Statistické zpracování dat
4.přednáška

Mgr. Radmila Krkošková, Ph.D.

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ

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

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ

Jednoduchá regresní analýza

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Obsah přednášky

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ

- Co je regresní analýza (RA - jednoduchá, vícenásobná, lineární, nelineární)
- Rozdíl mezi RA a ANOVA
- Co je podstatou jednoduché lineární RA (bodový diagram, regresní přímka, regresní koeficienty, přiléhavost - koeficient determinace, testy hypotéz)

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Obsah přednášky

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ

- Co je podstatou jednoduché **ne**lineární RA (základní typy nelinearity, Törnquistovy křivky)
- Kdy RA nemá smysl?
- Aplikace na příkladech z ekonomické oblasti (marketingový výzkum, průměrné fixní náklady, Phillipsaova křivka aj.)

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Závislosti mezi kvantitativními statistickými znaky

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ

- Problém závislosti 2 znaků řeší jednoduchá regresní analýza (lineární a nelineární)
- **Příklad**: *Závislost zisku z prodeje výrobku na výdajích za reklamu*
- Východiskem je vždy grafické znázornění
- Mírami závislosti jsou regresní koeficienty, resp. koeficienty determinace (a korelace)
- Někdy je výhodné využít z kvantitativních dat pouze ordinální informaci (tj. uspořádání) a aplikovat ANOVA
- Míry asociace mezi více znaky řeší vícenásobné regresní a korelační metody

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Příklad – výdaje na reklamu

ANOVA JRA

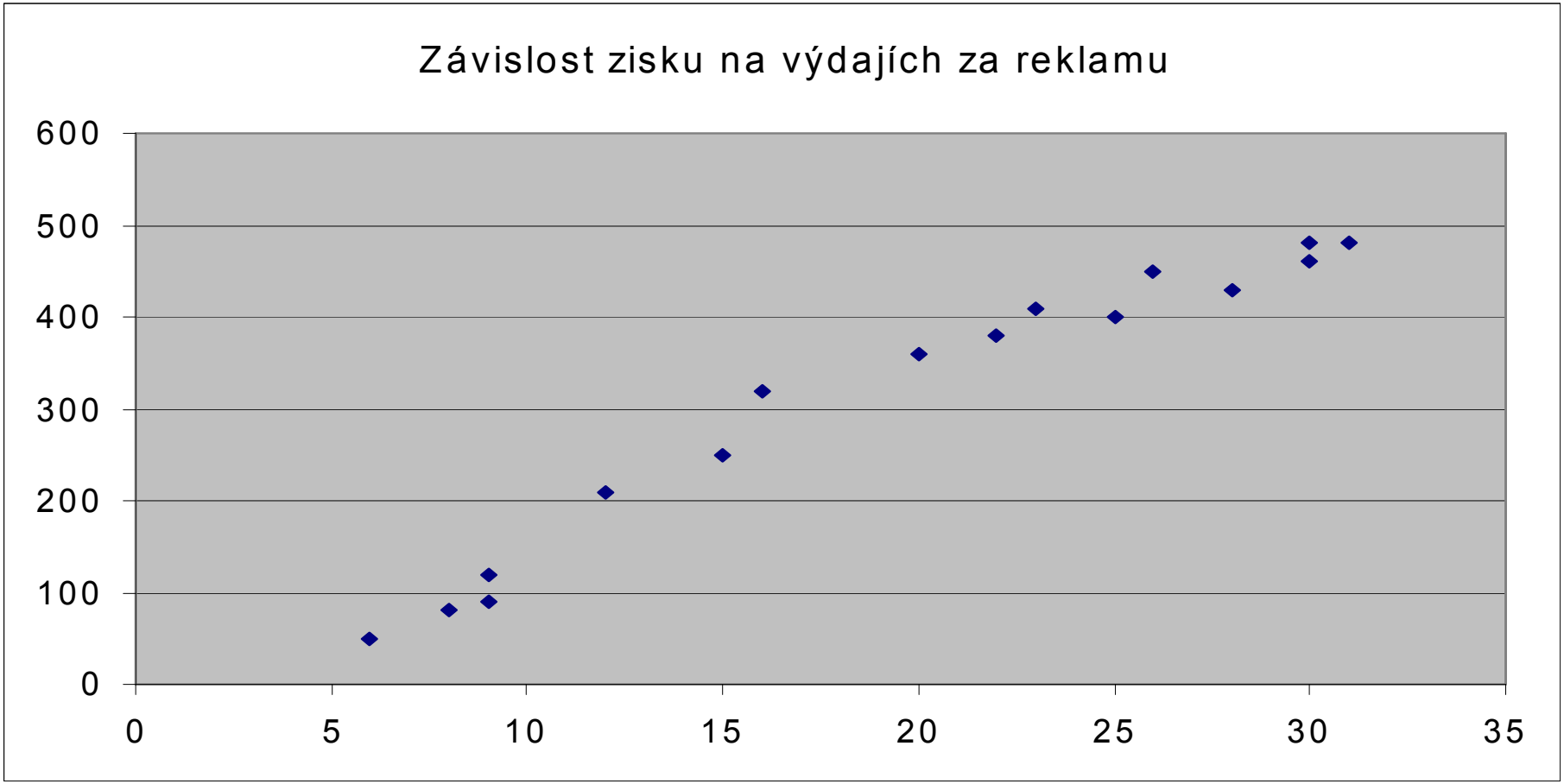
č. firmy	Výdaje na reklamu	Výdaje na reklamu	Zisk
1	malé	6	50
2	malé	8	80
3	malé	9	90
4	malé	9	120
5	středně velké	12	210
6	středně velké	15	250
7	středně velké	16	320
8	středně velké	20	360
9	středně velké	22	380
10	středně velké	23	410
11	velké	25	400
12	velké	26	450
13	velké	28	430
14	velké	30	460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

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ

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

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Příklad – grafické znázornění



Jednoduchá (jednorozměrná) lineární RA

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ

- Východiskem je vždy grafické znázornění
- Uspořádání bodů má tvar přímky, viz (B) nebo (C):

regresní přímka:

$$Y = B_0 + B_1 X$$

kritérium ← Y; prediktor ← X

regresní koeficienty: posunutí ← B_0; směrnice ← B_1

regresní model: $y_i = B_0 + B_1 x_i + \varepsilon_i, i = 1,2,\ldots n$

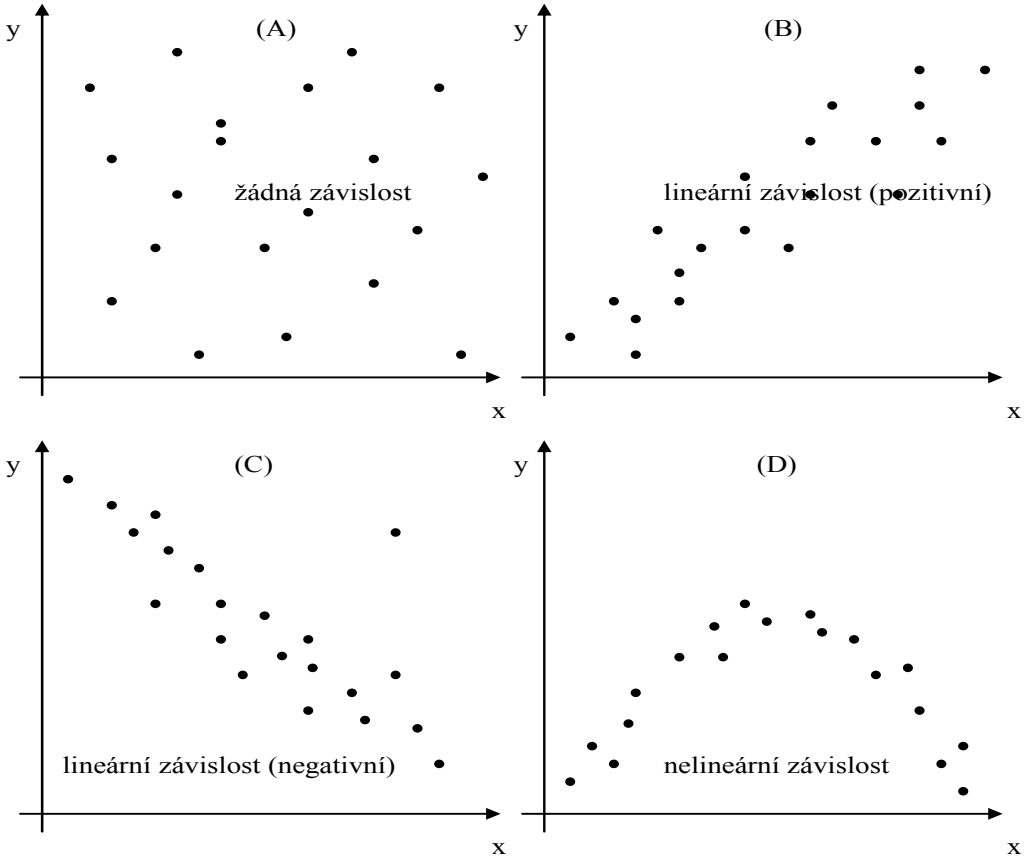
náhodná složka ← ε_i

- **Cíl**: nalezení nejlepších odhadů regresních koeficientů

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Bodový diagram (Scatter diagram)

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ



EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Metoda nejmenších čtverců

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ

(K. F. Gauss, 1777 – 1855)

- Data – body: $(x_1,y_1), (x_2,y_2),\ldots,(x_n,y_n)$
- Odhady regresních koeficientů B_0, B_1:

$$\sum_{i=1}^{n}\left(y_i-(b_1x_i+b_0)\right)^2 \rightarrow \min$$

$$b_1=\frac{\sum_{i=1}^{n}x_iy_i-n\bar{x}\,\bar{y}}{\sum_{i=1}^{n}x_i^2-n\bar{x}^2} \qquad b_0=\bar{y}-b_1\bar{x}$$

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Metoda nejmenších čtverců

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ

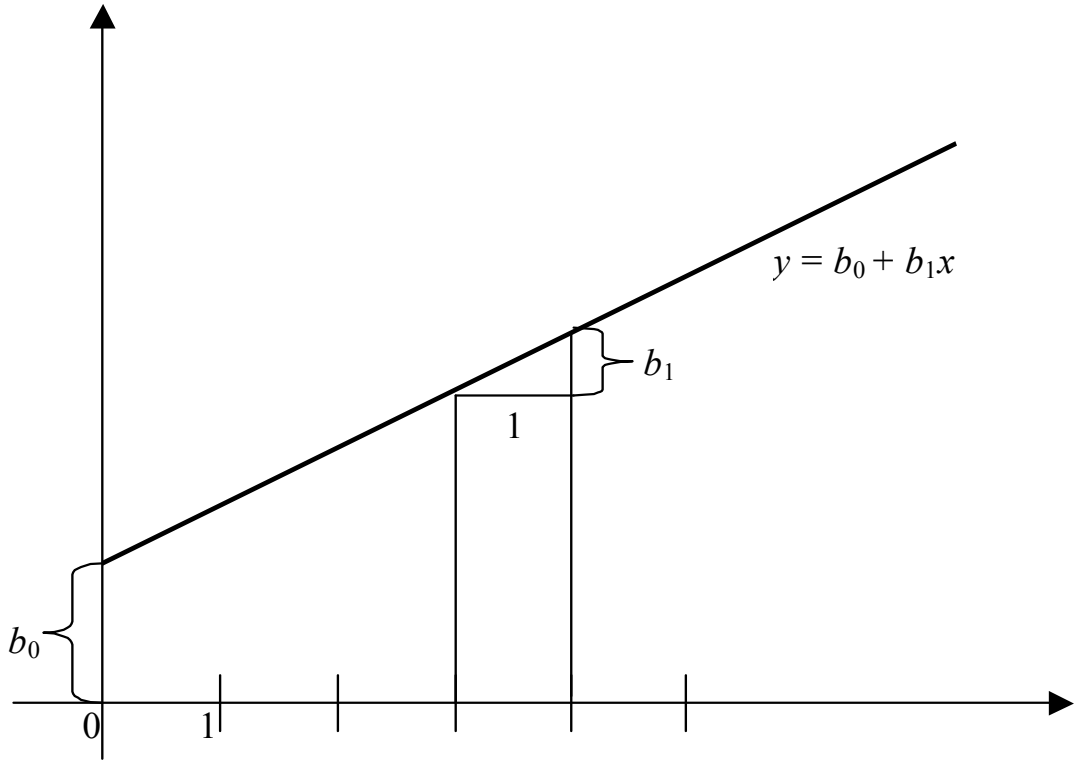
- Interpretace regresních koeficientů:

b_0 - úroveň kritéria y při nulové úrovni prediktoru x

b_1 - přírůstek kritéria y při jednotkovém přírůstku prediktoru x

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Regresní přímka



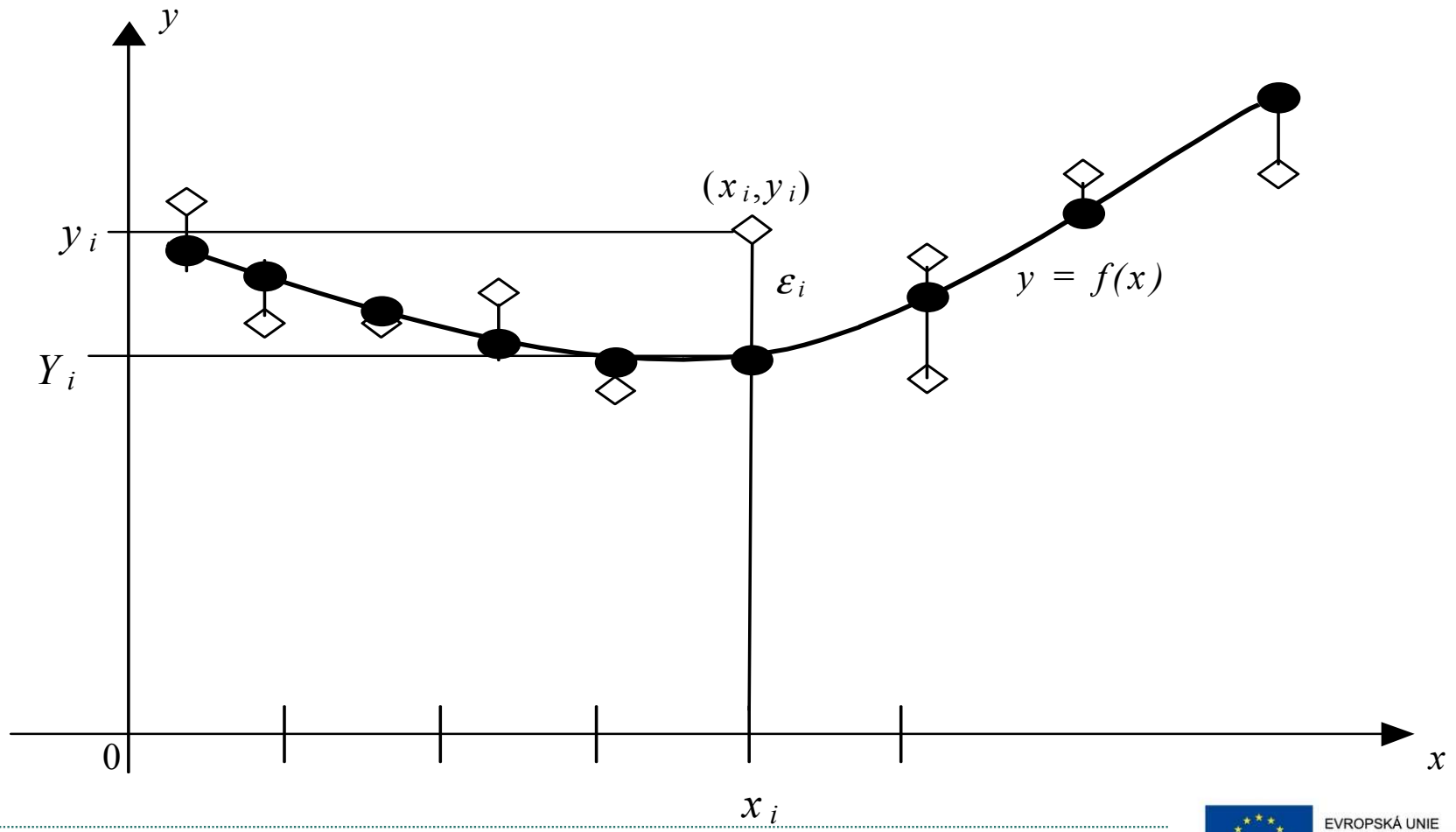
SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ

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

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Přiléhavost dat k regresní křivce

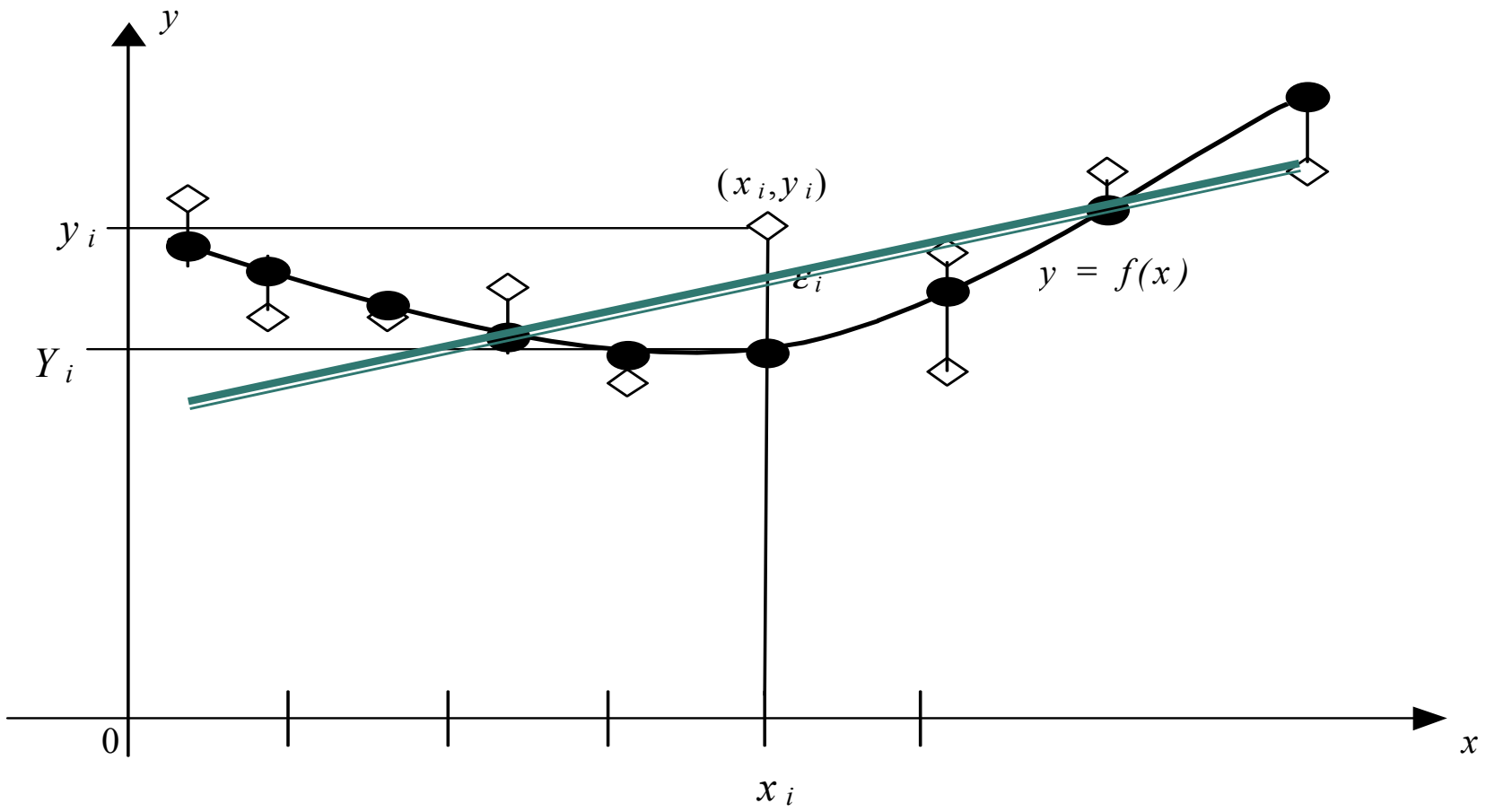
SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ



EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Přiléhavost dat k regresní přímce

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ



EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Přiléhavost regresní přímky k datům

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ

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

  Y_i - teoretické hodnoty („na regresní přímce“)

- Reziduální součet čtverců: $S_R = \sum_{i=1}^{n} (y_i - Y_i)^2$

- Celkový součet čtverců: $S_y = \sum_{i=1}^{n} (y_i - \bar{y})^2$

- Platí vztah: $S_y = S_T + S_R$

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Přiléhavost regresní přímky k datům

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ

Koeficient determinace –

míra přiléhavosti dat k regresní křivce:

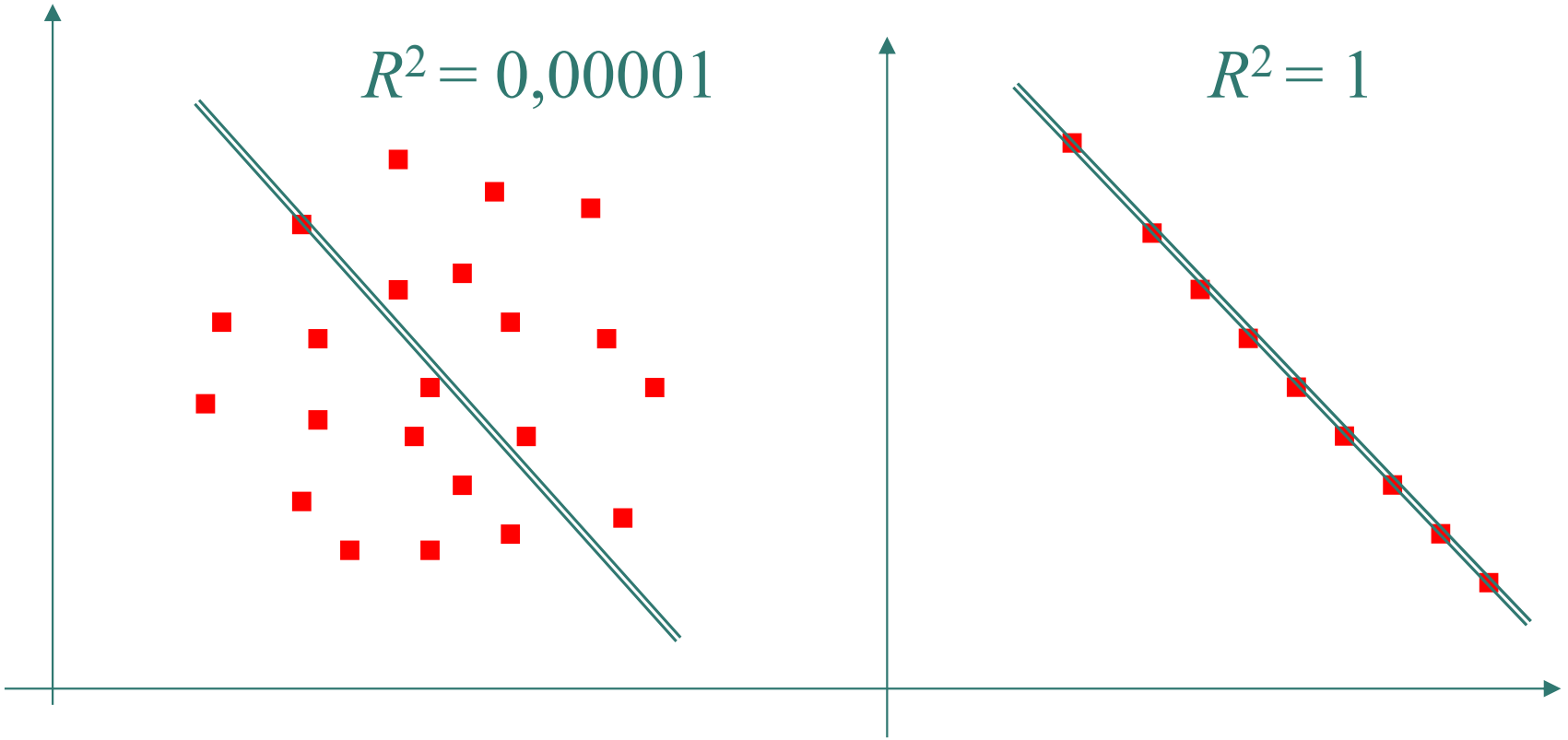
$$R^2 = \frac{S_T}{S_y} = 1 - \frac{S_R}{S_y}$$

- Platí: $0 \leq R^2 \leq 1$
- Pozor! R^2 má platnost pro libovolný typ regresní funkce!

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Extrémní hodnoty koeficientu determinace R^2

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ



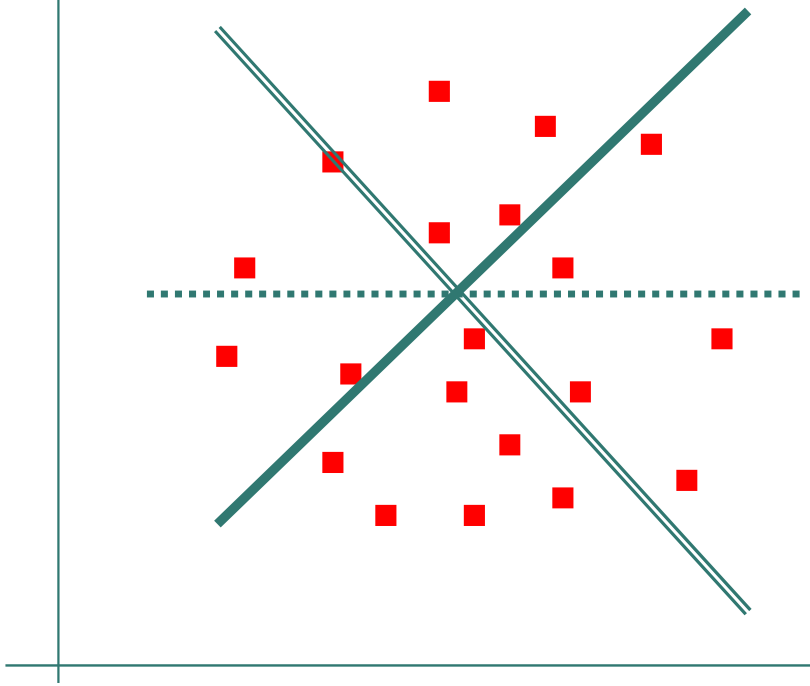
EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Jak jsou „výstižné“ regresní modely?

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ

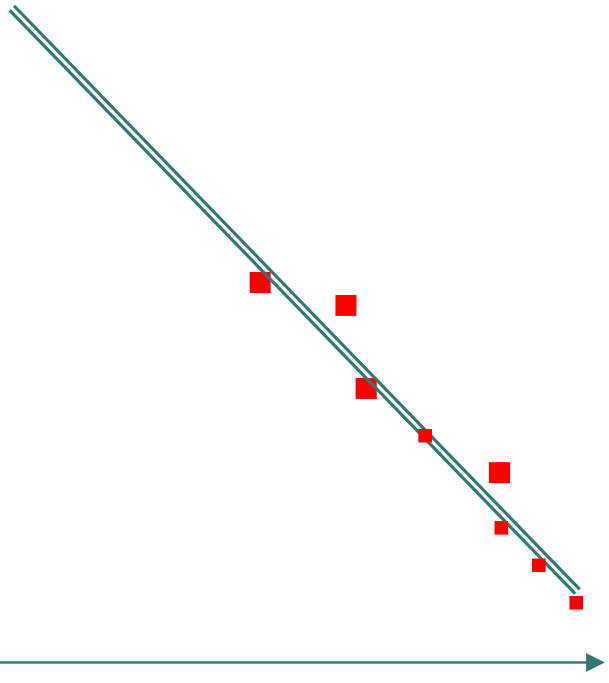
A)

Statisticky nevýznamný



B)

Statisticky významný



$$y = B_0 + B_1 x + \varepsilon$$

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Klasický jednoduchý lineární regresní model

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ

Předpoklady:

1. Vysvětlující proměnná X je nestochastická – vyplývá z povahy problému

2. **Střední hodnota náhodné chyby ε je 0**, tj.
   $E(\varepsilon) = 0$ – pro MNČ vždy splněno!

3. **Rozptyl náhodné chyby ε je konstantní**, tj.
   $Var(\varepsilon) = \sigma^2$ - test, např. Chi-kvadrát
   (**Homoskedasticta**)

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Klasický jednoduchý lineární regresní model

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ

Předpoklady:

4. **Náhodné chyby ε jsou nekorelované**, tj. **Autokorelace** = 0, tj. $Cov(\varepsilon_i, \varepsilon_j) = 0$ pro $i \neq j$ – test nulovosti korelačního koeficientu

5. **Náhodná chyba má normální rozdělení**, tj. $\varepsilon \sim N(0, \sigma^2)$ – test normality

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Testy hypotéz

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ

1. Testuje se hypotéza H_0: regresní koeficient = 0 - t-test
   (A) H_0: $B_0 = 0$, (B) H_0: $B_1 = 0$

2. Test současné nulovosti obou regresních koeficientů - F-test
   (v Excelu tzv. ANOVA)

3. Testuje se hypotéza H_0: koeficient determinace = 0 - t-test
   H_0: $R^2 = 0, j = 0,1$

4. Intervaly spolehlivosti regresních koeficientů

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Testy hypotéz – 1.TEST

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ

1. Testuje se hypotéza H_0: regresní koeficient = 0

(A) $H_0: B_0 = 0$, (B) $H_0: B_1 = 0$

Testové kritérium: $$T = \frac{b_j}{\sqrt{\frac{S_R}{n-2} h_j}}$$

$$h_0 = \frac{\sum x_i^2}{n\sum x_i^2 - \left(\sum x_i\right)^2} \qquad h_1 = \frac{n}{n\sum x_i^2 - \left(\sum x_i\right)^2}$$

Kritický obor: $$|T| > t_{\alpha/2}(n-2)$$

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Testy hypotéz – 3.TEST

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ

3. Testuje se hypotéza H_0: koeficient determinace = 0

H_0: $R^2 = 0, j = 0,1$

Testové kritérium: $$T = \frac{R\sqrt{n-2}}{\sqrt{1-R^2}}$$

$$R = \sqrt{1-\frac{S_R}{S_y}}$$

Kritický obor: $$T > t_\alpha(n-2)$$

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Příklad 1 – STUDIE – regresní rovnice

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ

Existuje (lineární) závislost počtu vypitých limonád (za týden) na věku?

Kriterium y - počet limonád / týden
Prediktor x - věk respondenta

Regresní rovnice: $y = 4{,}40 + 0{,}37x$

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Příklad 1 – STUDIE – testování hypotéz

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ

Hypotézy o statistické významnosti regresních koeficientů B_j a R^2:

H_0: koeficient = 0

$b_0 = 4{,}40$ (p-hodnota $= 0{,}48 > 0{,}05 \Rightarrow H_0$ nezamítáme)

$b_1 = 0{,}37$ (p-hodnota $= 0{,}16 \Rightarrow H_0$ nezamítáme)

Koeficient determinace (přiléhavost): $R^2 = 0{,}179$

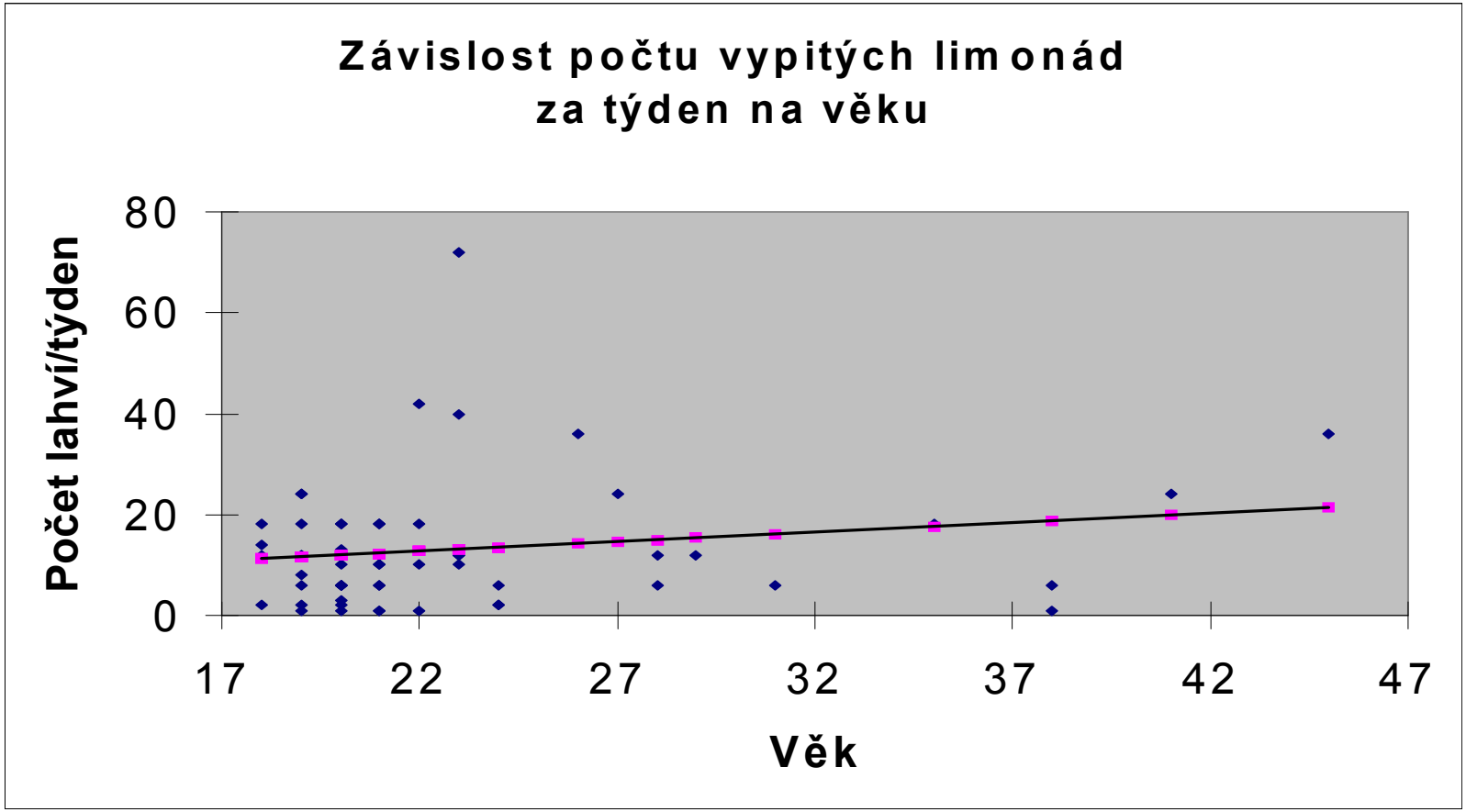
(p-hodnota $= 0{,}12 \Rightarrow H_0$ nezamítáme)

Závěr: **Regresní model není statisticky významný!**

***Jinak řečeno*:** **Neexistuje lineární závislost počtu vypitých limonád na věku!**

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Grafické znázornění



SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ
EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Příklad 2 – výdaje na reklamu

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ

Výdaje na reklamu	Zisk
6	50
8	80
9	90
9	120
12	210
15	250
16	320
20	360
22	380
23	410
25	400
26	450
28	430
30	460
30	480
31	480

Existuje (lineární) závislost zisku z prodeje výrobku na výdajích za reklamu?

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Příklad 2 - řešení

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ

Kritérium

y - zisk z prodeje daného výrobku (v mil. Kč/rok)

Prediktor

x - výdaje na reklamu (v mil. Kč/rok)

Regresní rovnice: $y = -24{,}88 + 17{,}32x$

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Příklad 2 - řešení

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ

Hypotézy o statistické významnosti regresních koeficientů a R^2:

b_0 = - 24,878 (p-hodnota = 0,27 $\Rightarrow$ H_0 nezamítáme)

b_1 = 17,316 (p-hodnota = 0,0000008 $\Rightarrow$ H_0 zamítáme)

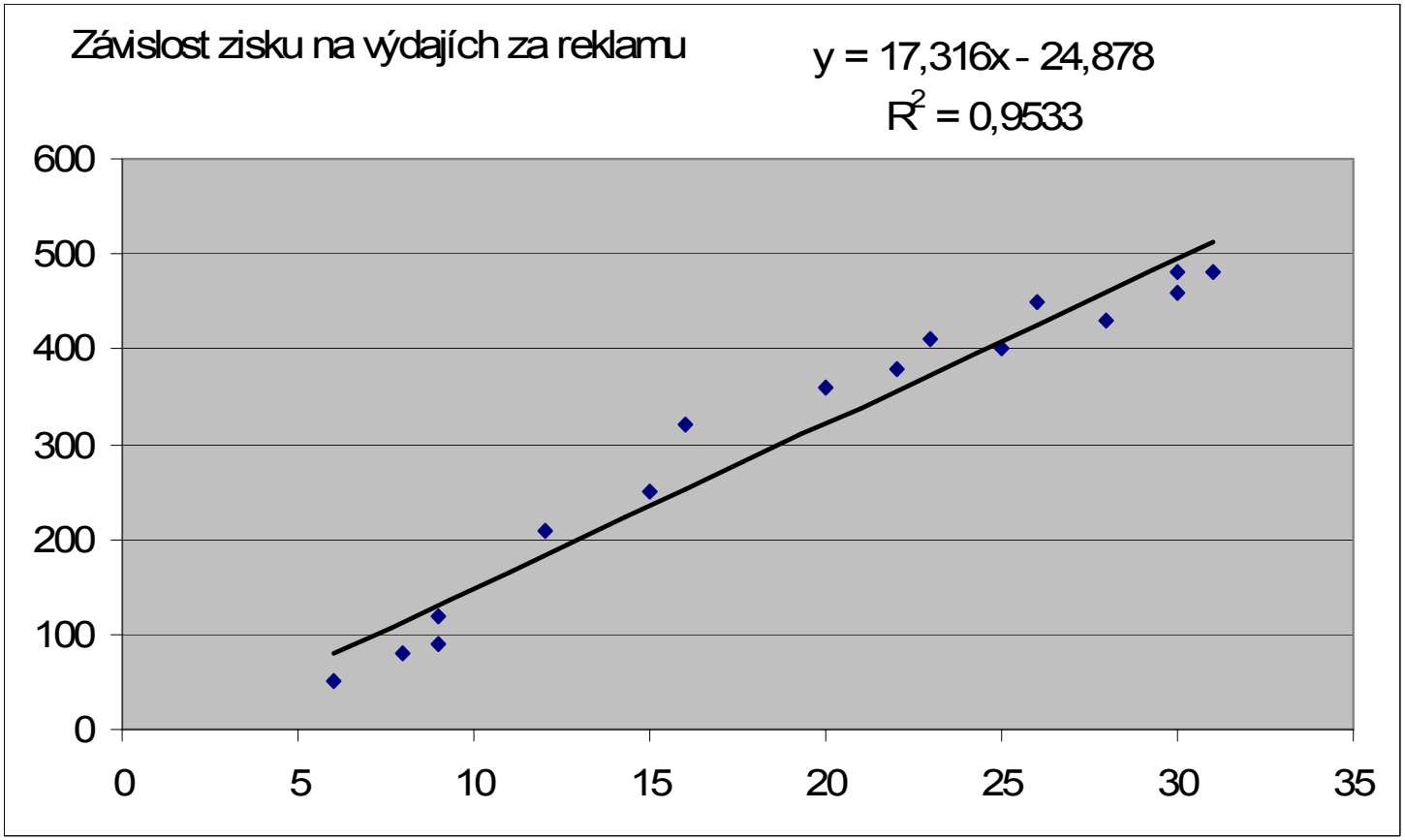
Koeficient determinace (přiléhavost): R^2 = 0,958
(p-hodnota = 0,00005 $\Rightarrow$ H_0 zamítáme)

Závěr: Existuje *silná* lineární závislost!

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Příklad 2 – grafické znázornění

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ



EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Příklad 2 – řešení v EXCELU

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ

Data → Analýza dat → Regrese…

Regresní statistika	
Násobné R	0,976
Hodnota spolehlivosti R	0,953
Nastavená hodnota spolehlivosti R	0,950
Chyba stř. hodnoty	34,574
Pozorování	16

ANOVA

	Rozdíl	*SS*	*MS*	*F*	*Významnost F*
Regrese	1	341758,78	341758,78	285,91	0,00
Rezidua	14	16734,97	1195,36		
Celkem	15	358493,75			

	Koeficienty	*Chyba stř. hodnoty*	*t stat*	*Hodnota P*	*Dolní 95%*	*Horní 95%*
Hranice	-24,878	21,643	-1,149	0,270	-71,298	21,541
Výdaje na reklamu	17,316	1,024	16,909	0,000	15,120	19,513

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Závěr přednášky

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ

Děkuji Vám za pozornost!!!

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY