

Vícečetná lineární regrese

Petr Ocelík

MVZn4003 Úvod do kvantitativních metod

Blok přednášek

1. Jednoduchá a vícečetná lineární regrese

2. Aplikace a diagnostika

1. opakování

2. interpretace kategorických prediktorů

3. předpoklady

Opakování

- Co je to regresní analýza?

korelace (r)

- Symetrická metoda
(nerozlišujeme ZP a prediktor)
- Měří sílu i směr vztahu
- O kolik se v průměru mění hodnota proměnné v závislosti na změně hodnot druhé?

regrese

- Asymetrická metoda
(rozlišujeme ZP a prediktor/y)
- Měří sílu i směr vztahu a **umožňuje predikci hodnot ZP** v závislosti na hodnotách prediktoru/ů.
- Jak velký je vliv prediktoru/ů na ZP?
- Jaká bude hodnota ZP při určité hodnotě prediktoru/ů?

Opakování

- Co nám říká p-hodnota?

Opakování

Regresní model: lineární regresní funkce

Jednoduchý lineární regresní model:

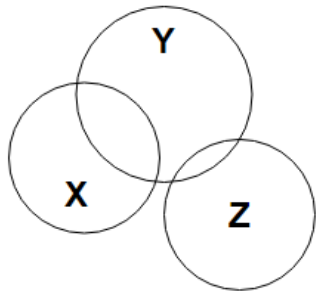
- $Y' = b_0 + b_1 * X$
- závislá proměnná = průsečík + sklon * prediktor

Vícečetný lineární regresní model:

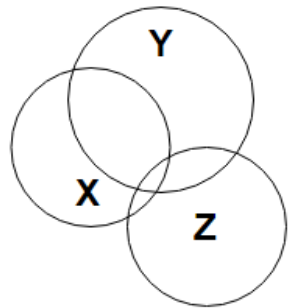
- $Y' = b_0 + b_1 * X + b_2 * Z$
- závislá proměnná = průsečík + sklon * prediktor X + sklon * prediktor Z

Někdy také:

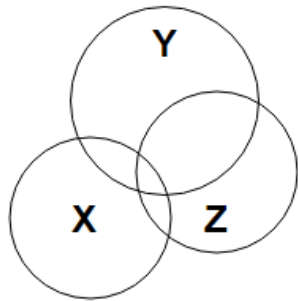
- $Y' = b_0 + b_1 * X_1 + b_2 * X_2$
- závislá proměnná = průsečík + sklon * prediktor X_1 + sklon * prediktor X_2



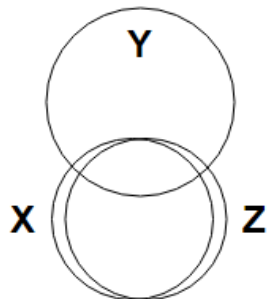
X a Z nekorelují → interpretace jako v jednoduché regresi
(X a Z jsou vzájemně nezávislé)



X koreluje s Y více než Z → účinek X je při kontrole vlivu Z
silnější než účinek Z při kontrole vlivu X



Z koreluje s Y více než X → účinek Z je při kontrole vlivu X
silnější než účinek X při kontrole vlivu Z



X a Z (téměř) dokonale korelují → hodnoty koeficientů se
budou blížit nekonečnu (neinterpretovatelné) →
problém **kolinearity**

- Příklad regrese spokojenosti s politickým systémem (**závislá proměnná Y**) v závislosti na socioekonomickém statusu (**X**) a důvěře v instituce (**Z**):

- $Y = 48.24 + 0.57*X + 0.67*Z$

	A	B	C
intercept	51.55* (0.86)	47.29* (1.94)	48.24* (1.64)
SES (X)	0.65* (0.16)	-	0.57* (0.15)
trust (Z)	-	0.92* (0.33)	0.67* (0.28)
R ²	0.36	0.20	0.46
Std. error in parentheses * = p < 0.05			

	A	B	C
intercept	51.55* (0.86)	47.29* (1.94)	48.24* (1.64)
SES (X)	0.65* (0.16)	-	0.57* (0.15)
důvěra (Z)	-	0.92* (0.33)	0.67* (0.28)
R2	0.36	0.20	0.46
* = p < 0.05; Std. error in parentheses			

Který prediktor má větší účinek (efekt) na ZP?

standardizovaný koeficient:

$$b_{\text{std}} = b * \frac{s(X)}{s(Y)}$$

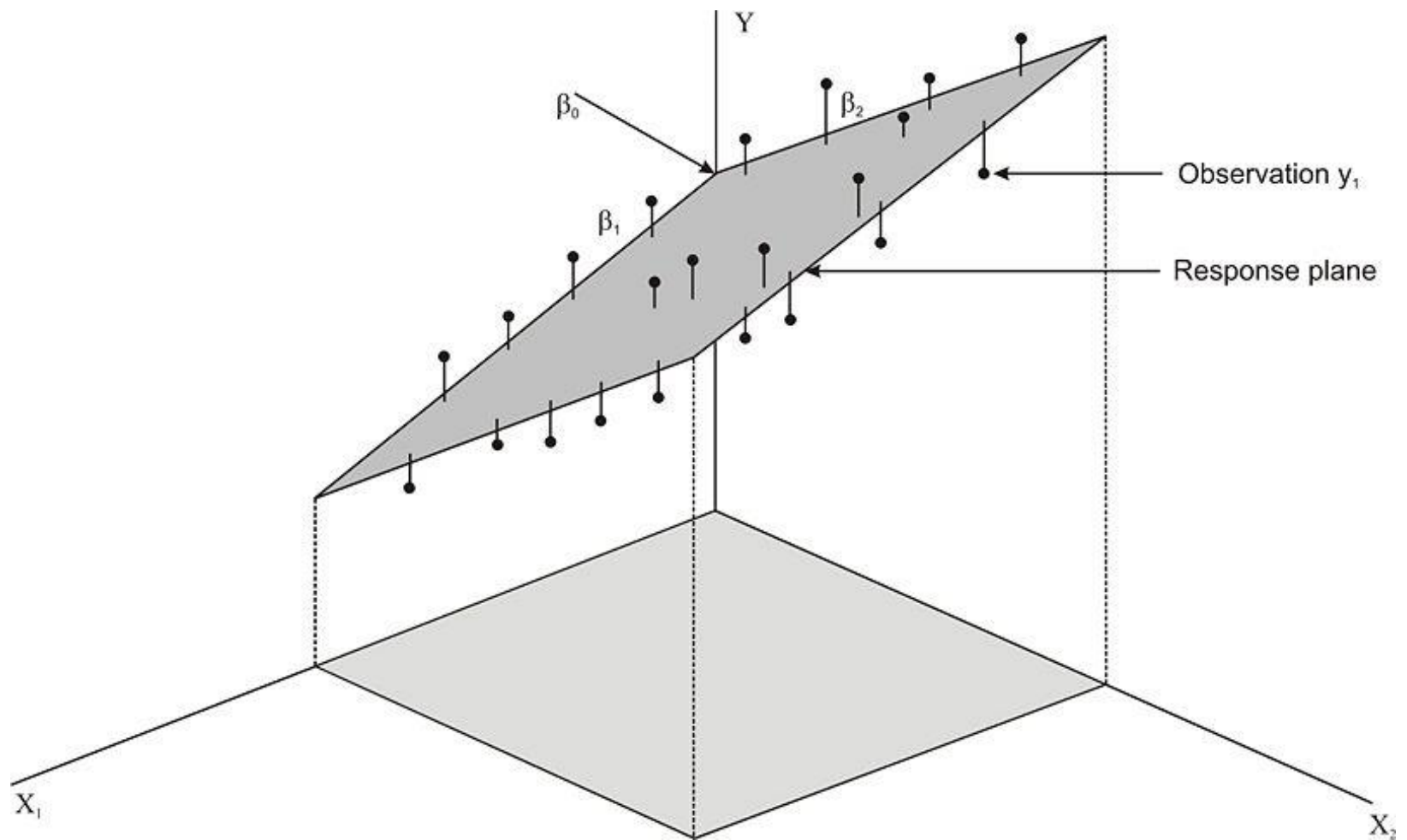
slabý: < 0.29

střední: 0.3-0.49

velký > 0.5

- Nutné zohlednit **jednotky, v nichž jsou proměnné měřeny!**
 - X (SES) = dlouhá škála; Z (důvěra) = krátká škála
- Pro porovnatelné hodnoty koeficientů **standardizujeme** skrze směrodatné odchylky ZP a daného prediktoru:
 - $b_{1\text{std}} = b_1 * (\text{std.dev}(X) / \text{std.dev}(Y)) = 0.57 * (5.496/6.017) = \mathbf{0.52 \text{ (SES)}}$
 → zvýšení hodnoty X o 1 std.dev. znamená zvýšení Y o 0.52 std.dev.
 - $b_{2\text{std}} = b_2 * (\text{std.dev}(Z) / \text{std.dev}(Y)) = 0.67 * (2.952/6.017) = \mathbf{0.33 \text{ (důvěra)}}$
 → zvýšení hodnoty Z o 1 std.dev. znamená zvýšení Y o 0.33 std.dev.
- Standardizované koeficienty se nazývají *beta*

Regresní rovina



- **Rozptyl vysvětlený** regresním model (**SSM**: sum of squares of model)

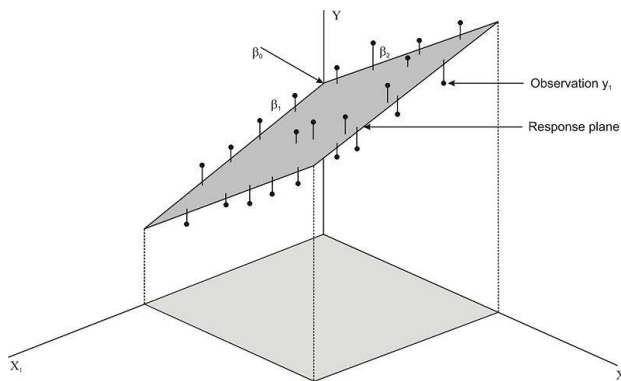
$$SSM = \Sigma(Y' - \text{mean}(Y))^2$$
- **Nevysvětlený rozptyl**: chyba modelu (**SSR**: sum of squares of residuals):

$$SSR = \Sigma(Y - Y')^2$$
- **Celkový rozptyl** (**SST**: total sum of squares) = SSM + SSR

$$SST = \Sigma(Y - \text{mean}(Y))^2$$

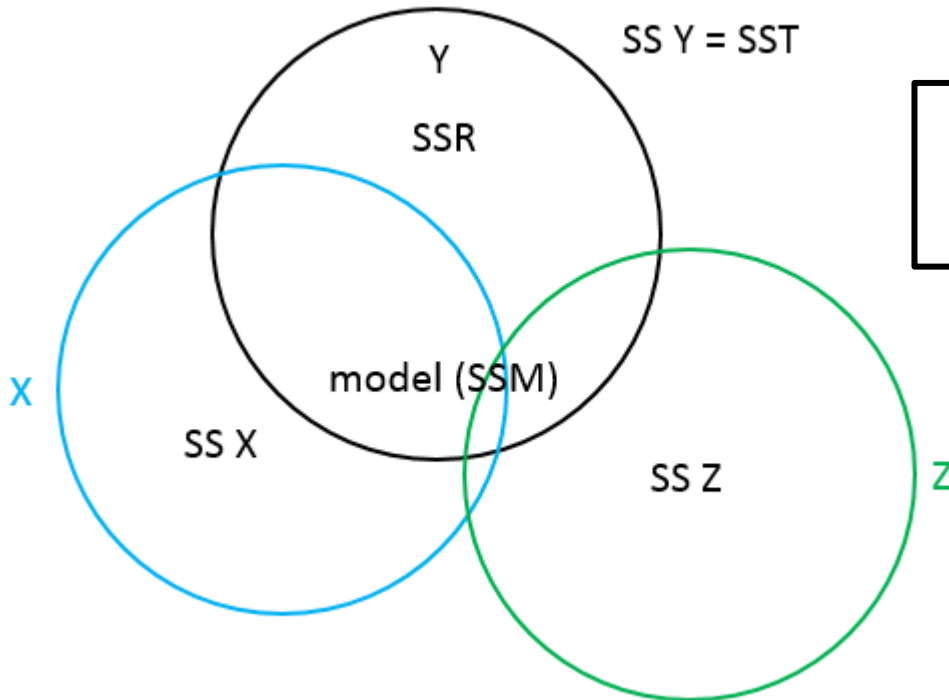
Y' = predikované hodnoty (rovina)
 Y = pozorované hodnoty (body)
 $\text{mean}(Y)$ = průměrná hodnota ZP

- Podíl **SSM/SST** ukazuje **(1)** těsnost regresního vztahu mezi ZP a prediktory a **(2)** přesnost predikce založené na regresním modelu (funkci)



Koeficient determinace

- KD (R^2) ukazuje podíl rozptylu Y vysvětleného regresním modelem (SSM) vůči celkovému rozptylu Y (SST) = SSM / SST
- $SST = SSM$ (vysvětlený rozptyl) + SSR (nevysvětlený rozptyl)

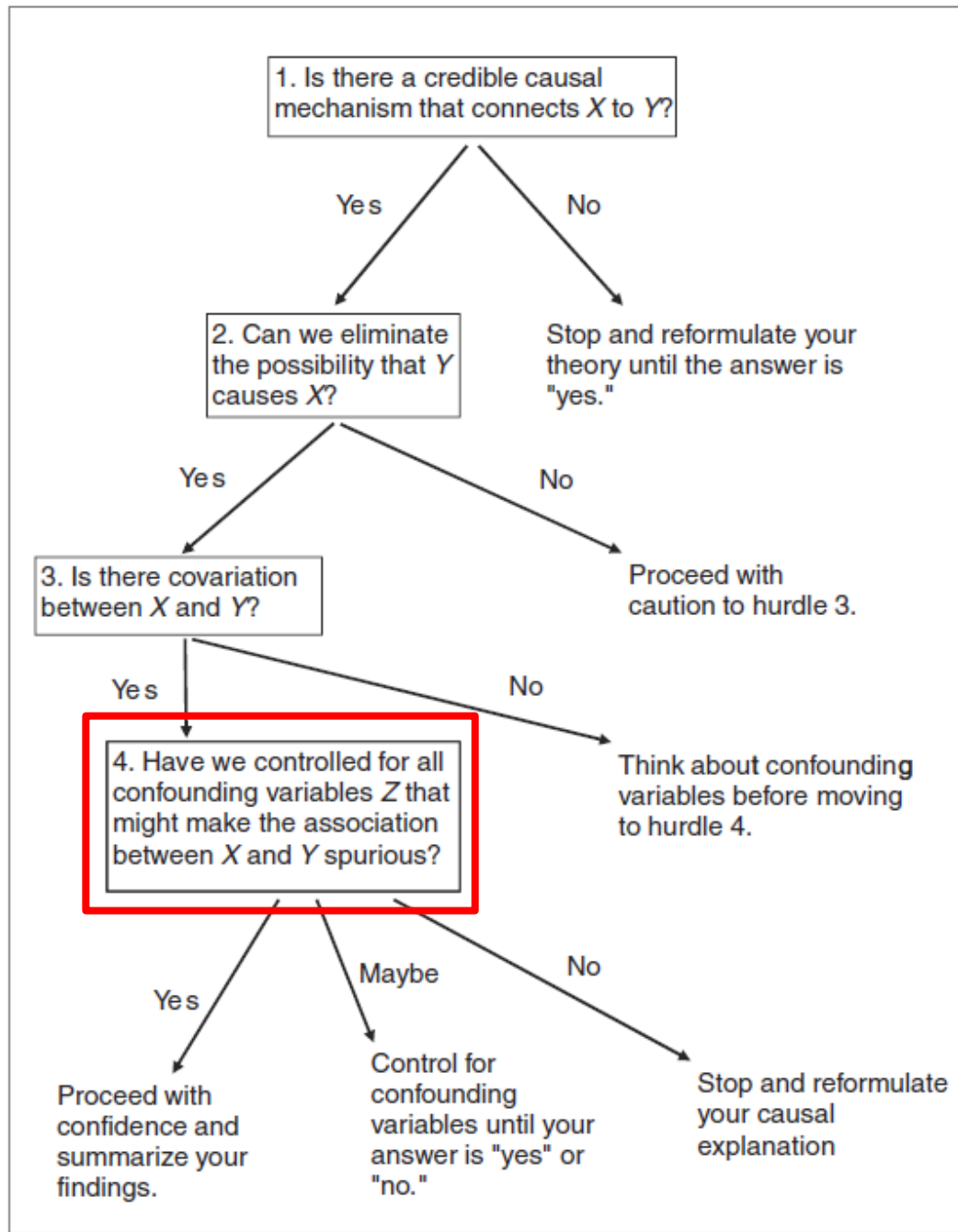


Pro vícečetnou regresi užíváme **upravené R^2** penalizující vyšší počet prediktorů

SSM	sum of squares of model
SSR	sum of squares of residuals
SST	sum of squares of total

Vícečetná lineární regrese

- Umožňuje predikci ZP v závislosti na hodnotách **více než jednoho prediktoru**
 - Jaká je predikovaná hodnota spokojenosti s pol. systémem (ZP Y), pokud bereme v potaz SES (prediktor X) a důvěru v instituce (prediktor Z)?
- Umožňuje odhadovat **dílčí (parciální) účinky (efekty) jednotlivých prediktorů** na ZP
 - Jak se změní hodnota spokojenosti s pol. systémem (ZP Y), pokud se hodnota prediktoru X změní o jednotku při kontrole prediktoru Z?
 - Jak se změní hodnota spokojenosti s pol. systémem (ZP Y), pokud se hodnota SES (X) změní o jednotku při kontrole důvěry (Z)?

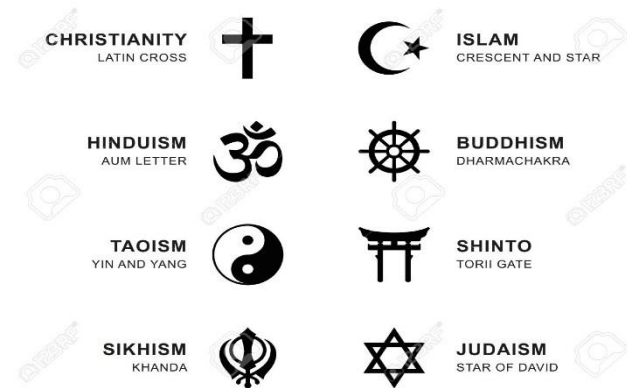


Interpretace účinků kategorických prediktorů

Kategorické proměnné

- **Nominální:** definovány pouze **kvalitativní odlišností**

- homogenita kategorií
- vzájemná výlučnost kategorií
- úplnost kategorií



- **Pořadové:** definovány **kvalitativní odlišností** a **pořadím kategorií**

- homogenita kategorií
- vzájemná výlučnost kategorií
- úplnost kategorií
- **pořadí kategorií**



Kategorické proměnné jako prediktory

- Kategorické proměnné (KP) jsou v regresním modelu reprezentovány tzv. **indikátorovými (dummy) proměnnými**
- **Indikátorová proměnná:** pouze hodnoty 0 nebo 1 indikující ne/přítomnost daného kategorického účinku
- Počet indikátorových proměnných = $k - 1$, kde k je počet kategorií proměnné



→ účinky KP jsou definovány $k - 1$ koeficienty

Kategorické proměnné jako prediktory

- Počet indikátorových proměnných = $k - 1$, kde k je počet kategorií proměnné
- Např.: **kulturně-geografický charakter (X)** = ne-Jih (0), Jih (1) jako prediktor **(Y) Obamova volebního výsledku (obama2012)**



$$Y = 50.049 - 5.918 * X_{\text{Jih}}$$

- **Jaká bude predikce pro ne-jihní státy a pro jihní státy?**

Kategorické proměnné jako prediktory

- Počet indikátorových proměnných = $k - 1$, kde k je počet kategorií proměnné
- Např.: **kulturně-geografický charakter (X)** = ne-Jih (0), Jih (1) jako prediktor **(Y) Obamova volebního výsledku (obama2012)**



$$Y = 50.049 - 5.918 * X$$

- **Jaká bude predikce pro ne-jihní státy a pro jihní státy?**
- Predikce pro **ne-jihní státy**: $Y = 50.049 - 5.918 * 0 = 50.049$
- Predikce pro **jihní státy**: $Y = 50.049 - 5.918 * 1 = 44.131$

Kategorické proměnné jako prediktory

- Např.: **kulturně-geografický charakter (X)** = ne-Jih (0), Jih (1) jako prediktor (**Y**) **Obamova volebního výsledku** (obama2012) s kontrolou pro (**Z**) **podíl městského obyvatelstva** (urban)

$$Y = 31.543 - 3.998 * X + 0.250 * Z$$

- Predikce pro **ne-jížní státy** s kontrolou pro **urban**:

$$Y = 31.543 - 3.998 * 0 + 0.250 * 1 = 31.793$$

- Predikce pro **jižní státy** s kontrolou pro **urban**:

$$Y = 31.543 - 3.998 * 1 + 0.250 * 1 = 27.795$$

Kategorické proměnné jako prediktory

$$Y = 31.543 - 3.998 * X + 0.250 * Z$$

- Predikce pro **ne-jížní státy** s kontrolou pro **urban**:

$$Y = 31.543 - 3.998 * 0 + 0.250 * 1 = 31.793$$

- Predikce pro **jížní státy** s kontrolou pro **urban**:

$$Y = 31.543 - 3.998 * 1 + 0.250 * 1 = 27.795$$

- **Dále: hodnoty kategorií mění hodnotu průsečíku, sklon se nemění:**

- Predikce pro **ne-jížní státy** s kontrolou pro **urban**:

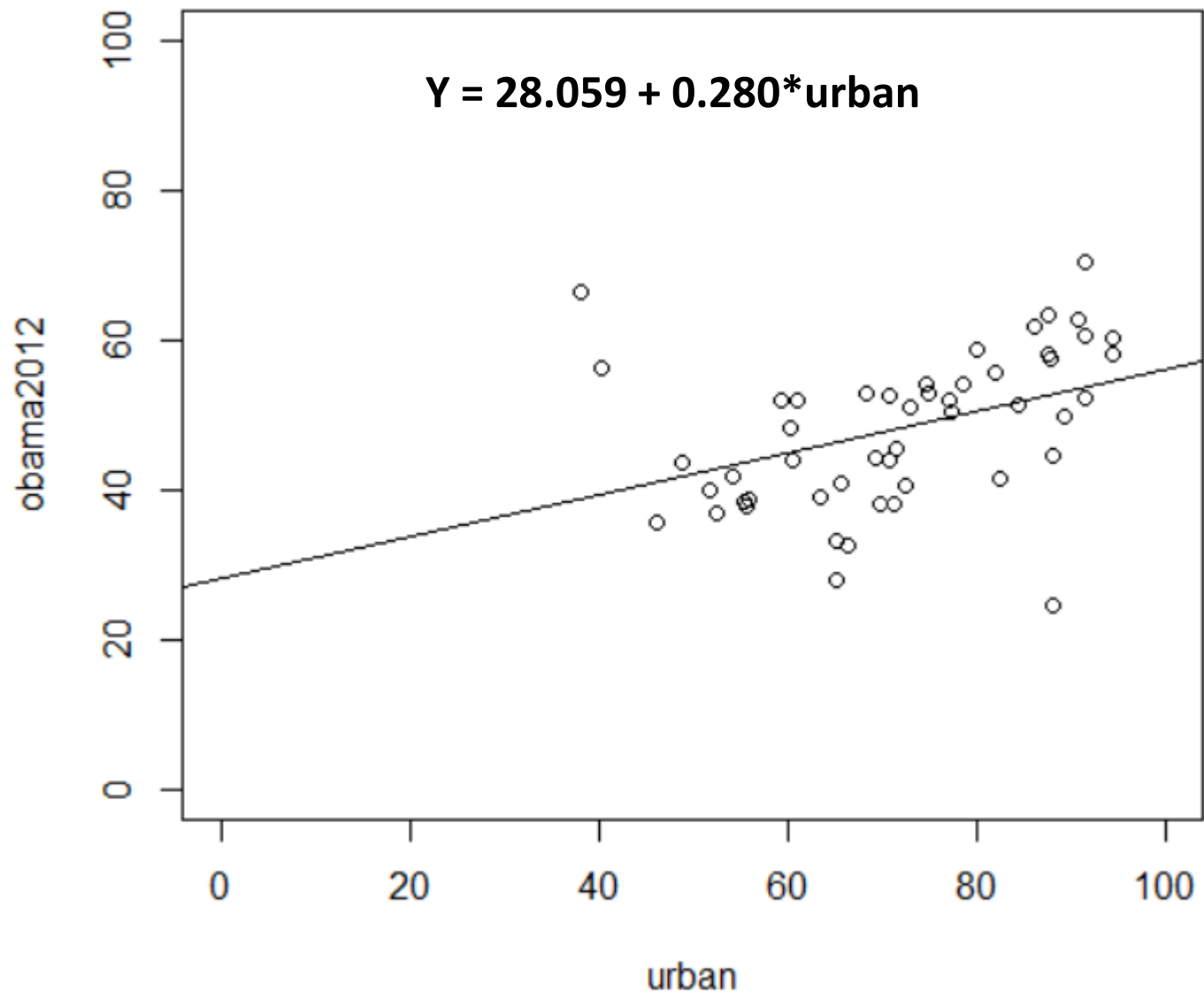
$$Y = 31.543 - \cancel{3.998 * 0} + 0.250 * 1 = 31.793$$

- Predikce pro **jížní státy** s kontrolou pro **urban**:

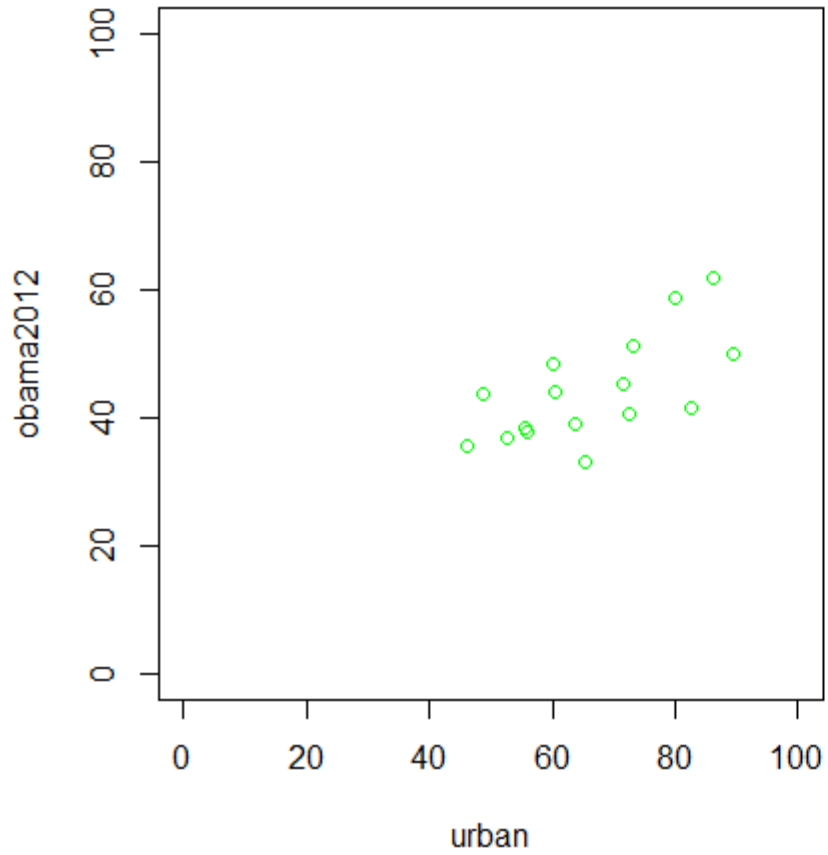
$$Y = 27.545 + 0.250 * 1 = 27.795$$



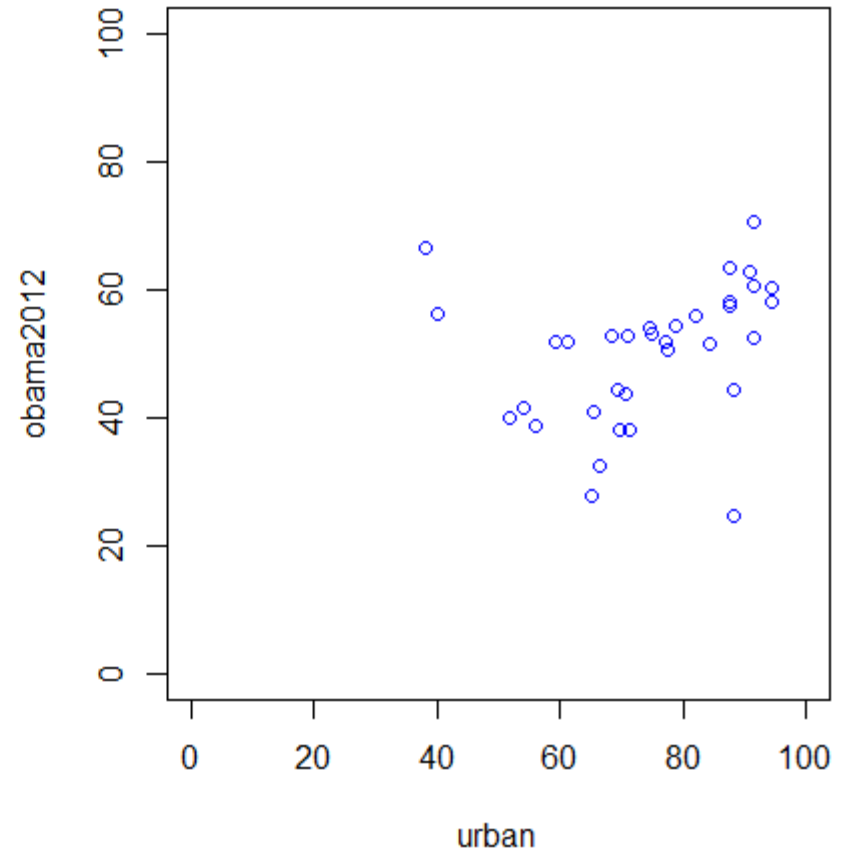
všechny státy



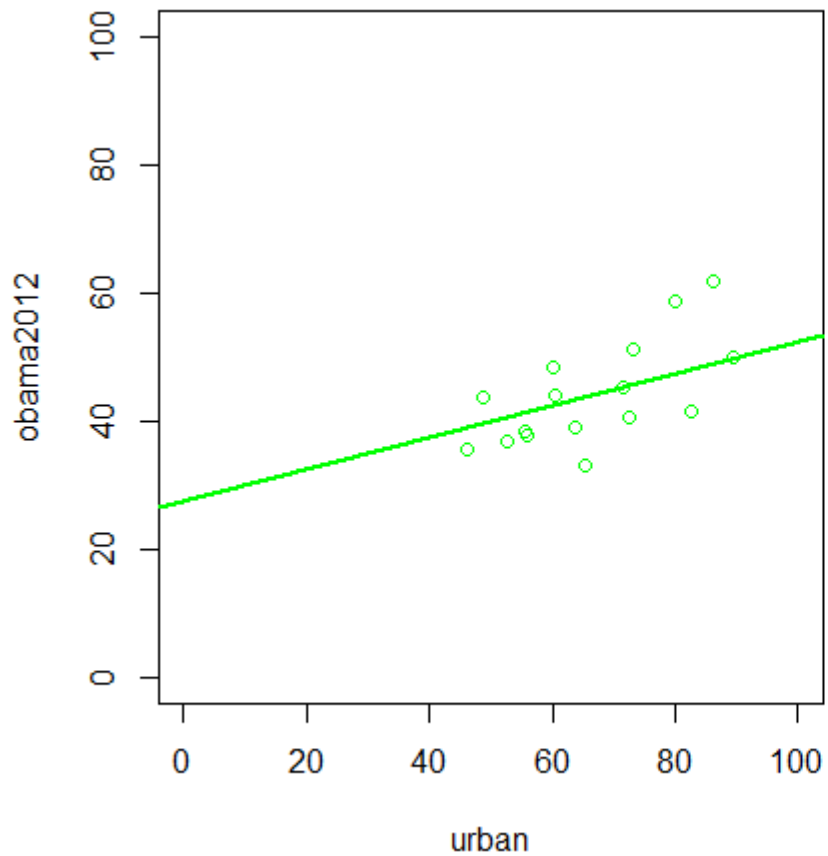
jižní státy



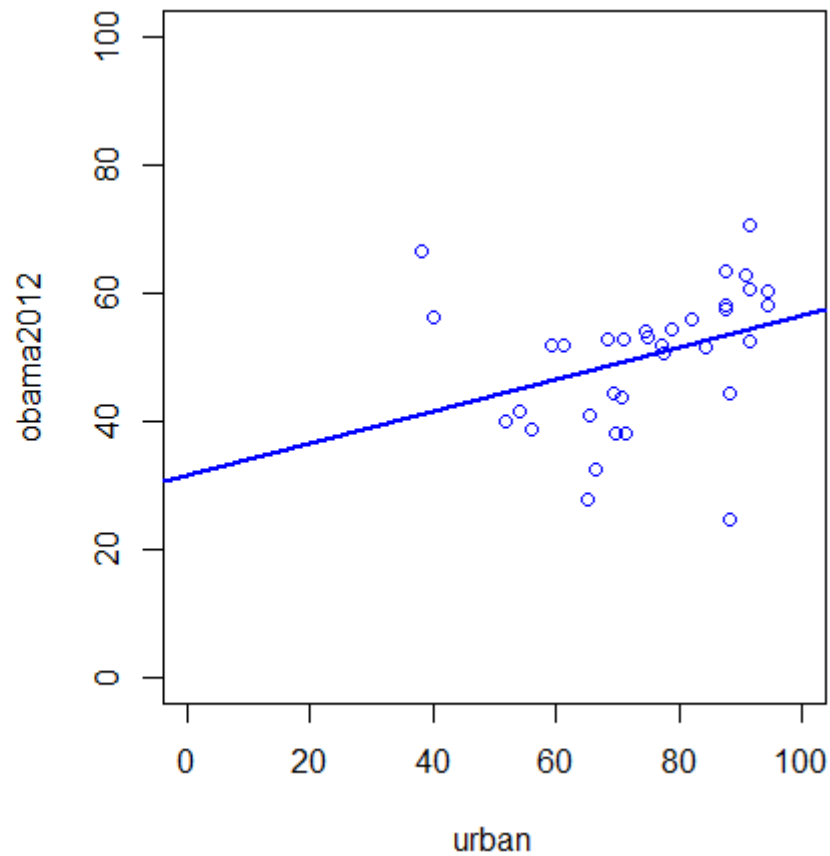
nejižní státy



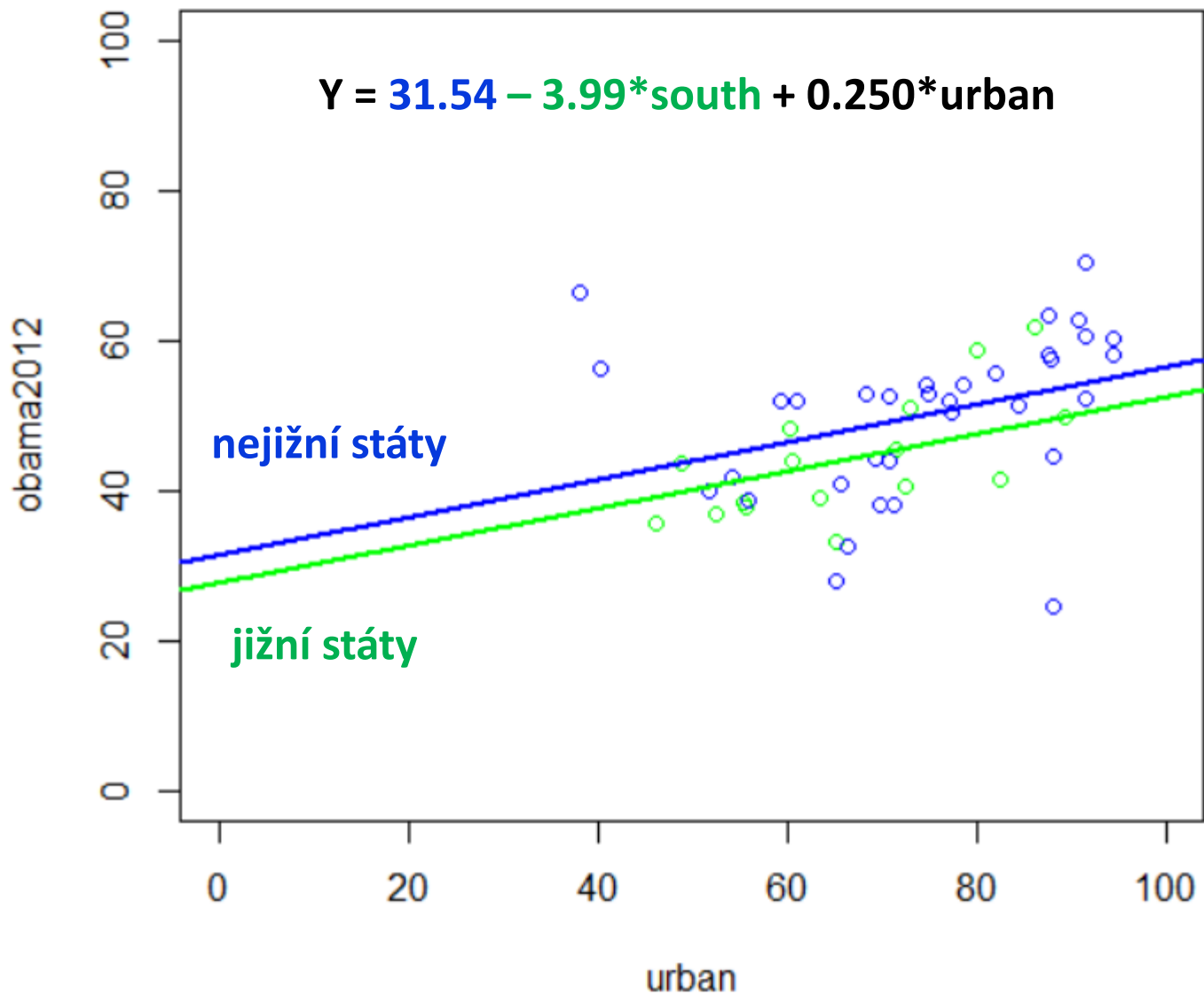
jižní státy



nejižní státy



Posun průsečíku: Jih = 27.55, ne-Jih = 31.54



Řazené kategorické proměnné

- Např.: **podpora práv gayů (X) = vysoká, střední, nízká**; s kontrolou pro (Z) podíl městského obyvatelstva (urban)
- Konstrukce indikátorových proměnných: **k - 1**
 - Referenční kategorie (průsečík) = **vysoká podpora** (nízká i střední podpora = 0)
 - Indikátorová proměnná 1 = **střední podpora** (nízká podpora = 0)
 - Indikátorová proměnná 2 = **nízká podpora** (střední podpora = 0)



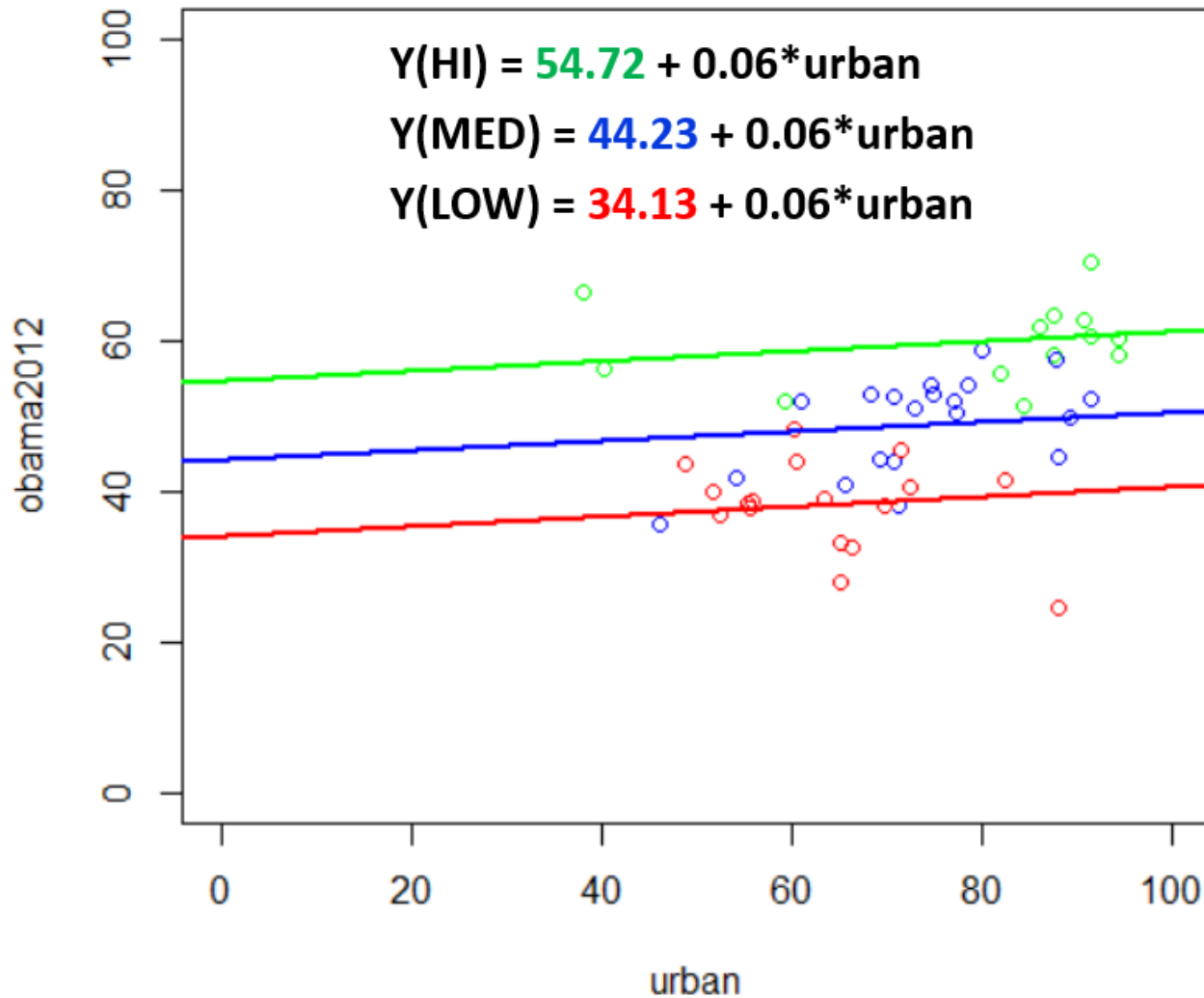
$$Y = 54.72 - 20.59 * \text{gay_supp_LOW} - 10.49 * \text{gay_supp_MED} + 0.06 * \text{urban}$$

$$Y(\text{HI}) = 54.72 + 0.06 * \text{urban}$$

$$Y(\text{MED}) = 44.23 + 0.06 * \text{urban} ; 54.72 - 10.49 = 44.23$$

$$Y(\text{LOW}) = 34.13 + 0.06 * \text{urban} ; 54.72 - 20.59 = 34.13$$

$$Y = 54.72 - 20.59 * \text{gay_supp_LOW} - 10.49 * \text{gay_supp_MED} + 0.06 * \text{urban}$$



Shrnutí

- Interpretace účinků spojitých a kategorických prediktorů se liší
- Spojité prediktory: **jednotková změna** hodnoty prediktoru **definuje sklon** regresní přímky
- Kategorické prediktory: **změna úrovně (kategorie)** prediktoru **definuje průsečík** regresní přímky
- V případě kategorických prediktorů vždy stanovujeme **referenční kategorii**
- **Počet indikátorových proměnných** zachycujících ostatní kategorie prediktoru = **# kategorií (k) - 1**

Předpoklady

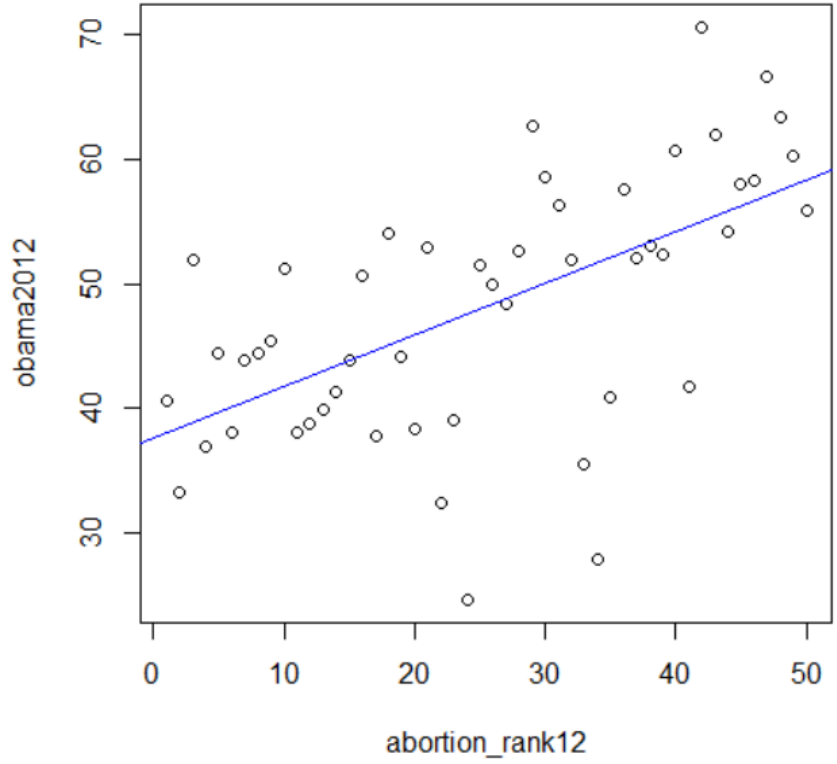
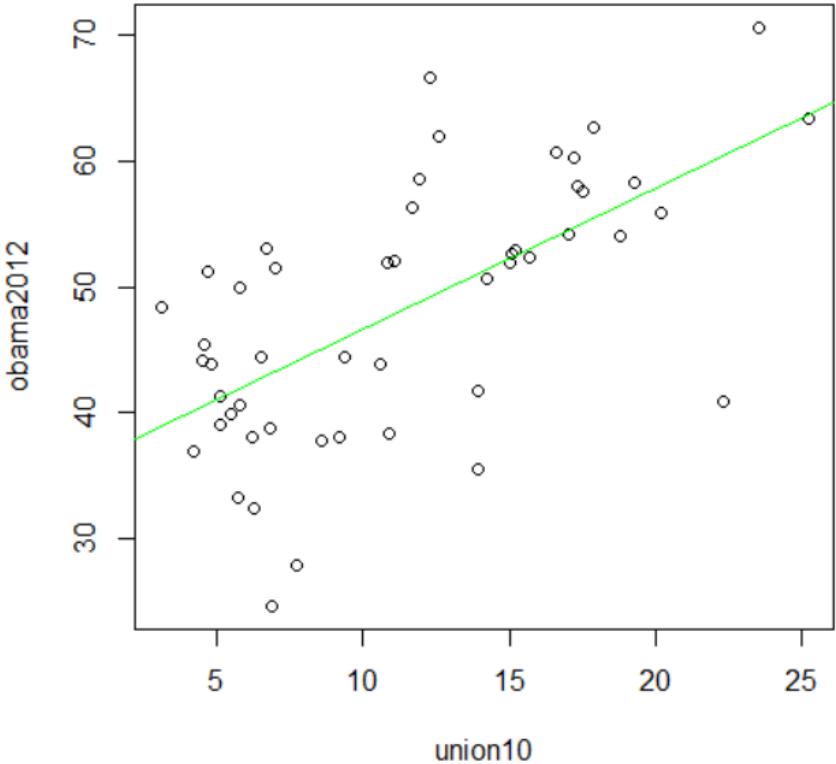
Předpoklady

- Spojitá závislá proměnná
- Nenulový rozptyl prediktorů
- Nezávislost pozorování (náhodný výběr)
- Lineární vztah mezi prediktory a závislou proměnnou
- Normální rozdělení reziduí
- Homoskedasticita (nezávislost rozptylu ZP)
- *Kolinearita prediktorů není vysoká*
- *Absence vlivných pozorování*

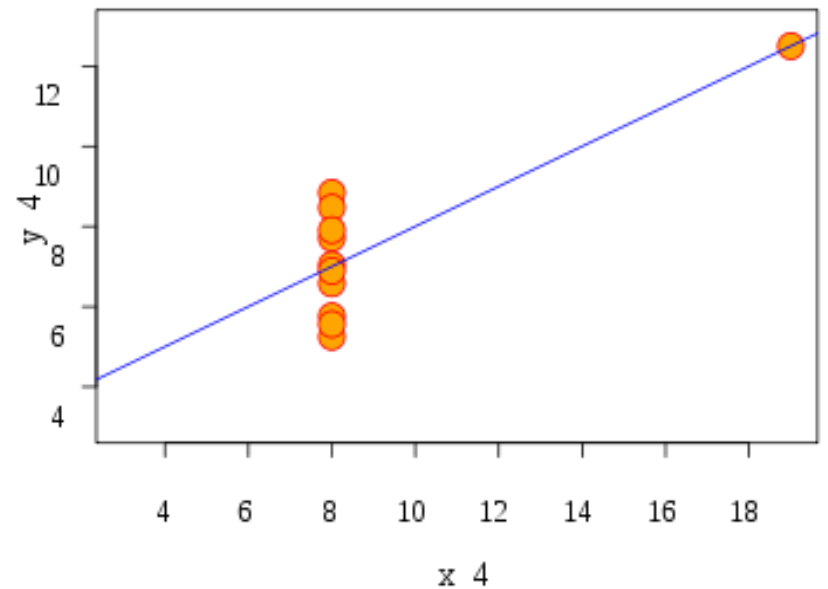
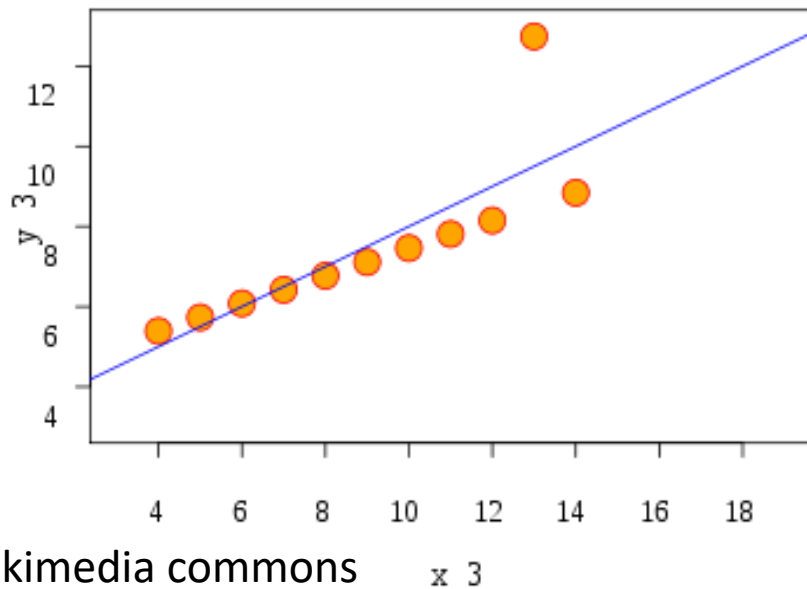
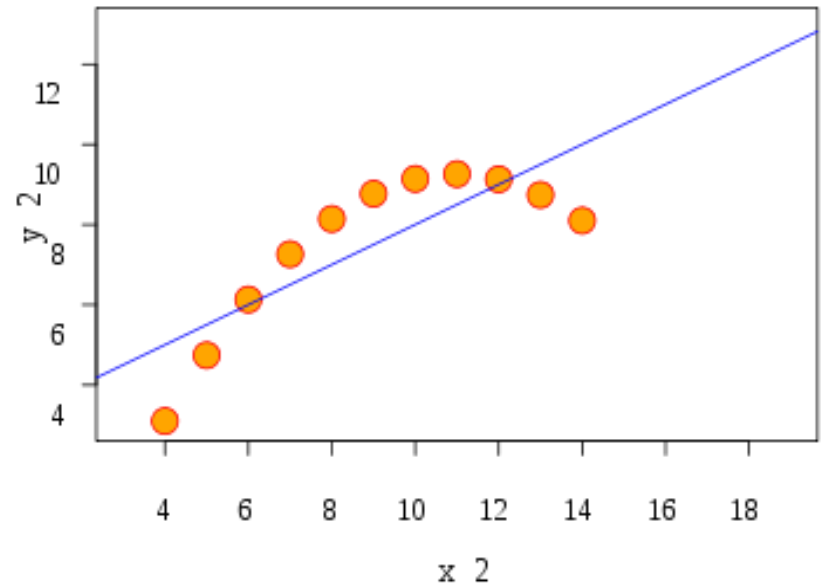
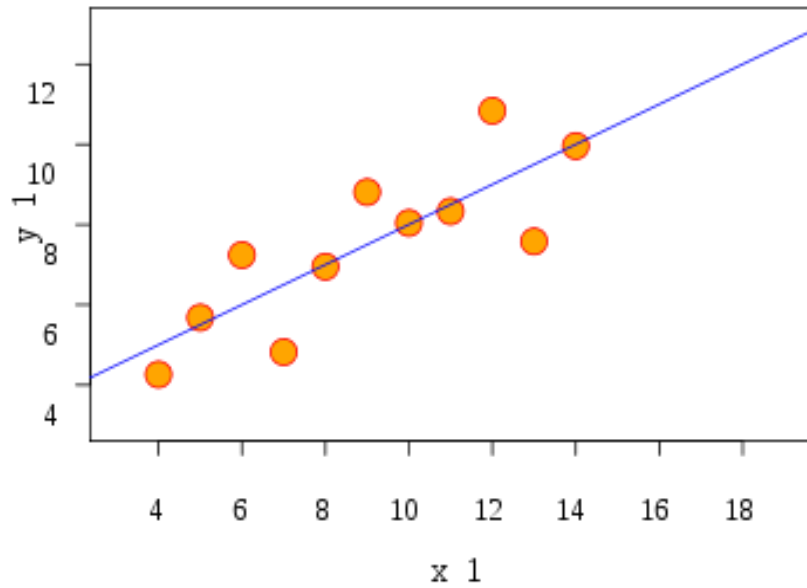
Předpoklady

- Spojitá závislá proměnná
- Nenulový rozptyl prediktorů
- Nezávislost pozorování (náhodný výběr)
- **Lineární vztah mezi prediktory a závislou proměnnou**
- Normální rozdělení reziduí
- Homoskedasticita (nezávislost rozptylu)
- *Kolinearita prediktorů není vysoká*
- *Absence vlivných pozorování*

Lineární vztah X and Y



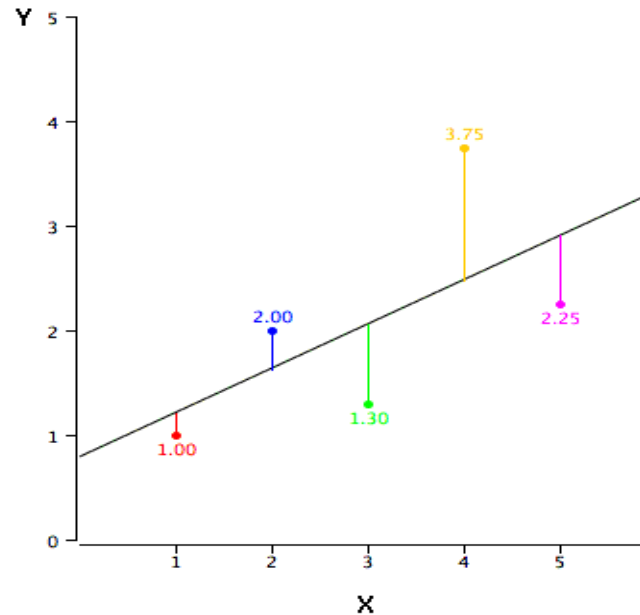
Anscombe's quartet



Předpoklady

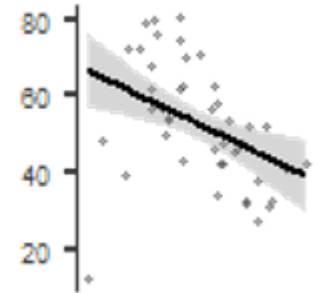
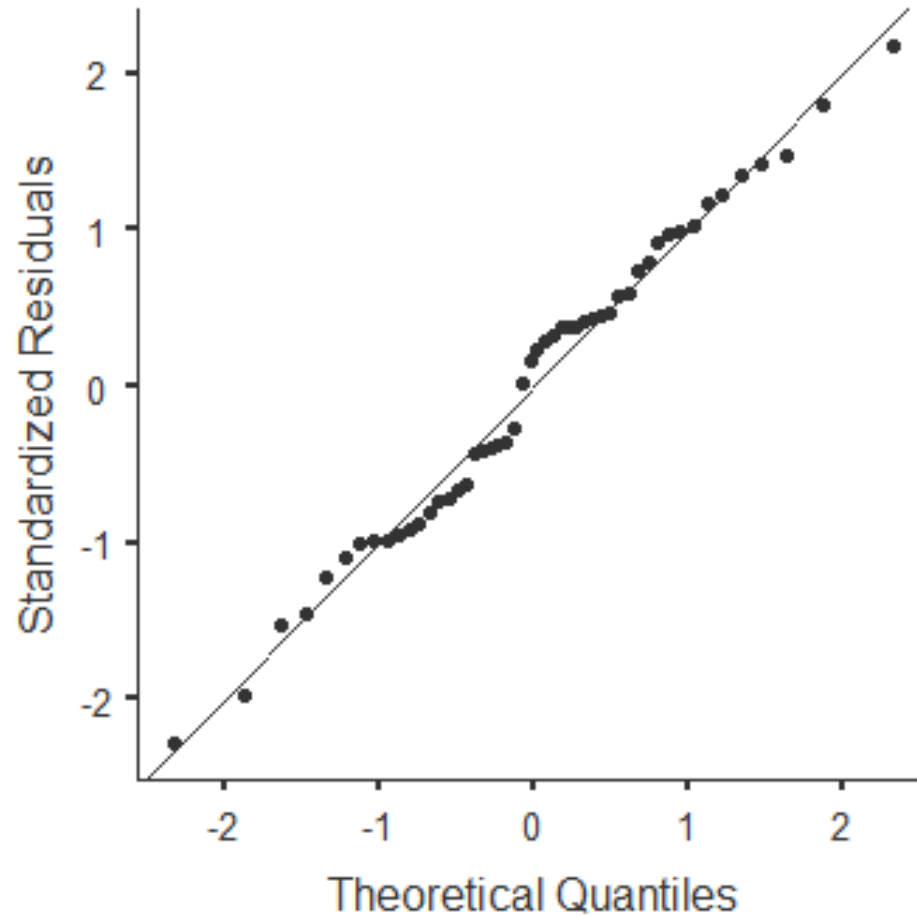
- Spojitá závislá proměnná
- Nenulový rozptyl prediktorů
- Nezávislost pozorování (náhodný výběr)
- Lineární vztah mezi prediktory a závislou proměnnou
- **Normální rozdělení reziduí**
- Homoskedasticita (nezávislost rozptylu)
- *Kolinearita prediktorů není vysoká*
- *Absence vlivných pozorování*

Rezidua



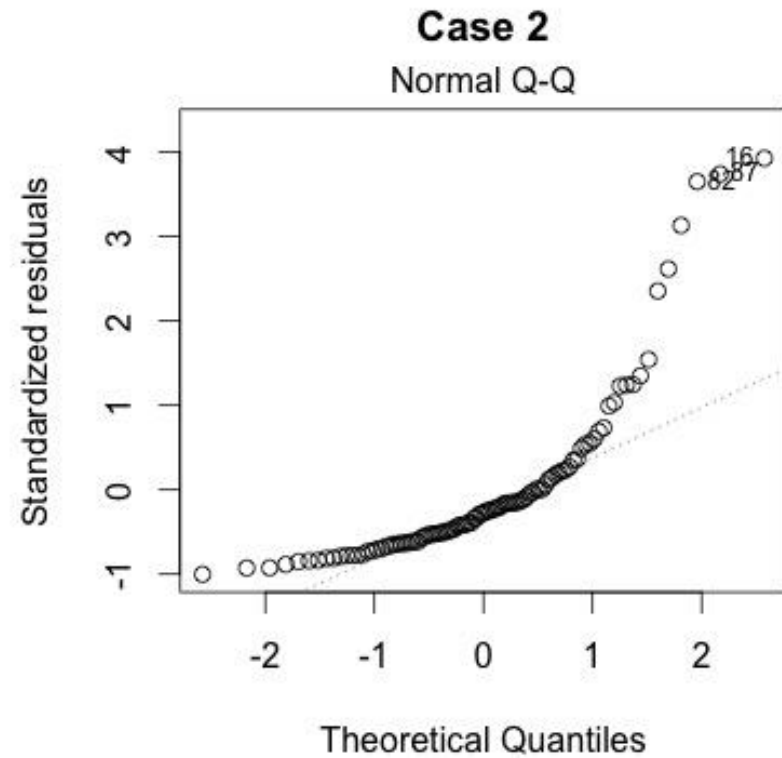
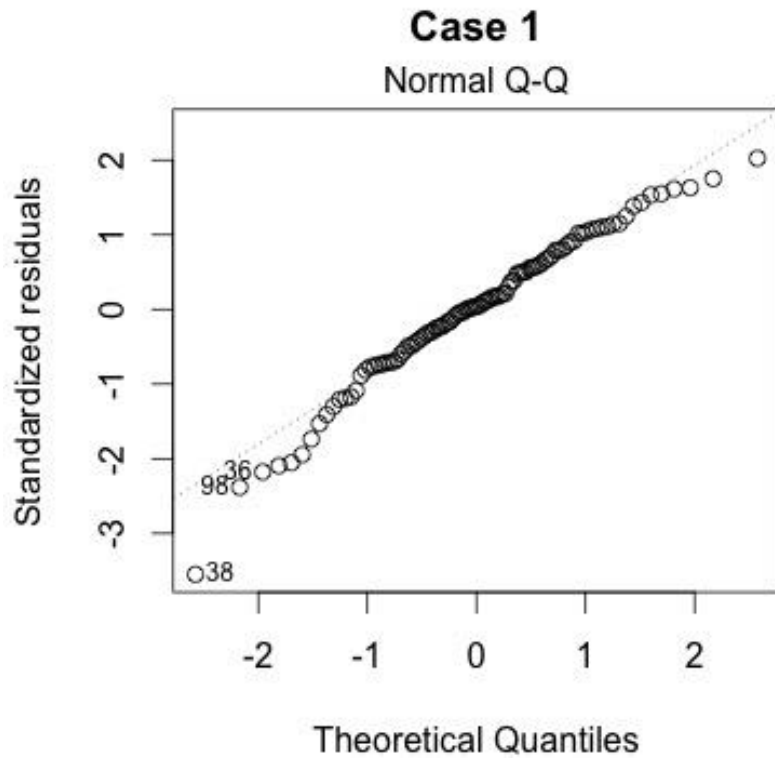
- **Rezidua:** rozdíl mezi predikovanou a pozorovanou hodnotou ZP (Y) → vyjádřena ve škále ZP
- **Standardizovaná** rezidua: jsou dělena **jejich** směrodatnou odchylkou → nulová střední hodnota a jednotkový rozptyl
- **Studentizovaná** rezidua: jsou dělena **jejich** směrodatnou odchylkou nezahrnující i-té pozorování → při normálním rozdělení reziduí sledují Studentovo t-rozdělení

Kvantilový graf (Q-Q plot)



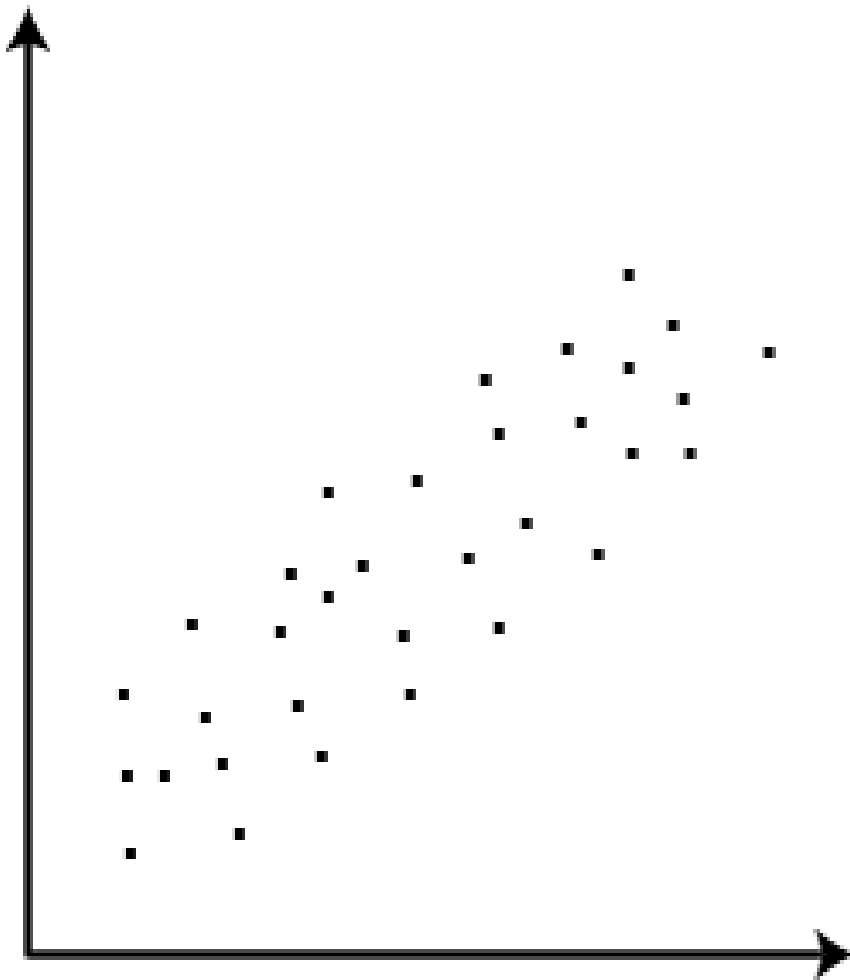
relig_prot x
obama2012

Kvantilový graf

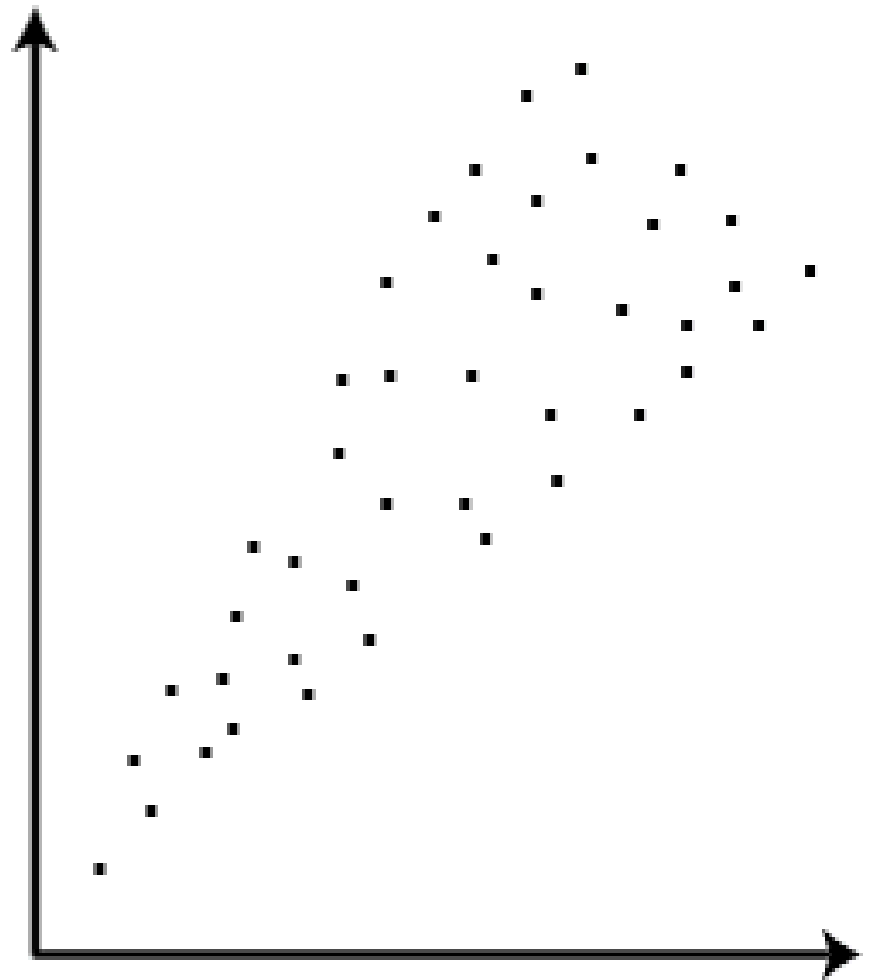


Předpoklady

- Spojitá závislá proměnná
- Nenulový rozptyl prediktorů
- Nezávislost pozorování (náhodný výběr)
- Lineární vztah mezi prediktory a závislou proměnnou
- Normální rozdělení reziduí
- **Homoskedasticita (nezávislost rozptylu ZP)**
- *Kolinearita prediktorů není vysoká*
- *Absence vlivných pozorování*



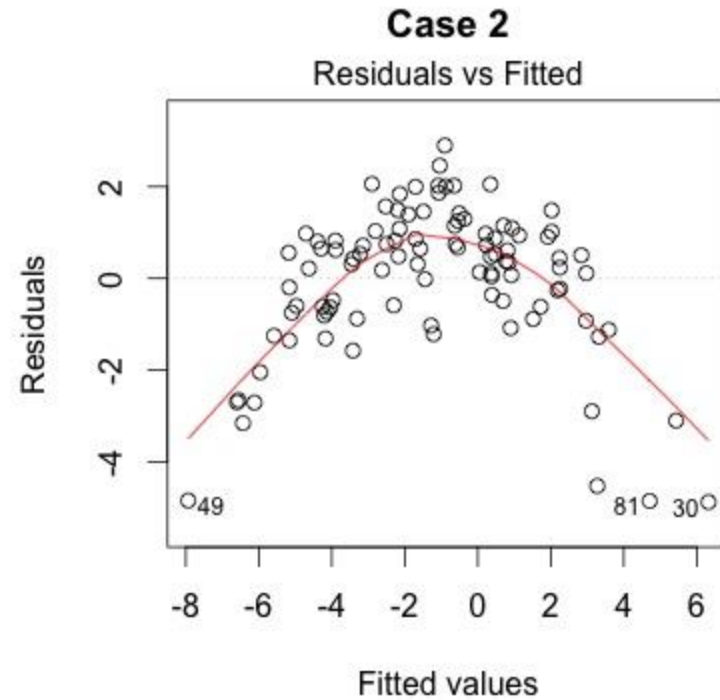
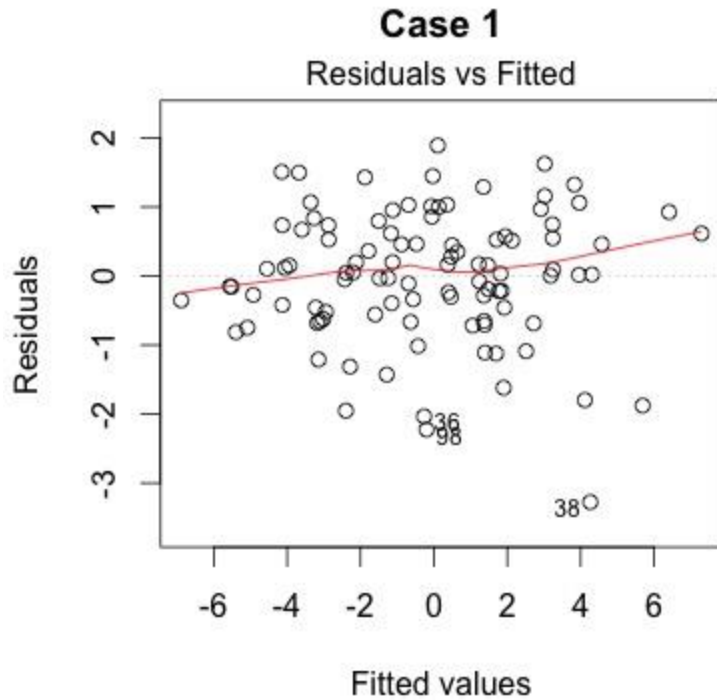
Homoscedasticity



Heteroscedasticity



Reziduální graf

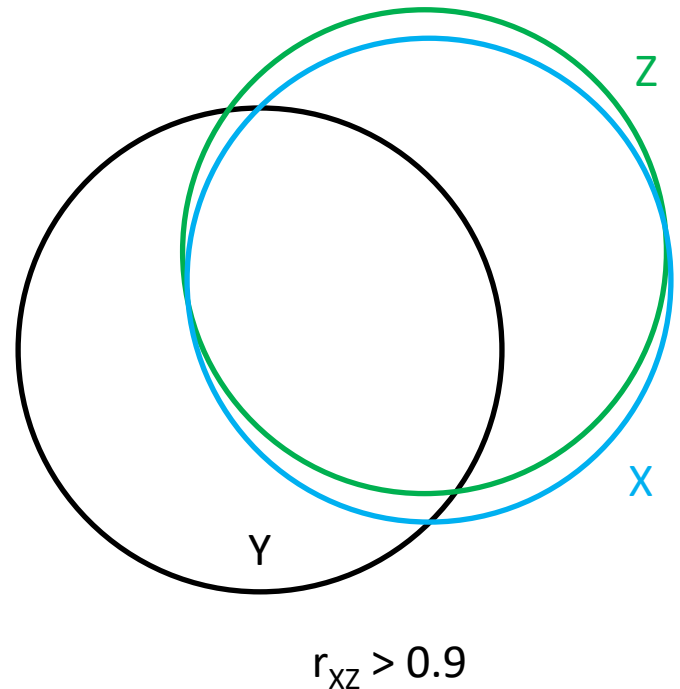
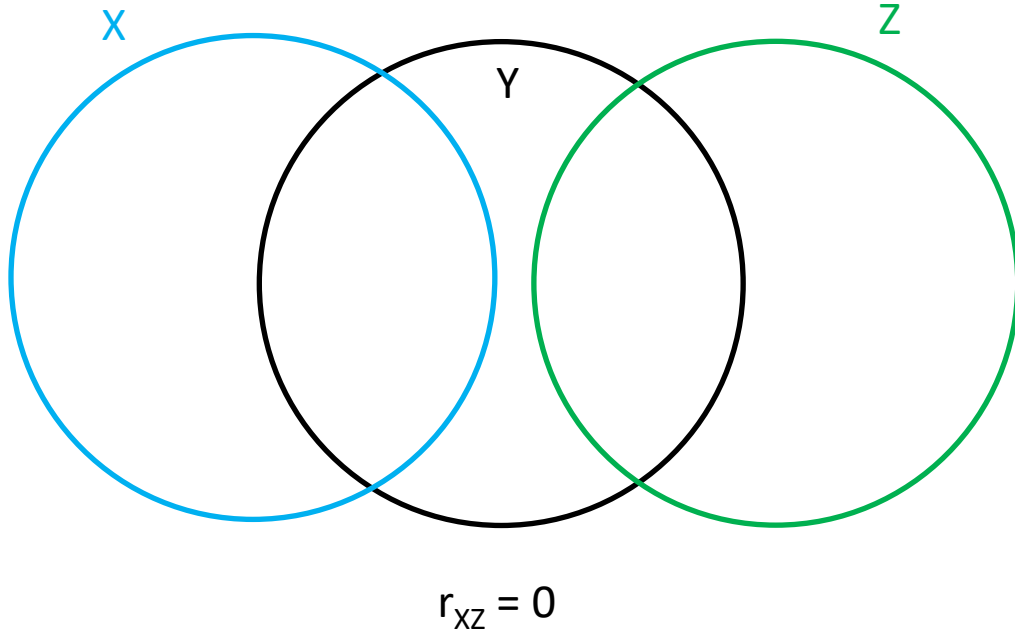


Předpoklady

- Spojitá závislá proměnná
- Nenulový rozptyl prediktorů
- Nezávislost pozorování (náhodný výběr)
- Lineární vztah mezi prediktory a závislou proměnnou
- Normální rozdělení reziduí
- Homoskedasticita (nezávislost rozptylu)
- ***Kolinearita prediktorů není vysoká***
- *Absence vlivných pozorování*

Kolinearita

- Kolinearita (multikolinearita) = dva nebo více prediktorů je silně korelováno



- Před odhadem regresního modelu: **korelační matice** prediktorů
- Po odhadu regresního modelu: **diagnostika VIF a tolerance**

Kolinearita

- Před odhadem regresního modelu: **korelační matice** prediktorů

Correlation Matrix

		obama2012	relig_prot	ba_or_more	union10
obama2012	Pearson's r	—			
	p-value	—			
relig_prot	Pearson's r	-0.413	—		
	p-value	0.003	—		
ba_or_more	Pearson's r	0.630	-0.628	—	
	p-value	< .001	< .001	—	
union10	Pearson's r	0.630	-0.507	0.400	—
	p-value	< .001	< .001	0.004	—

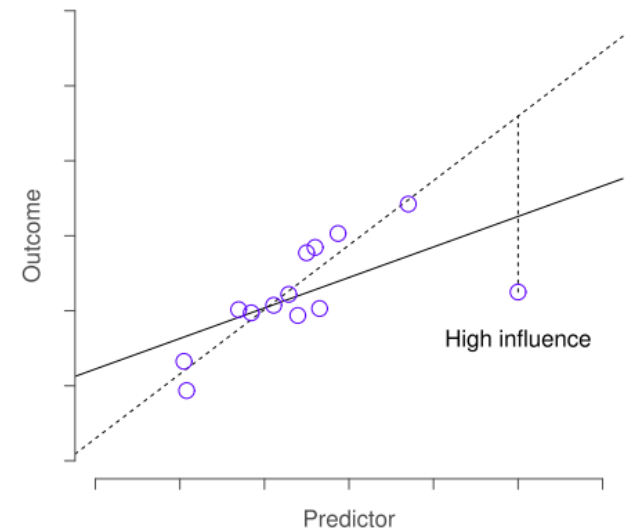
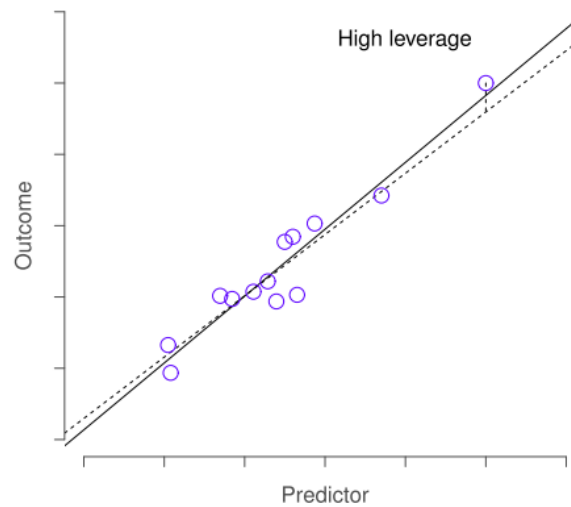
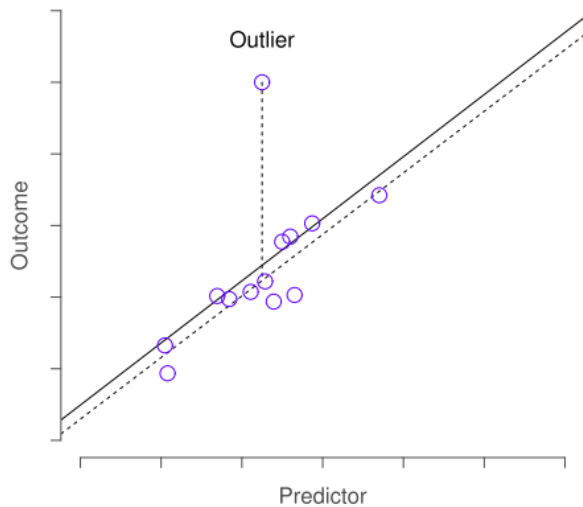
Kolinearita

- **Tolerance:** kolik rozptylu daného prediktoru je vysvětleno ostatními prediktory?
 - Regresní model, kde je prediktor z původního modelu vysvětlován zbývajícím prediktory
 - **Tolerance = $1 - R^2$** , tj. **nevysvětlený rozptyl daného prediktoru**
 - Hrubé pravidlo: tolerance < 0.1 (někdy 0.2) indikuje závažnou kolinearitu
- **Inflační faktor rozptylu (VIF):** je obrácenou hodnotou tolerance, tj. $VIF = 1/\text{tolerance}$
 - Tj. hodnota vyšší než $1/0.1 = 10$ indikuje závažnou kolinearitu

	VIF	Tolerance
prcapinc	1.70	0.587
relig_high	1.70	0.587

Vlivná pozorování

- Pozorování s **velkým vlivem na výsledky** regresního modelu



Vlivná pozorování: Cookova vzdálenost

- Pozorování s **velkým vlivem na výsledky** regresního modelu
- **Cookova vzdálenost:**
 1. výpočet regresního modelu bez konkrétního pozorování
 2. **rozdíl** mezi predikovanou hodnotou s pozorováním a bez pozorování
 3. výpočet proveden pro všechna pozorování
- Hrubé pravidlo: **Cookova vzdálenost > 1** indikuje vlivné pozorování

Cook's Distance

Mean	Median	SD	Range	
			Min	Max
0.0494	0.00657	0.219	3.34e-7	1.55

Závěrem

- Lineární regresi využíváme pro **predikci hodnot spojité proměnné (ZP)** v závislosti na hodnotách jednoho či více prediktorů
- Zpravidla nás zajímá, do jaké míry **specifický prediktor ovlivňuje hodnotu ZP při kontrole ostatních prediktorů** → **vícečetná regrese**
- Vícečetná regrese **umožňuje testování kauzálních vztahů** při předpokladu úspěšné kontroly ostatních nezbytných podmínek
- Jako každá metoda, lineární regrese spočívá na řadě **předpokladů**, které je nutné kontrolovat



	abort_ran...	abortion_...	or_more	cig_tax12	cig_tax1	
1	Less restr	35	26.6	2.000	HiTax	
2	Mid	20	22.0	0.425	LoTax	
3	More restr	4	18.9	1.150	MidTax	
4	More restr	5	25.6	2.000	HiTax	
5	Less restr	49	29.9	0.870	MidTax	
6	Mid	25	35.9	0.840	MidTax	
7	Less restr	45	35.6	3.400	HiTax	
8	Mid	30	28.7	1.600	MidTax	
9	Mid	26	25.3	1.339	MidTax	
10	More restr	9	9.9	27.5	0.370	LoTax
11	Less restr	42	9.9	29.6	3.200	HiTax
12	Less restr	37	7.4	25.1	1.360	MidTax
13	Mid	22	7.5	23.9	0.570	LoTax
14	Less restr	36	11.7	30.6	1.980	HiTax
15	More restr	7	8.1	22.5	0.995	MidTax
16	More restr	11	10.2	29.5	0.790	LoTax
17	More restr	17	8.5	21.0	0.600	LoTax
18	More restr	1	6.9	21.4	0.360	LoTax
19	Less restr	40	16.4	38.2	2.510	HiTax
20	Less restr	43	16.0	35.7	2.000	HiTax
21	Mid	31	9.6	26.9	2.000	HiTax
22	Mid	18	9.4	24.6	2.000	HiTax
23	Mid	28	10.3	31.5	1.600	MidTax
24	More restr	8	9.5	25.2	0.170	LoTax
25	More restr	15	7.1	19.6	0.680	LoTax
26	Less restr	41	8.3	27.4	1.700	MidTax
27	Mid	27	8.8	26.5	0.450	LoTax
28	More restr	12	6.7	25.8	0.440	LoTax
29	More restr	6	8.8	27.4	0.640	LoTax
30	Mid	32	11.2	32.0	1.680	MidTax
31	Less restr	46	12.9	34.5	2.700	HiTax
32	Less restr	38	10.4	25.3	1.660	MidTax
33	Less restr	39	7.6	21.8	0.800	LoTax

Correlation Matrix

Linear Regression

Logistic Regression

2 Outcomes

Binomial

N Outcomes

Multinomial

Ordinal Outcomes



Linear Regression



a gay_pomcyz
 a gay_policy_con
 a gay_support
 a gay_support3
 a gb_win00
 a gb_win04
 gore00
 gun_dealer
 gun_murder10
 gun_rank_rev
 gunlaw_rank
 gunlaw_scale
 hispanic04
 hispanic08

Dependent Variable
 obama2012

Covariates
 ba_or_more

Factors
 a gunlaw_rank3_rev

> Model Builder

> Reference Levels

> Assumption Checks

▾ Model Fit

Fit Measures

- R
 R²
 Adjusted R²

- AIC
 BIC
 RMSE

> Model Coefficients

> Estimated Marginal Means

Overall Model Test

- F test

Linear Regression

Model Fit Measures

Model	R	R ²	Adjusted R ²	Overall Model Test			
				F	df1	df2	p
1	0.744	0.553	0.524	19.0	3	46	< .001

Model Coefficients - obama2012

Predictor	Estimate	SE	t	p
Intercept	31.305	7.780	4.02	< .001
ba_or_more	0.606	0.286	2.12	0.040
gunlaw_rank3_rev:				
Mid – Fewer restr	4.460	2.645	1.69	0.099
More restr – Fewer restr	12.967	3.289	3.94	< .001

References

- [1] The jamovi project (2019). *jamovi*. (Version 1.0) [Computer Software]. Retrieved from <https://www.jamovi.org>.
- [2] R Core Team (2018). *R: A Language and environment for statistical computing*. [Computer software]. Retrieved from <https://cran.r-project.org/>.

Linear Regression



Dependent Variable
 → obama2012

Covariates
 → ba_or_more

Factors
 → gunlaw_rank3_rev

Variable
 gunlaw_rank3_rev

Reference Level
 Fewer restr

Linear Regression

Model Fit Measures

Model	R	R ²	Adjusted R ²	Overall Model Test			
				F	df1	df2	p
1	0.744	0.553	0.524	19.0	3	46	< .001

Model Coefficients - obama2012

Predictor	Estimate	SE	t	p
Intercept	31.305	7.780	4.02	< .001
ba_or_more	0.606	0.286	2.12	0.040
gunlaw_rank3_rev:				
Mid - Fewer restr	4.460	2.645	1.69	0.099
More restr - Fewer restr	12.967	3.289	3.94	< .001

reference: fewer restrictive = 31.305
mid restrictive = 31.305 + 4.460
more restrictive = 31.305 + 12.967

References

- [1] The jamovi project (2019). *jamovi*. (Version 1.0) [Computer Software]. Retrieved from <https://www.jamovi.org>.
- [2] R Core Team (2018). *R: A Language and environment for statistical computing*. [Computer software]. Retrieved from <https://cran.r-project.org/>.

> Assumption Checks

> Model Fit

> Model Coefficients

> Estimated Marginal Means



Linear Regression



Dependent Variable

→ obama2012

Covariates

←

Factors

→ south

Variables in list:

- permit
- pop_18_24
- pop_18_24_10
- prcapinc
- region
- relig_import
- religiosity
- reppct_m
- rtw
- secularism
- secularism3
- seniority_sen2
- to_0004

> Model Builder

∨ Reference Levels

Variable	Reference Level
south	Nonsouth

> Assumption Checks

> Model Fit

> Model Coefficients

> Estimated Marginal Means

Linear Regression

Model Fit Measures

Model	R	R ²
1	0.271	0.0735

Model Coefficients - obama2012

Predictor	Estimate	SE	t	p
Intercept	47.09	1.52	31.05	< .001
south:				
South - Nonsouth	-5.92	3.03	-1.95	0.057

Linear Regression

Model Fit Measures

Model	R	R ²
1	0.271	0.0735

Model Coefficients - obama2012

Predictor	Estimate	SE	t	p
Intercept	47.09	1.52	31.05	< .001
south:				
Nonsouth - South	5.92	3.03	1.95	0.057

References

- [1] The jamovi project (2019). *Jamovi*. (Version 1.0) [Computer Software]. Retrieved from <https://www.jamovi.org>.
- [2] R Core Team (2018). *R: A Language and environment for statistical computing*. [Computer software]. Retrieved from <https://cran.r-project.org/>.

Linear Regression



permit
 pop_18_24
 pop_18_24_10
 prcapinc
 region
 relig_import
 religiosity
 reppct_m
 rtw
 secularism
 secularism3
 seniority_sen2
 to_0004

Dependent Variable
 obama2012

Covariates

Factors
 south

> Model Builder

▼ Reference Levels

Variable	Reference Level
<input type="checkbox"/> south	South

> Assumption Checks

> Model Fit

> Model Coefficients

> Estimated Marginal Means

Linear Regression

Model Fit Measures

Model	R	R ²
1	0.271	0.0735

Model Coefficients - obama2012

Predictor	Estimate	SE	t	p
Intercept	47.09	1.52	31.05	<.001
south:				
South - Nonsouth	-5.92	3.03	-1.95	0.057

Linear Regression

Model Fit Measures

Model	R	R ²
1	0.271	0.0735

Model Coefficients - obama2012














Predictor	Estimate	SE	t	p
Intercept	47.09	1.52	31.05	<.001
south:				
Nonsouth - South	5.92	3.03	1.95	0.057

References

- [1] The jamovi project (2019). *jamovi*. (Version 1.0) [Computer Software]. Retrieved from <https://www.jamovi.org>.
- [2] R Core Team (2018). *R: A Language and environment for statistical computing*. [Computer software]. Retrieved from <https://cran.r-project.org/>.

Linear Regression





-  abort_rank3
-  abortion_rank12
-  adv_or_more
-  cig_tax12
-  cig_tax12_3
-  conserv_advantage
-  conserv_public
-  dem_advantage
-  govt_worker
-  gun_rank3
-  gun_rank11
-  gun_scale11
-  hr_cons_rank11


Dependent Variable

→  obama2012

Covariates

→  relig_prot

 urban

 ba_or_more

Factors

→

> Model Builder

> Reference Levels

∨ Assumption Checks

Assumption Checks

- Autocorrelation test
- Collinearity statistics
- Q-Q plot of residuals
- Residual plots

Data Summary

- Cook's distance

> Model Fit

> Model Coefficients

Linear Regression

Model Fit Measures

Model	R	R ²
1	0.640	0.410

Model Coefficients - obama2012

Predictor	Estimate	SE	t	p
Intercept	5.6082	14.7492	0.380	0.706
relig_prot	0.0232	0.1102	0.211	0.834
urban	0.0972	0.0977	0.994	0.325
ba_or_more	1.2647	0.3229	3.917	<.001

Data Summary

Cook's Distance

Mean	Median	SD	Range	
			Min	Max
0.0457	0.00536	0.162	6.40e-6	1.03














Assumption Checks

Collinearity Statistics

	VIF	Tolerance
relig_prot	1.96	0.510
urban	1.56	0.640
ba_or_more	1.72	0.582

Linear Regression



-  abort_rank3
-  abortion_rank12
-  adv_or_more
-  **cig_tax12**
-  cig_tax12_3
-  conserv_advantage
-  conserv_public
-  dem_advantage
-  govt_worker
-  gun_rank3
-  gun_rank11
-  gun_scale11
-  hr_cons_rank11

Dependent Variable

Covariates

Factors

Assumption Checks

Data Summary

- Autocorrelation test
- Collinearity statistics
- Q-Q plot of residuals
- Residual plots

- Cook's distance

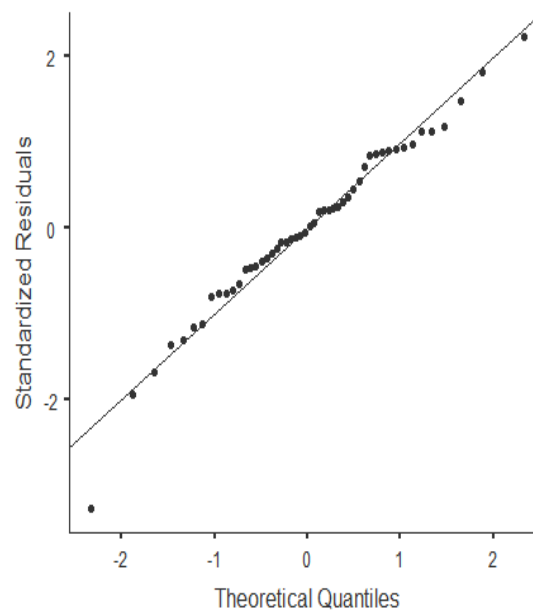
Assumption Checks

Collinearity Statistics

	VIF	Tolerance
relig_prot	1.96	0.510
urban	1.56	0.640
ba_or_more	1.72	0.582

[3]

Q-Q Plot



Residuals Plots

Linear Regression

- abort_rank3
- abortion_rank12
- adv_or_more
- cig_tax12
- cig_tax12_3
- conserv_advantage
- conserv_public
- dem_advantage
- govt_worker
- gun_rank3
- gun_rank11
- gun_scale11
- hr_cons_rank11

Dependent Variable
 obama2012

Covariates
 relig_prot
 urban
 ba_or_more

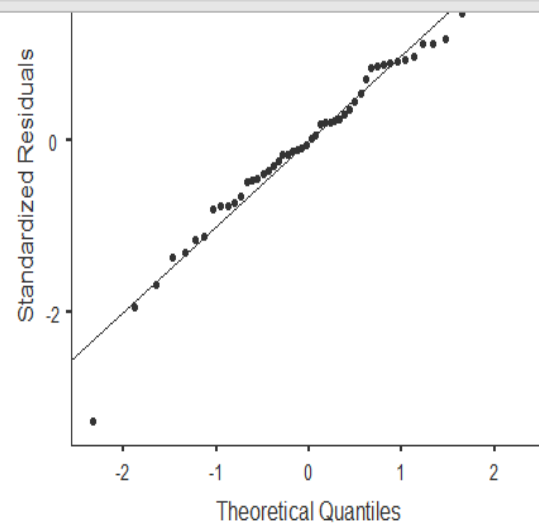
Factors

- > Model Builder
- > Reference Levels
- ▼ Assumption Checks

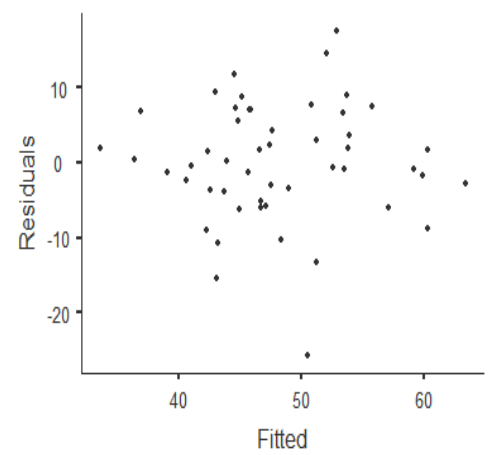
- Assumption Checks**
- Autocorrelation test
 - Collinearity statistics
 - Q-Q plot of residuals
 - Residual plots

- Data Summary**
- Cook's distance

- > Model Fit
- > Model Coefficients



Residuals Plots



Linear Regression



Dependent Variable

→ obama2012

Covariates

→ relig_prot
urban
ba_or_more

Factors

→

- abort_rank3
- abortion_rank12
- adv_or_more
- cig_tax12
- cig_tax12_3
- conserv_advantage
- conserv_public
- dem_advantage
- govt_worker
- gun_rank3
- gun_rank11
- gun_scale11
- hr_cons_rank11

Model Builder

Reference Levels

Assumption Checks

Assumption Checks

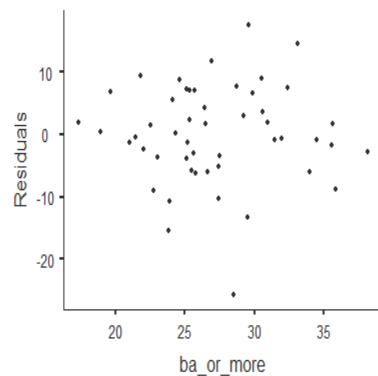
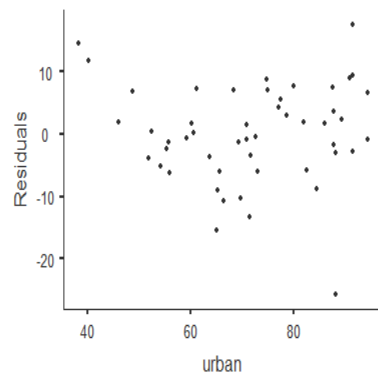
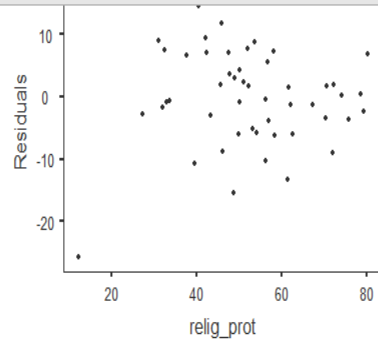
Data Summary

- Autocorrelation test
- Collinearity statistics
- Q-Q plot of residuals
- Residual plots
- Cook's distance

Model Fit

Model Coefficients

Estimated Marginal Means



Seminář