

## Cvičení 7.: Jednoduchá lineární regrese

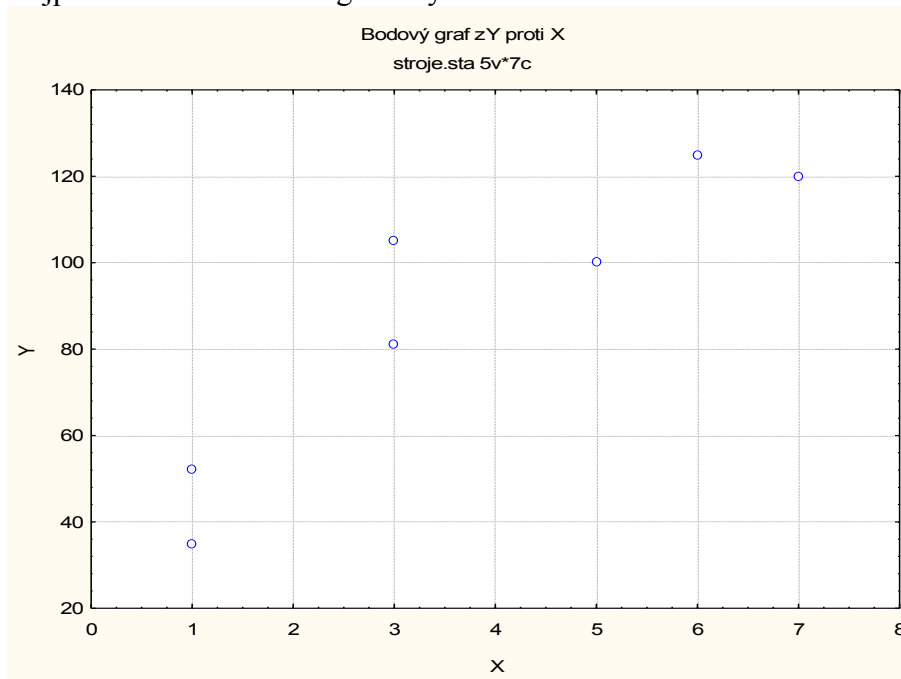
**Vzorový příklad:** U sedmi náhodně vybraných strojů v určitém podniku se zjišťovalo stáří stroje v letech (proměnná X) a týdenní náklady v Kč na údržbu stroje (proměnná Y). Data: (1,35), (1,52), (3,81), (3,105), (5,100), (6,125), (7, 120)

Data znázorníte graficky. Vyzkoušejte následující čtyři modely:

$y = \beta_0 + \beta_1 x$ ,  $y = \beta_0 + \beta_1 \sqrt{x}$ ,  $y = \beta_0 + \beta_1 \log_{10} x$ ,  $y = \beta_0 + \beta_1 1/x$ . Vyberte ten model, který poskytuje nejvyšší index determinace. Určete regresní odhad týdenních nákladů pro stroj starý čtyři roky.

### Řešení:

Nejprve data znázorníme graficky:



Datový soubor s proměnnými X a Y doplníme o proměnné SQRTX, LOGX a INVX. Hodnoty proměnné SQRTX resp. LOGX resp. INVX získáme tak, že do Dlouhého jména napíšeme =sqrt(x) resp. =Log10(x) resp. =1/x.

	1	2	3	4	5
	X	Y	SQR	LOG	INV
1	1	35	1	0	1
2	1	52	1	0	1
3	3	81	1,732	0,477	0,333
4	3	105	1,732	0,477	0,333
5	5	100	2,236	0,699	0,2
6	6	125	2,449	0,778	0,167
7	7	120	2,646	0,845	0,143

Regresní analýzu provedeme tak, že roli nezávisle proměnné bude hrát proměnná X, pak SQRTX, LOGX a nakonec INVX.

Model s proměnnou X:

		Výsledky regrese				
		R= ,91004028 R2= ,8				
		F(1,5)=24,099 p<,004				
N=		Be	Sm	B	Sm	t(Úrc
		be		B		B
Abs				39,	11,	3,4 0,0
X	0,9	0,1	13,	2,	4,9	0,0

Model s proměnnou SQRTX:

		Výsledky regrese				
		R= ,93923698 R2= ,8				
		F(1,5)=37,433 p<,001				
N=		Be	Sm	B	Sm	t(Úrc
		be		B		B
Abs				-0,	15,	-0,0 0,9
SQ	0,9	0,1	48,	7,	6,1	0,0

Model s proměnnou LOGX:

		Výsledky regrese				
		R= ,95349135 R2= ,9				
		F(1,5)=50,033 p<,000				
N=		Be	Sm	B	Sm	t(Úrc
		be		B		B
Abs				44,	7,	5,9 0,0
LOG	0,9	0,1	93,	13,	7,0	0,0

Model s proměnnou INVX

		Výsledky regrese				
		R= ,94282234 R2= ,8				
		F(1,5)=40,010 p<,001				
N=		Be	Sm	B	Sm	t(Úrc
		be		B		B
Abs				12,	7,	16, 0,0
INV	-0,9	0,1	-8,	13,	-6,	0,0

Vidíme, že nejvyšší index determinace poskytuje model s proměnnou LOGX:  $ID^2 = 90,9\%$ . Má také nejmenší směrodatnou chybu odhadu.

Určíme regresní odhad týdenních nákladů pro stroj starý 4 roky v modelu s nezávisle proměnnou LOGX. Nejprve vypočteme  $\log(4) = 0,602$

Pro výpočet predikované hodnoty zvolíme Rezidua/předpoklady/předpovědi Předpovědi závisle proměnné X: 0,602 OK. Ve výstupní tabulce je hledaná hodnota označena jako Předpověď.

Předpovězené hodnoty (stroje.sta)			
proměnné: Y			
Proměnná	B-váží	Hodnota	B-váží * Hodnota
LOGX	93,23472	0,602000	56,1273
Abs. člen			44,6457
Předpověď			100,7730
-95,0%LS			88,9277
+95,0%LS			112,6184

Bodový odhad je 100,77 Kč. Vidíme, že s pravděpodobností aspoň 0,95 budou týdenní náklady na údržbu stroje starého 4 roky činit minimálně 88,93 Kč a maximálně 112,62 Kč.

Nakonec znázorníme data se všemi čtyřmi regresními křivkami. K původnímu datovému souboru s proměnnými X,Y přidáme 4 nové proměnné PREDIKCE1, ..., PREDIKCE4. Do Dlouhých jmen těchto proměnných napíšeme příslušné regresní rovnice, tj.

$$=39,44444+13,14957*x$$

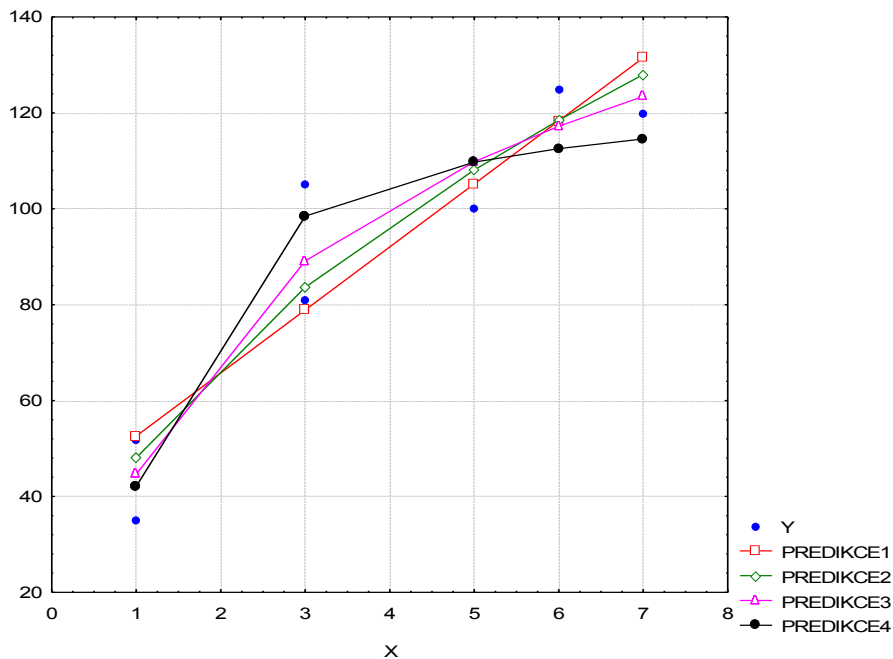
$$=-0,4776+48,55972*\sqrt{x}$$

$$=44,64571+93,23472*\log x$$

$$=126,6192-84,4832*\ln x$$

	1	2	3	4	5	6	7	8	9
	X	Y	SC	LC	INPR	PR	PR	PR	PR
		35		40		52	48	44	42
		45		40		52	48	44	42
		80	1,70	40,3		78	83	89	98
		110	1,70	40,3		78	83	89	98
	4	120	2,00	40		104	108	109	109
	6	120	2,00	70		114	118	117	112
	7	120	2,00	80		130	127	123	114

Obrázek vytvoříme pomocí vícenásobného bodového grafu.



### **Příklad k samostatnému řešení**

V r. 2010 bylo u studentů MU provedeno dotazníkové šetření, které se týkalo údajů o mobilech. Na dotazník odpovědělo 67 respondentů.

Dotazník:

1. Jaká je značka Vašeho mobilního telefonu?

- a) LG
- b) Nokia
- c) Samsung
- d) Sony Ericsson
- e) jiná
- f) nemám mobilní telefon

2. Jaký operační systém Váš mobilní telefon používá?

- a) Android
- b) Symbian
- c) Windows Mobile
- d) jiný
- e) žádný

3. Na pětibodové stupnici vyjádřete spokojenost s funkcemi svého mobilu

- a) naprostá spokojenost
- b) spokojenost
- c) neutrální postoj
- d) nespokojenost
- e) naprostá nespokojenost

4. Máte v úmyslu si do konce roku 2010 pořídit jiný mobil?

- a) ano
- b) ne

5. Jaká je barva Vašeho mobilu?

- a) bílá
- b) černá
- c) stříbrná
- d) jiná

6. Uveďte hmotnost svého mobilu (v g)

7. Uveďte výšku svého mobilu (v mm)

8. Uveďte šířku svého mobilu (v mm)

9. Uveďte hloubku svého mobilu (v mm)

10. Jaké je Vaše pohlaví?

- a) muž
- b) žena

Získaná data jsou uložena v souboru dotazník\_mobily.sta.

Modelujte závislost objemu mobilu na jeho hmotnosti pomocí regresní přímky.

- a) Vypočtete odhady regresních parametrů, napište rovnici regresní přímky a interpretujte její parametry. Do dvourozměrného tečkového diagramu zakreslete regresní přímku s 95% pásem spolehlivosti a 95% predikčním pásem.
- b) Najděte odhad rozptylu, proveďte celkový F-test a rovněž dílčí t-testy o významnosti regresních parametrů..
- c) Najděte 95% intervaly spolehlivosti pro regresní parametry a zjistěte relativní chyby odhadů regresních parametrů.
- d) Vypočtete index determinace a interpretujte ho. Vypočtete rovněž střední absolutní procentuální chybu predikce (MAPE).
- e) Proveďte analýzu reziduí.