STATISTIKA

8. PREZENTACE

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ

Téma přednášky:
a) regresní analýza,
b) lineární regrese,
c) metoda nejmenších čtverců,
d) koeficient determinace.

Mgr. Radmila Krkošková, Ph.D.

Jaké a k čemu jsou metody stanovení závislosti

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ

- závislostí 1. **kvantitativního** znaku na 2. **kvantitativním** znaku (nebo více kvantitativních znacích) - **regresní a korelační analýza**
- závislost dvou znaků - **jednoduchá regresní analýza** (**jednoduchá korelační analýza**)

Jaké a k čemu jsou metody stanovení závislosti

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ

- závislost znaku na více znacích - **vícenásobná regresní analýza**

- znalost závislostí umožňuje:
  předvídat chování (prognózovat, predikovat) závislé veličiny

Příklad – Zisk z reklamy

nezávislá - *závislá* veličina (proměnná)

Firma č.	Výdaje na reklamu (tis. Kč)	Zisk z prodeje (10 tis. Kč)
1	6	5
2	8	8
3	9	9
4	9	12
5	12	21
6	15	25
7	16	32
8	20	36
9	22	51
10	23	59

Jednoduché regresní modely

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ

$$y = f(x) + \varepsilon$$

závisle proměnná

reziduum

regresní funkce **nezávisle** proměnná

Lineární regresní funkce:

$$f(x) = \beta_0 + \beta_1 x$$

Jednoduché regresní modely

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ

Parabolická regresní funkce :

$$f(x) = \beta_0 + \beta_1 x \ + \beta_2 x^2$$

Exponenciální regresní funkce :

$$f(x) = \beta_0 \beta_1^x$$

Logaritmická regresní funkce:

$$f(x) = \beta_0 + \beta_1 \log x$$

Jednoduchá lineární regrese

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ

- výběr párových hodnot:
  $(y_1, x_1), (y_2, x_2), (y_3, x_3),\ldots,(y_n, x_n)$

- 2 způsoby získání dat:

(A) hodnoty nezávisle proměnné x_i se předem pevně zvolí a k nim se „změří“ příslušné hodnoty y_i

(B) hodnoty (y_i, x_i) se „změří“ na n náhodně zvolených jednotkách základního souboru

Jednoduchá lineární regrese

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ

Soubor párových hodnot se geometricky znázorní v rovině **bodovým grafem:**

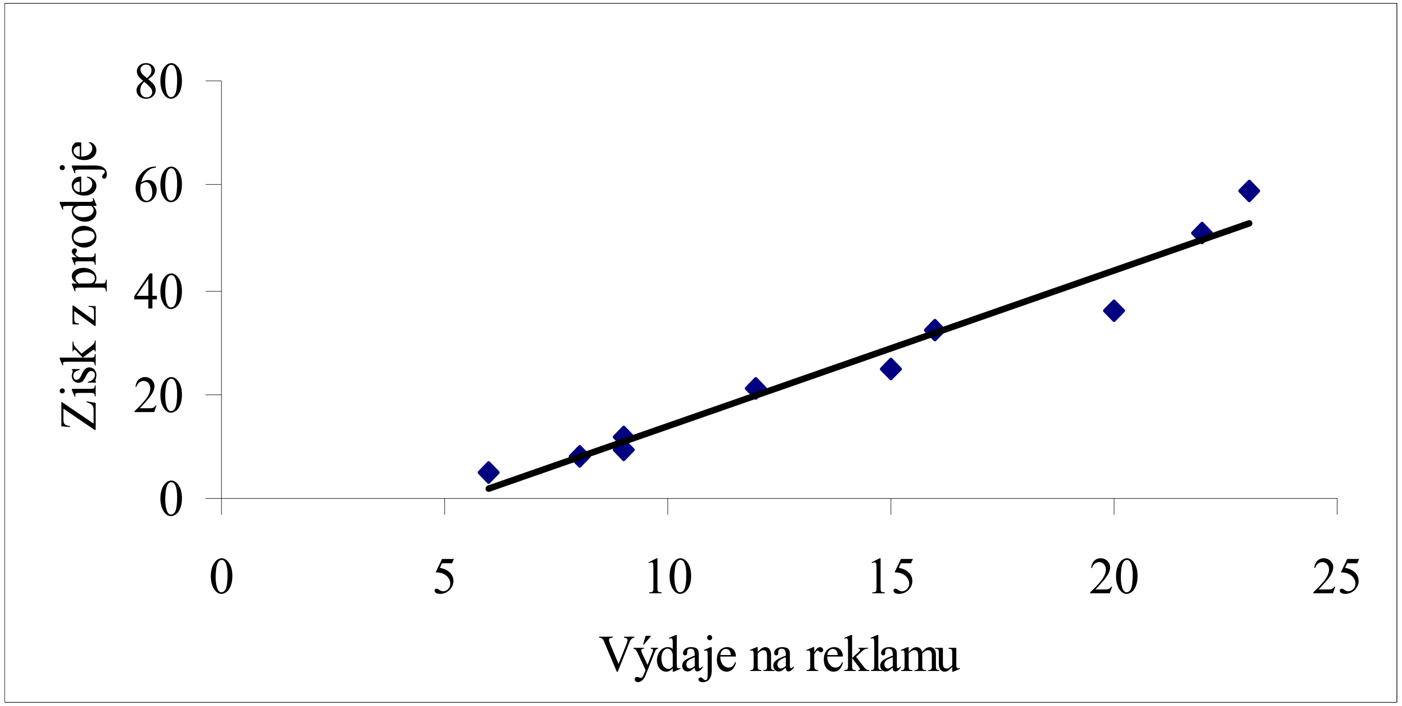
reziduum

JLR model: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad i = 1,2,\ldots,n$

regresní koeficienty a jejich odhady $\boldsymbol{b}_0$, $\boldsymbol{b}_1$

Příklad: Zisk z reklamy (grafické znázornění)

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ



Příklad: Výdaje na reklamu

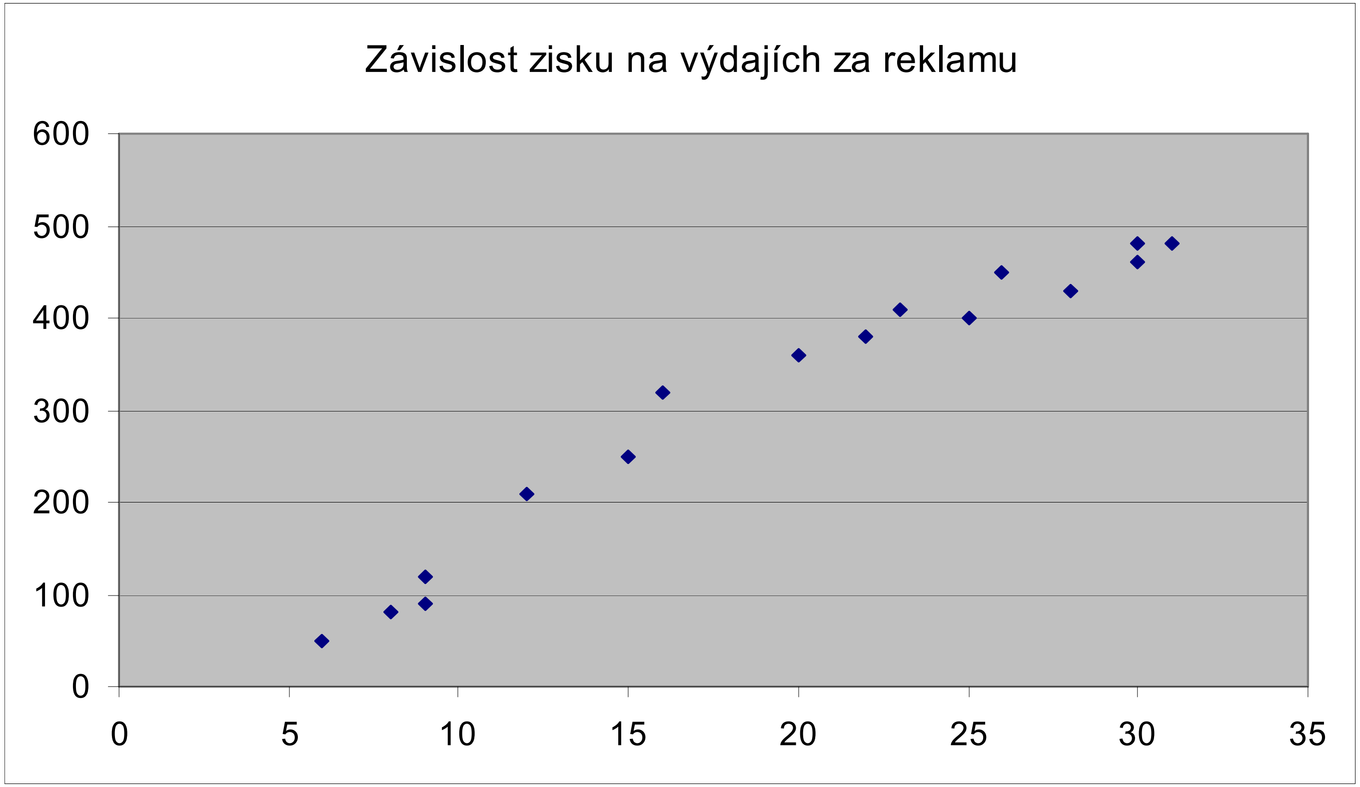
JRA

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ

č. firmy	Výdaje na reklamu	Výdaje na reklamu	Zisk
1	malé	6	50
2	malé	8	80
3	malé	9	90
4	malé	9	120
5	středně velké	12	210
6	středně velké	15	250
7	středně velké	16	320
8	středně velké	20	360
9	středně velké	22	380
10	středně velké	23	410
11	velké	25	400
12	velké	26	450
13	velké	28	430
14	velké	30	460
15	velké	30	480
16	velké	31	480

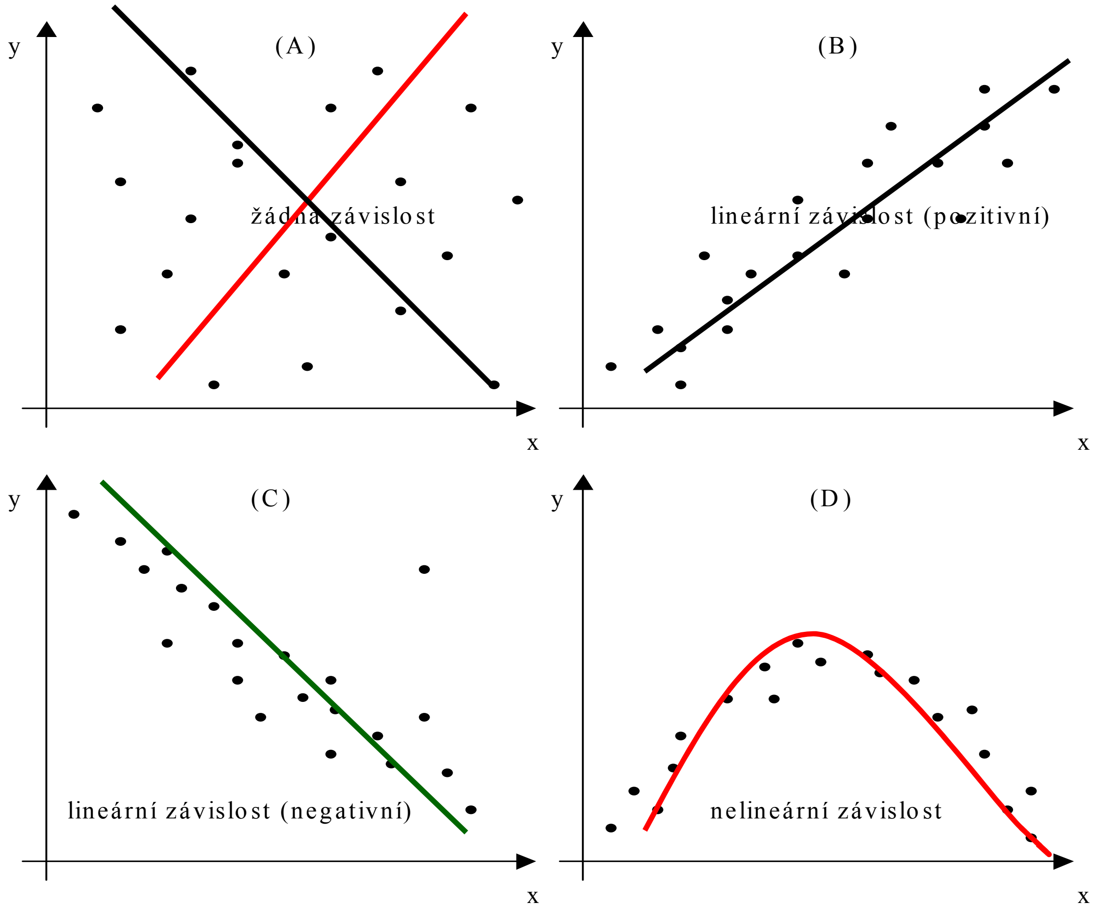
Příklad: grafické znázornění

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ



Bodový diagram (Scatter diagram)

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ



Metoda nejmenších čtverců

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ

Idea MNČ: minimalizovat reziduální součet čtverců:

$$S_R = \sum_{i=1}^{n} (y_i - Y_i)^2 = \sum_{i=1}^{n} (y_i - (b_0 + b_1 x_i))^2$$

Příklad: Zisk z reklamy

$$b_1 = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{x^2} - \overline{x}^2} = \frac{462{,}1 - 14 \cdot 25{,}8}{230 - 14^2} = \frac{100{,}9}{34} = 2{,}97$$

$$b_0 = \overline{y} - b_1 \overline{x} = 25{,}8 - 2{,}97 \cdot 14 = -15{,}78$$

Regresní funkce: $\boxed{Y = -15{,}78 + 2{,}97x}$

Příklad: Zisk z reklamy – ruční výpočty

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ

i	x_i	y_i	x_i^2	x_iy_i	Y_i	$(Y_i - \bar{y})^2$	$(y_i - \bar{y})^2$
1	6	5	36	30	2,04	565,21	432,64
2	8	8	64	64	7,98	318,22	316,84
3	9	9	81	81	10,95	221,15	282,24
4	9	12	81	108	10,95	221,15	190,44
5	12	21	144	252	19,86	35,62	23,04
6	15	25	225	375	28,77	8,61	0,64
7	16	32	256	512	31,74	34,84	38,44
8	20	36	400	720	43,62	315,88	104,04
9	22	51	484	1122	49,56	562,08	635,04
10	23	59	529	1357	52,53	711,60	1102,24
Součet	140	258	2300	4621	258	2994,3	3125,6
Průměr	14	25,8	230	462,1			

Předpoklady lineárního modelu

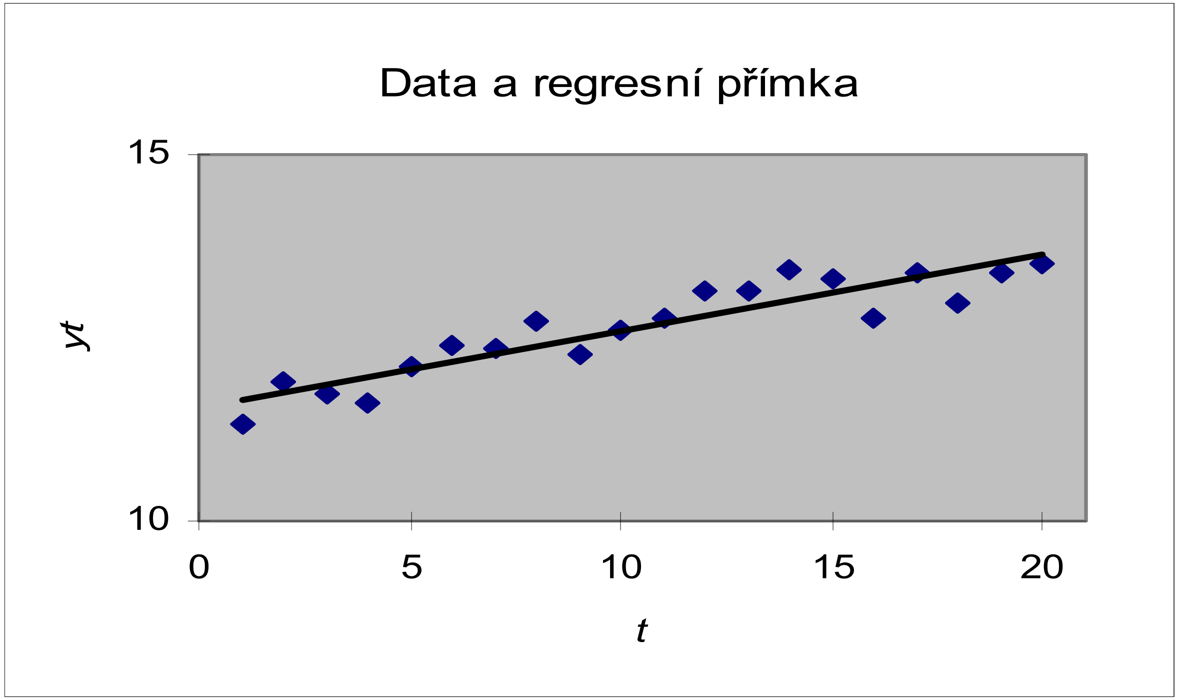
1. Hodnoty vysvětlující proměnné x_i se volí předem, ***nejsou*** to tedy náhodné veličiny.

2. Náhodné složky (rezidua) ε_i mají **normální rozdělení** pravděpodobnosti se střední hodnotou 0 a (neznámým) konstantním rozptylem σ^2 - tzv. **homoskedasticita**

3. Náhodné složky jsou **nekorelované**, tj.

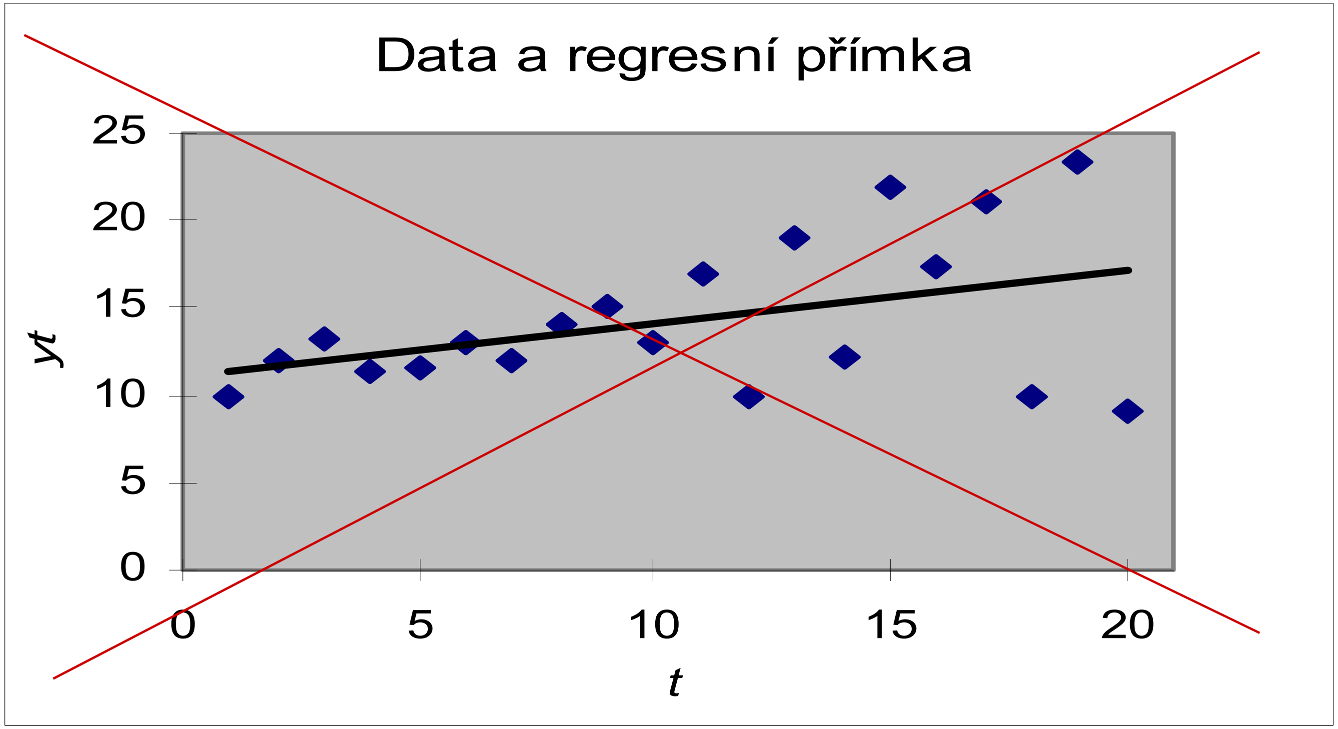
$\rho(\varepsilon_i , \varepsilon_j) = 0$ pro každé $i \neq j$, $i,j = 1,2,\ldots,n$.
(ρ - **korelační koeficient**)

Předpoklady lineárního modelu - jsou splněny



Předpoklady lineárního modelu – nejsou splněny

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ



Koeficient determinace R^2

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ

Koeficient determinace charakterizuje **přiléhavost dat k regresnímu modelu** (číslo mezi 0 a 1):

$$\boxed{R^2 = \frac{S_T}{S_y} = \frac{S_y - S_R}{S_y} = 1 - \frac{S_R}{S_y}}$$

$$S_y = S_R + S_T$$

- **teoretický součet čtverců**: $S_T = \sum_{i=1}^{n} \left(Y_i - \bar{y}\right)^2$

- **reziduální součet čtverců** $S_R = \sum_{i=1}^{n} \left(y_i - Y_i\right)^2$

Koeficient determinace R^2 - upravený

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ

Pro malé soubory: $$R_{adj}^2 = 1 - \left(1 - R^2\right)\frac{n-1}{n-2}$$

Výpočet koeficientu determinace

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ

Závislost zisku z prodeje na velikosti nákladů na reklamu:

$$R^2 = \frac{S_T}{S_y} = \frac{2994{,}3}{3125{,}6} = 0{,}958 \qquad R^2_{adj} = 0{,}953$$

Koeficient korelace (odmocnina koeficientu determinace)

$$R = 0{,}979 \qquad R_{adj} = 0{,}979$$

Závěr přednášky

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ

Děkuji Vám za pozornost !!!