu závisle proměnné
  - ▶ Jak různé vlastnosti obcí ovlivňují to, kolik procent hlasů v ní kandidát dostane
  - ▶ Testování teorie
  - ▶ Musíme mít nějaký předpoklad, jaké proměnné, proč a jak by měly ovlivňovat závisle proměnnou
  - ▶ Proč = mechanismus
  - ▶ Jak = hypotéza



# Podmínky použití

---

- ▶ Pro prostorovou analýzu můžeme některé jinak důležité podmínky ignorovat
  - ▶ Normalita závisle proměnné
- ▶ Některé musíme ignorovat
  - ▶ Nezávislost pozorování
- ▶ Podmínky
  - ▶ Jedna závisle proměnná
  - ▶ Předpoklad lineárního vztahu
  - ▶ **Nezávislost nezávisle proměnných mezi sebou**
  - ▶ **Opakem je multikolinearita**



# Specifika v prostorové analýze

---

## ▶ Nezávislost pozorování

- ▶ Často narušeno
- ▶ V blízkých lokalitách často podobné hodnoty

## ▶ Normalita závisle proměnné

- ▶ Velmi důležitá zejména pro hodnoty inferenční statistiky
- ▶ V analýze zahrnující všechny případy není nutná taková přísnost
- ▶ Rozdělení by se ale normálnímu mělo alespoň přibližovat

## ▶ Multikolinearita

- ▶ Častý problém

## ▶ Nestacionarita

- ▶ V různých místech mohou být vztahy mezi proměnnými

---

▶ různé

# Teorie

---

## ▶ Teorie konfliktních linií

- ▶ Strany vznikly, protože některé sociální skupiny chtěly být reprezentovány
- ▶ Strany jako odraz sociální struktury
- ▶ Vlastníci x pracující
- ▶ Církev x stát
- ▶ Město x venkov
- ▶ Centrum x periferie

## ▶ Politická socializace

- ▶ Prostředí může ovlivňovat předávané hodnoty
  - ▶ Např. sudety
- 



# Geografické efekty

---

- ▶ Sousedský efekt
- ▶ Efekt nákazy
- ▶ Efekt sporného bodu



# Kontextuální vs kompozitní efekty

---

## ▶ Kompozitní efekt

- ▶ Strana má v místě se silným zastoupením skupiny vysokou podporu, protože členové skupiny stranu volí
- ▶ Např. katolíci a KDU-ČSL

## ▶ Kontextuální efekt

- ▶ Přítomnost skupiny vytváří příhodný kontext pro to, aby stranu volili jiní lidé
- ▶ Např. efekt nezaměstnanosti na podporu levicových stran
- ▶ Nezaměstnaní se velice často neúčastní voleb
- ▶ Vysoká nezaměstnanost aktivizuje zaměstnance pocitově ohrožené nezaměstnaností



# Proměnné

---

- ▶ Výběr proměnných je určen teorií
- ▶ V procentech – jinak zkoumáte jen efekt velikosti obce
- ▶ Správně spočítaná procenta
  - ▶ Jinak se měří nezaměstnanost a jinak religiozita
- ▶ Proměnné, které měří různé věci
- ▶ Proměnné, které mají smysl
  - ▶ - teorie





# Problém se senátními volbami

---

- ▶ Je kandidát totéž co strana?
  - ▶ Korelace s podporou strany
  - ▶ Vysoká korelace: ok
  - ▶ Nízká korelace: předpoklady o podpoře kandidáta nemohou být totožné jako předpoklady o podpoře strany
    - ▶ Sousedský efekt?
    - ▶ Koalice stran
    - ▶ Představitel nějakého „křídla“ strany?
    - ▶ Kampaň?
    - ▶ Něco jiného?



# Základ: „jednoduchá regrese“

---

- ▶ Závisle proměnná: podpora kandidáta
- ▶ Nezávisle proměnné: indikátory konfliktních linií
  
- ▶ Příklad: podpora Czernina v obvodu jičín
- ▶ Np:
  - ▶ vlastníci/pracující: nezaměstnanost
  - ▶ Město/venkov: velikost obce (dummy)
  - ▶ Navíc – okres, podíl důchodců



# Co regrese dělá

---

- ▶ **Odhad parametrů** přímky (při 1 nezávisle proměnné), roviny (při 2) či nadroviny (při více)
- ▶ Parametry: **sklon** (pro každou proměnnou) a **konstanta** (jedna pro celý model)
- ▶ Parametry popisují vztah mezi nezávisle a závisle proměnnou
- ▶ Hodnota závisle proměnné (y) = konstanta (a) + sklon(b)\*hodnota nezávisle proměnné (x)
- ▶  $y = a + b \cdot x$
- ▶  $y = a + b_1 \cdot x + b_2 \cdot x + b_3 \cdot x + \dots$



# Co nám výpočet poskytne?

---

- ▶ R-square ( česky index determinace)
  - ▶ Ukazuje jak dobře model sedí na data
- ▶ Parametry
  - ▶ Unstandardized beta (nestandardizovaný beta koeficient)
    - ▶ Standardizovaný beta koeficient
  - ▶ Constant (konstanta)
- ▶ Hodnoty signifikance



	B	Beta	VIF
konstanta	33.62		
nezaměstano st	-0.51	-0.09	1.48
vš	0.50	0.16	1.56
nad 65	-0.03	-0.01	1.14
okres jičín	-17.13	-0.75	1.25
R	0.60		

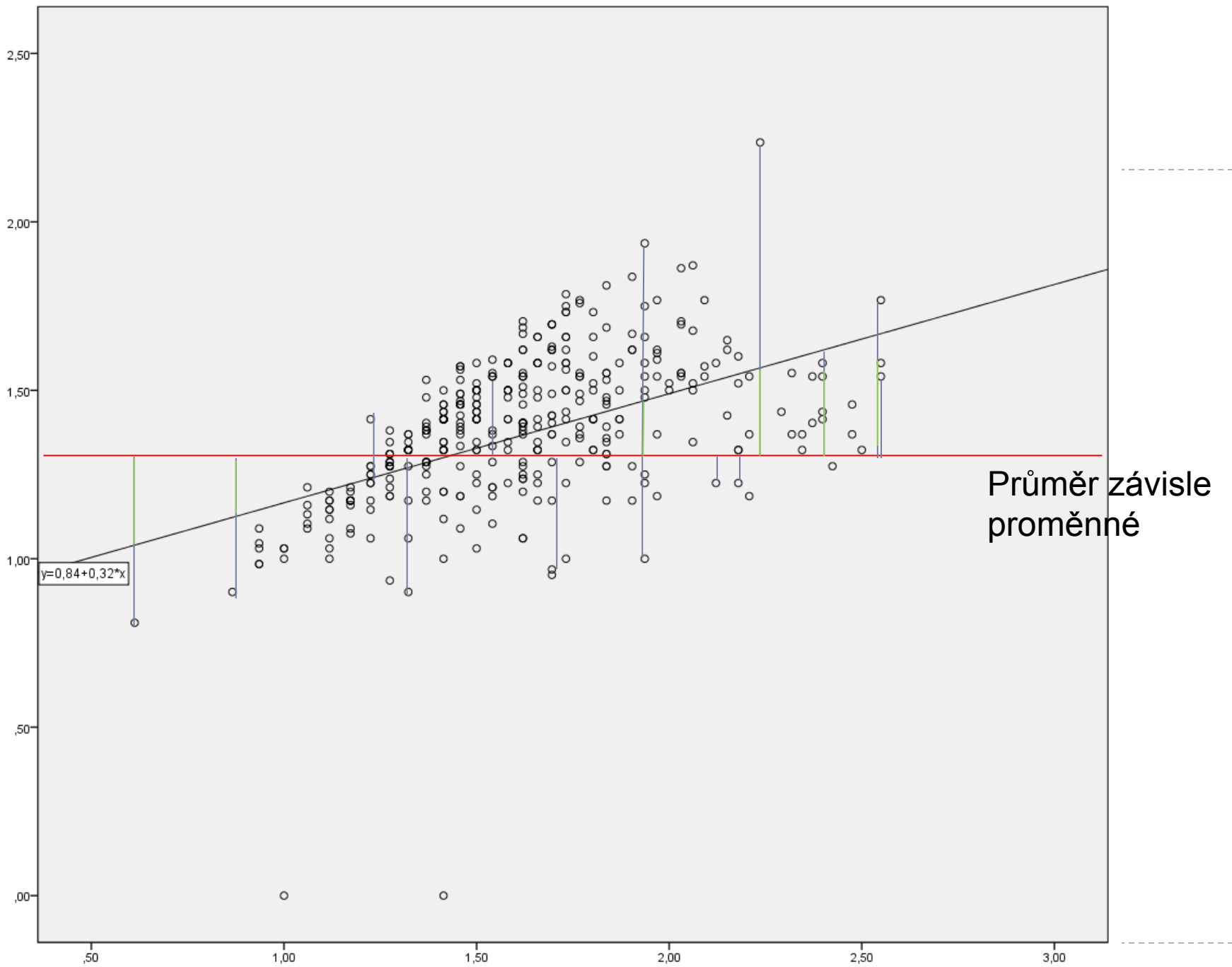


# Co je to R-square?

---

- ▶ Ukazuje, kolik procent rozptylu závisle proměnné je vysvětleno přidáním nezávisle proměnných
- ▶ Původní rozptyl je vypočten jako suma kvadratických odchylek mezi průměrem a jednotlivými hodnotami závisle proměnné
- ▶ „nový“ rozptyl je vypočten jako suma odchylek od regresní přímky/roviny
- ▶ Rozdíl mezi původním a novým rozptylem vydělený původní variabilitou = R-square
- ▶ Čím víc proměnných, tím nižší R-square
  - ▶ Řešeno pomocí adjusted R-square





$y = 0,84 + 0,32 * x$

Průměr závisle  
proměnné

# Nestandardizovaný Beta koeficient

---

- ▶ efekt nezávisle proměnné na závisle proměnnou
- ▶ **„o kolik se změní hodnota závisle proměnné, pokud se hodnota nezávisle proměnné změní o jednotku“** pokud vše ostatní zůstává shodné
- ▶ Různé proměnné se mohou změnit o různý počet jednotek
  - ▶ Pro srovnání síly proměnných v modelu – standardizovaný koeficient beta ( jakou změnu v počtu směrodatných odchylek závisle proměnné způsobí změna o směrodatnou odchylku nezávisle proměnné)

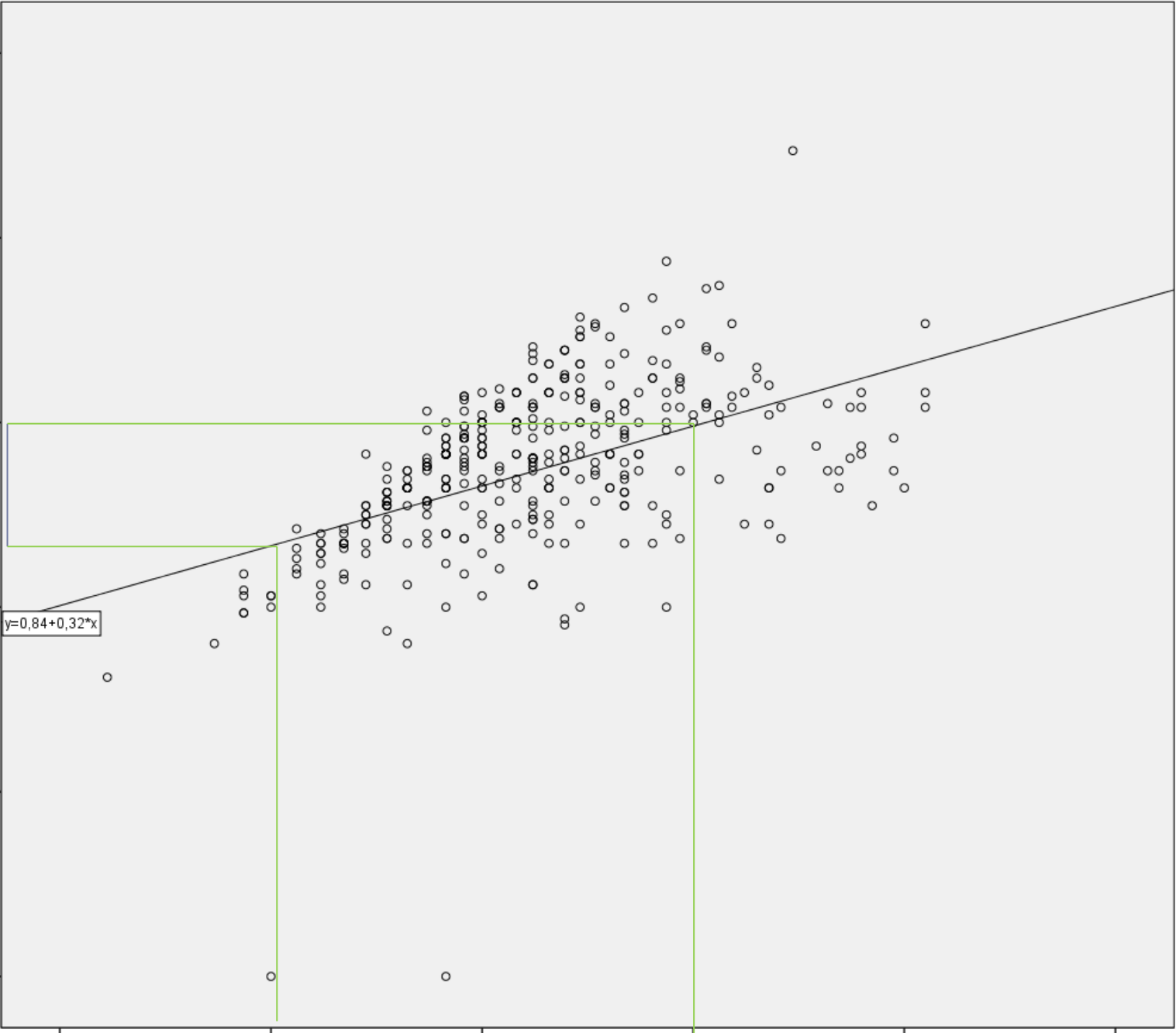




Y

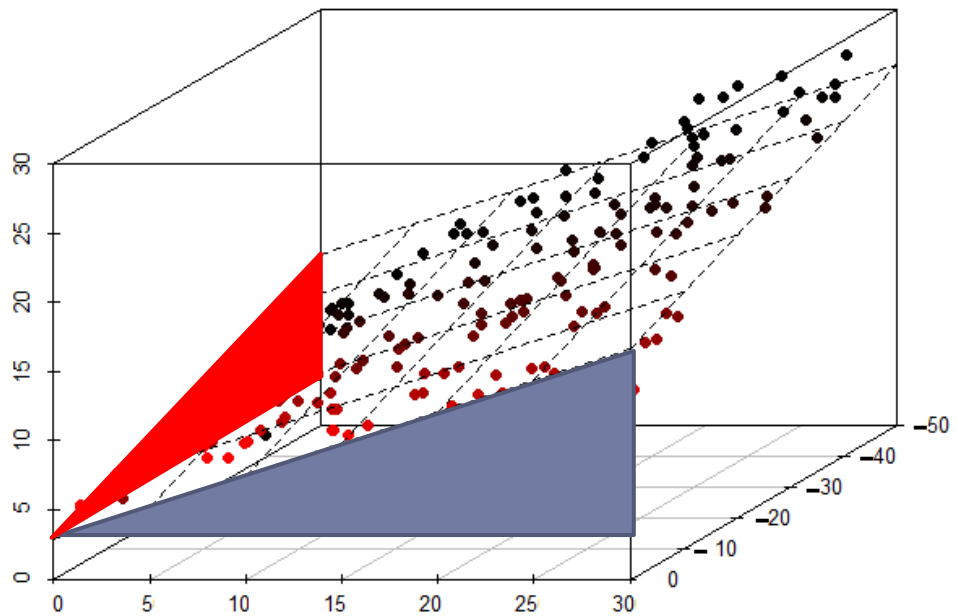
2,50  
2,00  
1,50  
1,00  
,50  
,00

$y=0,84+0,32*x$



,50 1,00 1,50 2,00 2,50 3,00

X



# Interpretace efektu dummy proměnné

---

- ▶ Podpora Kandidáta je v obcích nad 1500 obyvatel o 0,3procentního bodu nižší než v obcích do 1500 obyvatel
- ▶ Nebo též
- ▶ Pokud je vše ostatní shodné, pak rozdíl v podpoře kandidáta mezi vesnicí (obce do 1500 obyvatel) a maloměsty/městy (obce nad 1500 obyvatel) je 0,3 procentního bodu. V menších obcích je podpora vyšší.



- 
- ▶ **Kategorie K se liší o  $\pm XX$  od referenční kategorie** (pokud je vše ostatní shodné)
  - ▶ Interpretace musí obsahovat:
    - ▶ Identifikaci kategorie proměnné, ke které je koeficient vztažen
    - ▶ Identifikaci referenční kategorie
    - ▶ Informaci o velikosti rozdílu



# Interpretace efektu kardinální proměnné

---

- ▶ Pokud je v obci A o 1 pb vyšší podíl vš obyvatelstva než v obci B a vše ostatní je shodné, pak v obci A je podpora kandidáta o 0,5 pb vyšší
- ▶ Nebo též
- ▶ S růstem podílu obyvatel s VŠ vzděláním o 1 pb (pokud vše ostatní zůstává shodné) podpora kandidáta roste o 0,5 pb
- ▶ Lze násobit
  - ▶ Pokud podíl VŠ obyvatelstva vzroste o 2 pb, pak zisk kandidáta vzroste o 1 pb

