

5.2 Lineárny trend

Uvažujme **lineárnu regresnú funkciu** (regresnú priamku).
Nech $(t_1, y_1), \dots, (t_n, y_n)$ je postupnosť bodov v rovine.

Priamku $Tr_t = a_0 + a_1 t$ nazveme regresnou priamkou (lineárnym trendom), ak pre všetky $a_0, a_1 \in R$ platí

$$\frac{\partial S(a_0, a_1)}{\partial a_0} = 0,$$
$$\frac{\partial S(a_0, a_1)}{\partial a_1} = 0,$$

$$\text{kde } S(a_0, a_1) = \sum_{i=1}^n (y_i - a_0 - a_1 t_i)^2.$$

Konštanty a_0, a_1 sú riešením sústavy rovníc

$$a_0 n + a_1 \sum_{i=1}^n t_i = \sum_{i=1}^n y_i$$
$$a_0 \sum_{i=1}^n t_i + a_1 \sum_{i=1}^n (t_i)^2 = \sum_{i=1}^n t_i y_i$$

Ak chceme získať predpoveď očakávanej hodnoty v čase t_{n+1} , dosadíme túto hodnotu do nájdenej rovnice priamky teda

$$y_p = Tr_{n+1} = a_0 + a_1 t_{n+1}.$$

Získaná predpoveď v čase je tzv. bodová predpoveď. Často sa stretávame s požiadavkou, aby výsledkom uvedenej extrapolácie (predpovede) bola virohodnejšia hodnota, a to predpoveď intervalu, v ktorom môžeme s udanou pravdepodobnosťou očakávať vývoj budúcej hodnoty daného časového radu. V tomto prípade hovoríme o tzv. pravdepodobnostnej intervalovej predpovedi.

Pre prípad trendovej priamky, tvar intervalovej predpovede budúcej hodnoty je

$$Tr_{n+1} - t_{1-\alpha/2}(n-2) \cdot s \cdot g_p < y_p < Tr_{n+1} + t_{1-\alpha/2}(n-2) \cdot s \cdot g_p$$

Kde $t_{1-\alpha/2}(n-2)$ je $(1-\alpha/2) \cdot 100\%$ kvantil t -studentovho rozdelenia pravdepodobnosti o $n-2$ stupňoch voľnosti

$$s = \sqrt{\frac{\sum_{i=1}^n y_i^2 - \sum_{i=1}^n Tr_i^2}{n-2}}, \quad g_p = \sqrt{1 + \frac{1}{n} + \frac{(t_{n+1} - \bar{t})^2}{\sum_{i=1}^n t_i^2 - n \cdot (\bar{t})^2}}, \quad \bar{t} = \frac{1}{n} \sum_{i=1}^n t_i.$$

Pri použití Excelu dostaneme na hladine významnosti α horný a dolný odhad koeficientov a_0, a_1 .

Poznámka Niekedy používame váženú metódu najmenších štvorcov

$$S(A_0, A_1) = \sum_{i=1}^n w_i (y_i - A_0 - A_1 t_i)^2,$$

kde w_i sú váhy odpovedajúce jednotlivým bodom časového radu.

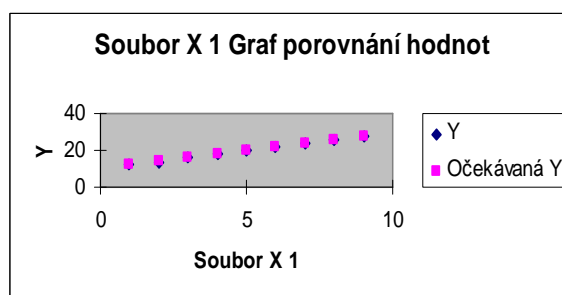
Poznámka Pri aproximácii funkcie viacerých premenných zaradíme do modelu len tie, ktoré významne zlepšia regresný odhad. Počet premenných sa snažíme minimalizovať tak, aby model nebol zbytočne zložitý. Jednou z metód, ktorá nám umožní výber najlepšej podmnožiny premenných, je metóda postupnej regresie (stepwise), ktorá spočíva v postupnom pridávaní premenných do modelu dovtedy, kým nedostaneme prijateľný odhad.

Príklad 5.1 Uvedieme určenie lineárneho trendu pomocou metódy najmenších štvorcov pre časový rad daný tabuľkou.

x	1	2	3	4	5	6	7	8	9
y	12,1	13,8	16,05	17,94	20,05	21,93	24,13	26,1	28

Riešenie pomocou EXCELU:

t	y	Očakávaná Trt	Rezidua
1	12,1	11,96778	0,132222
2	13,8	13,97861	-0,17861
3	16,05	15,98944	0,060556
4	17,94	18,00028	-0,06028
5	20,05	20,01111	0,038889
6	21,93	22,02194	-0,09194
7	24,13	26,04361	0,097222
8	26,1	24,03278	0,056389
9	28	28,05444	-0,05444
VÝSLEDEK			
	Koeficienty	Dolní 95%	Horní 95%
Hranice	9,956944	9,770736	10,14315
Soubor X 1	2,010833	1,977743	2,043923



Teda trend je $Tr_t = 9,956944 + 2,010833 \cdot t$.

Riešenie pomocou MAPLEu:

```
> with(stats):fit[leastsquare]([t,Trt]]([1,2,3,4,5,6,7,8,9],
[12.1,13.8,16.05,17.94,20.05,21.93,24.13,26.1,28]));
```

$$Trt = 9.956944444 + 2.010833333t$$

alebo

```
> with(stats):Tvalues:=[1,2,3,4,5,6,7,8,9]:
```

Yvalues:=[12.1,13.8,16.05,17.94,20.05,21.93,24.13,26.1,28]:

> Trt:= fit[leastsquare[[t,y], y=a*t+b, {a,b}]]([Tvalues, Yvalues]);

$$Trt := y = 2.010833333t + 9.956944444$$

Riešenie pomocou MATLABu:

>> x=[1,2,3,4,5,6,7,8,9];y=[12.1,13.8,16.05,17.94,20.05,21.93,24.13,26.1,28];

>> Trt=polyfit(x,y,1)

Trt =

2.0108 9.9569

Príklad 5.2 Pomocou metódy najmenších štvorcov aproximujme funkciu zadanú tabuľkou lineárnou funkciou a odhadnime hodnotu v bode $t = 6$.

t	1	1	2	2	2	3	4	4	4	5
y	2,1	1,9	1,1	1,2	0,9	-2,1	-7,1	-7,3	-6,9	-14,3

Riešenie:

Lineárna funkcia má tvar $Tr_t = a_0 + a_1 \cdot t$, konštanty a_0, a_1 sú riešením sústavy rovníc:

$$a_0 n + a_1 \sum_{i=1}^n t_i = \sum_{i=1}^n y_i$$

$$a_0 \sum_{i=1}^n t_i + a_1 \sum_{i=1}^n (t_i)^2 = \sum_{i=1}^n t_i y_i$$

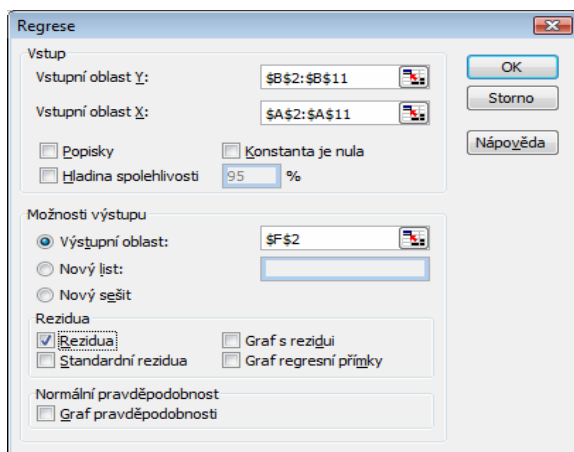
$$a_0 \cdot 10 + a_1 \cdot 28 = -30,5$$

$$a_0 \cdot 28 + a_1 \cdot 96 = -152,6$$

$$a_0 = 7,6409$$

$$a_1 = -3,8182$$

Iný spôsob výpočtu (pomocou EXCEL-u):



Použijeme Excel príkazy
Nástroje - Analýza dát – Regresia

Pole Y– označíme hodnoty y

Pole X – označíme hodnoty t

Výstupná oblasť – ľubovoľné prázdne okno

Výstup z EXCEL-u:

VÝSLEDEK

<i>Regresní statistika</i>	
Násobné R	0,9596
Hodnota	
spolehlivosti R	0,9208
Nastavená hodnota	
spolehlivosti R	0,9108
Chyba stř. hodnoty	1,6614
Pozorování	10

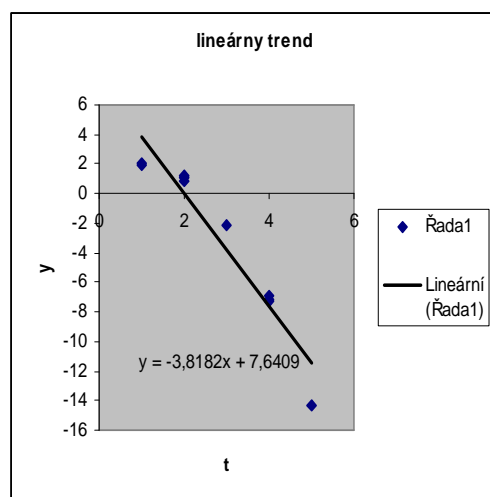
ANOVA

					Významnost
	<i>Rozdíl</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>F</i>
Regrese	1	256,5818	256,5818	92,9510	0,0000
Rezidua	8	22,0832	2,7604		
Celkem	9	278,6650			

	<i>Koeficienty</i>	<i>Chyba</i> <i>stř. hodnoty</i>	<i>t stat</i>	<i>Hodnota</i> <i>P</i>	<i>Dolní 95%</i>	<i>Horní 95%</i>
Hranice	7,6409	1,2271	6,2270	0,0003	4,8113	10,4705
Soubor X 1	-3,8182	0,3960	-9,6411	0,0000	-4,7314	-2,9049

REZIDUA

<i>Pozorování</i>	<i>Očekávaná</i> <i>Y</i>	<i>Rezidua</i>
1	3,8227	-1,7227
2	3,8227	-1,9227
3	0,0045	1,0955
4	0,0045	1,1955
5	0,0045	0,8955
6	-3,8136	1,7136
7	-7,6318	0,5318
8	-7,6318	0,3318
9	-7,6318	0,7318
10	-11,4500	-2,8500



Hľadaná krivka má tvar: $Tr_t = 7,6409 - 3,8182 \cdot t$

Hodnoty s pravdepodobnosťou $(1-\alpha) \cdot 100\%$ ležia medzi priamkami:

$$4,8113 - 4,7314t \quad \text{a} \quad 10,4705 - 2,9049t$$

Hodnota pre $t = 6$:

$$Tr_6 = 7,6409 - 3,8182 \cdot 6 = -15,2683$$