

## Korelačná a regresná analýza

**Príklad** V tabuľke sú uvedené výsledky meraní výkonu [kW] a otáčok [min<sup>-1</sup>] benzínového motora:

Otáčky	2000	2500	3000	3500	4000
Výkon	29	43	55	64	71

Vypočítajme: a) výberovú modifikovanú kovarianciu  $k_{xy}^*$ , b) výberový korelačný koeficient  $r_{xy}$ , c) odhadnime regresné koeficienty polynómu 1. a 2. stupňa pre vyjadrenie závislosti výkonu motora od jeho otáčok a na hladine významnosti  $\alpha = 0,1$  rozhodnime, ktorý model je výhodnejší.

**Riešenie:** a) Výberovú modifikovanú kovarianciu vypočítame podľa vzťahu

$$k_{xy}^* = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}), \text{ kde } \bar{x} \text{ a } \bar{y} \text{ sú výberové priemery a } n \text{ je rozsah náhodného výberu } V_n.$$

```
>> x=[2000:500:4000];y=[29,43,55,64,71];  
>> KOVxy=cov(x,y),kxy=KOVxy(1,2)
```

```
KOVxy =  
1.0e+005 *  
6.2500 0.1313  
0.1313 0.0028           % kovariančná matica
```

```
kxy =  
13125           % výberová modifikovaná kovariancia
```

b) Výberový korelačný koeficient vypočítame podľa vzťahu  $r_{xy} = \frac{k_{xy}^*}{s_x^* s_y^*}$ , kde  $s_x^*$ ,  $s_y^*$  sú modifikované výberové smerodajné odchýlky náhodného výberu  $V_n$ .

```
>> KORxy=corrcoef(x,y),rxy=KORxy(1,2)
```

```
KORxy =  
1.0000 0.9907  
0.9907 1.0000           % korelačná matica
```

```
rxy =  
0.9907           % výberový korelačný koeficient
```

c)

Aproximáciu pomocou polynómu n-tého stupňa dostaneme použitím **polyfit(x,y,n)**.

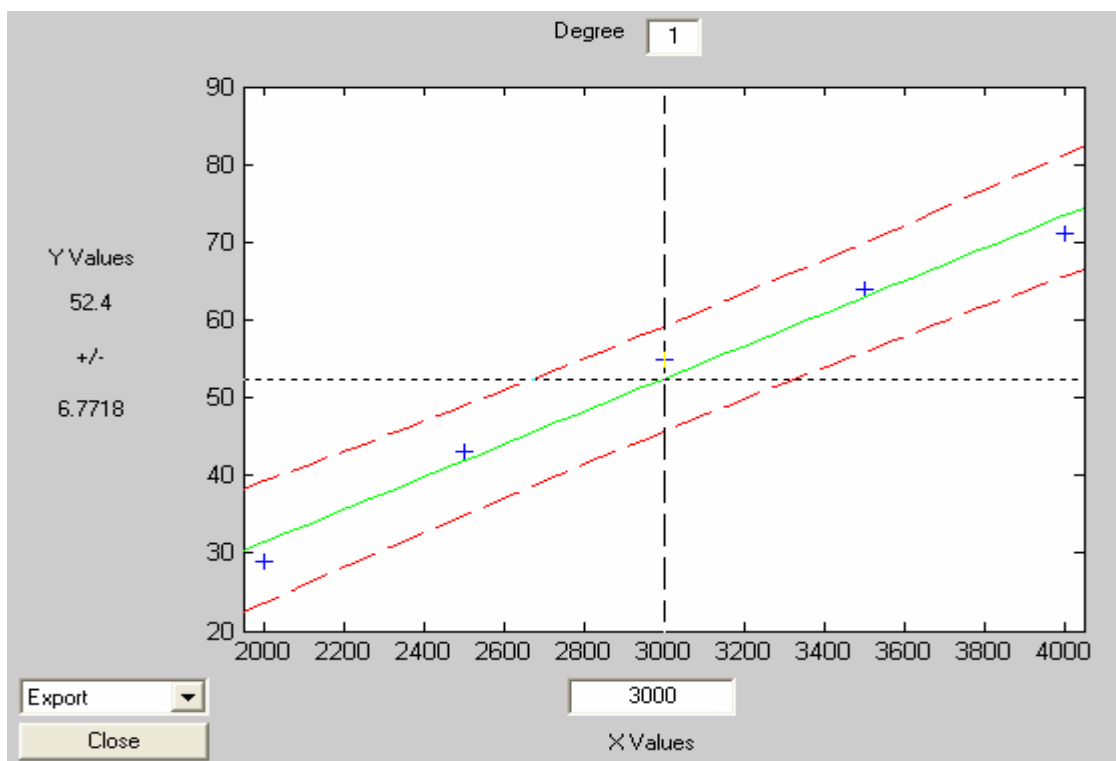
```
>> x=[2000:500:4000];y=[29,43,55,64,71];polyfit(x,y,1)
```

```
ans =  
0.0210 -10.6000
```

Teda hľadaná lineárna funkcia je  $f(x) = 0,0210x - 10,6000$  (koeficienty sú zaokrúhlené na 4 desatinné miesta).

**Funkcia polytool(x,y,n,alpha)** zobrazí 100(1-alpha)%-ný interval spoľahlivosti pre predpovedané hodnoty pomocou polynómu n-tého stupňa.

```
>> polytool(x,y,1,0.1)
```



Teda napríklad  $f(3000) = 52,4 \pm 6,7718$ .

`>> format long;polyfit(x,y,2)`

ans =

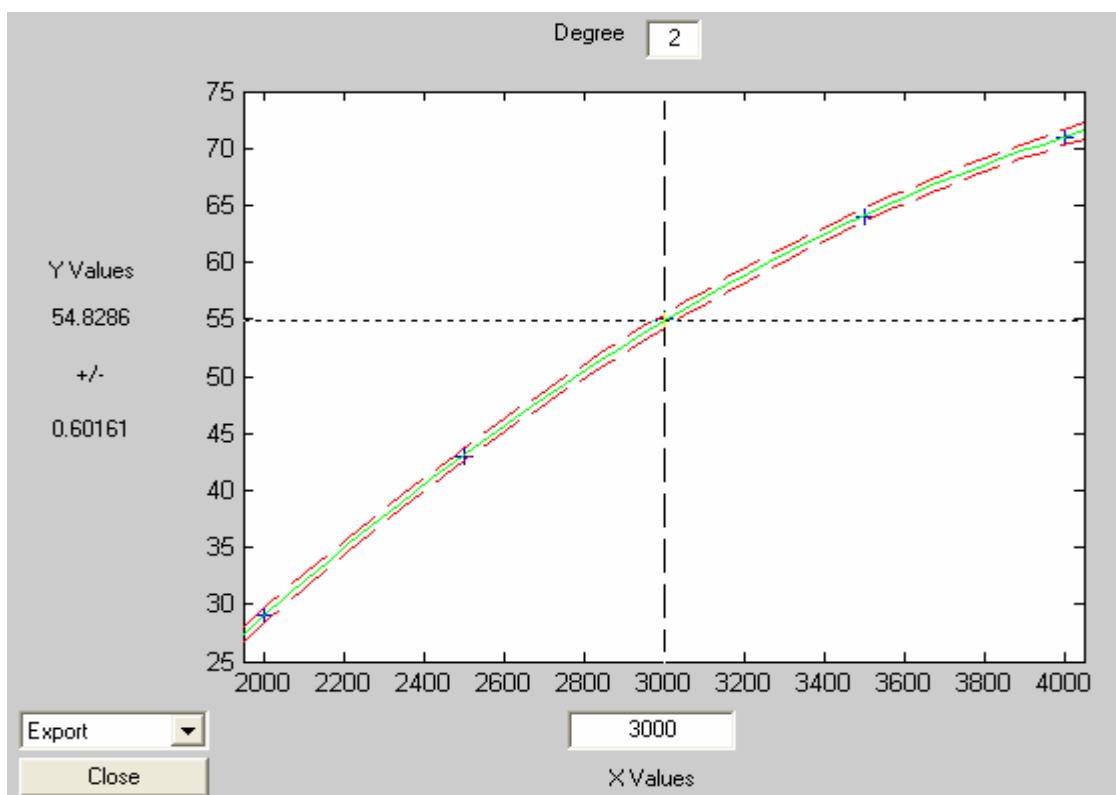
-0.00000485714286 0.05014285714286 -51.88571428571460

Teda  $f(x) = -0,00000485714286x^2 + 0,05014285714286x - 51,88571428571460$ .

Ak nepoužijeme **format long**, dostaneme v tomto prípade nekorektný výsledok

$f(x) = -0,0000x^2 + 0,0501x - 51,8857$ .

`>> polytool(x,y,2,0.1)`



To znamená, že  $f(3000) = 54,8286 \pm 0,60161$ , čo je podstatne lepšia aproximácia ako pomocou lineárnej funkcie.