

Vícečetná lineární regrese

Petr Ocelík

MVZn4003 Úvod do kvantitativních metod

5. ledna 2020

Blok přednášek

1. Jednoduchá lineární regrese
2. Vícečetná lineární regrese
3. Diagnostika a aplikace

Blok přednášek

1. Jednoduchá lineární regrese
- 2. Vícečetná lineární regrese**
 1. intuice
 2. regresní model
 3. interpretace výsledků
3. Diagnostika a aplikace

(Lineární) regrese

- Regrese je **metoda predikce** (odhadu) hodnot **závislé proměnné** (Y) na základě hodnot jedné (X) či více **nezávislých proměnných** (prediktorů).
- Lineární regrese odhaduje lineární vztah
- Jednoduchá lineární regrese: pouze jeden prediktor
- **Vícečetná lineární regrese**: více než jeden prediktor

- Kellstedt & Whitten (2013) příklad regrese volebního zisku inkumbenta (**závislá proměnná Y**) v závislosti na změně HDP/pc (growth) během předchozího roku (**prediktor X**).

- $Y = 51.55 + 0.65 * X$

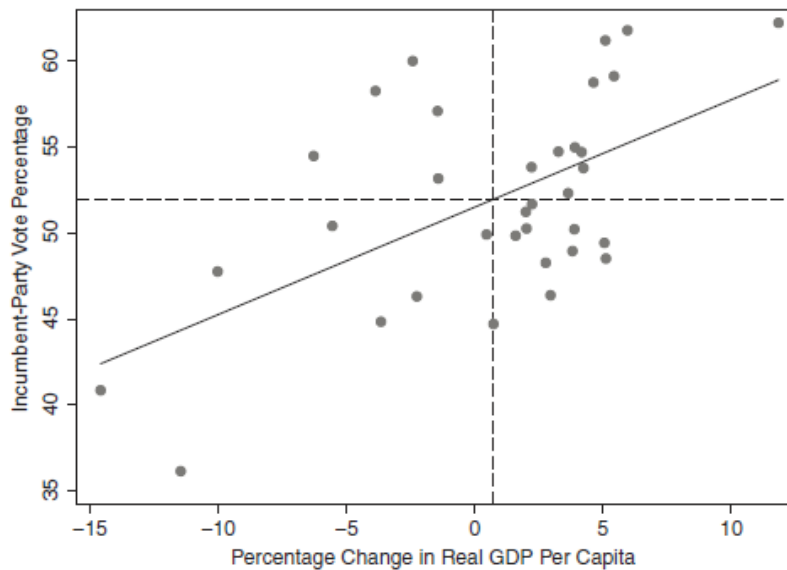


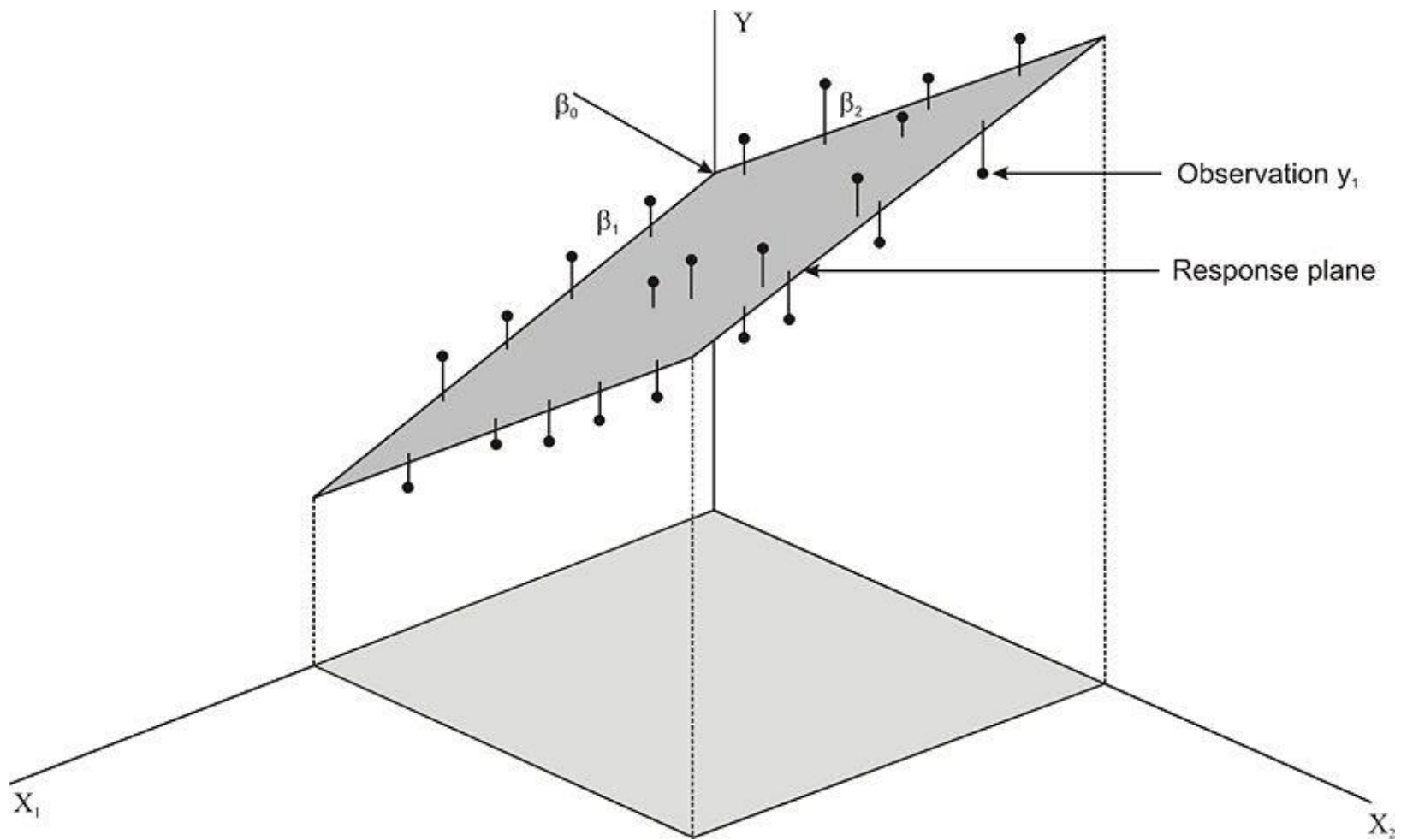
Figure 8.3. OLS regression line through scatter plot with mean-delimited quadrants.

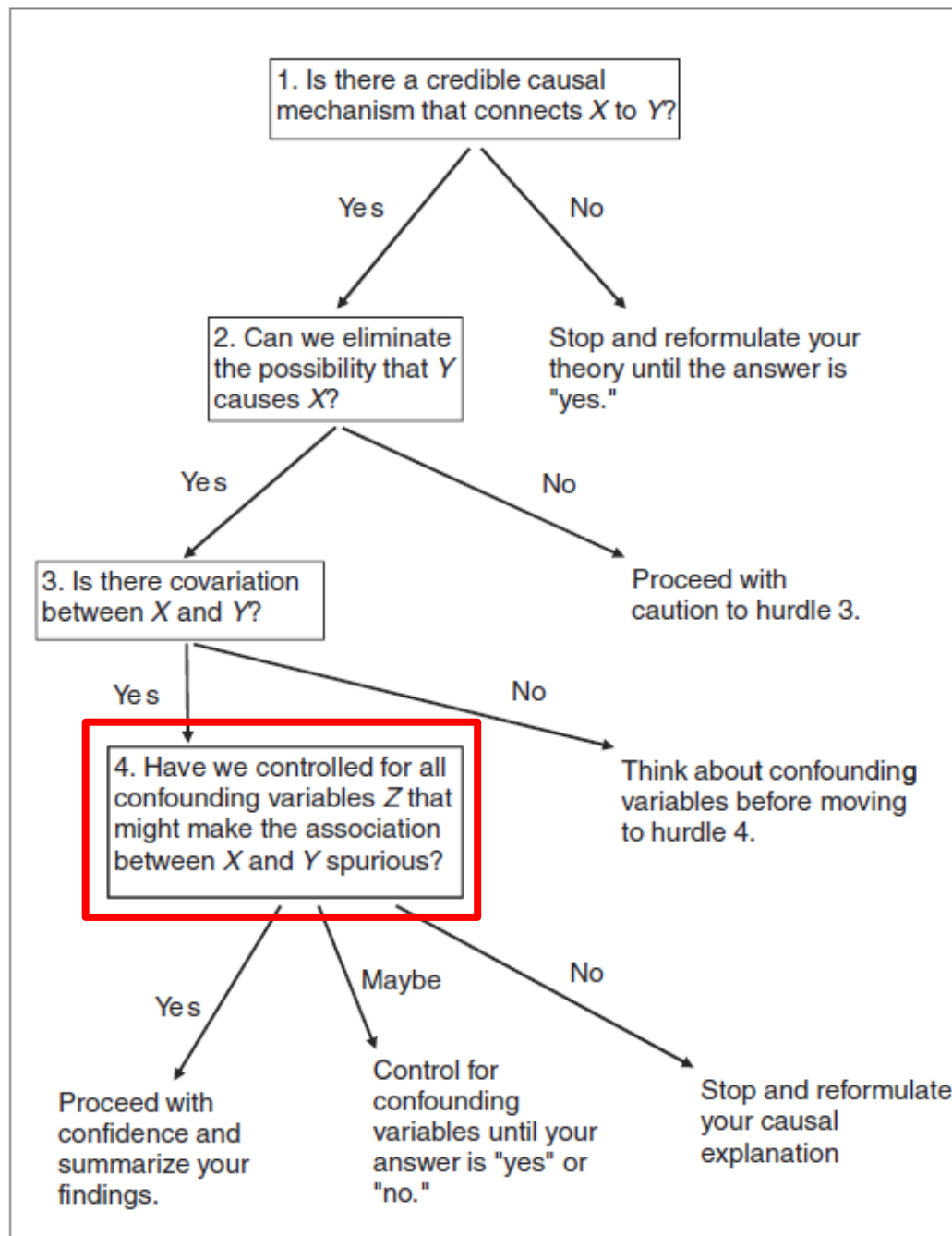
	A
intercept	51.55* (0.86)
growth (X)	0.65* (0.86)
R ²	0.36
Std. error in parentheses * = p < 0.05	

- Kellstedt & Whitten (2013) příklad regrese volebního zisku inkumbenta (**závislá proměnná Y**) v závislosti na:
 - **(1)** změně HDP/pc (growth) během minulého roku (**prediktor X**) a
 - **(2) dobrých zprávách (good news) o ekonomice** pro poslední kvartál (**prediktor 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)
growth (X)	0.65* (0.16)	-	0.57* (0.15)
good news (Z)	-	0.92* (0.33)	0.67* (0.28)
R ²	0.36	0.20	0.46
Std. error in parentheses * = p < 0.05			

Regresní rovina





Vícečetná lineární regrese

- Umožňuje predikci ZP v závislosti na hodnotách **více než jednoho prediktoru**
 - Jaká je predikovaná hodnota volebního zisku inkumbenta (ZP Y), pokud bereme v potaz ekonomický růst (prediktor X) a dobré zprávy o ekonomice (prediktor Z)?
- Umožňuje odhadovat **dílčí (parciální) účinky (efekty) jednotlivých prediktorů** na ZP
 - Jak se změní hodnota volebního zisku inkumbenta (ZP Y), pokud se hodnota prediktoru X změní o jednotku (% bod) při kontrole prediktoru Z?
 - Jak se změní hodnota volebního zisku inkumbenta (ZP), pokud se hodnota prediktoru Z změní o jednotku (jeden kvartál) při kontrole prediktoru X?

Regresní model: lineární regresní funkce

- **Lineární regresní funkce** má tvar:

- $Y' = b_0 + b_1 * X$

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

Y' = predikovaná hodnota ZP

protože teoretickou funkci pouze odhadujeme na základě vzorku prostřednictvím empirické funkce, odhad nutně zahrnuje nevysvětlený rozptyl ZP (chyba modelu)

- **Rozlišení pro populaci a vzorek:**

- $Y = \beta_0 + \beta_1 X + \varepsilon$; teoretická regresní funkce (populace)

- $Y' = b_0 + b_1 X$; empirická regresní funkce (vzorek)

- b_0 a b_1 jsou odhady populačních parametrů $\beta_0 + \beta_1$

- Tj.: b_0 a b_1 počítáme ze vzorku, Y a X známe (data ze vzorku)

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:

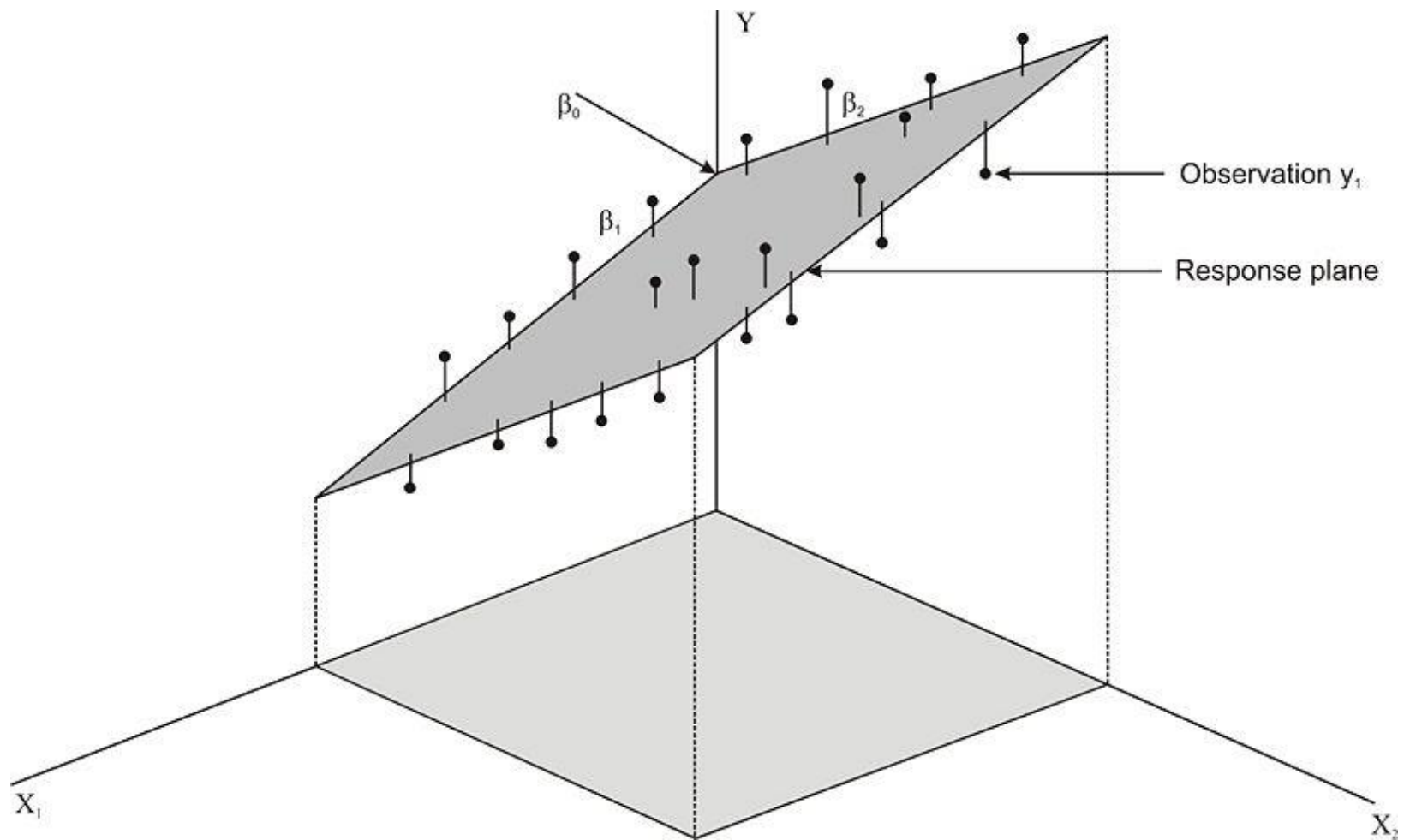
- $Y' = b_0 + b_1 * X_1 + b_2 * X_2$
- závislá proměnná = průsečík + sklon * prediktor X_1 + sklon * prediktor X_2

Jak odhadujeme regresní model?

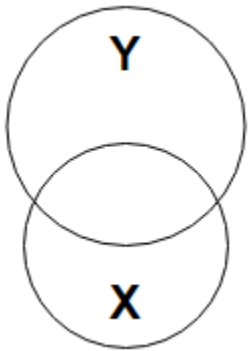
Metoda nejmenších čtverců

- **Metoda nejmenších čtverců (Ordinary Least Squares):** odhad koeficientů, průsečíku a sklonů, v lineárním regresním modelu
- **Minimalizuje vertikální vzdálenosti** mezi pozorovanými hodnotami (Y) a predikovanými hodnotami Y (Y').
- **Reziduum je chyba predikce** daná rozdílem mezi pozorovanou (Y) a predikovanou hodnotou (Y') ZP
- **Reziduum = $(Y - Y')$** ; součet reziduí $\sum(Y - Y') = 0$
- Kvadratické reziduum = $(Y - Y')^2$
- **Součet kvadratických reziduí $\sum(Y - Y')^2$**
- **OLS minimalizuje součet kvadratických reziduí: $Y' = \min \sum(Y - Y')^2$**

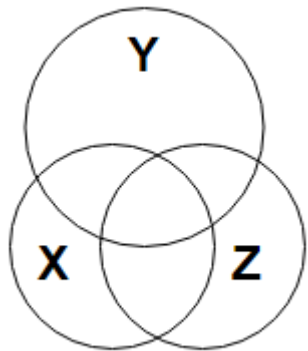
Regresní rovina



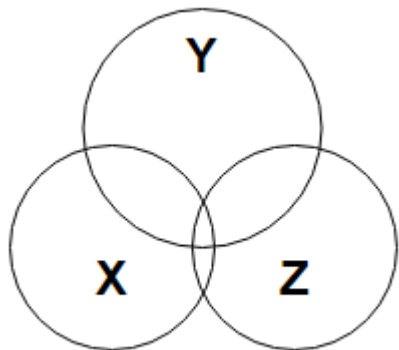
Interpretace koeficientů sklonu: intuice



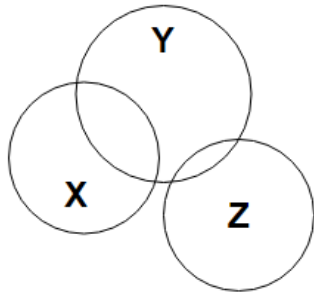
jednoduchý model s jedním prediktorem



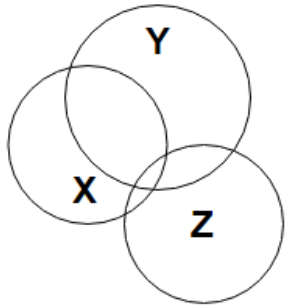
model se dvěma silněji korelovanými prediktory s menšími **dílčími účinky** (efekty)



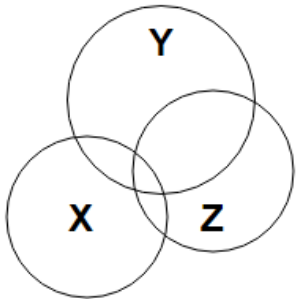
model se dvěma slaběji korelovanými prediktory s většími **dílčími účinky** (efekty)



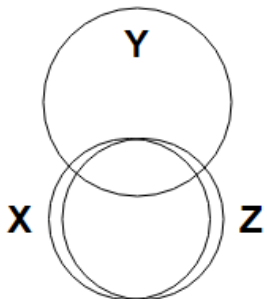
X a Z nekorelují → interpretace jako v jednoduché regresi (X a Z jsou vzájemně nezávislé)



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 nekonečnou (neinterpretovatelné) → problém **kolinearity**

Proč je to důležité?

Příklad: jeden prediktor

- **Predikce příjmu (Y) na základě pohlaví (X)**
 - X: muž (0), žena (1)

$Y' = b_0 + b_1 * X$; příjem = průsečík + sklon * pohlaví

- Výsledek odhadu: **$Y' = 20000 + 1500 * X$**
- **Jaký je predikovaný příjem pro muže a jaký pro ženy?**

Příklad: jeden prediktor

- **Predikce příjmu (Y) na základě pohlaví (X)**
 - X: muž (0), žena (1)

$Y' = b_0 + b_1 * X$; příjem = průsečík + sklon * pohlaví

- Výsledek odhadu: **$Y' = 20000 + 1500 * X$**
- Predikce pro muže (X = 0): **$Y' = 20000 + 1500 * 0 = 20\ 000$**
- Predikce pro ženy (X = 1): **$Y' = 20000 + 1500 * 1 = 21\ 500$**

Příklad: jeden prediktor

- **Predikce příjmu (Y) na základě pohlaví (X)**
 - X: muž (0), žena (1)

$Y' = b_0 + b_1 * X$; příjem = průsečík + sklon * pohlaví

- Výsledek odhadu: **$Y' = 20000 + 1500 * X$**
- Predikce pro muže (X = 0): $Y' = 20000 + 1500 * 0 = 20\ 000$
- Predikce pro ženy (X = 1): $Y' = 20000 + 1500 * 1 = 21\ 500$
- **Co když jsme opomněli jiný důležitý prediktor?** (Nápady?)

Příklad: statistická kontrola

- Predikce příjmu (Y) na základě pohlaví (X) a **vzdělání (Z)**
 - X: muž (0), žena (1); Z: # roků dokončeného vzdělání

$Y' = b_0 + b_1 * X_1 + b_2 * Z$; příjem = průsečík + sklon * pohlaví + sklon * vzdělání

- Výsledek odhadu: **$Y' = 18500 - 1500 * X + 2000 * Z$**
- Predikce pro muže (X = 0): **$Y' = 18500 - 1500 * 0 + 2000 * 1 = 20\ 500$**
- Predikce pro ženy (X = 1): **$Y' = 18500 - 1500 * 1 + 2000 * 1 = 19\ 000$**

→ **při kontrole vzdělání se efekt pohlaví otáčí**

- Kellstedt & Whitten (2013) příklad regrese volebního zisku inkumbenta (**závislá proměnná Y**) v závislosti na:
 - **(1)** změně HDP/pc (growth) během minulého roku (**prediktor X**) a
 - **(2) dobrých zprávách (good news) o ekonomice** pro poslední kvartál (**prediktor Z**)

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

Který prediktor má větší účinek (efekt) na ZP?

	A	B	C
intercept	51.55* (0.86)	47.29* (1.94)	48.24* (1.64)
growth (X)	0.65* (0.16)	-	0.57* (0.15)
good news (Z)	-	0.92* (0.33)	0.67* (0.28)
R2	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)
growth (X)	0.65* (0.16)	-	0.57* (0.15)
good news (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:

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

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

Interpretace účinků

- Jaké jsou účinky prediktorů **union10** a **abortion_rank12**?
 - **union10**: % pracovníků v odborech
 - **abortion_rank12**: ranking 50 států (1 = nejméně, 50 = nejvíce)

Model Fit Measures

Model	R	R ²	Adjusted R ²
1	0.673	0.452	0.429

Model Coefficients - obama2012

Predictor	Estimate	SE	t	p
Intercept	33.926	2.537	13.37	< .001
union10	0.764	0.252	3.03	0.004
abortion_rank12	0.218	0.100	2.17	0.035

Model Fit Measures

Model	R	R ²	Overall Model Test			
			F	df1	df2	p
1	0.673	0.452	19.4	2	47	< .001

Model Coefficients - obama2012

Predictor	Estimate	SE	t	p	Stand. Estimate
Intercept	33.926	2.537	13.37	< .001	
union10	0.764	0.252	3.03	0.004	0.430
abortion_rank12	0.218	0.100	2.17	0.035	0.308

Descriptives

	obama2012	abortion_rank12	union10
N	50	50	50
Missing	0	0	0
Mean	48.2	25.5	11.4
Median	50.2	25.5	10.9
Standard deviation	10.3	14.6	5.79
Minimum	24.7	1	3.10
Maximum	70.5	50	25.2

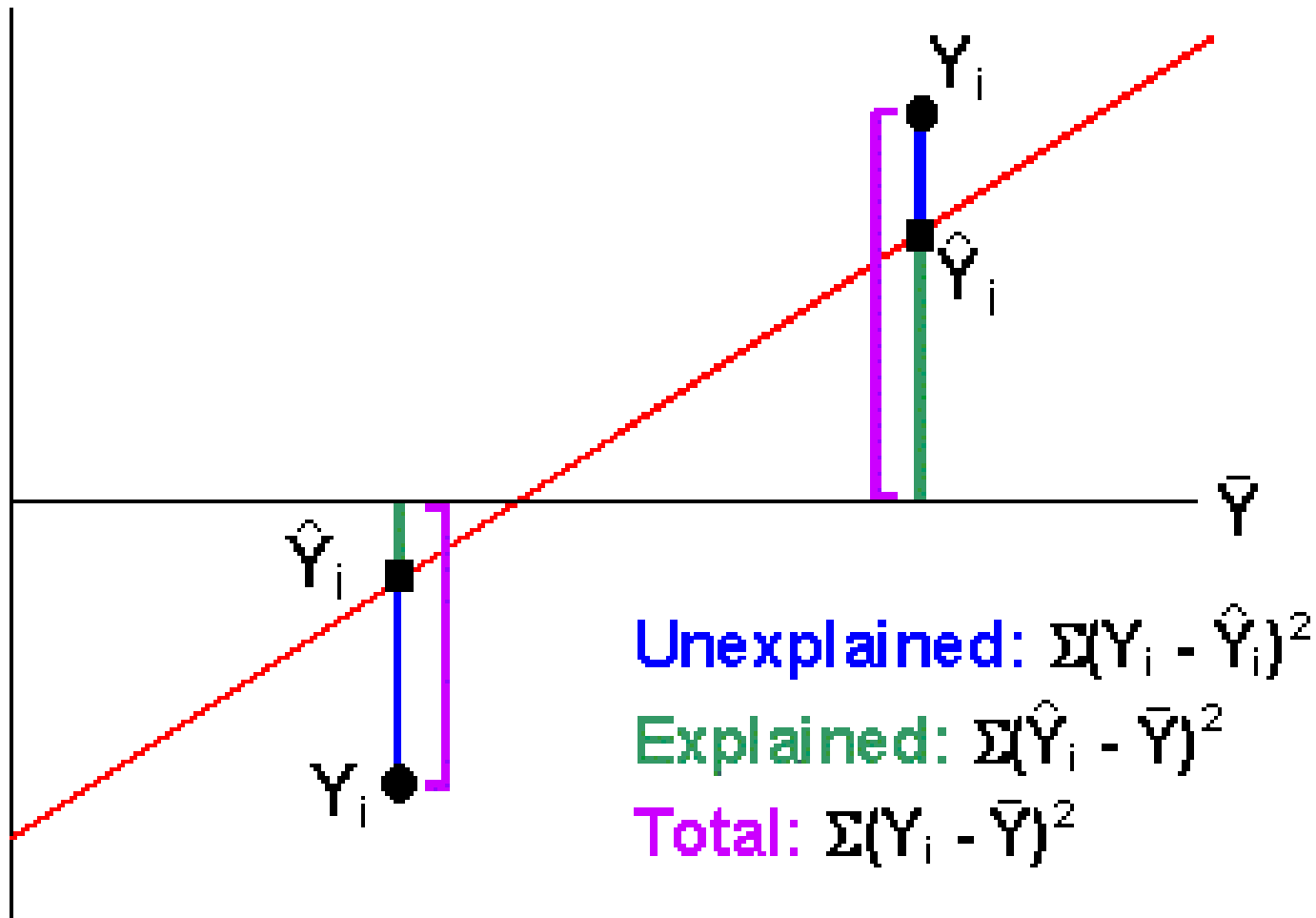
- **Beta váhy** (Stand. Estimate) ukazují, že účinek **union10** je cca o 40 % vyšší než účinek **abortion_rank12**

vše při předpokladu kontroly ostatních prediktorů v modelu

- **Beta pro union10:** 0.430, tj. zvýšení o 1 std.dev. = zvýšení ZP o 0.430 std.dev.
- Zvýšení **union10** o **5.8 p.b.** zvýší **obama2012** o $0.43 \cdot 10.3 = 4.43$ p.b.
- **Beta pro abortion:** 0.308, tj. zvýšení o 1 std.dev. = zvýšení ZP o 0.308 std.dev.
- Zvýšení **abortion** o **14.5 míst** zvýší **obama2012** o $0.308 \cdot 10.3 = 3.17$ p.b.

Jak dobře regresní model vysvětluje
rozptyl závislé proměnné?

Celkový rozptyl = nevysvětlený rozptyl + vysvětlený rozptyl



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

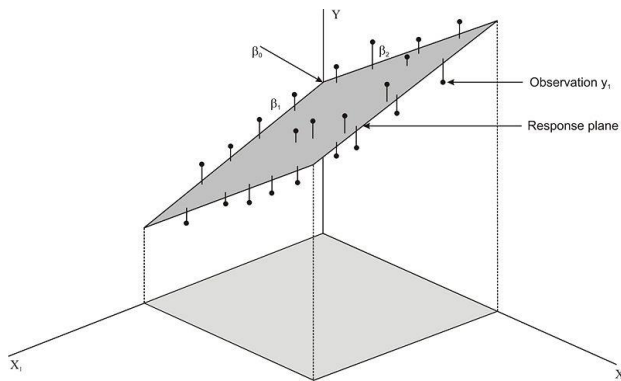
$$SSM = \Sigma(Y' - \text{mean}(Y))^2$$
- **Nevysvětlený rozptyl**: chyba modelu (**SSR**: sum of squares of residuals):

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

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

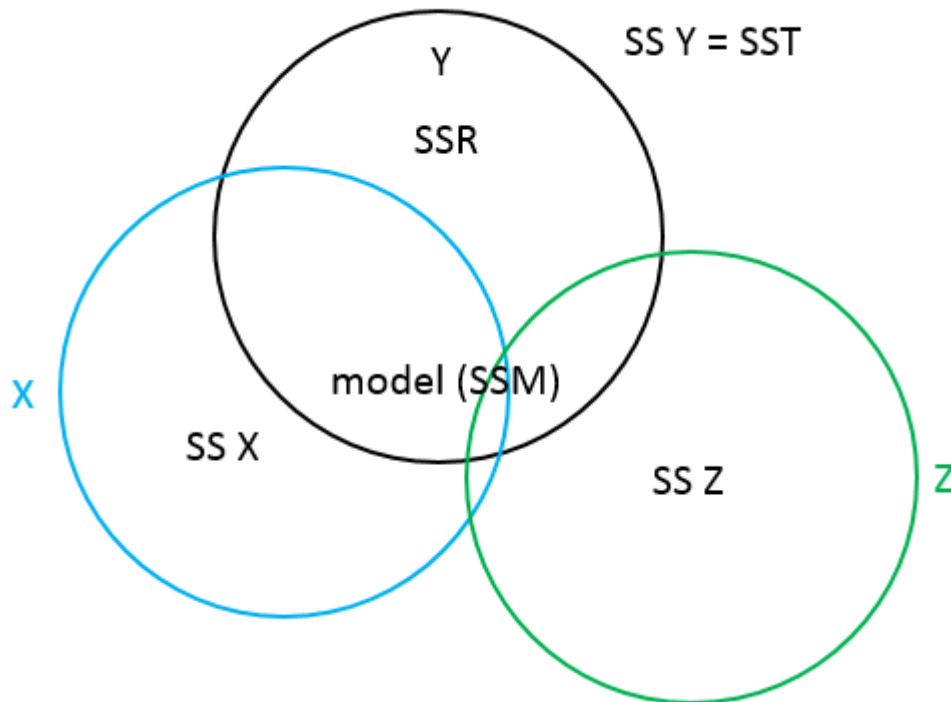
Y' = predikované hodnoty (rovina)
 Y = pozorované hodnoty (body)
 $\text{mean}(Y)$ = průměrná hodnota ZP

- 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) = $\mathbf{SSM / SST}$
- $\mathbf{SST = SSM}$ (vysvětlený rozptyl) + \mathbf{SSR} (nevysvětlený rozptyl)



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

Upravené R^2

- **Vyšší počet prediktorů $\cong R^2$**
- Zvýšením počtu prediktorů se snižuje **parsimonie modelu**
- **Upravené R^2 (adjusted R^2) (1)** penalizuje za počet prediktorů (p) a **(2)** za nižší počet pozorování (n)

$$R^2_{\text{adj}} = 1 - \frac{((1 - R^2) * (n - 1))}{(n - p - 1)}$$

Model Fit Measures

Model	R	Overall Model Test					
		R ²	Adjusted R ²	F	df1	df2	p
1	0.786	0.618	0.602	38.1	2	47	< .001

R² udává, že model vysvětluje 62 % rozptylu ZP (obama2012), upravené R² udává 60 %.

Model Coefficients - obama2012

Predictor	Estimate	SE	t	p	Stand. Estimate
Intercept	22.585	5.843	3.87	< .001	
gay_policy_con:					
Yes - No	-10.764	2.059	-5.23	< .001	-0.518
ba_or_more	0.901	0.215	4.19	< .001	0.415

vyšší R² znamená nižší čítenel

$$R^2_{adj} = 1 - \frac{((1 - R^2) * (n - 1))}{(n - p - 1)}$$

větší počet prediktorů (p) znamená nižší jmenovatel

$$R^2_{adj} = 1 - \frac{((1 - 0.618) * (50 - 1))}{(50 - 2 - 1)} = 0.602$$

Jsou odhady statisticky významné?

Celý model: testovací statistika F

- Abychom zjistili, zda se regresní model jako celek významně liší od nuly, uijeme **F-test**
- **F-test**: závisí hodnota ZP na lineární kombinaci prediktorů?
- H0: předpoklad, že všechny koeficienty $\beta = 0$
- H0: $\beta_1 = \beta_2 = \dots = \beta_k = 0$; HA: $\beta \neq 0$
- $F = \frac{\text{vysvětlený rozptyl}}{\text{nevysvětlený rozptyl}} ; F = \frac{R^2 * (n - p - 1)}{(1 - R^2) * p}$
 - n = velikost vzorku
 - p = počet prediktorů
 - R² = koeficient determinace
- F rozdělení s k – 1 a n – 2 stupni volnosti

F-hodnota regresního modelu je **testovací statistikou**

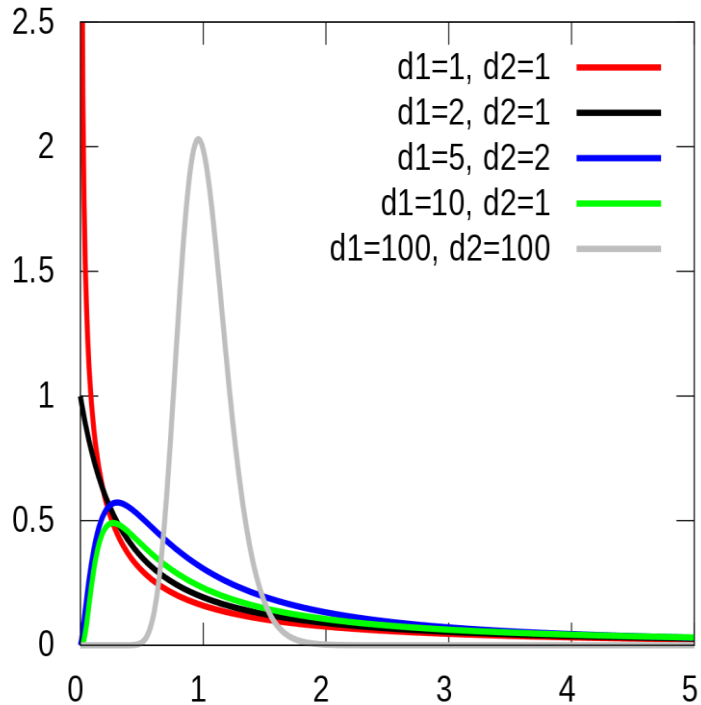


TABLE E

F critical values (continued)

		Degrees of freedom in the numerator								
		1	2	3	4	5	6	7	8	9
Degrees of freedom in the denominator										
	<i>p</i>									
18	.100	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00
	.050	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
	.025	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93
	.010	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60
	.001	15.38	10.39	8.49	7.46	6.81	6.35	6.02	5.76	5.56
19	.100	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98
	.050	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
	.025	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88
	.010	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52
	.001	15.08	10.16	8.28	7.27	6.62	6.18	5.85	5.59	5.39
20	.100	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96
	.050	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
	.025	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84
	.010	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46
	.001	14.82	9.95	8.10	7.10	6.46	6.02	5.69	5.44	5.24
21	.100	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95
	.050	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
	.025	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80
	.010	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40
	.001	14.59	9.77	7.94	6.95	6.32	5.88	5.56	5.31	5.11
22	.100	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93
	.050	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
	.025	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76
	.010	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35
	.001	14.38	9.61	7.80	6.81	6.19	5.76	5.44	5.19	4.99
23	.100	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92
	.050	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
	.025	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73
	.010	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30
	.001	14.20	9.47	7.67	6.70	6.08	5.65	5.33	5.09	4.89
24	.100	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91
	.050	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
	.025	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70
	.010	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26
	.001	14.03	9.34	7.55	6.59	5.98	5.55	5.23	4.99	4.80
25	.100	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89
	.050	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
	.025	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68
	.010	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22
	.001	13.88	9.22	7.45	6.49	5.89	5.46	5.15	4.91	4.71
26	.100	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88
	.050	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
	.025	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65
	.010	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18
	.001	13.74	9.12	7.36	6.41	5.80	5.38	5.07	4.83	4.64
27	.100	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87
	.050	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
	.025	5.63	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.63
	.010	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15
	.001	13.61	9.02	7.27	6.33	5.73	5.31	5.00	4.76	4.57

Model Fit Measures

Model	R	R ²	Adjusted R ²	Overall Model Test			
				F	df1	df2	p
1	0.786	0.618	0.602	38.1	2	47	< .001

Model se statisticky významně odlišuje od nuly; při dané úrovni testu tedy odmítáme H₀: β₁ = B_k = 0

Model Coefficients - obama2012

Predictor	Estimate	SE	t	p	Stand. Estimate
Intercept	22.585	5.843	3.87	< .001	
gay_policy_con:					
Yes - No	-10.764	2.059	-5.23	< .001	-0.518
ba_or_more	0.901	0.215	4.19	< .001	0.415

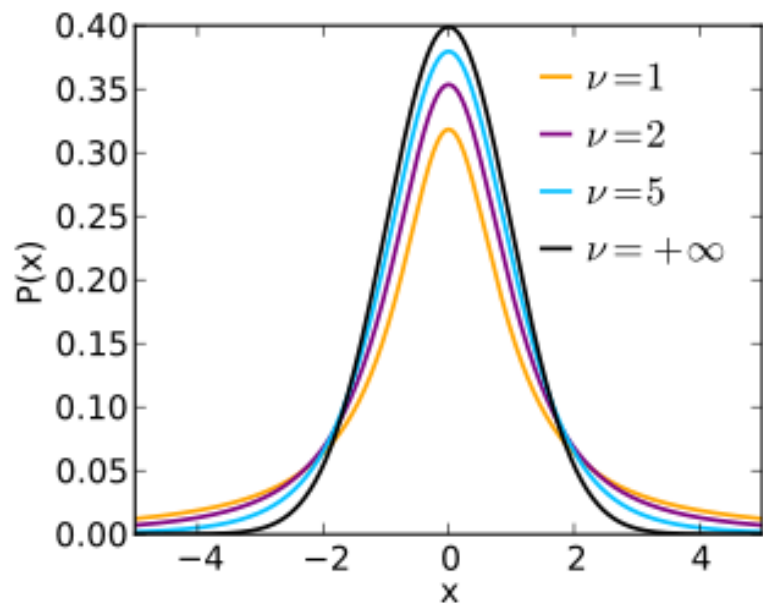
$$F = \frac{\text{vysvětlený rozptyl}}{\text{nevysvětlený rozptyl}} ; F = \frac{R^2 * (n - p - 1)}{(1 - R^2) * p}$$

n = velikost vzorku
p = počet prediktorů
R² = koeficient determinace

$$F = \frac{R^2 * (n - p - 1)}{(1 - R^2) * p} = \frac{0.6183 * (50 - 2 - 1)}{(1 - 0.6183) * 2} = \frac{29.0601}{0.7634} = \mathbf{38.07}$$

Koeficienty: testovací statistika t

- Abychom zjistili, zda se koeficient b významně liší od nuly, použijeme **t-test** pro regresní koeficient
- *Je hodnota b významně odlišná od předpokládaného populačního průměru 0? $H_0: \beta = 0$; $H_A: \beta \neq 0$*
- $t = \frac{\text{signal}}{\text{noise}}$; $t = \frac{b - \beta}{SE(b)}$; t rozdělení s $n - 2$ stupni volnosti
 - b = koeficient sklonu
 - β = předpokládaný populační koeficient ($\beta = 0$)
 - $SE(b)$ = std. chyba koeficientu (viz Handl 2009)
- **t-hodnota** regresního koeficientu b je **testovací statistikou**



	0.005	0.01	Area in One Tail 0.025	0.05	0.10
Degrees of Freedom	0.01	0.02	Area in Two Tails 0.05	0.10	0.20
1	63.657	31.821	12.706	6.314	3.078
2	9.925	6.965	4.303	2.920	1.886
3	5.841	4.541	3.182	2.353	1.638
4	4.604	3.747	2.776	2.132	1.533
5	4.032	3.365	2.571	2.015	1.476
6	3.707	3.143	2.447	1.943	1.440
7	3.499	2.998	2.365	1.895	1.415
8	3.355	2.896	2.306	1.860	1.397
9	3.250	2.821	2.262	1.833	1.383
10	3.169	2.764	2.228	1.812	1.372
11	3.106	2.718	2.201	1.796	1.363
12	3.055	2.681	2.179	1.782	1.356
13	3.012	2.650	2.160	1.771	1.350
14	2.977	2.624	2.145	1.761	1.345
15	2.947	2.602	2.131	1.753	1.341
16	2.921	2.583	2.120	1.746	1.337
17	2.898	2.567	2.110	1.740	1.333
18	2.878	2.552	2.101	1.734	1.330
19	2.861	2.539	2.093	1.729	1.328
20	2.845	2.528	2.086	1.725	1.325
21	2.831	2.518	2.080	1.721	1.323
22	2.819	2.508	2.074	1.717	1.321
23	2.807	2.500	2.069	1.714	1.319

Model Fit Measures

Model	R	R ²	Adjusted R ²	Overall Model Test			
				F	df1	df2	p
1	0.786	0.618	0.602	38.1	2	47	< .001

Model se statisticky významně odlišuje od nuly; při dané úrovni testu tedy odmítáme $H_0: \beta_1 = \beta_k = 0$

Model Coefficients - obama2012

Predictor	Estimate	SE	t	p	Stand. Estimate
Intercept	22.585	5.843	3.87	< .001	
gay_policy_con: Yes – No	-10.764	2.059	-5.23	< .001	-0.518
ba_or_more	0.901	0.215	4.19	< .001	0.415

Koeficient průsečíku i koeficienty sklonu jsou statisticky významné; při dané úrovni testu tedy odmítáme $H_0: \beta = 0$

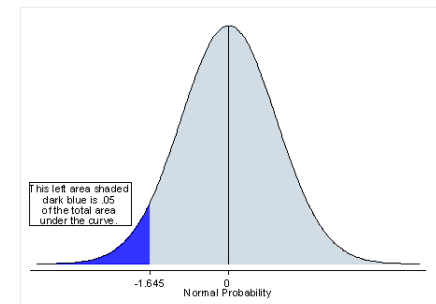
- **Koeficient** (Estimate) = odhady hodnot koeficientů průsečíku a sklonu
- **Beta koeficienty** (Stand. Estimate) = std. odhady koeficientů sklonu
- **Std. chyba** (SE) = přesnost odhadu koeficientů (viz Handl 2009)
- **t-hodnota** (t) = testovací statistika koeficientů; $t = \text{koeficient} / \text{SE}$
- **p-hodnota** (p) = pravděpodobnost, že pozorujeme danou, nebo extrémnější, hodnotu testovací statistiky t při platnosti $H_0: \beta = 0$

Shrnutí

- 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 (následující přednáška)

Vícečetná lineární regrese v Jamovi

- **Teoretický rámec:** volební chování je do určité míry určeno sociokulturními štěpnými liniemi (Norris & Inglehart 2019; Lipset & Rokkan 1967)
- **H0:** Podíl protestantů **relig_prot** (prediktor X) neovlivňoval volební podporu Obamy **obama2012** (závislá proměnná Y); $\beta_1 \geq 0$
- **HA:** Podíl protestantů **relig_prot** (prediktor X) snižoval volební podporu Obamy **obama2012** (závislá proměnná Y); $\beta_1 < 0$
- **Které další proměnné by mohly výsledek Obamy ovlivňovat?**



Vícečetná lineární regrese v Jamovi

- **Teoretický rámec:** volební chování je do určité míry určeno sociokulturními štěpnými liniemi (Norris & Inglehart 2019; Lipset & Rokkan 1967)
- **H0:** Podíl protestantů **relig_prot** (prediktor X) neovlivňoval volební podporu Obamy **obama2012** (závislá proměnná Y); $\beta_1 \geq 0$
- **HA:** Podíl protestantů **relig_prot** (prediktor X) snižoval volební podporu Obamy **obama2012** (závislá proměnná Y); $\beta_1 < 0$
- **Které další proměnné by mohly výsledek Obamy ovlivňovat?**
 - socioekonomický status (per capita income: prcapinc)
 - podíl městského obyvatelstva (urban)
 - konzervativnost politiky vůči gayům (gay_policy_con)



Exploration



T-Tests



ANOVA



Regression



Frequencies



Factor



R



Modules

	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



obama_win12 → obama2012

pop2000

pop2010

pop2010_hun_thou

popchn0010

popchnpct

pot_policy

prochoice

prolife

relig_cath

relig_high

relig_low

religiosity3

Covariates

relig_prot

Factors

> Model Builder

> Reference Levels

> Assumption Checks

Model Fit

Fit Measures

- R
- R²
- Adjusted R²
- AIC
- BIC
- RMSE

Overall Model Test

- F test

Linear Regression

Model Fit Measures

Model	R	R ²	Overall Model Test			
			F	df1	df2	p
1	0.413	0.170	9.86	1	48	0.003

Model Coefficients - obama2012

Predictor	Estimate	SE	t	p
Intercept	63.195	4.9721	12.71	< .001
relig_prot	-0.287	0.0914	-3.14	0.003

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/>.

	nader00	obama08	oba
1	10.067	37.89	McC
2	1.100	38.74	McC
3	1.456	38.86	McC
4	2.979	44.91	McC
5	3.818	60.94	Obe
6	5.251	53.66	Obe
7	4.416	60.59	Obe
8	2.536	61.91	Obe
9	1.635	50.91	Obe
10	0.517	46.90	McC
11	5.877	71.85	Obe
12	2.233	53.93	Obe
13	2.450	35.91	McC
14	2.188	61.86	Obe
15	0.843	49.87	Obe
16	3.366	41.57	McC
17	1.502	41.15	McC
18	1.160	39.93	McC
19	6.421	61.80	Obe
20	2.655	61.92	Obe
21	5.696	57.71	Obe
22	1.989	57.33	Obe
23	5.195	54.06	Obe
24	1.632	49.24	McC
25	0.817	43.00	McC
26	5.946	47.11	McC
27		49.70	Obe
28	3.291	44.47	McC
29	3.521	41.60	McC
30	3.901	54.13	Obe
31	2.967	57.14	Obe
32	3.550	56.91	Obe
33	2.464	55.15	Obe
34	3.577	62.88	Obe
35	2.505	51.38	Obe
36		34.35	McC
37	5.043	56.75	Obe
38	2.104	54.47	Obe
39	6.124	62.86	Obe
40	1.461	44.90	McC
41		44.75	McC
42	0.953	41.79	McC

Linear Regression

Dependent Variable

obama2012

Covariates

relig_prot
urban
prcapinc

Factors

gay_policy_con

Model Builder

Reference Levels

Assumption Checks

Model Fit

Fit Measures

R

R²

Adjusted R²

AIC

BIC

RMSE

Model Coefficients

Estimated Marginal Means

Results

Linear Regression

Model Fit Measures

Model	R	R ²	Adjusted R ²	Overall Model Test			
				F	df1	df2	p
1	0.771	0.594	0.558	16.5	4	45	< .001

Model Coefficients - obama2012

Predictor	Estimate	SE	t	p
Intercept *	14.9954	12.4388	1.206	0.234
relig_prot	0.0979	0.0919	1.065	0.293
urban	0.0640	0.0844	0.758	0.453
gay_policy_con:				
Yes - No	-11.5345	2.4015	-4.803	< .001
prcapinc	8.78e-4	2.84e-4	3.093	0.003

* Represents reference level

References

- [1] The jamovi project (2020). *jamovi*. (Version 1.6) [Computer Software]. Retrieved from <https://www.jamovi.org>.
- [2] R Core Team (2020). *R: A Language and environment for statistical computing*. (Version 4.0) [Computer software]. Retrieved from <https://cran.r-project.org>. (R packages retrieved from MRAN snapshot 2020-08-24).



Data

Analyses

Edit



Exploration



T-Tests



ANOVA



Regression



Frequencies



Factor

Linear Regression



Linear Regression configuration panel:

- Dependent Variable:** obama2012
- Covariates:** relig_prot, urban, prcapinc
- Factors:** gay_policy_con
- Available Variables:** indpct_m, kerry04, libpct_m, mccain08, modpct_m, nader00, obama08, obama_win08, over64, permit, pop_18_24, pop_18_24_10, region

Model Builder

Reference Levels

Assumption Checks

Model Fit

Model Coefficients

Omnibus Test

 ANOVA test

Estimate

 Confidence intervalInterval %

Standardized Estimate

 Standardized estimate Confidence intervalInterval %

Estimated Marginal Means

Results

Linear Regression

Model Fit Measures

Model	R	R ²	Adjusted R ²	Overall Model Test			
				F	df1	df2	p
1	0.771	0.594	0.558	16.5	4	45	< .001

Model Coefficients - obama2012

Predictor	Estimate	SE	t	p	95% Confidence Interval		
					Stand. Estimate	Lower	Upper
Intercept *	14.9954	12.4388	1.206	0.234			
relig_prot	0.0979	0.0919	1.065	0.293	0.1408	-0.125	0.407
urban	0.0640	0.0844	0.758	0.453	0.0926	-0.154	0.339
gay_policy_con:							
Yes - No	-11.5345	2.4015	-4.803	< .001	-1.1212	-1.591	-0.651
prcapinc	8.78e-4	2.84e-4	3.093	0.003	0.3798	0.132	0.627

* Represents reference level

References

[1] The jamovi project (2020). *jamovi*. (Version 1.6) [Computer Software]. Retrieved from <https://www.jamovi.org>.

[2] R Core Team (2020). *R: A Language and environment for statistical computing*. (Version 4.0) [Computer software]. Retrieved from MRAN snapshot 2020-08-24).



Data

Analyses

Edit



Exploration



T-Tests



ANOVA



Regression



Frequencies



Factor

Descriptives



- evo2012
- evr2012
- gay_policy
- gay_policy2
- gay_policy_con
- gay_support
- gay_support3
- gb_win00
- qb_win04

Variables

- obama2012
- prcapinc

Split by

 Frequency tables

 Statistics

Sample Size

 N Missing

Percentile Values

 Cut points for equal groups

 Percentiles

Dispersion

 Std. deviation Minimum
 Variance Maximum
 Range S. E. Mean

Central Tendency

 Mean
 Median
 Mode
 Sum

Distribution

 Skewness
 Kurtosis

Normality

 Shapiro-Wilk

 Plots

Results

Linear Regression

Model Fit Measures

Model	R	R ²	Adjusted R ²	Overall Model Test			
				F	df1	df2	p
1	0.771	0.594	0.558	16.5	4	45	< .001

Model Coefficients - obama2012

Predictor	Estimate	SE	t	p	Stand. Estimate	95% Confidence Interval	
						Lower	Upper
Intercept *	14.9954	12.4388	1.206	0.234			
relig_prot	0.0979	0.0919	1.065	0.293	0.1408	-0.125	0.407
urban	0.0640	0.0844	0.758	0.453	0.0926	-0.154	0.339
gay_policy_con:							
Yes - No	-11.5345	2.4015	-4.803	< .001	-1.1212	-1.591	-0.651
prcapinc	8.78e-4	2.84e-4	3.093	0.003	0.3798	0.132	0.627

* Represents reference level

Descriptives

Descriptives

	obama2012	prcapinc
N	50	50
Missing	0	0
Mean	48.2	31951
Median	50.2	31331
Standard deviation	10.3	4449
Minimum	24.7	24630
Maximum	70.5	45398

Seminář