

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
 - Kategorické prediktory, kontrasty
 - Interakce-moderace
 - Mediace
 - Odhad potřebné velikosti vzorku
 - Lineární regrese pomocí GENLIN
-

MLR: Interpretace regresních koeficientů

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

- **B_i ; b_i** vyjadřuje nárůst Y' při nárůstu X_i o jednu jednotku; v jednotkách Y , při kontrole všech ostatních prediktorů (\approx semiparciální korelace); jedinečný přínos
 - K porovnání síly prediktoru v různých skupinách, modelech, vzorcích
 - **β_i ; b_i^* ; **BETA**** vyjadřuje nárůst Y' při nárůstu X_i o 1; jsou-li X_i i Y standardizovány, při kontrole všech ostatních prediktorů (\approx semiparciální korelace); jedinečný přínos
 - k porovnání prediktorů mezi sebou v rámci jednoho modelu
 - k porovnání různě operacionalizovaného prediktoru v různých modelech
 - ukazatel velikosti účinku
 - **b_0** – obtížně interpretovatelný průsečík ... leda by prediktory byly **centrované**
 - V různých modelech nemusí být vliv prediktoru stejný
-

Centrování – usnadnění interpretace

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

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

- $PREDIKTOR_C = PREDIKTOR - M_{PREDIKTOR}$

- *Následným podělením 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.
```



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. Depre	-,194	-,059
----------	-------	-------

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

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

$$\text{Je-li } MUZ=0, \text{ pak } DEP' = 2,064 - 0,126 * 0 = 2,064$$

- Při pouze jednom prediktoru udává b_0 průměr skupiny kódované 0

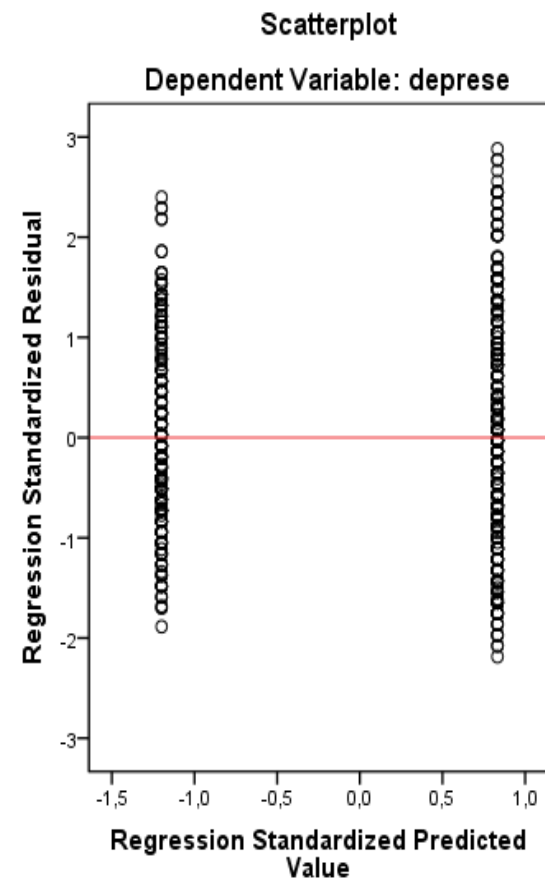
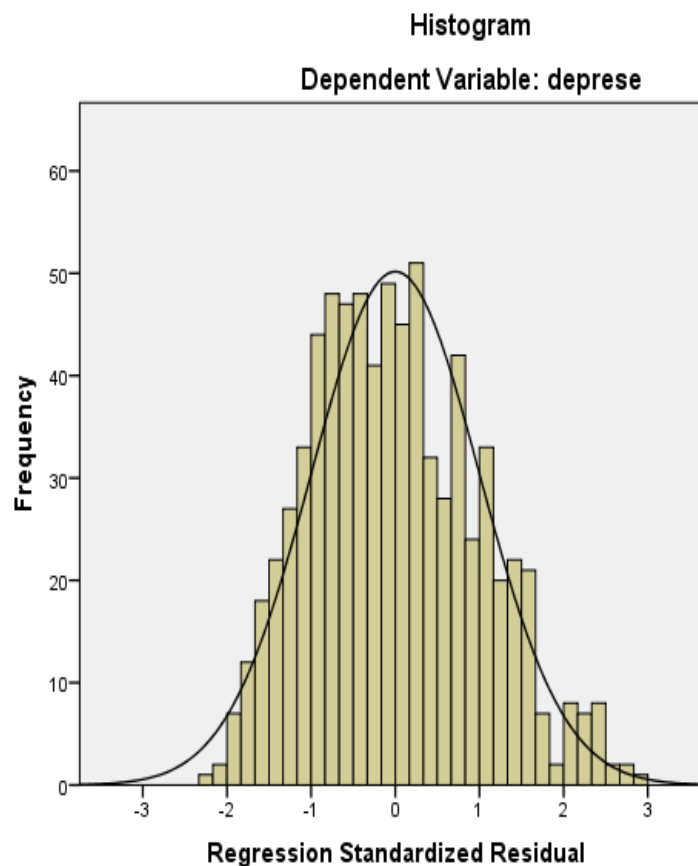
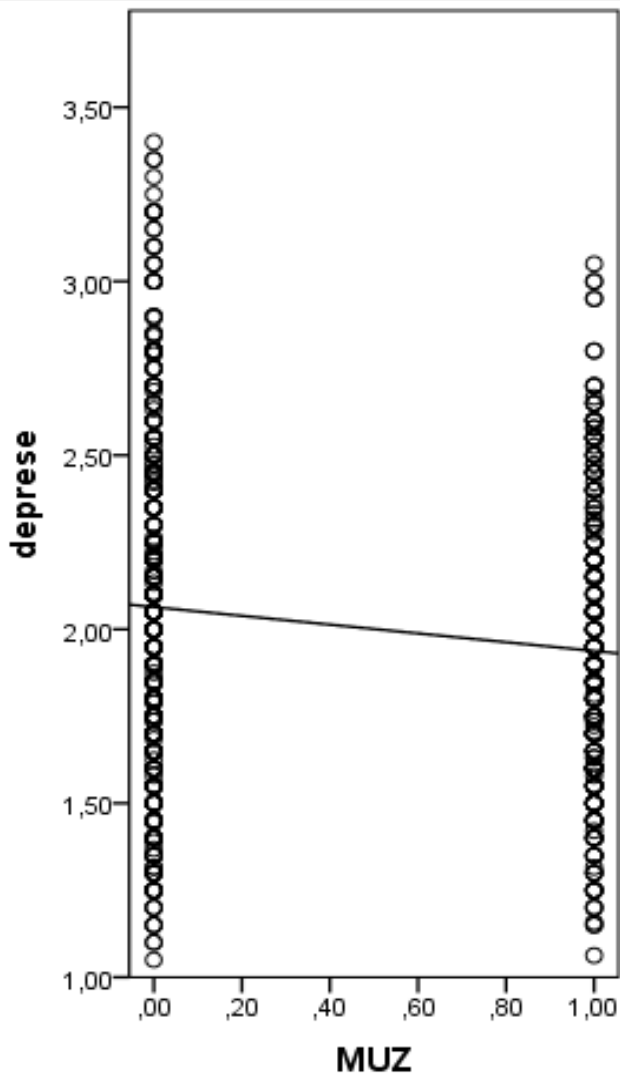
$$\text{Je-li } MUZ=1, \text{ pak } DEP' = 2,064 - 0,126 * 1 = 1,938$$

- 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
	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
	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 Coefficients	t	Sig.
		B	Std. Error	Beta		
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

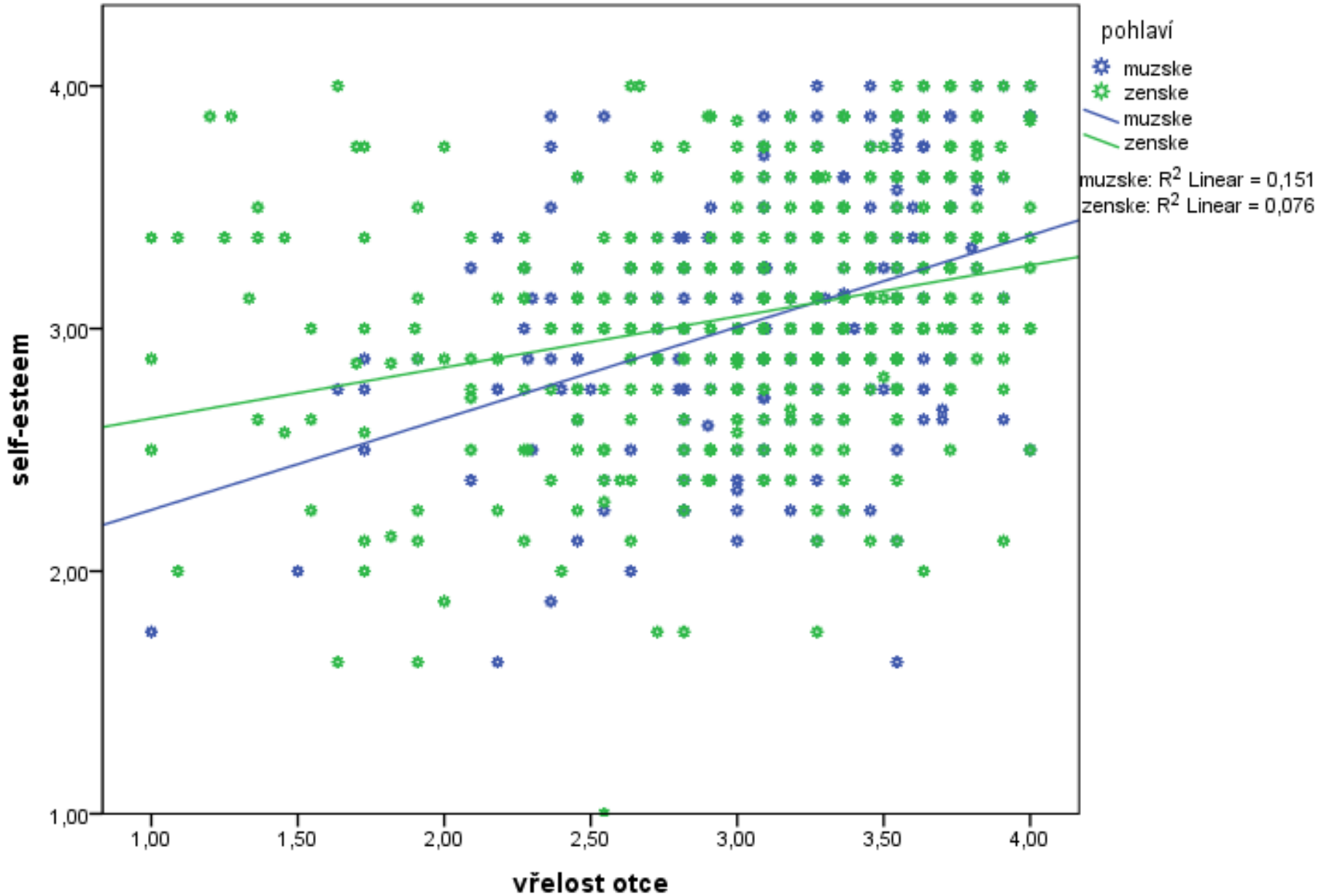
- $Y = b_0 + b_{A1}X_{A1} + b_{A2}X_{A2} + \dots + b_mX_m + e$
 - Indikátorové kódování
 - b_{Ai} udává rozdíl průměrných hodnot Y mezi indikovanou skupinou a referenční skupinou; sig b_{Ai} znamená sig rozdílu
 - b_{Ai} udává o kolik nám členství ve skupině zvyšuje/snižuje predikovanou hodnotu oproti referenční skupině
 - b_0 udává (při absenci jiných prediktorů) průměr Y v referenční skupině
 - Testy koeficientů jsou vlastně t-testy rozdílu průměrů mezi indikovanou a referenční skupinou
-

Jiné než indikátorové kódování

- Dummy proměnné nám umožňují srovnávat průměry skupin
 - 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)
-



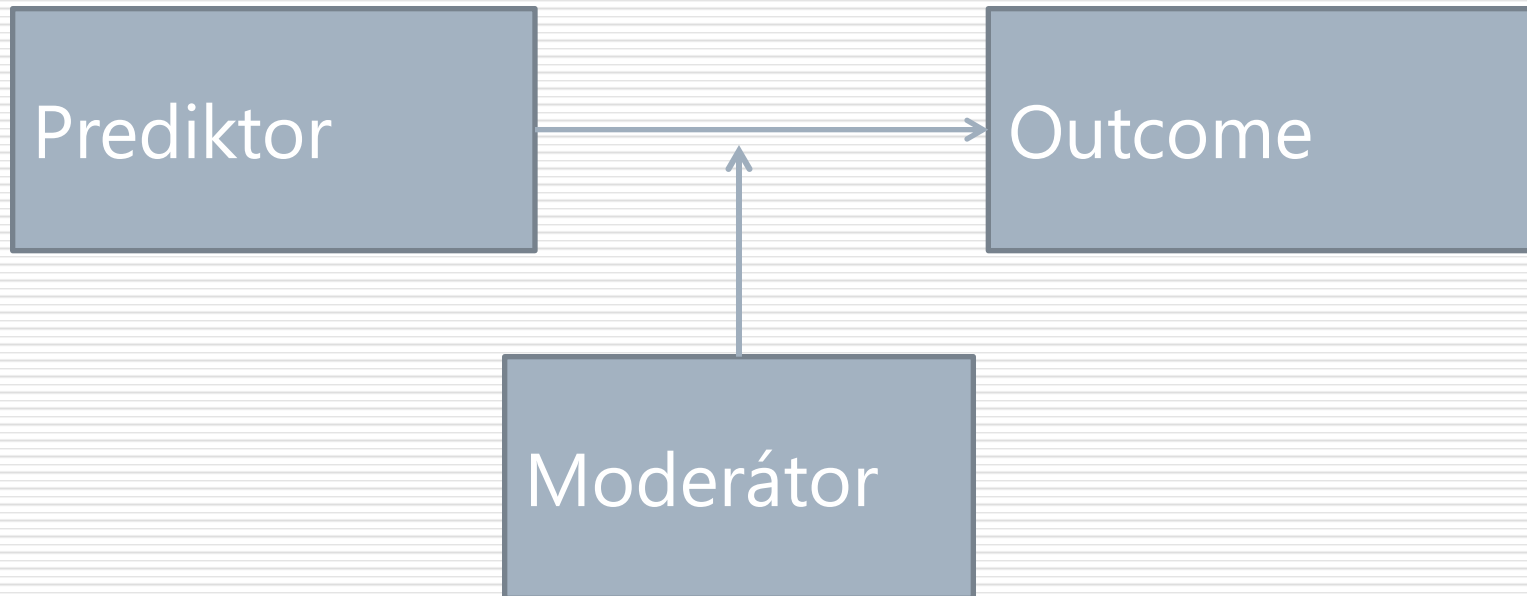
Moderace(interakce)



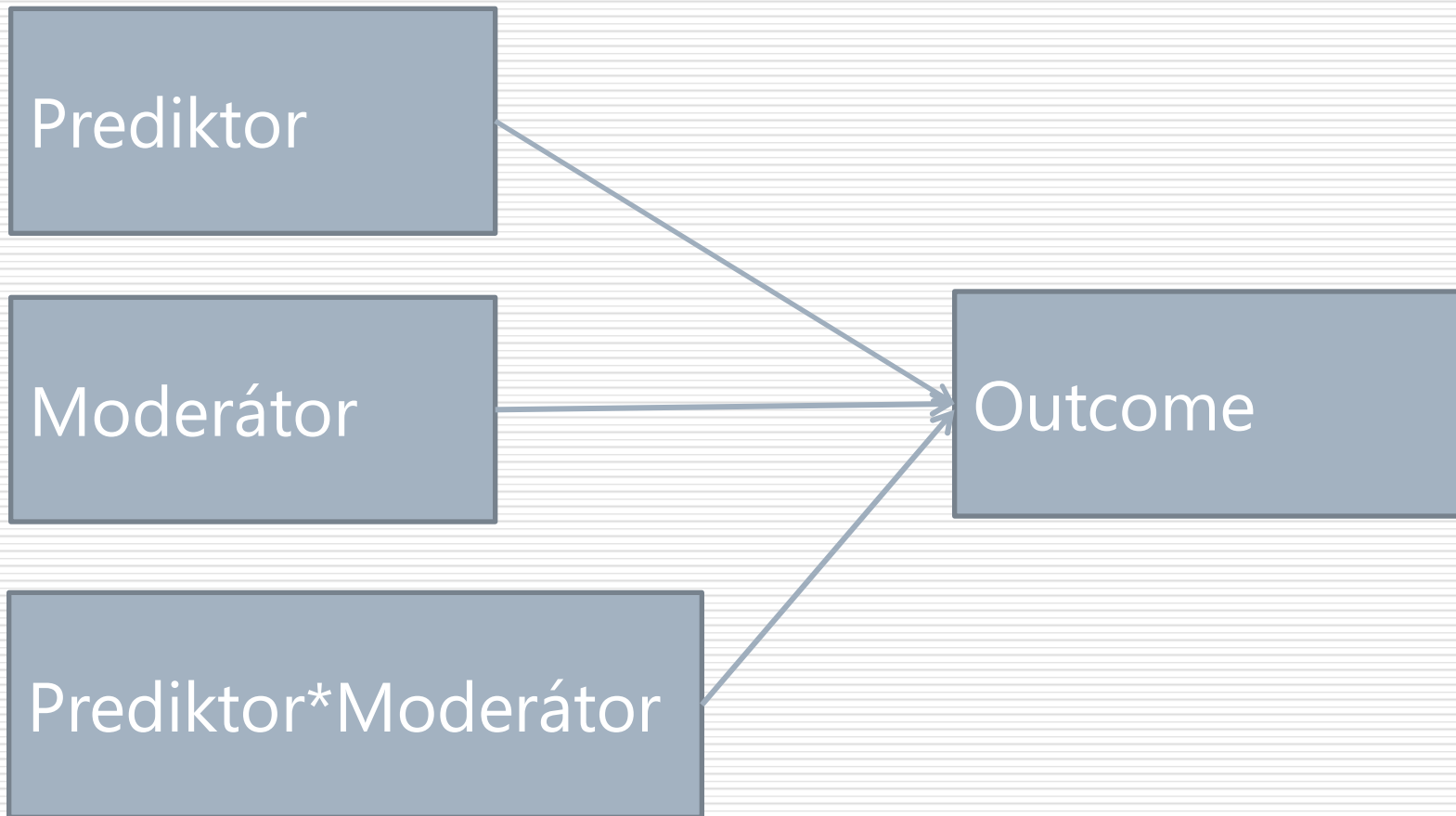
Moderace(interakce)

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

Moderace jako součást lineárního modelu



Moderace jako součást lineárního modelu



Moderace jako součást lineárního modelu

- Závislou tak predikujeme pomocí
 - prediktoru
 - moderátoru
 - 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	Beta		
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

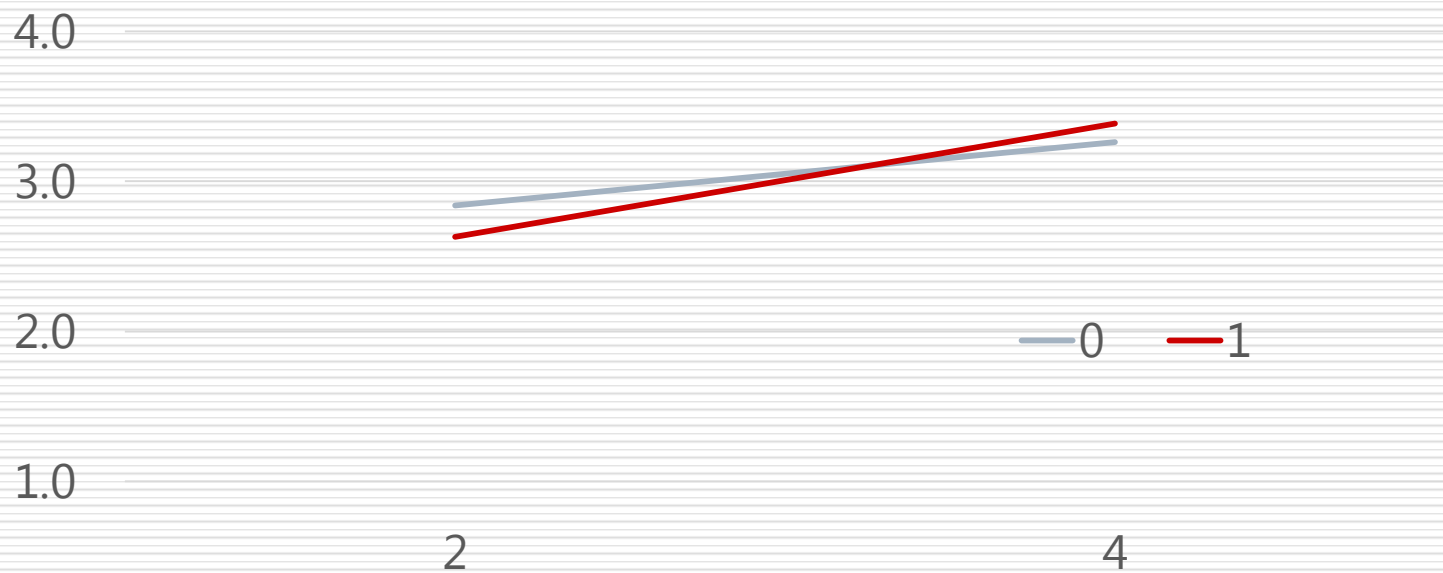
$$SELFE' = 2,4 + 0,2warm_o - 0,5MUZ + 0,2VOxMUZ$$

Interpretace

- Je-li MUZ=0, pak $SELFE' = 2,4 + 0,2warm_o$
 - Je-li MUZ=1, pak $SELFE' = 2,4 + 0,2warm_o - 0,5 + 0,2(warm_o * MUZ)$
 $SELFE' = 2,4 + 0,2warm_o - 0,5 + 0,2warm_o$
 $SELFE' = 2,4 + 0,4warm_o - 0,5$
 - 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)



<i>Coefficients^a</i>						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2,419	0,107		22,707	0,000
	warm_o	0,211	0,034	0,255	6,156	0,000
	vřelost otce					
	MUZ	-0,543	0,203	-0,554	-2,677	0,008
	VOxMUZ	0,167	0,064	0,543	2,592	0,010
a. Dependent Variable: selfe self-esteem						
		vřelost nízká	vřelost vysoká			
		2	4			
pohlaví	0	2,840	3,261			
	1	2,630	3,385			

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

Závěrečné poznámky k moderaci

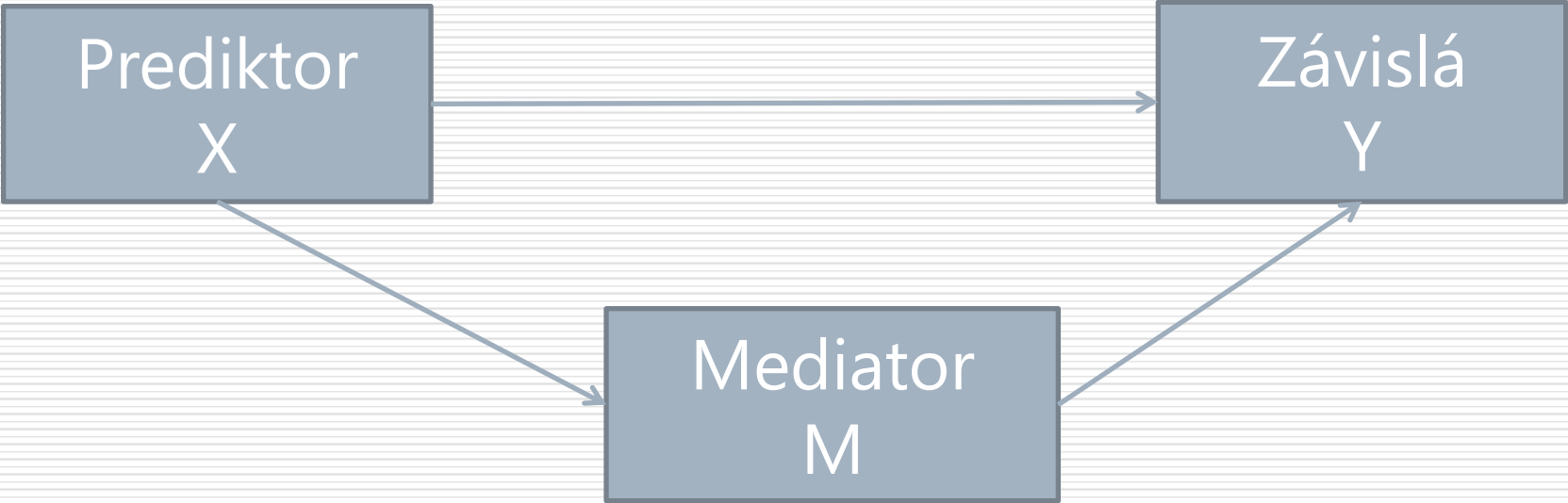
- Moderátor může být i spojitá proměnná.
 - Pak je dobré ji centrovat (popř. i prediktor), aby byla interpretace jednodušší (b_x při $M=0$)
 - Obecnějším termínem pro moderaci je **interakce** – může být mezi všemi typy proměnných a realizuje se stejně, tj. násobením
 - Interakce může být i vyššího řádu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



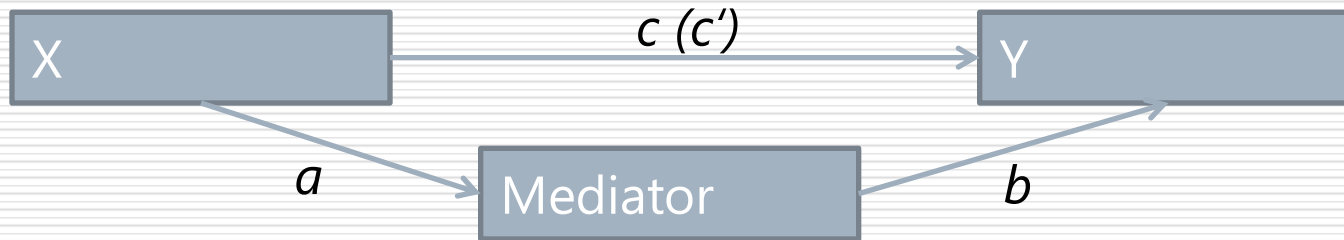
MEDIACE

- ❑ 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

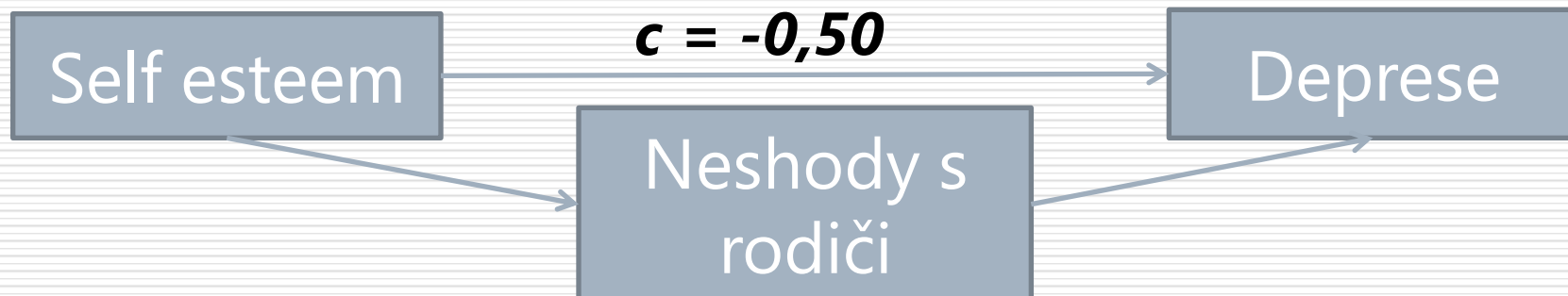


MEDIACE (Baron-Kenny, moderně)



1. X predikuje Y : regr. koef. c
 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') klesne (ideálně na 0)
 5. Nepřímý efekt X na Y (přes M) ($a \cdot b$) se statisticky významně liší od 0 – Sobelův test
$$z = \frac{ab}{\sqrt{(b^2 SE_a^2) + (a^2 SE_b^2)}}$$
-

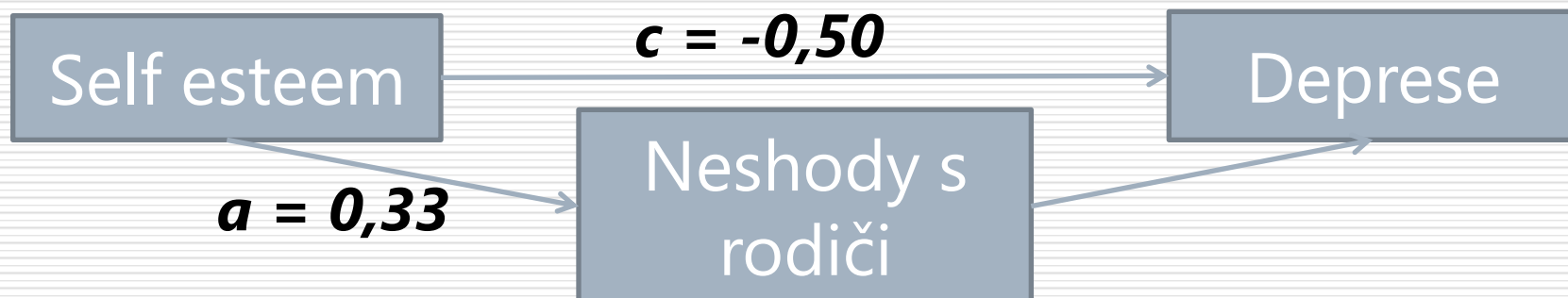
Mediace - příklad



1. Regrese DEP na SE

	B	Std. Error	Beta	t
(Constant)	3,553	,094		37,6
selfe self-esteem	-,504	,030	-,517	-16,5

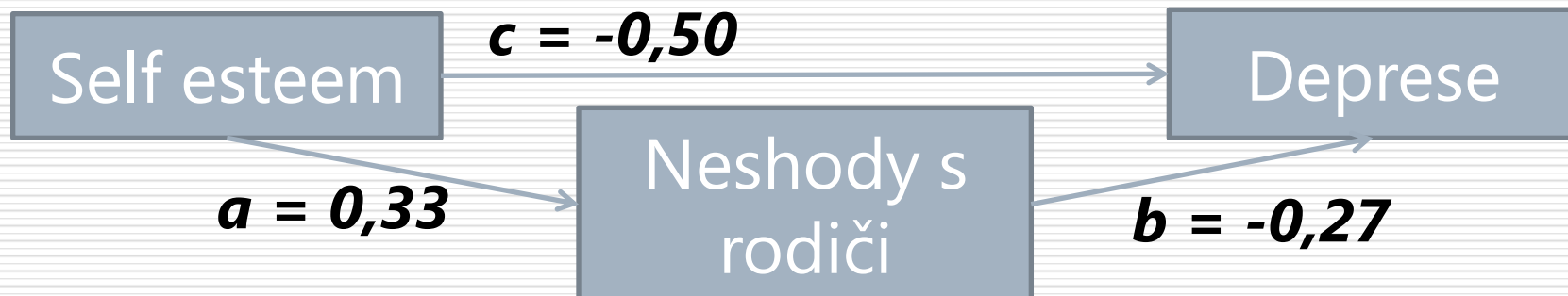
Mediace - příklad



2. Regrese Neshod na SE

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

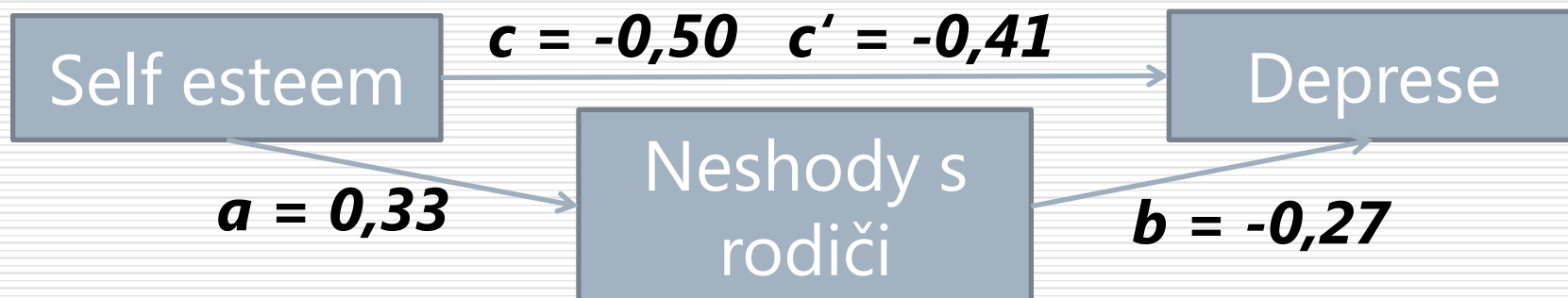
Mediace - příklad



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 (O)	-,270	,024	-,338	-11

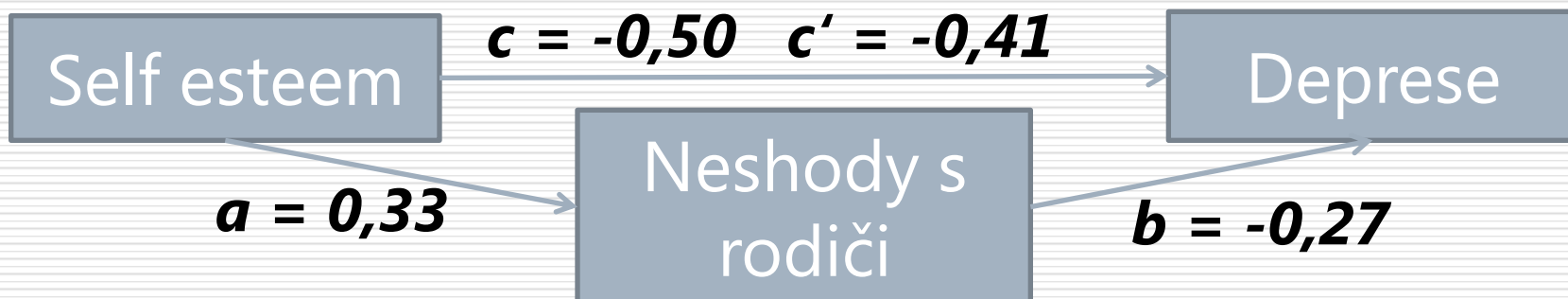
Mediace - příklad



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

Mediace – závěrečné poznámky

- Mediační model je explicitně kauzální – je třeba umět obhájit směr šipek.
 - Intervaly spolehlivosti 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 two normal distributions: a solid red curve for the null hypothesis (H0) centered at 0, and a dashed blue curve for the alternative hypothesis (H1) centered at approximately 3.3969515. A vertical green line at t = 2.039514 indicates the critical value. The area under the H0 curve to the right of the critical value is shaded and labeled $\frac{\alpha}{2}$. The area under the H1 curve to the left of the critical value is shaded and labeled β .

Input Parameters:

Parameter	Value
Tail(s)	Two
Determine =>	
Slope H1	0.4941850
α err prob	0.05
Power (1 - β err prob)	0.90
Slope H0	0
Std dev σ_x	0.481
Std dev σ_y	0.467

Output Parameters:

Parameter	Value
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

Parameter	Value
Correlation ρ	0.509
Std dev σ_x	0.481
Std dev σ_y	0.467
Slope H1	0.494185

Buttons: Calculate, Calculate and transfer to main window, Close.

$$H_0: R^2 = 0$$

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

Test family: F tests
Statistical test: Linear multiple regression: Fixed model, R² deviation from zero
Type of power analysis: A priori: Compute required sample size - given α , power, and effect size

Input Parameters

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

Output Parameters

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

X-Y plot for a range of values Calculate

R²

From correlation coefficient

Squared multiple correlation ρ^2

From predictor correlations

Number of predictors

Squared multiple correlation ρ^2

Specify matrices

Calculate Effect size f²

Calculate and transfer to main window

Close

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

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

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

ΔR^2

$1 - R^2$

From variances

Variance explained by special effect	0.1
Residual variance	0.8

Direct

Partial R ²	0.1111111
------------------------	-----------

Calculate

Effect size f ²	0.125
----------------------------	-------

Calculate and transfer to main window

Close

$H_0: b_i=0$

- 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)

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

Determine =>	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 Section

From variances

Variance explained by predictor	0.05
Residual variance	0.6

Direct

Partial R^2	0.07692308
---------------	------------

Calculate Effect size f^2 0.08333333

Calculate and transfer to main window

Close

Síla testu a velikost vzorku v MLR

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

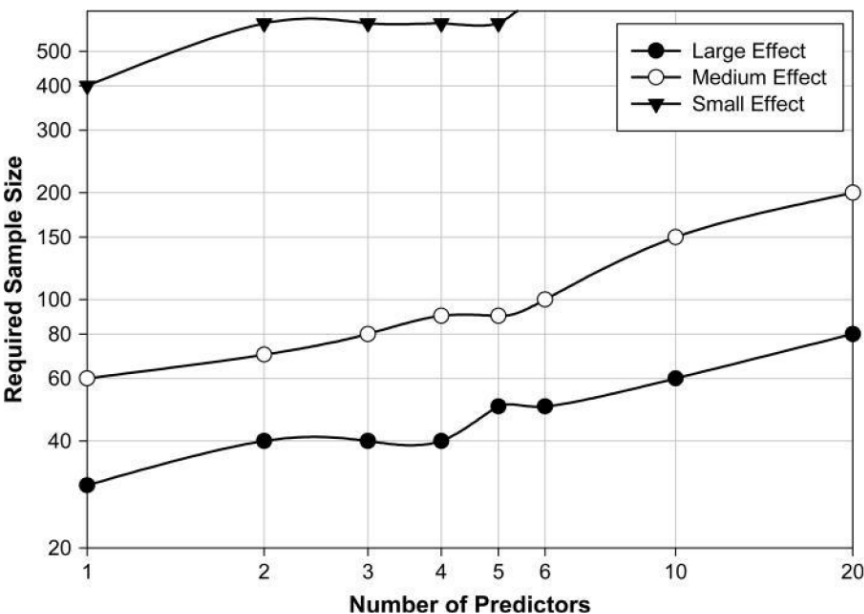


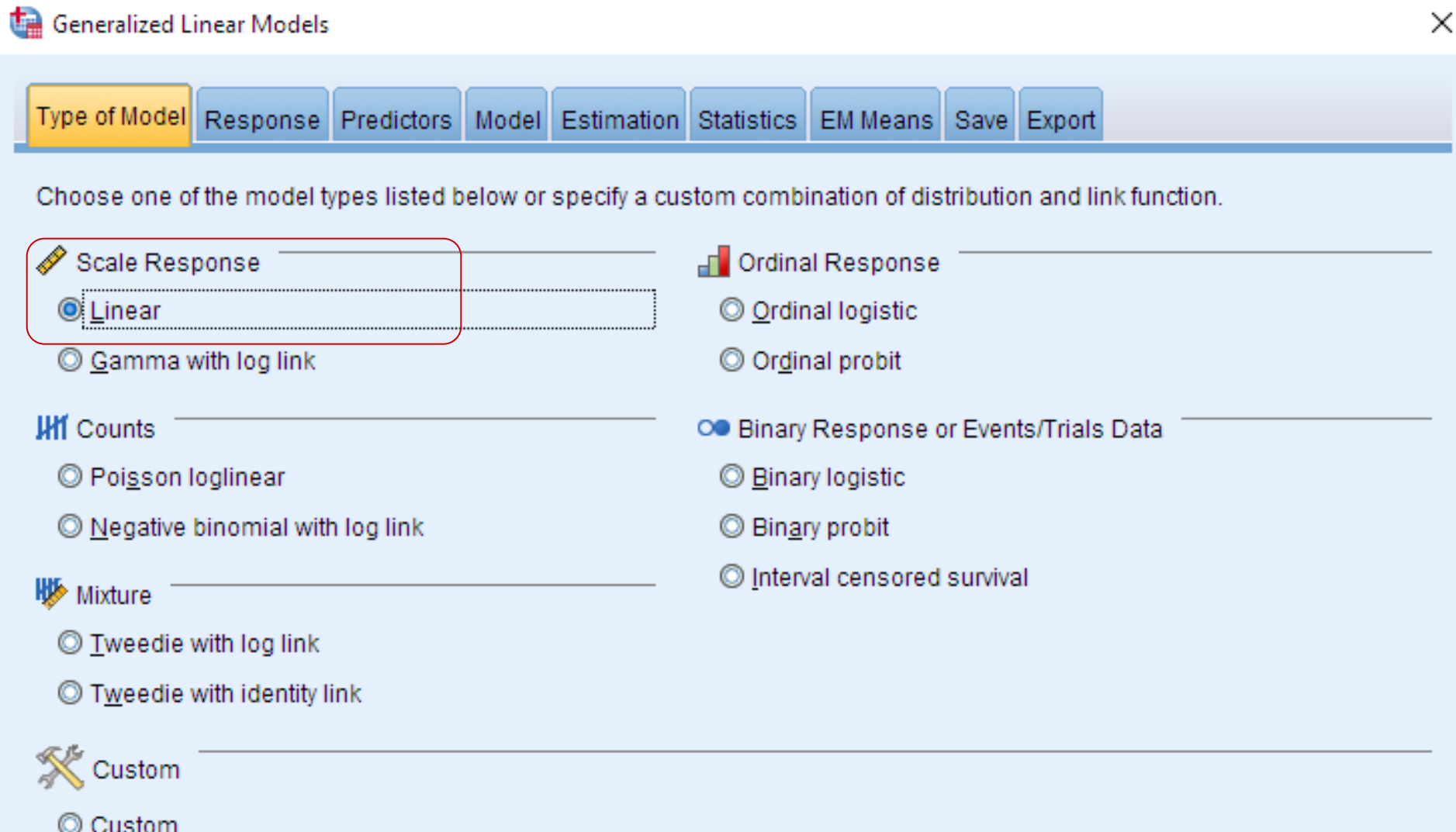
TABLE 5 Minimum R^2 That Can Be Found Statistically Significant with a Power of .80 for Varying Numbers of Independent Variables and Sample Sizes

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

 Mixture

- Tweedie with log link
- Tweedie with identity link

 Custom

- Custom

Binary Response or Events/Trials Data

- Binary logistic
- Binary probit
- Interval censored survival

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 - potlacení emoci [ro_...
- coping - rodina - potlacení [ro_pot]
- coping - rodina - aktivní coping [ro_akt]
- coping - rodina - hledání sociální opo...
- coping - škola - potlacení emoci [sk_...
- coping - škola - potlacení [sk_pot]
- coping - škola - aktivní coping [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 [deppozi]

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 narození [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]
- spolubydlicí [spolub99]
- jaký jiný spolubydlicí [spolubv2]

Factors:

pohlaví [pohlavi]

KATEGORICKÉ



Options...

**VOLBA
REFERENČNÍ K.**

Covariates:

vřelost otce [warm_o]

METRICKÉ



Offset

Variable



Offset Variable:

Fixed value

Value:



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.

Change in parameter estimates

Minimum:

1E-006

Type:

Absolute

Change in log-likelihood

Absolute

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

Tolerance level: ,0001

PŘESNĚJŠÍ, POMALEJŠÍ

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
- General estimable functions
- Iteration history
 - Print Interval: 1
- Lagrange multiplier test of scale parameter or negative binomial ancillary parameter

KONTRASTY



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
<p>ZDE MOŽNOST ZVOLIT KONTRASTY PRO KATEGORICKÉ S VÍCE KATEGORIEMI</p>		

Scale

Compute means for response

Compute means for linear predictor

Adjustment for Multiple Comparisons:



Hrátky s prediktory

Prediktory lze do modelu vložit všechny najednou, jednotlivě, nebo po skupinkách

Porovnáváme tak vlastně mnoho modelů lišících se zahrnutými prediktory.

- Vše najednou = ENTER
 - Postupně po jednom = FORWARD
 - Vše a postupně ubírat = BACKWARD
 - Po blocích, blockwise = ENTER + další blok
-

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
 - 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
 - kontrola intervenujících proměnných
 - Minimalizace chyby 1. typu
 - Podle výzkumné relevance
 - Od ústředních po „co kdyby“; maximalizace síly
-

Obvyklý postup regresní analýzy

- Na základě teoretických rozvah stanovíme různé modely, jejichž srovnání je potenciálně zajímavé
 - Nejjednodušší srovnání je u hierarchických modelů, kdy je jeden model plně vnořen do následujícího – to umožňuje testovat inkrement R^2
 - Až v druhé řadě se zabýváme jednotlivými regresními koeficienty v modelu, který je nejúplnější/nejllepší
-



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**
 - ❑ ...leđa by šlo o zjevnou chybu v datech či vzorku
 - ❑ ...leđa by nám šlo výhradně o zpřesnění predikce (nikoli o testy hypotéz)
-

Daignostika 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!!

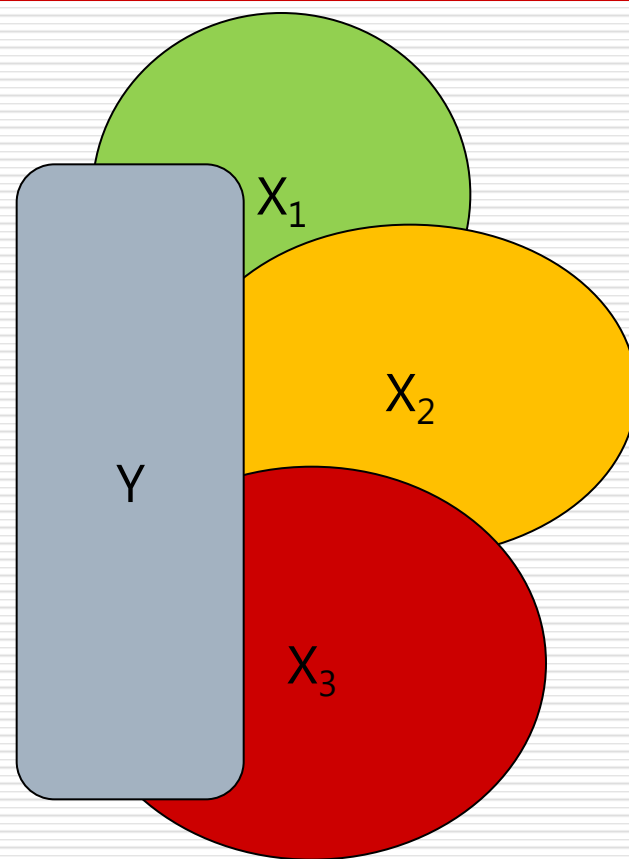
Diagnostika 3: Předpoklady regrese

- ❑ 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
-



MLR: Shoda modelu s daty: R^2

- Část rozptylu Y vysvětleného dohromady všemi prediktory
- Predikční síla sady prediktorů
- Ukazatel velikosti účinku
- R : Mnohonásobná (mutiple) korelace
- Vždy nadhodnocuje >> při replikaci vychází nižší R^2
 - shrinkage correction – Adjusted (upravené) R^2
 - Wherry (SPSS, Statistica) –kdybychom model dělali z cenzových dat
 - cross-validation
 - Stein (Field) – očekávané R^2 při replikaci
 - split-sample analýza



Úkol

1. Rozpracujte model predikující sebehodnocení dospívajících z minulého zadání

- Z původních prediktorů nechejte jen vřelost otce.
- Do modelu zařadte moderaci efektu vřelosti otce pohlavím.
- Ověřte, že moderační efekt zůstává, i když statisticky kontrolujeme kohortu a vzdělání otce

2. Z proměnných Long vymyslete jednoduchý mediační model. Zkuste kriticky zdůvodnit kauzalitu a stanovit velikost nepřímého efektu.

Z analýz sepište zprávu v souladu s konvencemi.

Odevzdejte do pondělí – do 14 hodin.