

PSY252

Statistická analýza dat v psychologii II

Přednáška 3

Pokročilá využití regresního modelu

Mnohonásobná lineární regrese

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k + e$$

Osnova

- Centrování prediktorů
 - Vkládání prediktorů po blocích - porovnávání modelů
 - Kategorické prediktory, kontrasty
 - Interakce-moderace
 - Mediace
 - Odhad potřebné velikosti vzorku
 - Lineární regrese pomocí GENLIN
-

Interpretace Y' a regresních koeficientů

$$Y' = b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k$$

Y' predikovaná hodnota je **podmíněný průměr Y** , $M(Y|X_i)$, tj. průměr lidí s danými hodnotami X_i

- Kdybychom sehnali náhodný vzorek lidí, kteří mají stejnou hodnotu u všech prediktorů, pak by Y' měla být rovna průměru jejich Y a jejich rozptyl je roven reziduálnímu rozptylu ... pokud jsou předpoklady modelu splněny.
- V tomto smyslu je predikovaná hodnota očekávanou hodnotou $E(Y|X_i)$

B_i ; b_i vyjadřuje nárůst Y' při nárůstu X_i o jednu jednotku; v jednotkách Y , pokud by ostatní prediktory zůstaly nezměněné (\approx semiparciální r); jedinečný přínos

β_i ; **BETA** vyjadřuje nárůst Y' při nárůstu X_i o 1SD v násobcích SD Y pokud by ostatní prediktory zůstaly nezměněné (\approx semiparciální r)

B_0 ; b_0 – obtížně interpretovatelný průsečík ... leda by prediktory byly **centrované**

V různých modelech nemusí být vliv prediktoru stejný

Centrování – usnadnění interpretace

- Průsečík = predikovaná hodnota Y' , když má prediktor hodnotu 0. Hodnota 0 mnohdy není smysluplnou hodnotou prediktoru....

C: Transformace prediktoru, aby **0** byla průměrem

- $$\text{PREDIKTOR}_C = \text{PREDIKTOR} - M_{\text{PREDIKTOR}}$$

- *Následným dělením SD bychom dospěli k z-skórum. Tím bychom ale přišli o měřítko.*

- Průsečík pak udává predikci pro člověka, který má průměrnou hodnotu všech prediktorů

Příklad centrování

□ deprese ~ selfe, effi

```
COMPUTE selfec=selfe-3.058621.
```

```
COMPUTE effic=effi-2.743662.
```

```
EXECUTE.
```

```
REGRESSION /DEPENDENT deprese  
/METHOD=ENTER selfe effi.
```

```
REGRESSION /DEPENDENT deprese  
/METHOD=ENTER selfec effic.
```





Možnosti práce s modely

1. Odhadneme model, který jsme plánovali.
 2. Odhadneme řadu modelů, s postupně se rozšiřující sadou prediktorů
 - hierarchická regrese
 3. Necháme nějaký algoritmus vybrat nejlepší sadu prediktorů z dostupných
 - stepwise algoritmy
-

Hierarchická lineární regrese

- Bloková, se sadami (sets) prediktorů
 - Prediktory vkládáme po skupinách (popř. jednotlivě) v teoreticky zdůvodněném pořadí
 - Teoreticky zdůvodněné pořadí umožňuje rozdělit rozptyl Y na smysluplné části (variance partitioning)
 - Změna pořadí prediktorů změní velikost těch částí
 - Smysluplnost se odvozuje právě od toho, kterými prediktory je rozptyl Y vysvětlen.
 - Zajímá nás schopnost sady prediktorů vylepšit model
 - Srovnání různých oblastí vlivu na zkoumaný jev
 - Zkoumání inkrementální validity
-

Obvyklá řazení bloků

Dle času, kauzální priority

- Př. od dispozičním k situačním...

Od známých k neznámým vlivům

- Stat. kontrola intervenujících proměnných

Podle výzkumné relevance

- Od ústředních po „co kdyby“; maximalizace síly
-

Srovnávání modelů po blocích

- Po vložení bloku sledujeme, zda došlo k nárůstu R^2 oproti předchozímu modelu
 - Test signifikance nárůstu R^2
 - Je-li A 1. blok prediktorů a B 2., $H_0: R^2_{Y.AB} - R^2_{Y.A} = 0$
 - Testová statistika
$$F = \frac{(R^2_{Y.AB} - R^2_{Y.A}) / k_B}{(1 - R^2_{Y.AB}) / (n - k_A - k_B - 1)}$$
 - $df_{\text{čitatel}} = k_B$ (počet proměnných v bloku B)
 - $df_{\text{jmenovatel}} = n - k_A - k_B - 1$
 - SPSS to počítá zaškrtnutím „R squared change“
-

Upozornění

- I když je přidávání bloků v SPSS provedeno automaticky jedním příkazem, nezapomínejme, že jde o sérii samostatných (obyčejných) regresních modelů.
 - Hierarchická regrese tedy není zvláštním typem regrese, regresního modelu.

 - Uvnitř jednotlivých modelů nezáleží na pořadí prediktorů.
 - Můžeme také říci, že uvnitř bloků nezáleží na pořadí.
-

Interpretace v hierarchické regresí

Přírůstek R^2

- Pokles není možný, minimem je nulový nárůst.
- Často nás zajímá, jestli je vskutku nějaký nárůst (H_0 není jen formální)

Změna B_i

- Pokud byl prediktor v dřívějším bloku, zajímá nás jeho změna
 - Možný je pokles i nárůst, vč. změny znaménka, nově přidané prediktory mění význam B_i
 - Velikost změn souvisí s tím, jak moc spolu prediktory korelují
 - **Pokles** – nově přidané prediktory asi vysvětlují Y tou částí svého rozptylu, kterou sdílí s prediktorem, jehož B poklesl
 - **Nárůst** – nově přidané prediktory asi vysvětlují část rozptylu prediktoru, jehož B narostl, a tato část rozptylu vůbec nesouvisí s Y .
 - Efekt suprese, „Simpsonův paradox“,
-

Příklad hierarchické regrese se supresí.

- Fieldovy modelky.
-



Použití kategorických prediktorů

Prediktorem může být i kategorická proměnná

- dichotomická
 - na dichotomie transformovaná
-

Příklad s pohlavím

- Změna kódování na 1=muž, 0=žena (MUŽ)

```
RECODE POHLAVI (1=1) (2=0) INTO MUZ.
```

```
EXECUTE.
```

```
FREQ MUZ.
```

```
T-TEST GROUPS=pohlavi(1 2) /VARIABLES=deprese.
```

- Regrese s proměnnou MUŽ jako jediným prediktorem

```
REGRESSION /DEPENDENT deprese /METHOD=ENTER muz.
```

Group Statistics

	pohlaví	N	Mean	Std. Deviation	Std. Error Mean
deprese	muzske	309	1,9376	,41479	,02360
	zenske	445	2,0638	,49520	,02347

t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
					Lower	Upper
-3,673	752	,000	-,12619	,03436	-,19363	-,05874

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2,064	,022		93,836	,000
	MUZ	-,126	,034	-,133	-3,673	,000

a. Depe	-,194	-,059
---------	-------	-------

Je-li kategorický prediktor kódovaný 0/1...

$$DEP' = 2,064 - 0,126 * MUZ$$

Je-li $MUZ=0$, pak $DEP' = 2,064 - 0,126 * 0 = 2,064$

- Při pouze jednom prediktoru udává b_0 průměr skupiny kódované 0

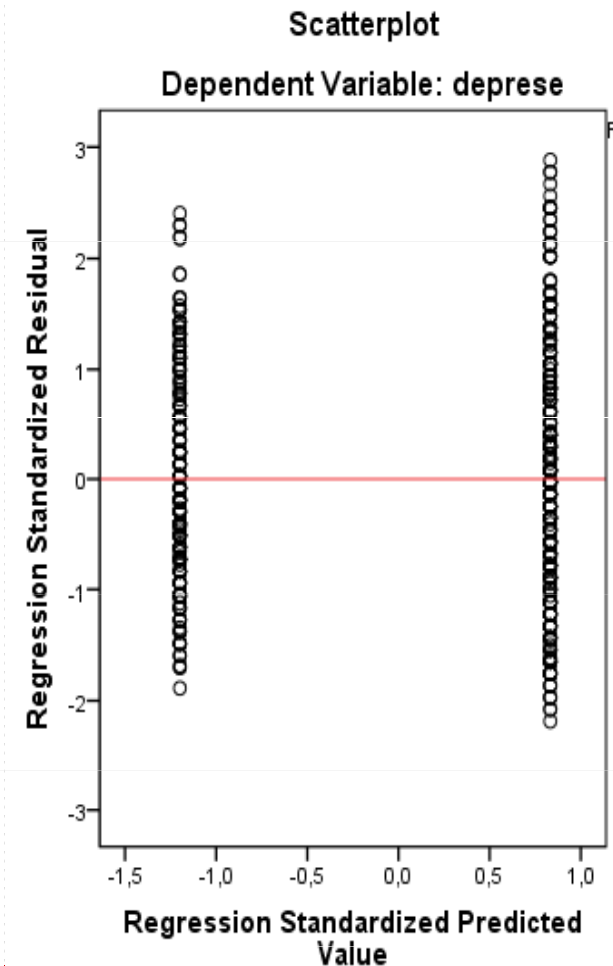
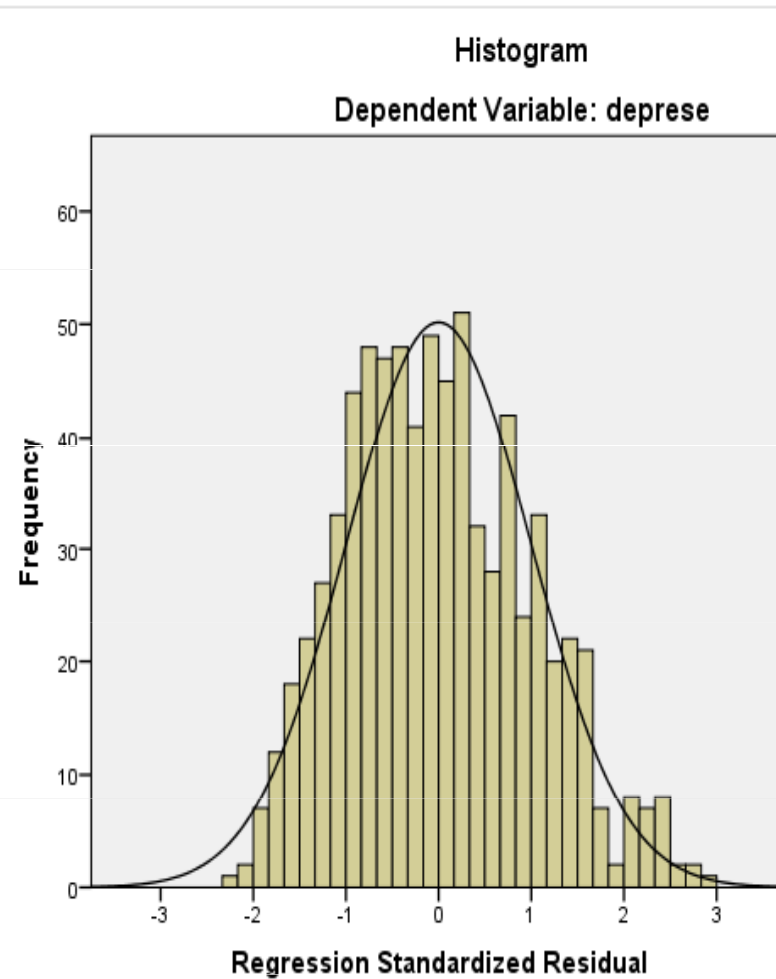
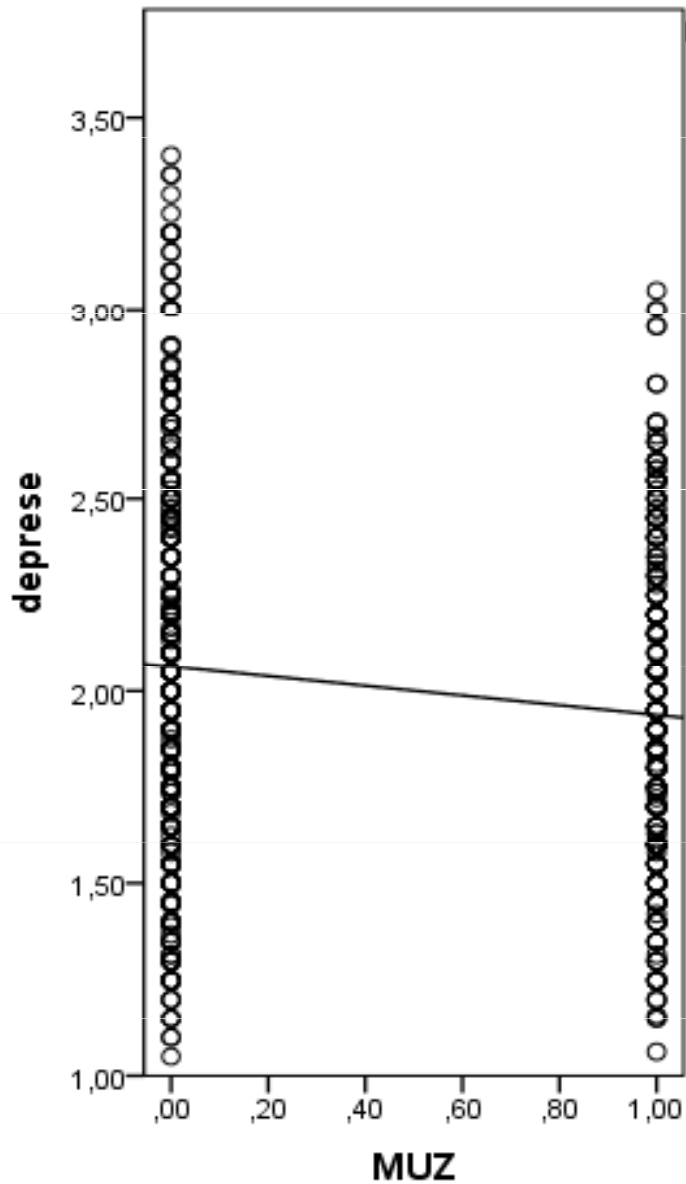
Je-li $MUZ=1$, pak $DEP' = 2,064 - 0,126 * 1 = 1,938$

- Průměr skupiny kódované 1 je roven $b_0 + b_{MUZ}$
- ... jeho regresní koeficient udává, o kolik se liší průměr skupiny kódované 1 od skupiny kódované 0.

- Při kódování 2/0 by udával polovinu rozdílu....
-

Předpoklady regrese stále platí

(kinda)



Kategorické prediktory s k hodnotami

Dummy coding -> dummy variables

- $k-1$ dichotomických proměnných nesoucích informaci obsaženou v původní kategorické p.
 - **Indikátorové kódování** (indicator coding)
 - Pro každou hodnotu vyjma jedné vytvoříme proměnnou, kde bude mít respondent 1, pokud tuto hodnotu má, jinak 0
 - Zbylá hodnota = **referenční kategorie**, i.e. ten, kdo má všechny dummies = 0
 - Zahrnutí všech $k-1$ dummy proměnných mezi prediktory
-

vzdel_ma vzdelani matky

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 základní	29	3,8	3,9	3,9
	2 vyučena	200	26,0	26,9	30,8
	3 střední (maturita)	321	41,8	43,2	74,0
	4 vysokoškolské	193	25,1	26,0	100,0
	Total	743	96,7	100,0	
Missing	0	25	3,3		
Total		768	100,0		

Vzdělání matky	Původní kód	Indikátorové kódování – 3 nové dummy proměnné		
		vm_zak	vm_vyu	vm_str
základní	1	1	0	0
vyučena	2	0	1	0
střední	3	0	0	1
vysokoškolské	4	0	0	0

vm_zak					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	714	93,0	96,1	96,1
	1,00	29	3,8	3,9	100,0
	Total	743	96,7	100,0	
Missing	System	25	3,3		
Total		768	100,0		

vm_vyu					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	543	70,7	73,1	73,1
	1,00	200	26,0	26,9	100,0
	Total	743	96,7	100,0	
Missing	System	25	3,3		
Total		768	100,0		

vm_str					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	422	54,9	56,8	56,8
	1,00	321	41,8	43,2	100,0
	Total	743	96,7	100,0	
Missing	System	25	3,3		
Total		768	100,0		

vzdel_ma	vm_zak	vm_vyu	vm_str
3	0	0	1
1	1	0	0
3	0	0	1
4	0	0	0
4	0	0	0
3	0	0	1
3	0	0	1
4	0	0	0
3	0	0	1
3	0	0	1
3	0	0	1
3	0	0	1
4	0	0	0
1	1	0	0
3	0	0	1
4	0	0	0
4	0	0	0
4	0	0	0
3	0	0	1
3	0	0	1
4	0	0	0
4	0	0	0
2	0	1	0

recode vzdel_ma (1=1) (2=0) (3=0) (4=0) into vm_zak.

recode vzdel_ma (1=0) (2=1) (3=0) (4=0) into vm_vyu.

recode vzdel_ma (1=0) (2=0) (3=1) (4=0) into vm_str.

execute.

vzdel_ma vzdelani matky	Mean	N	Std. Deviation
1 základní	2,1306	29	,43803
2 vyučena	2,0414	193	,45984
3 střední (maturita)	2,0071	319	,47061
4 vysokoškolské	1,9523	191	,46245
Total	2,0067	732	,46532

REGRESSION /DEPENDENT deprese
 /METHOD=ENTER vm_zak vm_vyu vm_str.

Coefficients^a

Model		Unstandardized Coefficients		Standardized	t	Sig.
		B	Std. Error	Coefficients		
1	(Constant)	1,952	,034		58,093	,000
	vm_zak	,178	,093	,075	1,927	,054
	vm_vyu	,089	,047	,085	1,881	,060
	vm_str	,055	,042	,058	1,290	,198

Jsou-li dummy kódované 0/1...

$$DEP' = 1,95 + 0,18vm_zak + 0,09vm_vyu + 0,05vm_str$$

Je-li **VŠ**, pak $DEP' = \mathbf{1,95} + 0,18*0 + 0,09*0 + 0,05*0$

- Při jednom kat. prediktoru udává b_0 průměr **referenční** skupiny (kódované 0 ve všech dummy)

Je-li **SŠ**, pak $DEP' = \mathbf{1,95} + 0,18*0 + 0,09*0 + 0,05*\mathbf{1} = 2,00$

- Průměr skupiny dané dummy je roven $b_0 + b_H$
 - ... jejich regresní koeficienty udávají, o kolik se liší průměr skupiny dané dummy proměnnou od referenční skupiny.
-

Interpretace regresních koeficientů dummy proměnných

□ $Y = b_0 + b_{A1}X_{A1} + b_{A2}X_{A2} + \dots + b_mX_m + e$

□ Indikátorové kódování

- b_{Ai} udává rozdíl průměrných hodnot Y mezi indikovanou skupinou a referenční skupinou, *ceteris paribus*
- b_{Ai} udává o kolik nám členství ve skupině zvyšuje/snižuje predikovanou hodnotu oproti referenční skupině
- b_0 udává (při absenci jiných prediktorů) průměr Y v referenční skupině

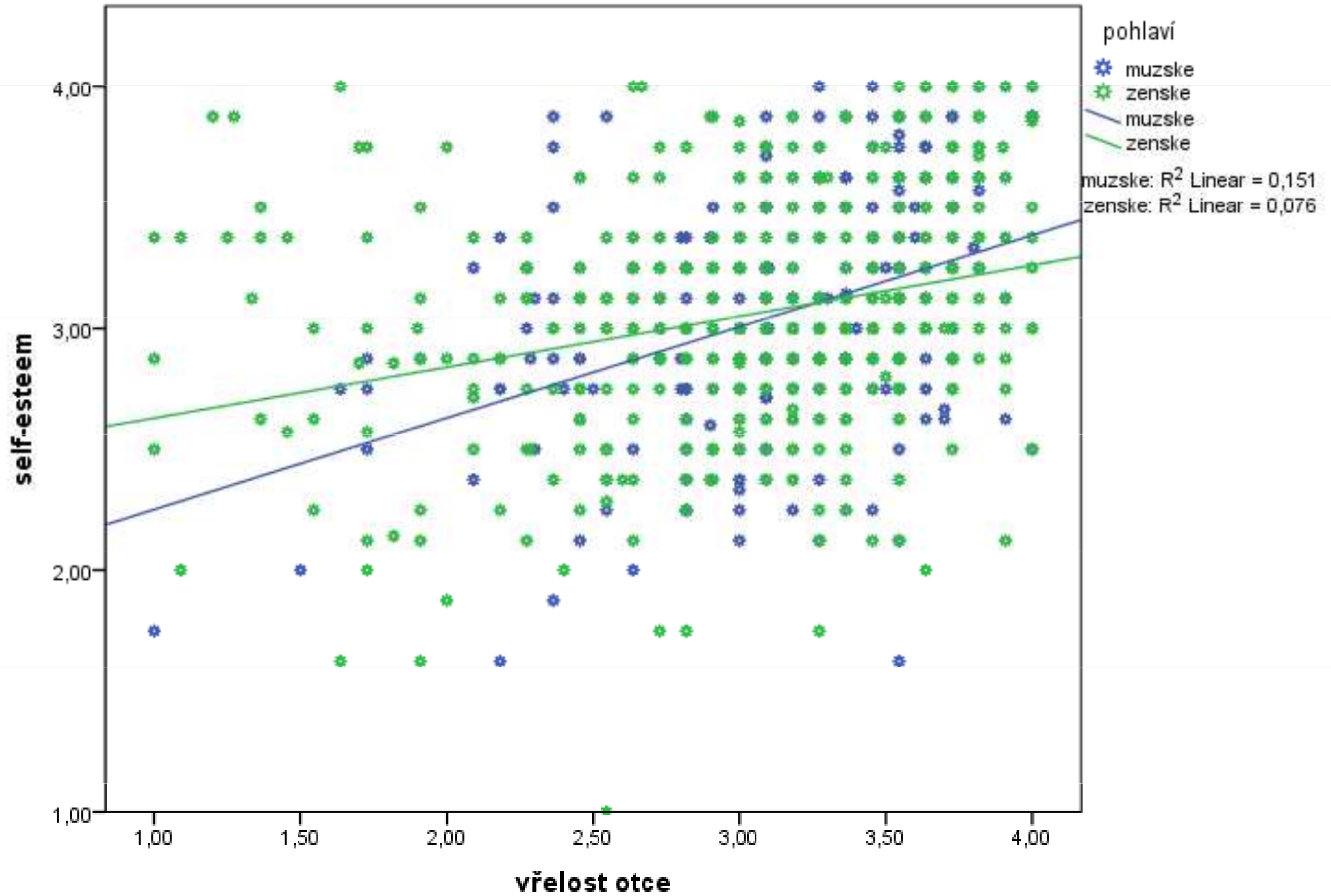
□ Testy koeficientů jsou ekvivalentní t-testům rozdílu průměrů mezi indikovanou a referenční skupinou

Jiné než indikátorové kódování

- Dummy proměnné nám umožňují srovnávat průměry skupin – ale dostaneme jen některá srovnání – **volba referenční skupiny!**
 - Máme-li více než 2 skupiny, roste množství srovnání, které bychom mohli chtít učinit.
 - Př. VŠ vs nižší vzdělání matky? SŠ vs VYU?
 - Tomu odpovídá i široká paleta způsobů jimiž můžeme vytvořit dummy proměnné.
 - **KONTRASTY** (contrasts)
-



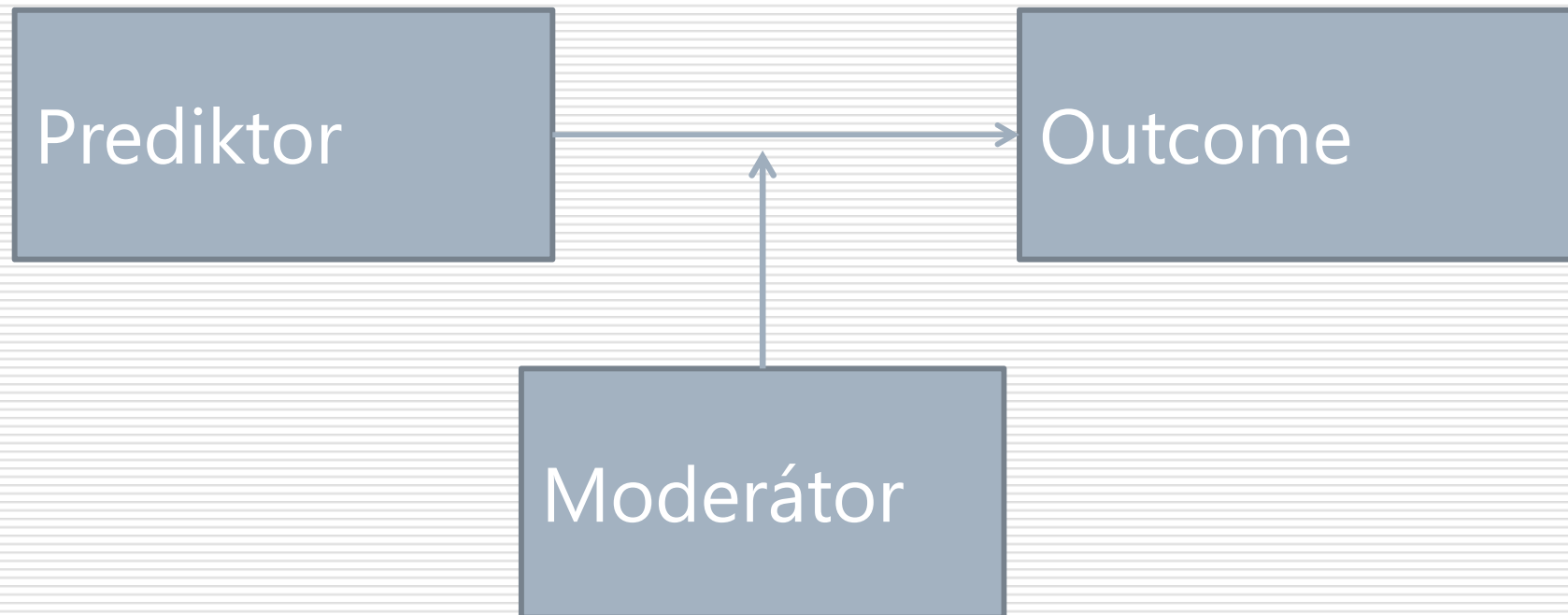
Moderace(interakce)



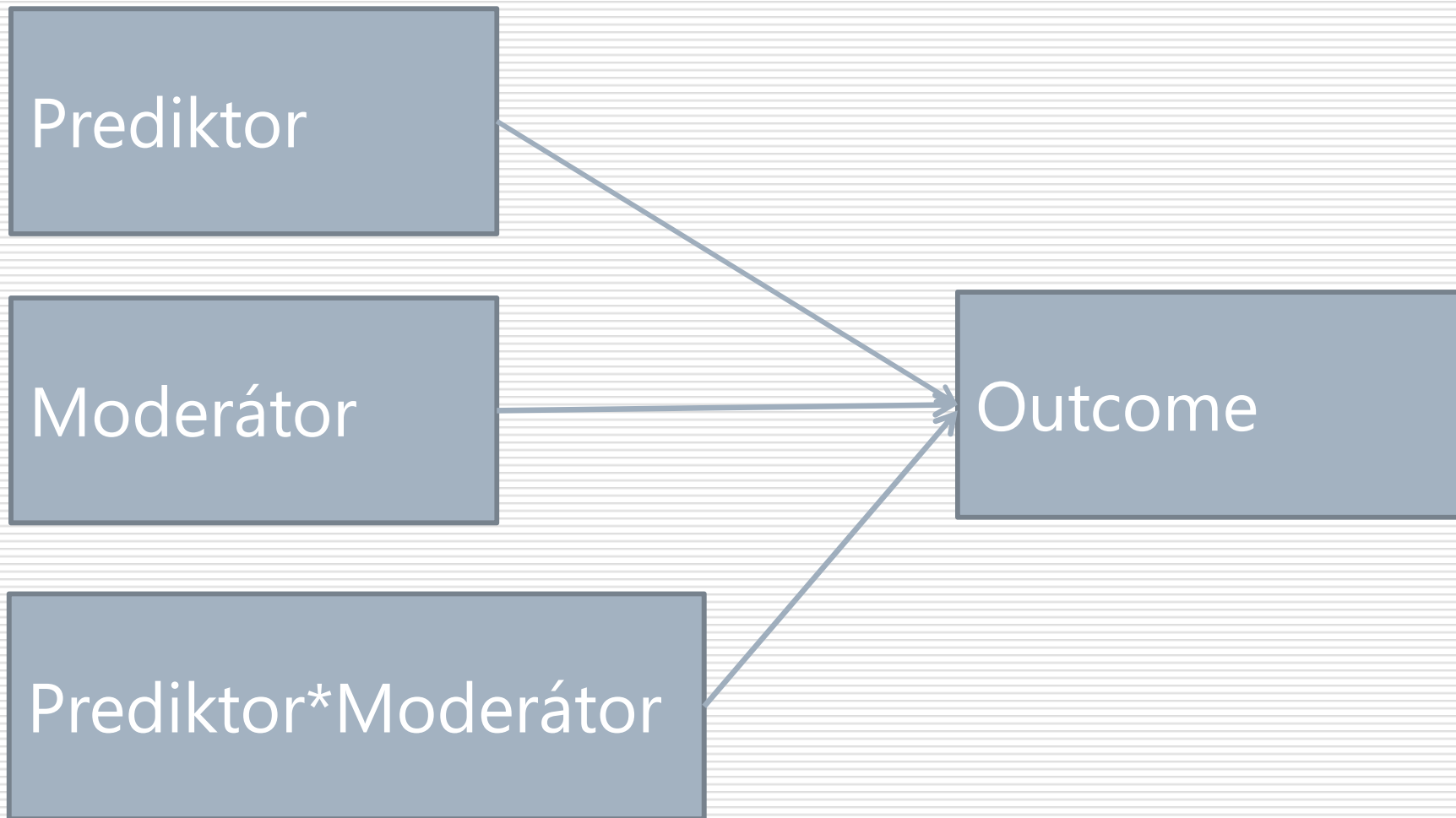
Moderace(interakce)

- Moderací se rozumí situace, kdy vztah mezi prediktorem a outcomem je jiný v různých skupinách vymezených kategorickou proměnnou – MODERÁTOREM
 - Historicky moderace = tlumení vztahu
-

Moderace konceptuálně



Moderace jako součást lineárního modelu



Moderace jako součást lineárního modelu

- Závislou tak predikujeme pomocí
 - prediktoru
 - moderátoru (nelze ho vynechat!)
 - interakčního členu, který si jako novou proměnnou spočítáme jako násobek prediktoru a moderátoru

- $$Y = b_0 + b_X X + b_M M + b_{XM} XM + e$$

Příklad moderace

- Závislá: Selfe – self-esteem
- Prediktor: warm_o – vřelost otce
- Moderátor: pohlaví

```
RECODE POHLAVI (1=1) (2=0) INTO MUZ.  
COMPUTE VOXMUZ=warm_o*MUZ.  
EXECUTE.
```

```
REGRESSION /DEPENDENT selfe  
           /METHOD=ENTER warm_o MUZ VOXMUZ.
```

Model		Unstandardized Coefficients		Standardized	t	Sig.
		B	Std. Error	Coefficients		
1	(Constant)	2,419	,107		22,707	,000
	warm_o vřelost otce	,211	,034	,255	6,156	,000
	MUZ	-,543	,203	-,554	-2,677	,008
	VOxMUZ	,167	,064	,543	2,592	,010

a. Dependent Variable: selfe self-esteem

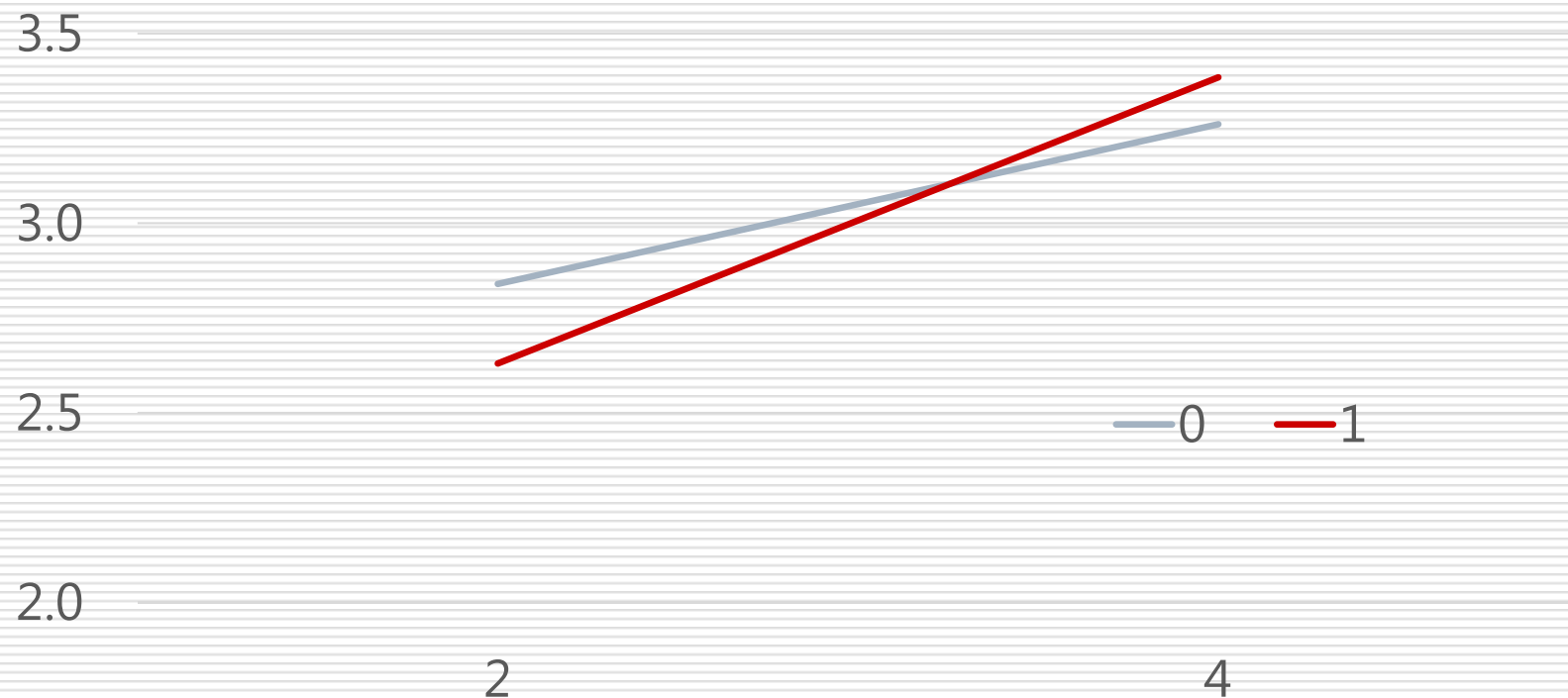
$$\text{SELFE}' = 2,4 + 0,2\text{warm}_o - 0,5\text{MUZ} + 0,2\text{VOxMUZ}$$

Interpretace

- Je-li $\text{MUZ}=0$, pak $\text{SELFE}' = 2,4 + 0,2\text{warm}_o$
 - Je-li $\text{MUZ}=1$, pak $\text{SELFE}' = 2,4 + 0,2\text{warm}_o - 0,5 + 0,2(\text{warm}_o * \text{MUZ})$
 $\text{SELFE}' = 2,4 + 0,2\text{warm}_o - 0,5 + 0,2\text{warm}_o$
 $\text{SELFE}' = 2,4 + 0,4\text{warm}_o - 0,5$
 - b_X udává efekt X na Y ve skupině, kde $M=0$
 - b_{MX} udává, o kolik je efekt X na Y větší, když $M=1$
- Obecně: b_{MX} udává, o kolik se mění b_X při jednotkové změně M

Moderace - zobrazení

- Simple slopes – regresní přímky pro vybrané hodnoty moderátoru (všechny, -1SD, M, 1SD)



-
- alternativně online tvořítka na grafy jako např. <https://www.desmos.com/calculator>
 - nebo R
-

Závěrečné poznámky k moderaci

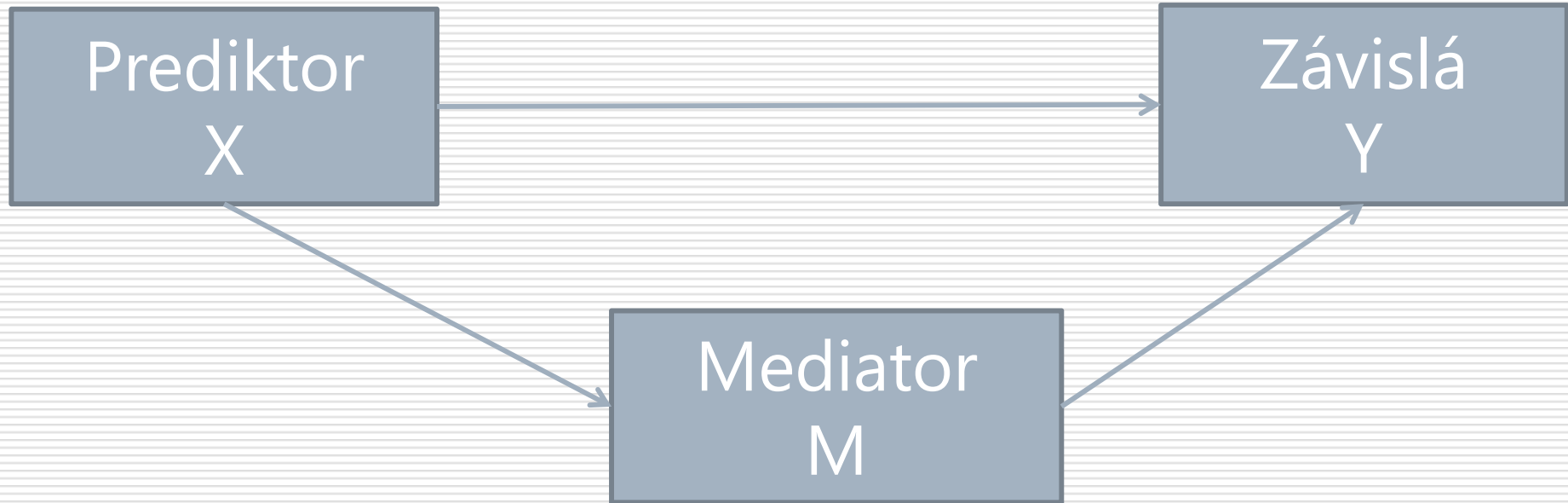
- Moderátor může být i spojitá proměnná.
 - Pak je dobré ji centrovat (popř. i prediktor), aby byla interpretace jednodušší (b_x při $M=0$) a aby se snížila kolinearita
 - Obecnějším termínem pro moderaci je **interakce** – může být mezi všemi typy proměnných a realizuje se stejně, tj. násobením
 - Interakce může být i vyššího řádu – 3 i více proměnných. Prudce rostou nároky na interpretaci.
 - **PROCESS** – plugin do SPSS usnadňující odhadování modelů s moderací (a dalších modelů) – viz Field
-



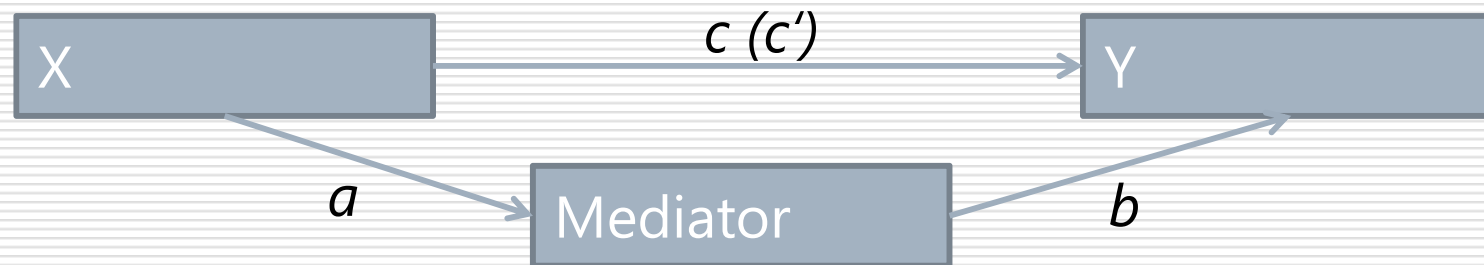
MEDIACE

- ❑ MODERACE a MEDIACE jsou prototypickým zapojením třetí proměnné do vztahu mezi dvěma proměnnými
 - ❑ MODERÁTOR je obvykle kategorická proměnná, která mění (historicky snižuje-moderuje) těsnost vztahu mezi X a Y
 - ❑ **MEDIÁTOR je proměnná, skrze níž se odehrává vztah mezi X a Y. Vztah mezi X a Y je pouze zdánlivý, protože X ve skutečnosti ovlivňuje Mediátor a Mediátor následně ovlivňuje Y.**
 - ❑ Terminologii a statistiku v tomto směru ustavili před 25 lety Baron a Kenny, <http://davidakenny.net/kenny.htm>
-

MEDIACE



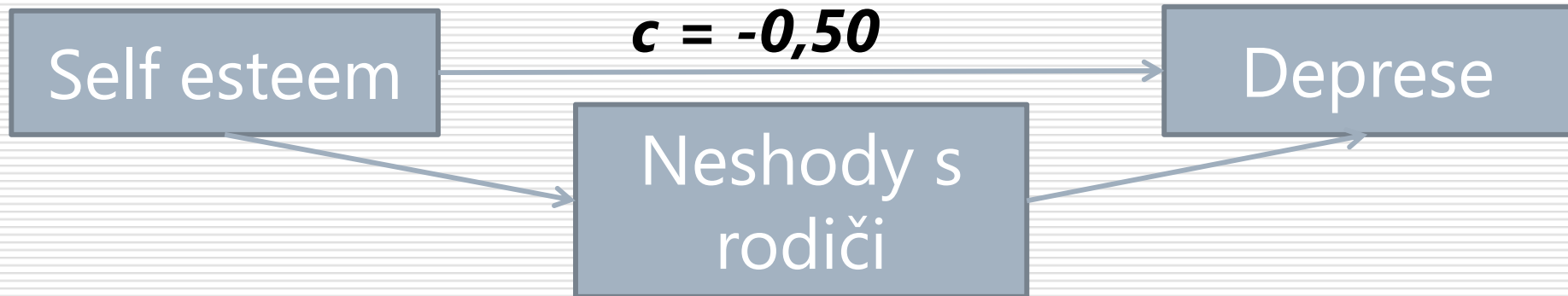
MEDIACE (Baron-Kenny, moderně)



1. X predikuje Y : regr. koef. **c** (**celkový efekt**)
2. X predikuje M : regr. koef. **a**
3. M predikuje Y , je-li X kontrolována regr. koef. **b**
4. Efekt X na Y po zařazení M (**c'** , **přímý efekt**) klesne (ideálně na 0)
5. **Nepřímý efekt** X na Y (přes M) (**ab**) se statisticky významně liší od 0 – Sobelův test

$$z = \frac{ab}{\sqrt{(b^2SE_a^2) + (a^2SE_b^2)}}$$

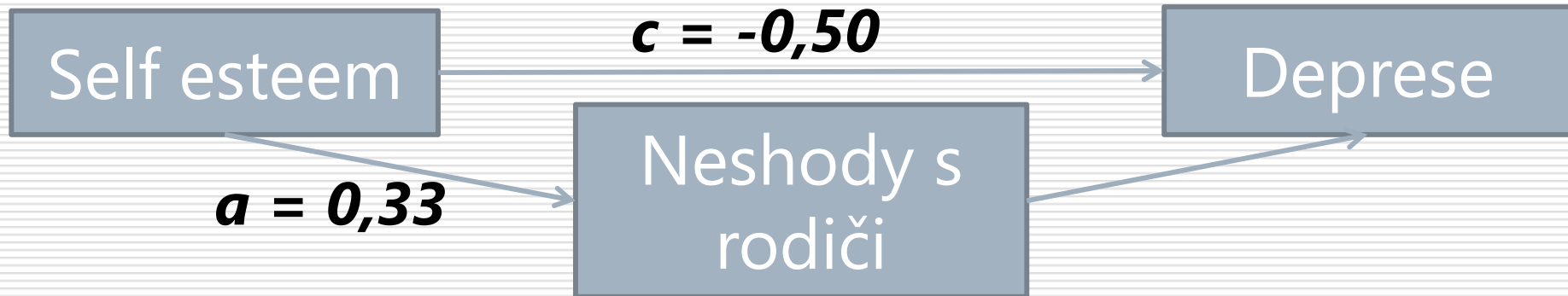
Mediace - příklad



1. Regrese DEP na SE

	B	Std. Error	Beta	t
(Constant)	3,553	,094		37,8
selfe self-esteem	-,504	,030	-,517	-16,5

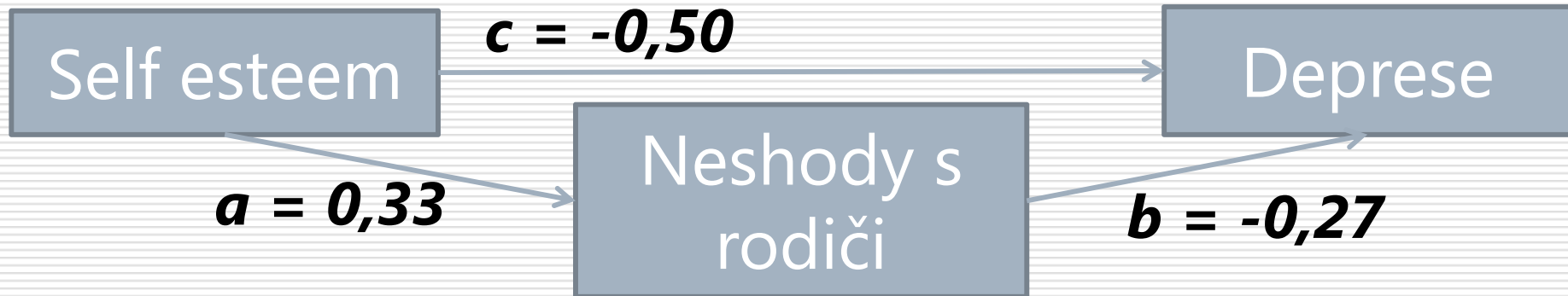
Mediace - příklad



2. Regrese Neshod na SE

	B	Std. Error	Beta	t
(Constant)	2,113	,132		16,0
selfe self-esteem	,328	,043	,271	7,6

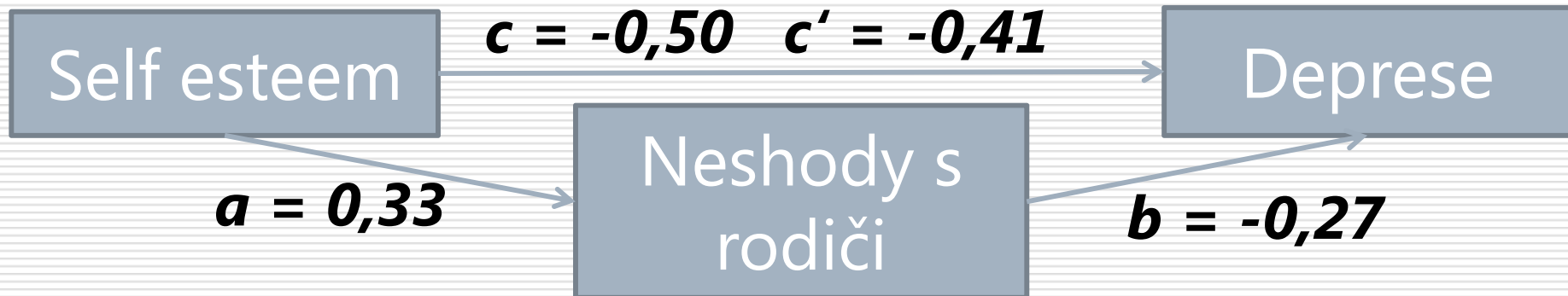
Mediace - příklad



3. Regrese DEP na Neshody a SE

	B	Std. Error	Beta	
(Constant)	4,105	,102		40
selfe self-esteem	-,409	,030	-,419	-13
neshody Neshody s rodiči (0)	-,270	,024	-,338	-11

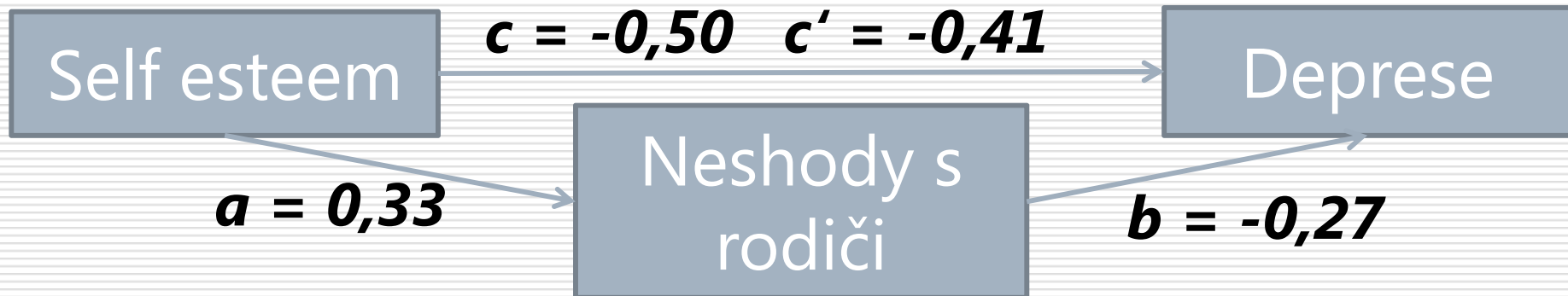
Mediace - příklad



4. Snížení c ?

	B	Std. Error	Beta	
(Constant)	4,105	,102		40
selfe self-esteem	-,409	,030	-,419	-13
neshody Neshody s rodiči (O)	-,270	,024	-,338	-11

Mediace - příklad



5. Velikost nepřímého efektu

□ Velikost = $a \cdot b = 0,33 \cdot 0,27 = 0,09$

□ Část. standardizovaná = $0,09 / 0,47 = 0,04$

□ $z = 0,09 / \text{odm}(0,27^2 \cdot 0,043^2 + 0,33^2 \cdot 0,024^2) = 6,3,$

$p < 0,001$

$$z = \frac{ab}{\sqrt{(b^2 SE_a^2) + (a^2 SE_b^2)}}$$

Mediace – příklad - PROCESS

<http://www.processmacro.org/index.html>

***** PROCESS Procedure for SPSS Release 2.16.3 *****

Written by Andrew F. Hayes, Ph.D. www.afhayes.com

Model = 4

Y = deprese

X = selfe

M = neshody

Sample size

748

Outcome: neshody

Model Summary

R	R-sq	MSE	F	df1	df2	p
,272	,074	,315	59,770	1,000	746,000	,000

Model

	coeff	se	t	p	LLCI	ULCI
constant	2,099	,133	15,805	,000	1,839	2,360
selfe	,332	,043	7,731	,000	,247	,416

Velikost mediačního (nepřímého) účinku

Záleží na tom, co chceme komunikovat

- Absolutní: Nestandardizovaná i standardizovaná **$a*b$**
 - efekt jednotkové změny prediktoru v jednotkách závislé (násobíme-li β , je to v SD)
 - Relativní vůči celkovému účinku **$a*b/c$** či vůči přímému účinku **$a*b/c'$**
 - intuitivní, ale nestabilní, zrádné kvůli možné změně znamének
 - R^2 vysvětlené mediačním efektem (Field 11.11)
 - Podíl rozptylu Y vysvětlený tou částí rozptylu, kterou X a M sdílí.
-

Mediace – závěrečné poznámky

- Mediační model je **explicitně kauzální** – je třeba umět obhájit směr šipek.
 - Intervaly spolehlivosti (i SE) na nepřímé efekty jsou spolehlivěji stanovovány bootstrapem (PROCESS)
 - Tato podoba mediačního modelu je mezistupněm mezi regresí a SEM modely
-



POWER ANALÝZA REGRESNÍHO MODELU

□ G*Power - <http://www.gpower.hhu.de/>

The screenshot displays the G*Power 3.1.9.2 software interface. The main window shows a graph of central and noncentral distributions. The central distribution is a solid red curve centered at 0, and the noncentral distribution is a dashed blue curve shifted to the right. A vertical green line marks the critical t value at 2.03951. The area under the noncentral curve to the left of the critical t is shaded blue and labeled β . The area under the central curve to the right of the critical t is shaded red and labeled $\alpha/2$.

The interface includes the following sections:

- Test family:** t tests
- Statistical test:** Linear bivariate regression: One group, size of slope
- Type of power analysis:** A priori: Compute required sample size - given α , power, and effect size
- Input Parameters:**
 - Tail(s): Two
 - Slope H1: 0.4941850
 - α err prob: 0.05
 - Power ($1 - \beta$ err prob): 0.90
 - Slope H0: 0
 - Std dev σ_x : 0.481
 - Std dev σ_y : 0.467
- Output Parameters:**
 - Noncentrality parameter δ : 3.3969515
 - Critical t: 2.0395134
 - Df: 31
 - Total sample size: 33
 - Actual power: 0.9082429
- Input mode:** $\rho, \sigma_x, \sigma_y \Rightarrow$ slope
 - Correlation ρ : 0.509
 - Std dev σ_x : 0.481
 - Std dev σ_y : 0.467
 - Calculate: Slope H1 0.494185
 - Calculate and transfer to main window
 - Close

$$H_0: R^2 = 0$$

$$f^2 = R^2 / (1 - R^2)$$

The screenshot displays the G*Power software interface for a power analysis. The main window is titled "Test family" and "Statistical test". The "Test family" is set to "F tests" and the "Statistical test" is "Linear multiple regression: Fixed model, R² deviation from zero". The "Type of power analysis" is "A priori: Compute required sample size - given α , power, and effect size".

The "Input Parameters" section includes:

- Determine =>
- Effect size f^2 : 0.4285714
- α err prob: 0.05
- Power (1 - β err prob): 0.90
- Number of predictors: 6

The "Output Parameters" section includes:

- Noncentrality parameter λ : 20.5714272
- Critical F: 2.3297712
- Numerator df: 6
- Denominator df: 41
- Total sample size: 48
- Actual power: 0.9051158

A secondary window is open, showing the "From correlation coefficient" option selected. The "Squared multiple correlation ρ^2 " is set to 0.3. The "From predictor correlations" option is unselected, with "Number of predictors" set to 3 and "Squared multiple correlation ρ^2 " set to ?. A red box highlights the R^2 symbol, with an arrow pointing to the 0.3 value in the "Squared multiple correlation ρ^2 " field. The "Calculate" button is highlighted in blue, and the "Calculate and transfer to main window" button is also highlighted in blue. The "Close" button is visible at the bottom right.

$$H_0: \Delta R^2 = 0$$

$$f^2 = \Delta R^2 / (1 - R^2)$$

Test family: F tests

Statistical test: Linear multiple regression: Fixed model, R² increase

Type of power analysis: A priori: Compute required sample size - given α, power, and effect size

Input Parameters

Determine =>

Effect size f ²	0.15
α err prob	0.05
Power (1-β err prob)	0.95
Number of tested predictors	2
Total number of predictors	5

Output Parameters

Noncentrality parameter λ	?
Critical F	?
Numerator df	?
Denominator df	?
Total sample size	?
Actual power	?

N prediktorů v testovaném bloku

N prediktorů celkem

From variances

Variance explained by special effect: 0.1

Residual variance: 0.8

Direct

Partial R²: 0.111111

Calculate

Effect size f²: 0.125

Calculate and transfer to main window

Close

X-Y plot for a range of values

Calculate

$$H_0: b_i = 0$$

- Pokud nás zajímá jen 1 prediktor, je to jako testovat, zda jeho přidání v samostatném závěrečném bloku přidá nějaký R².
- Pokud nás zajímá k prediktorů, upravíme α na α/k a počítáme jako pro 1 prediktor (počítáme se stejnou velikostí účinku

The screenshot shows a software interface for calculating power in a linear regression model. It is divided into several sections:

- Test family:** t tests
- Statistical test:** Linear multiple regression: Fixed model, single regression coefficient
- Type of power analysis:** A priori: Compute required sample size - given α , power, and effect size
- Input Parameters:**
 - Tail(s): Two
 - Effect size f^2 : 0.0833333
 - α err prob: 0.05
 - Power ($1 - \beta$ err prob): 0.90
 - Number of predictors: 4
- Output Parameters:**
 - Noncentrality parameter δ : 3.2787186
 - Critical t: 1.9792801
 - Df: 124
 - Total sample size: 129
 - Actual power: 0.9020390
- Calculation Method:** From variances (selected)
 - Variance explained by predictor: 0.05
 - Residual variance: 0.6
- Direct Method:** Partial R²: 0.07692308
- Buttons:** Determine =>, Calculate, Calculate and transfer to main window, Close

Síla testu a velikost vzorku v MLR

Přibývá nový faktor síly testu: **množství prediktorů**

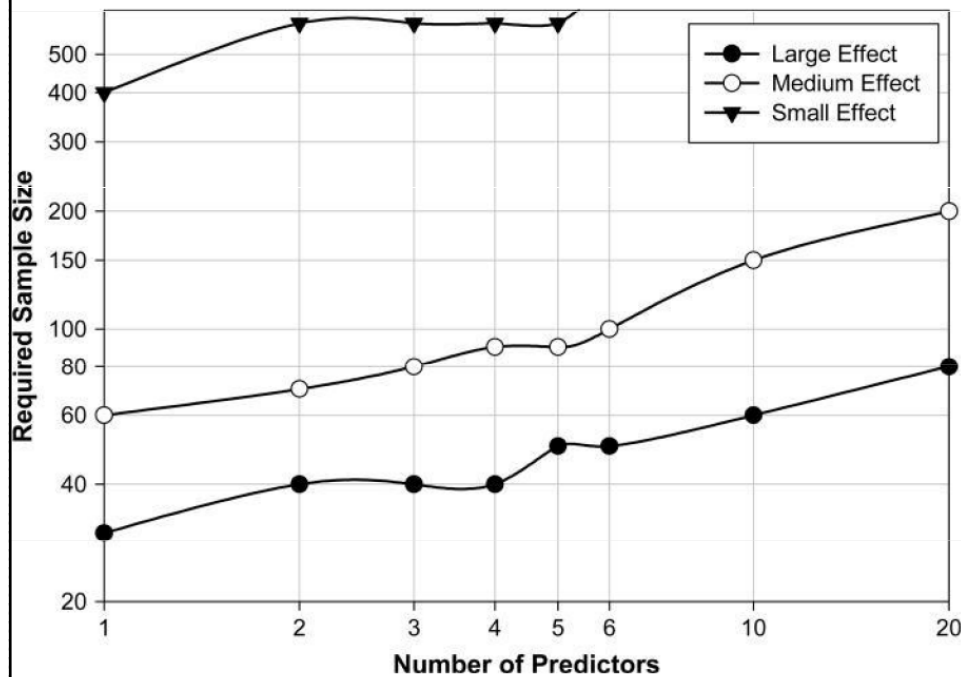


TABLE 5 Minimum R^2 That Can Be Found Statistically Significant with a Power of .80 for Varying Numbers of Independent Variables and Sample Sizes

Sample Size	Significance Level (α) = .01				Significance Level (α) = .05			
	No. of Independent Variables				No. of Independent Variables			
	2	5	10	20	2	5	10	20
20	45	56	71	NA	39	48	64	NA
50	23	29	36	49	19	23	29	42
100	13	16	20	26	10	12	15	21
250	5	7	8	11	4	5	6	8
500	3	3	4	6	3	4	5	9
1,000	1	2	2	3	1	1	2	2

Note: Values represent percentage of variance explained.
NA = not applicable.



Lineární regrese jako submodel generalizovaného lineárního modelu

Generalized Linear Models

Type of Model | Response | Predictors | Model | Estimation | Statistics | EM Means | Save | Export

Choose one of the model types listed below or specify a custom combination of distribution and link function.

- Scale Response
 - Linear
 - Gamma with log link
- Ordinal Response
 - Ordinal logistic
 - Ordinal probit
- Counts
 - Poisson loglinear
 - Negative binomial with log link
- Binary Response or Events/Trials Data
 - Binary logistic
 - Binary probit
 - Interval censored survival
- Mixture
 - Tweedie with log link
 - Tweedie with identity link
- Custom
 - Custom



Type of Model

Response

Predictors

Model

Estimation

Statistics

EM Means

Save

Export

Variables:

- vztah s rodiči [vztahsr]
- Individualismus [individ1]
- Individualismus 2 [individ2]
- Individualismus 3 [individ3]
- Negativní pocity - osamělost, strach, v...
- Negativní pocity - únava, stres [neg2]
- Negativní pocity - agresivní [neg3]
- coping - rodina - potlačení emoci [ro_...
- coping - rodina - potlačení [ro_pot]
- coping - rodina - aktivní koping [ro_akt]
- coping - rodina - hledání sociální opo...
- coping - škola - potlačení emoci [sk_...
- coping - škola - potlačení [sk_pot]
- coping - škola - aktivní koping [sk_akt]
- coping - škola - hledání sociální opor...
- optimismus [optim]
- zivotni spokojenost [ziv_sp]
- self-efficacy [effi]
- zdravotní potíže psychosomatického r...
- výkonová motivace - na úkol [vm_ukol]
- výkonová motivace - na sebe [vm_na...
- výkonová motivace - namáhavá-nech...
- deprese [deprese]
- deprese - negativní položky [depneg]
- deprese - pozitivní položky [depnoz]

Dependent Variable

Dependent Variable:

self-esteem [selfe]

Category order (multinomial only):

Ascending

Type of Dependent Variable (Binomial Distribution Only)

Binary

Reference Category...

Number of events occurring in a set of trials

Trials

Variable

Trials Variable:

Fixed value

Number of Trials:

Scale Weight

Scale Weight Variable:

Type of Model Response Predictors Model Estimation Statistics EM Means Save Export

Variables:

- id
- škola [skola]
- název třídy [trida]
- kohorta [kohorta]
- datum narozeni [dat_nar]
- vek v mesicich [vekm]
- vek v letech [vekr]
- věk v letech (zaokrouhleny) [vek]
- národnost [narodnos]
- známka z matematiky 1999 [mat99]
- známka z češtiny 1999 [cj99]
- známka z angličtiny 1999 [aj99]
- známka z němčiny 1999 [nj99]
- oblíbeny predmet [pr_oblib]
- neoblíbeny predmet [pr_neobl]
- ocekavane vzdelani [ocek_vzd]
- stav rodicu [stav_r99]
- pocet mladsich bratru [bratri_m]
- pocet mladsich sester [sestry_m]
- pocet starsich bratru [bratri_s]
- pocet starsich sester [sestry_s]
- pocet nevlastnich bratru [bratri_n]
- pocet nevlastnich sester [sestry_n]
- spolubydlíci [spolub99]
- jaký jiný spolubydlíci [spolubv2]

Factors:

pohlaví [pohlavi]

KATEGORICKÉ

VOLBA REFERENČNÍ K.

Options...

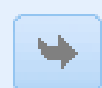
Covariates:

vřelost otce [warm_o]

METRICKÉ

Offset

Variable



Offset Variable:

[Empty text box]

Fixed value

Value:

[Empty text box]

Type of Model

Response

Predictors

Model

Estimation

Statistics

EM Means

Save

Export

Specify Model Effects

Factors and Covariates:

pohlavi
 warm_o

Build Term(s)

Type:

Interaction



Model:

pohlavi
warm_o
pohlavi*warm_o



**Které z prediktorů
zahrnout do modelu?
Chceme vytvořit i
interakční člen?**



Number of Effects in Model: 3

Build Nested Term

Term:

By *

(Within)

Add to Model

Clear

Type of Model Response Predictors Model Estimation Statistics EM Means Save Export

Parameter Estimation

Method:

Hybrid

Maximum Fisher Scoring Iterations:

1

Scale Parameter Method:

Maximum likelihood estimate

Value:

1

Covariance Matrix

 Model-based estimator Robust estimator Get initial values for parameter estimates from a dataset

Initial Values...

Nechat být.
Při podezření na
heteroskedascitu,
zvolit robustní
estimátor

Iterations

Maximum Iterations:

100



Check for separation of data points

Maximum Step-Halving:

5

Starting Iteration:

20

Convergence Criteria

At least one convergence criterion must be specified with a minimum greater than 0.

	Minimum:	Type:
<input checked="" type="checkbox"/> Change in parameter estimates	1E-006	Absolute
<input type="checkbox"/> Change in log-likelihood		Absolute
<input type="checkbox"/> Hessian convergence		Absolute

Type of Model

Response

Predictors

Model

Estimation

Statistics

EM Means

Save

Export

Model Effects

Analysis Type:

Type III

Confidence Interval Level (%):

95

Chi-square Statistics

 Wald Likelihood ratio

Confidence Interval Type

 Wald Profile likelihood

PŘESNĚJŠÍ, POMALEJŠÍ

Tolerance level: ,0001

Log-Likelihood Function:

Full

Bootstrap...

Print

 Case processing summary Descriptive statistics Model information Goodness of fit statistics Model summary statistics Parameter estimates Include exponential parameter estimates Covariance matrix for parameter estimates Correlation matrix for parameter estimates Contrast coefficient (L) matrices

KONTRASTY

 General estimable functions Iteration history

Print interval: 1

 Lagrange multiplier test of scale parameter or negative binomial ancillary parameter

Type of Model

Response

Predictors

Model

Estimation

Statistics

EM Means

Save

Export

Factors and Interactions:

M	Term
<input checked="" type="checkbox"/>	pohlavi



By *

Display Means for:

Term	Contrast	Reference Category
<div style="border: 2px solid red; border-radius: 20px; padding: 20px; margin: 10px auto; width: 80%;">ZDE MOŽNOST ZVOLIT KONTRASTY PRO KATEGORICKÉ S VÍCE KATEGORIEMI</div>		

Scale

 Compute means for response Compute means for linear predictor

Adjustment for Multiple Comparisons:





Diagnostika 1: Outliery a vlivné případy

Nemají některé případy příliš velký vliv na výsledky regrese?

- Outliery – mohou zvyšovat i snižovat b
 - **Rezidua** – případy s vysokými r. regrese predikuje nejhůř, standardizovaná, studentizovaná ± 3
 - **Vlivné případy** – případy, které nejvíc ovlivňují parametry
 - Co se stane s parametry regrese, když případ odstraníme?
 - DFBeta – rozdíl mezi parametrem s a bez, standardizované > 1
 - DFFit – rozdíl mezi predikovanou hodnotou a predikovanou hodnotou bez případu (adjustovanou)
 - Cookova vzdálenost > 1
 - Leverage $> 2(k+1)/n$, kde k = počet prediktorů, n = velikost vzorku
 - Případy s vysokými rezidui či vlivné případy **NEODSTRAŇUJEME**
 - ...leďa by šlo o zjevnou chybu v datech či vzorku
 - ...leďa by nám šlo výhradně o zpřesnění predikce (nikoli o testy hypotéz)
-

Diagnostika 2: Kolinearita

- Když 2 prediktory vysvětlují tutéž část variability závislé, jeden z nich je téměř zbytečný
- Komplikuje porovnávání síly preditorů
- Snižuje stabilitu odhadu parametrů
- V extrému (když lze jeden prediktor přesně vypočítat z ostatních) regresi úplně znemožňuje

- Korelace nad 0,9
- **Tolerance (= 1/VIF) cca pod 0,1**
- (VIF (= 1/tolerance) cca nad 10)

I při korelacích kolem 0,5 komplikuje interpretaci!!

Diagnostika 3: Předpoklady regrese

- ❑ Závislá alespoň intervalová, prediktory intervalové i kategorické
 - ❑ Absence vysoké kolinearity (žádné $r > 0,9$, tolerance $< 0,1$)
 - ❑ Neexistence intervenující proměnné, která by korelovala se závislou i prediktory
 - ❑ Homoskedascita (scatterplot ZRESID x ZPRED, parciální scatterplot)
 - ❑ Nezávislost reziduí (Durbin-Watson = 2)
 - ❑ Normálně rozložená rezidua (histogram, P-P)
 - ❑ Linearita vztahů
-

