

# Statistické zpracování dat 8.přednáška

Mgr. Radmila Krkošková, Ph.D.



**SLEZSKÁ  
UNIVERZITA**  
OBCHODNĚ PODNIKATELSKÁ  
FAKULTA V KARVINĚ



EVROPSKÁ UNIE  
Evropské strukturální a investiční fondy  
Operační program Výzkum, vývoj a vzdělávání



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



# Vícenásobná lineární regresní analýza (3)





## Fiktivní proměnné





- **Doposud** vysvětlující (nezávisle) proměnné:  
**kvantitativní (číselné)**

***Příklady:*** cena, příjem, náklady, výdaje na reklamu, hmotnost, teplota, vzdálenost aj.



- **Nyní** vysvětlující (nezávisle) proměnné:

**kvalitativní (kategoriální, binární) – fiktivní (dummy) proměnné**

***Příklady:*** pohlaví (M,Ž), národnost (CZ, SK), vzdělání (Z, SŠ, VŠ)

**Obvykle se kódují binárně: 0 nebo 1**

# Příklad - zadání



SLEZSKÁ  
UNIVERZITA  
OBCHODNĚ PODNIKATELSKÁ  
FAKULTA V KARVINĚ

Lze řešit též pomocí ANOVA (1 faktor)!

Zjistěte, zda má vzdělání (VŠ,SŠ) vliv na nástupní plat.

Data:

Nástupní plat	Vzdělání	VŠ_Binární
16,2	v	1
12,5	s	0
12	s	0
15,3	v	1
16	v	1
13,3	s	0
16,7	v	1
13	s	0
14	s	0
17	v	1

Regresní model:

$$Y_i = B_0 + B_1 D_i + u_i$$



# Příklad – řešení v Excelu

Řešení v Excelu: Nástroje → Analýza dat → Regrese:

$$\hat{Y} = 12,96 + 3,28D$$

	Koeficienty		ba stř. hodl	t stat	Hodnota P	Dolní 95%
Hranice	12,96	0,318748		40,65914	1,47369E-10	12,2249664
VŠ_Binární	3,28	0,450777		7,276323	8,58326E-05	2,24050546

$b_0$  - prům.plat SŠ

$b_1$  - nárůst

$b_0 + b_1$  - prům.plat VŠ

Regrese s jednou kvantitativní a jednou kvalitativní (binární) proměnnou se dvěma hodnotami

**Vysvětlující proměnné jak kvantitativní tak kvalitativní**

$$Y_i = B_0 + B_1 D_i + B_2 X_i + u_i$$



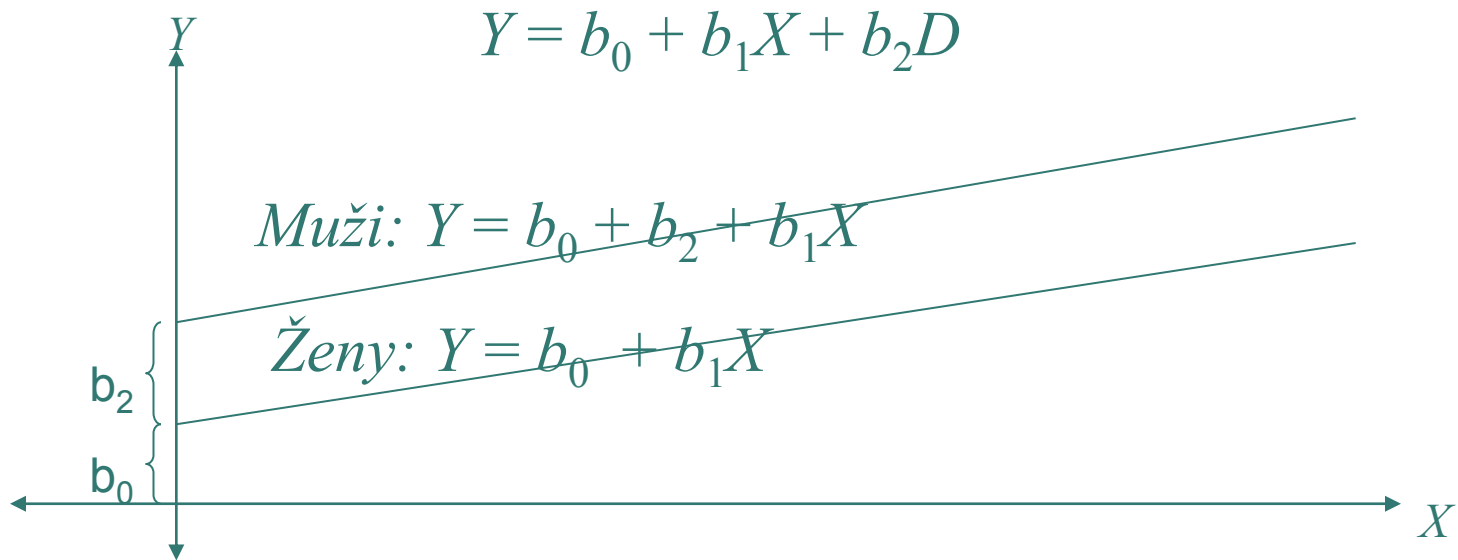
$$Y_i = B_0 + B_1 D_i + B_2 X_i + u_i$$

**Příklad:** Závislost platu VŠ učitele na délce praxe a pohlaví

$Y_i$  = měsíční plat VŠ učitele

$X_i$  = počet let praxe

$D_i$  = 1 jestliže M, 0 jestliže Ž



	<i>Koeficienty</i>	<i>Hodnota P</i>
$b_0$	12,97	7,34E-17
$b_1$	1,37	6,16E-14
$b_2$	3,33	6,13E-11

## Regrese s jednou kvantitativní a jednou kvalitativní proměnnou s více hodnotami

Kvalitativní proměnná nabývá 3 (nebo i více) hodnot:

$$Y_i = B_0 + B_1D_{1i} + B_2D_{2i} + B_3X_i + u_i$$

$$Y_i = B_0 + B_1 D_{1i} + B_2 D_{2i} + B_3 X_i + u_i$$

## *Příklad:*

Závislost výdajů rodiny na letní dovolenou na příjmech a vzdělání otce (VŠ, SŠ, Z)

# Fiktivní proměnné



SLEZSKÁ  
UNIVERZITA  
OBCHODNĚ PODNIKATELSKÁ  
FAKULTA V KARVÍNĚ

$$Y_i = B_0 + B_1 D_{1i} + B_2 D_{2i} + B_3 X_i + u_i$$

$Y_i$  = výdaje rodiny na letní dovolenou

$X_i$  = roční příjem rodiny

$D_{1i}$  = 1 jestliže SŠ, 0 jinak

$D_{2i}$  = 1 jestliže VŠ, 0 jinak



# Fiktivní proměnné

$$Y_i = B_0 + B_1 D_{1i} + B_2 D_{2i} + B_3 X_i + u_i$$

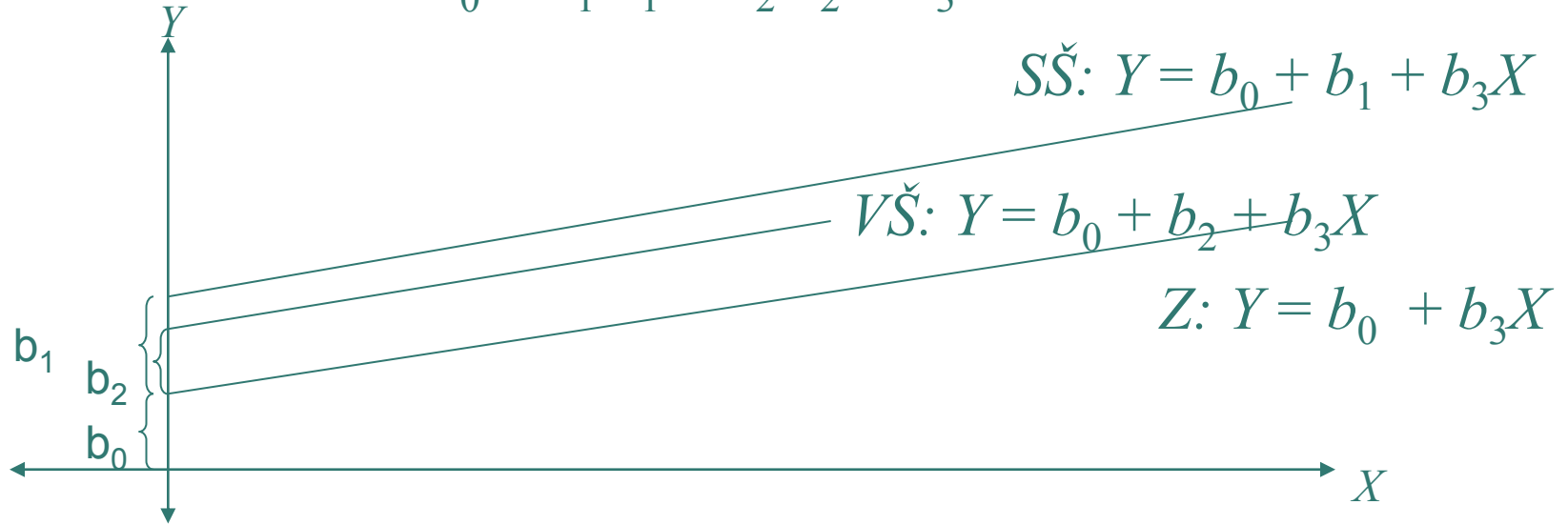
Výdaje na dov.	Roč.příjmy	SŠ	VŠ
60	400	0	1
39	310	1	0
18	180	0	0
19	190	0	0
72	470	0	1
33	270	1	0
31	260	1	0
17	170	0	0
64	430	0	1
79	490	0	1
15	150	0	0
31	250	1	0
36	290	1	0
20	200	0	0
62	410	0	1



**SLEZSKÁ  
UNIVERZITA**  
OBCHODNĚ PODNIKATELSKÁ  
FAKULTA V KARVINĚ



$$Y = b_0 + b_1D_1 + b_2D_2 + b_3X$$



# Kvalitativní proměnná s $K$ hodnotami

---



SLEZSKÁ  
UNIVERZITA  
OBCHODNĚ PODNIKATELSKÁ  
FAKULTA V KARVÍNĚ

$Z$  – *kvalitativní* proměnná s hodnotami

$z_1, z_2, \dots, z_{K-1}, z_K$

Lze ji nahradit  $k - 1$  binárními (fiktivními) proměnnými:





# Kvalitativní proměnná s $K$ hodnotami

---



**SLEZSKÁ  
UNIVERZITA**  
OBCHODNĚ PODNIKATELSKÁ  
FAKULTA V KARVINĚ

Lze ji nahradit  $k - 1$  binárními (fiktivními) proměnnými:

- $D_2 = 1$  jestliže  $Z = z_2$   
= 0 jinak
- $D_3 = 1$  jestliže  $Z = z_3$   
= 0 jinak
- .....  
•  $D_K = 1$  jestliže  $Z = z_K$   
= 0 jinak



# Regrese s proměnlivým sklonem

---



**SLEZSKÁ  
UNIVERZITA**  
OBCHODNĚ PODNIKATELSKÁ  
FAKULTA V KARVINĚ

- Kvalitativní (fiktivní) proměnné neovlivňují sklony v regresních modelech (rovnicích)
- **Příklad:** ve 2 obdobích (kvalitativní proměnná) může mít regresní přímka různý sklon



# Regrese s proměnlivým sklonem



SLEZSKÁ  
UNIVERZITA  
OBCHODNĚ PODNIKATELSKÁ  
FAKULTA V KARVINĚ

- **Příklad:** ve 2 obdobích (kvalitativní proměnná) může mít regresní přímka různý sklon

$$Y_t = B_0 + B_1X_t + B_2D_t + B_3X_tD_t + u_t$$

$Y_t$  = celkové úspory

$X_t$  = celkové příjmy

$D_t$  = 1 jestliže  $t$  je z prvního období, 0 - jinak



# Příklad – zadání



úspory	příjmy	období
57,7	831,8	0
66,3	894	0
61,4	981,6	0
89	1101,7	0
95,7	1210	0
104,6	1313,4	0
95,8	1451,4	0
93,7	1607,5	0
110,2	1812,4	0
118,1	2034	0
136,9	2258,5	0
159,4	2520,9	0
153,9	2670,8	1
130,6	2830,6	1
164,1	3008,7	1
125,4	3325,3	1
121,7	3531	1
104,2	3780	1

$$Y_t = B_0 + B_1 X_t + B_2 D_t + B_3 X_t D_t + u_t$$



# Příklad - výsledek



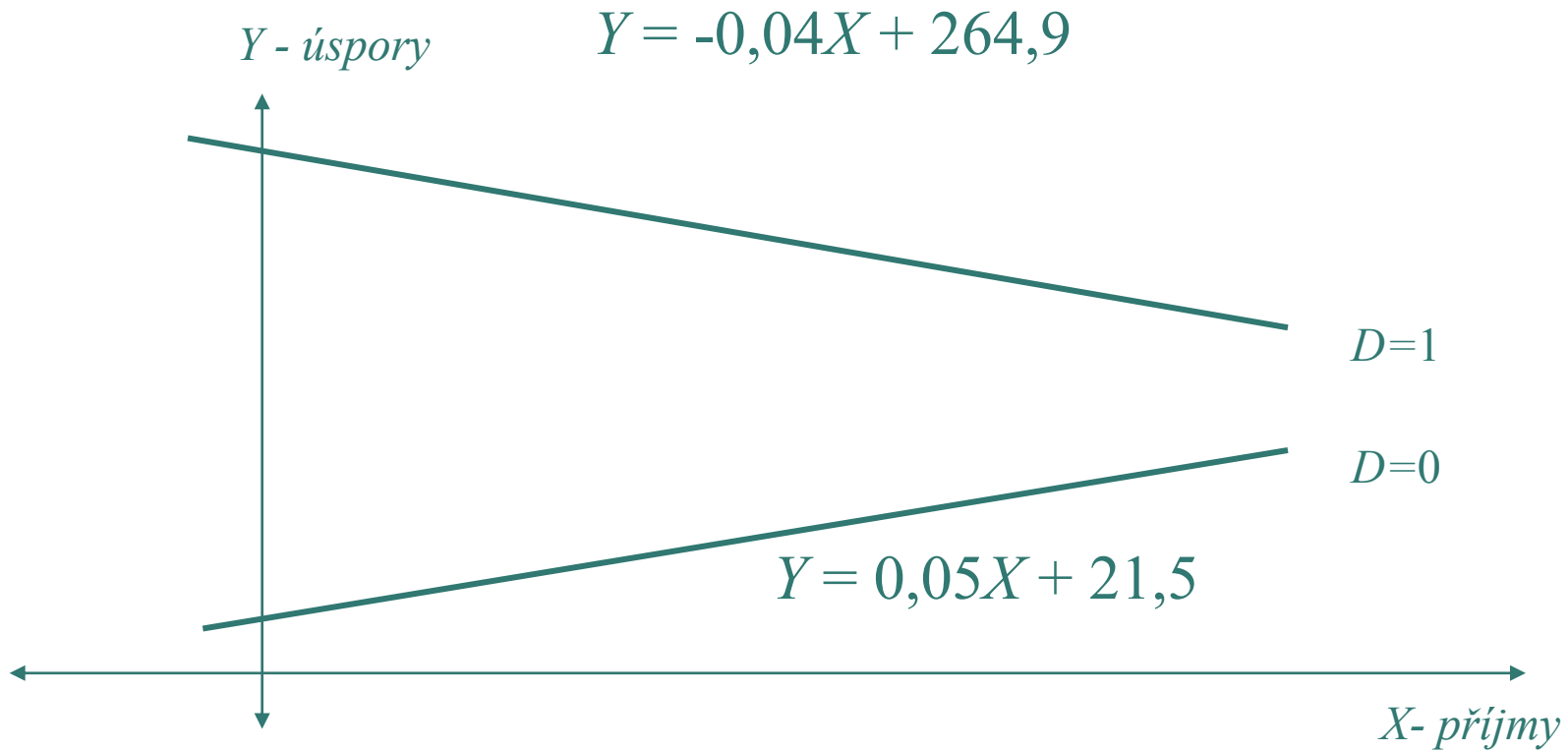
SLEZSKÁ  
UNIVERZITA  
OBCHODNĚ PODNIKATELSKÁ  
FAKULTA V KARVÍNĚ

	<i>Koeficienty Hodnota P</i>	
Hranice = $b_0$	21,5274	0,051002
příjmy = $b_1$	0,051532	1,13E-06
období = $b_2$	243,3986	3,2E-05
příjmy*období = $b_3$	-0,09276	9,34E-06

$$D = 1: Y_t = (b_0 + b_2) + (b_1 + b_3)X_t$$

$$D = 0: Y_t = b_0 + b_1X_t$$





# Shrnutí: fiktivní proměnné

---



**SLEZSKÁ  
UNIVERZITA**  
OBCHODNĚ PODNIKATELSKÁ  
FAKULTA V KARVINĚ

- Vysvětlující (nezávisle) proměnné: kvalitativní fiktivní (dummy) proměnné
- Kvalitativní hodnoty se kódují binárně s jednou nebo více bin. proměnnými



# Shrnutí: fiktivní proměnné

---



**SLEZSKÁ  
UNIVERZITA**  
OBCHODNĚ PODNIKATELSKÁ  
FAKULTA V KARVINĚ

- Kombinace více kvalitativních a kvantitativních nezávisle proměnných
- Fiktivními proměnnými lze modelovat proměnlivost sklonu regresní přímky ve více časových obdobích (data jsou časové řady)







- **Korelační analýza**  
(párový korelační koeficient)





- **Náhodná složka**  
(Durbin-Watsonův test autokorelace)



# Párový korelační koeficient



SLEZSKÁ  
UNIVERZITA  
OBCHODNĚ PODNIKATELSKÁ  
FAKULTA V KARVINĚ

- Měří míru lineární závislosti

- $$r = \frac{\overline{x \cdot y} - \bar{x} \cdot \bar{y}}{\sqrt{(\overline{x^2} - \bar{x}^2)(\overline{y^2} - \bar{y}^2)}}$$

- $r \in \langle -1; 1 \rangle$

- V Excelu funkce **CORREL**



# Testování korelačního koeficientu



SLEZSKÁ  
UNIVERZITA  
OBCHODNĚ PODNIKATELSKÁ  
FAKULTA V KARVINĚ

- $H_0: \rho = 0$  (není závislost),  
 $H_1: \rho \neq 0$  (existuje závislost).
- Testové kritérium:  $T = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$ .
- Kritický obor je dán kvantilem Studentova rozdělení (počet stupňů volnosti =  $n - 2$ ). V Excelu funkce T.INV ( $1-\alpha; n - 2$ ).

....

Při zjišťování závislosti reziduí se předpokládá model

$$\varepsilon_t = \rho \cdot \varepsilon_{t-1} + u_t,$$

kde  $\rho$  je tzv. koeficient autokorelace, který vyjadřuje těsnost závislosti náhodné složky na její předchozí hodnotě.

# Durbin-Watsonova statistika

---



**SLEZSKÁ  
UNIVERZITA**  
OBCHODNĚ PODNIKATELSKÁ  
FAKULTA V KARVINĚ

- $H_0: \rho = 0$  (není závislost),  
 $H_1: \rho \neq 0$  (existuje závislost).
- Testové kritérium:  $D = \frac{\sum(e_t - e_{t-1})^2}{\sum e_t^2}$ .
- $D \in \langle 0; 4 \rangle$





**Děkuji Vám za pozornost!!!**



# Durbin-Watsonova statistika



SLEZSKÁ  
UNIVERZITA  
OBCHODNĚ PODNIKATELSKÁ  
FAKULTA V KARVÍNĚ

V případě **nezávislosti** náhodných poruch se  $D$  pohybuje **okolo čísla 2**.

V případě **přímé závislosti** se  $D$  pohybuje kolem **0**  
a v případě **nepřímé závislosti** kolem **4**.

