Regresní analýza v Maple

Cílem tohoto materiálu není probrat problematiku regresní analýzy či kompletní problematiku práce se systémem počítačové algebry, systémem Maple.
Text pouze poukazuje na možnost využití statistiky a systému Maple pro odhalení trendu.

Příklad

V tabulce 1 jsou uvedeny roční tržby společnosti ABC za posledních 7 let. S využitím systému Maple vyrovnáme zadané tržby nejběžnějšími regresními funkcemi. Následně na základě reziduálního součtu čtverců a indexu determinace určíme nejvhodnější funkci pro vyrovnání zadaných tržeb a odhadneme vývoj tržeb v roce 8.

Tabulka 1: Přehled tržeb [v tis. Kč] společnosti ABC

Rok	1	2	3	4	5	6	7
Tržby	155625	172472	179589	186579	205421	214989	237937

Vytvořený soubor v Maple je k dispozici na adrese:
http://trvale.imatematika.cz/regression.mw

Práce v systému Maple

Nejprve inicializujeme knihovny, které budou použity:

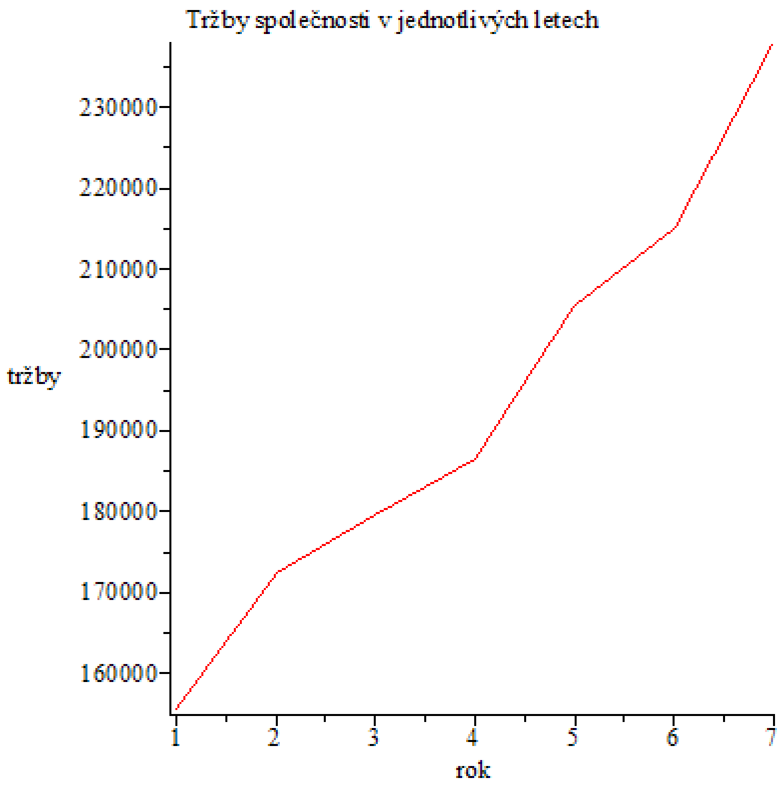
```
> with(Statistics) :
> with(plots) :
```

Definujeme nezávisle a závisle proměnné:

```
> X := Vector([1, 2, 3, 4, 5, 6, 7], datatype = float) :
> Y := Vector([155625, 172472, 179589, 186579, 205421, 214989,
    237937], datatype = float) :
```

Vykreslíme graf tržeb v jednotlivých letech:

```
> plot(X, Y, title = ["Tržby společnosti v jednotlivých letech"], labels
    = ["rok", "tržby [v tis. Kč]"])
```



Z grafu je na první pohled patrný rostoucí trend tržeb.

Zadaná data vyrovnáme nejběžnějšími regresními funkcemi.

Lineární regrese – výpočet regresní funkce:

> $linearni := LinearFit([1, x], X, Y, x)$

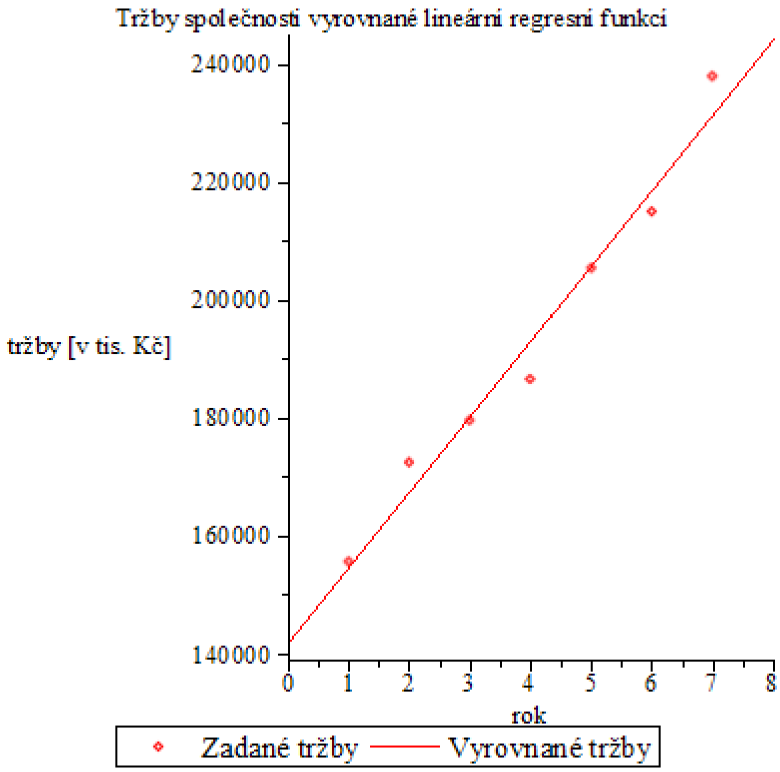
$$linearni := 1.42115714285714\ 10^5 + 12778.6428571429\ x$$

Graf – vyrovnání tržeb lineární regresní funkcí:

> $a := plot(\mathrm{X}, \mathrm{Y}, \mathrm{style} = \mathrm{point}, labels = [\text{"rok"}, \text{"tržby [v tis. Kč]"}], legend$
> $= \text{"Zadané tržby"}) :$

> $b := plot(linearni, x = 0\,..8, y = 140000\,..245000, title$
> $= [\text{"Tržby společnosti vyrovnané lineární regresní funkcí"}], legend$
> $= \text{"Vyrovnané tržby"}) :$

> $display(a, b)$



Kvadratická regrese – výpočet regresní funkce:

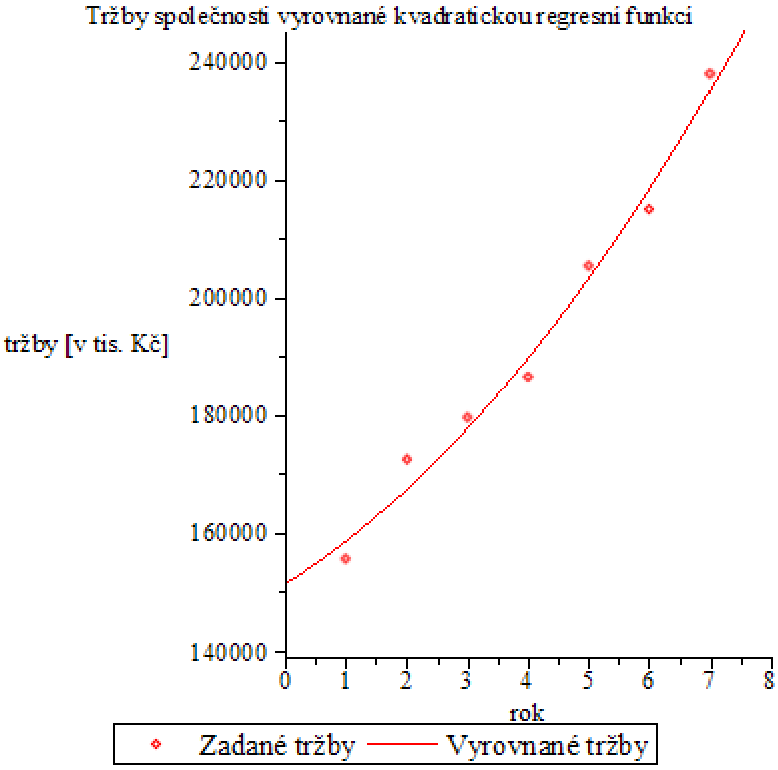
> $kvadraticka := PolynomialFit(2, X, Y, x)$

$$kvadraticka := 1.51610571428571\ 10^5 + 6448.73809523816\,x + 791.238095238088\,x^2$$

Graf – vyrovnání tržeb kvadratickou regresní funkcí:

> $b := plot(kvadraticka, x = 0\,..8, y = 140000\,..245000, title$ = ["Tržby společnosti vyrovnané kvadratickou regresní funkcí"], $legend$ = "Vyrovnané tržby") :

> $display(a, b)$



Exponenciální regrese – výpočet regresní funkce:

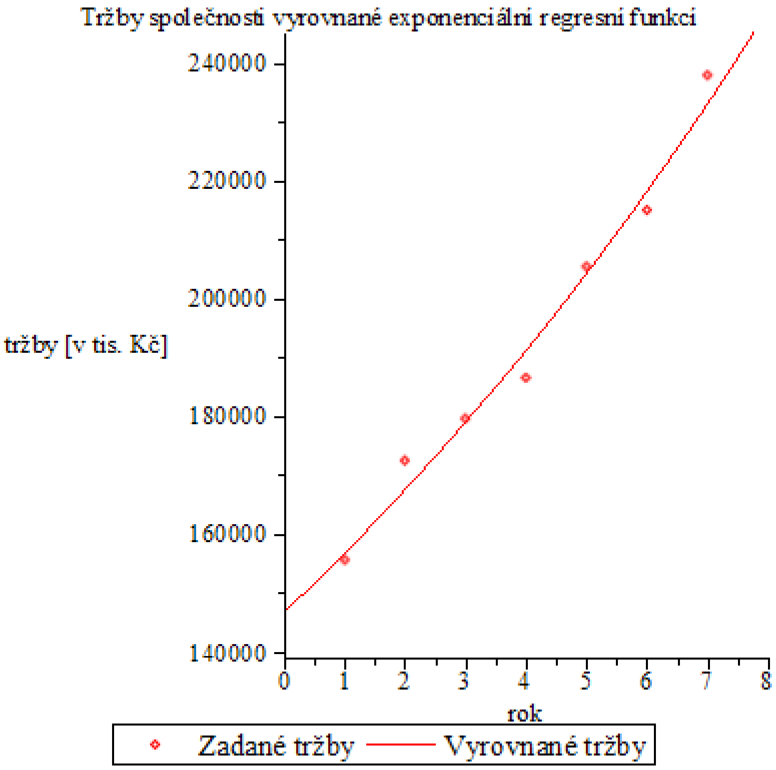
> $exponencialni := ExponentialFit(X, Y, x)$

$$exponencialni := 1.47064515668092\ 10^5\ e^{0.0660273050270541\ x}$$

Graf – vyrovnání tržeb exponenciální regresní funkcí:

> $b := plot(exponencialni, x = 0..8, y = 140000..245000, title$
> $= $ ["Tržby společnosti vyrovnané exponenciální regresní funkcí"],
> $legend = $ "Vyrovnané tržby") :

> $display(a, b)$



Logaritmická regrese – výpočet regresní funkce:

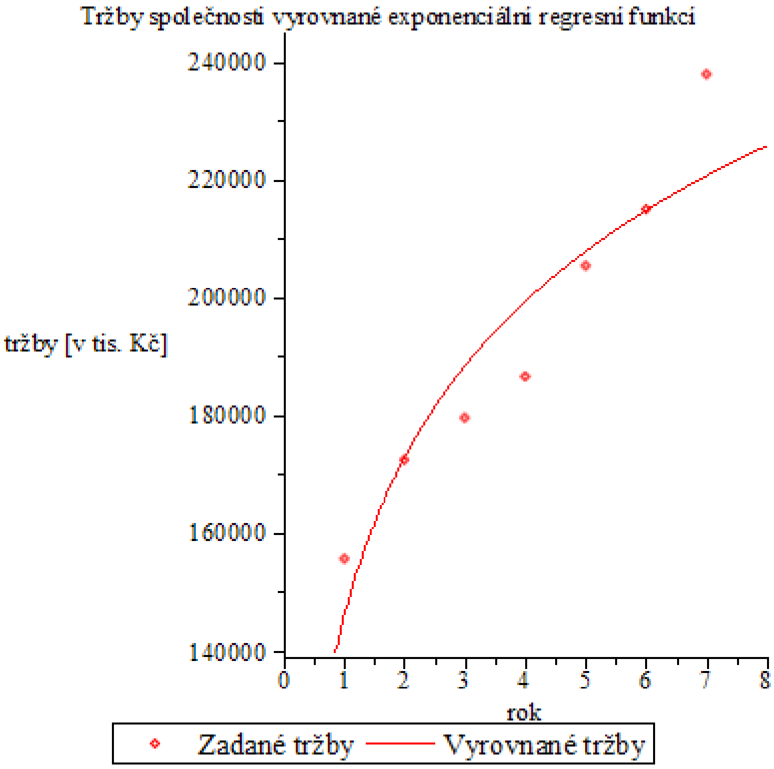
> $logaritmicka := LogarithmicFit(X, Y, x)$

$$logaritmicka := 1.46916234209107\ 10^5 + 38028.4133995251 \ln(x)$$

Graf – vyrovnání tržeb logaritmickou regresní funkcí:

> $b := plot(logaritmicka, x = 0..8, y = 140000..245000, title$
> $= [$"Tržby společnosti vyrovnané logaritmickou regresní funkcí"$],$
> $legend =$ "Vyrovnané tržby"$)$:

> $display(a, b)$



Procedura pro výpočet reziduálního součtu čtverců:

```
RSS := proc(a)
 local i, soucet;
i := 1;
 soucet := 0;
 while i ≤ numelems(X) do
 soucet := soucet + (Y[i]-subs(x=i, a))^2;
  i := i + 1;
 end do;
 return simplify(soucet);
  end proc:
```

Výpočet reziduálního součtu čtverců pro jednotlivé regresní funkce:

> $RSS(linearni)$

$$1.239097099\ 10^{8}$$

> $RSS(kvadraticka)$

$$7.132086110\ 10^{7}$$

> $RSS(logaritmicka)$

$$6.268592705\ 10^{8}$$

> $RSS(exponencialni)$

$$8.159611742\ 10^{7}$$

Nejmenší reziduální součet čtverců má použitá kvadratická regresní funkce, nejlépe tedy k zadaným datům přiléhá.

Pozn.: pro výpočet lze využít i volby output = residualsumofsquares. Je ovšem nutné poznamenat, že tato možnost se vztahuje na transformovaný model – nelze ji tedy využít pro exponenciální regresní funkci.

Ukázka výpočtu reziduálního součtu čtverců pomoci volby *output=residualsumofsquares***:**

> $LinearFit([1, x], X, Y, x, output = residualsumofsquares)$

$$1.23909709857142687\ 10^{8}$$

Protože reziduální součet čtverců není normován – nedá se zjistit jak „dobře“ regresní funkce vystihuje závislost mezi proměnnými, vypočteme tzv. index determinace.

Procedura pro výpočet indexu determinace:

```
indexDET := proc(a)
 local i, citatel, jmenovatel, prumer;
i := 1;
citatel := 0 ;
jmenovatel := 0;
prumer := 0;
for i from 1 by 1 to numelems(X) do
prumer := prumer + (Y[i]);
end do;
prumer := prumer / numelems(X);
i := 1;
 while i ≤ numelems(X) do
citatel := citatel + (Y[i]-subs(x = i, a))^2;
jmenovatel := jmenovatel + (Y[i] − prumer)^2;
 i := i + 1;
end do;
return simplify(1 - citatel/jmenovatel);
 end proc:
```

Výpočet indexu determinace pro jednotlivé regresní funkce:

> $indexDET(linearni)$

$$0.9736145267$$

> $indexDET(kvadraticka)$

$$0.9848128554$$

> $indexDET(exponencialni)$

$$0.9826248308$$

> $indexDET(logaritmicka)$

$$0.8665158803$$

Protože hodnota indexu determinace je u lineární, kvadratické i exponenciální regresní funkce blízká k jedné, jsou tyto funkce vhodné pro vyrovnání zadaných tržeb.

Odhad výše tržeb pro rok 8 pomocí kvadratické regresní funkce:

> $evalf(subs(x = 8, kvadraticka), 6)$

$$2.53840\ 10^5$$

Podle zvolené kvadratické regresní funkce jsou v roce 8 tržby společnosti ABC odhadnuty na 253 840tis. Kč.