

STATISTICA

Téma 8. Regresní a korelační analýza, regrese prostá

1) *Lineární i nelineární regrese prostá, korelace*

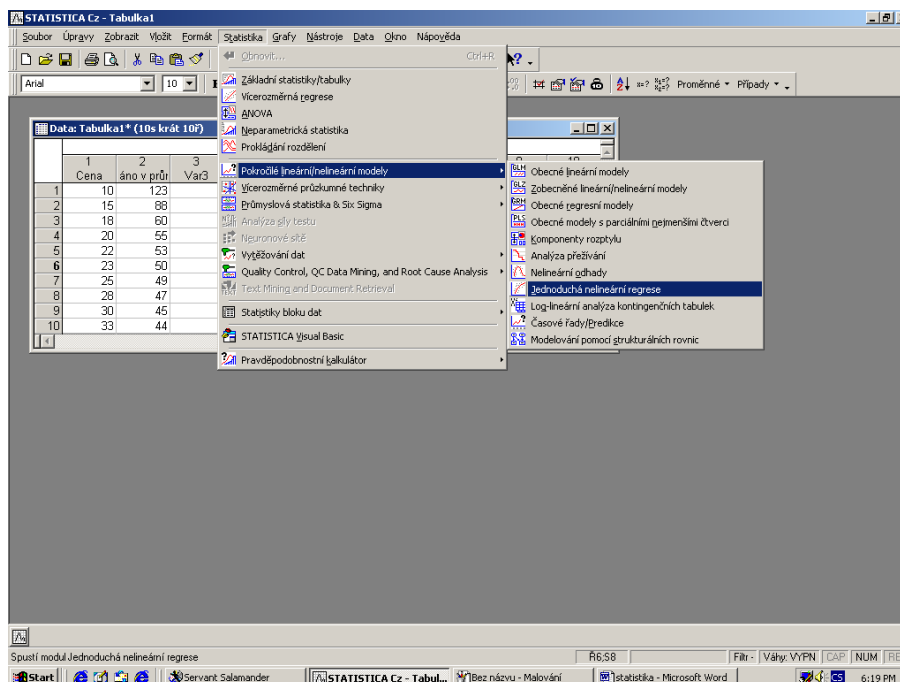
Naeditujeme data – viz obr. 1.

Obr. 1

	1	2	3	4	5	6	7	8	9	10
	Cena	ano v prur	Var3	Var4	Var5	Var6	Var7	Var8	Var9	Var10
1	10	123								
2	15	88								
3	18	60								
4	20	55								
5	22	53								
6	23	50								
7	25	49								
8	28	47								
9	30	45								
10	33	44								

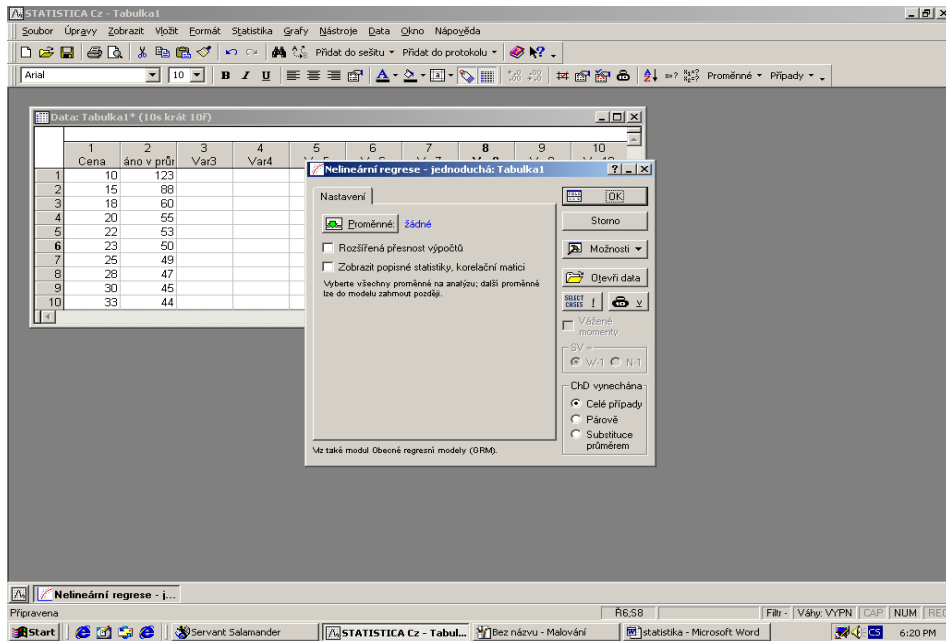
V menu *Statistika* zvolíme submenu *Pokročilé lineární/nelineární modely*. V něm nabídku na *Jednoduchá nelineární regrese* -viz obr. 2.

Obr. 2

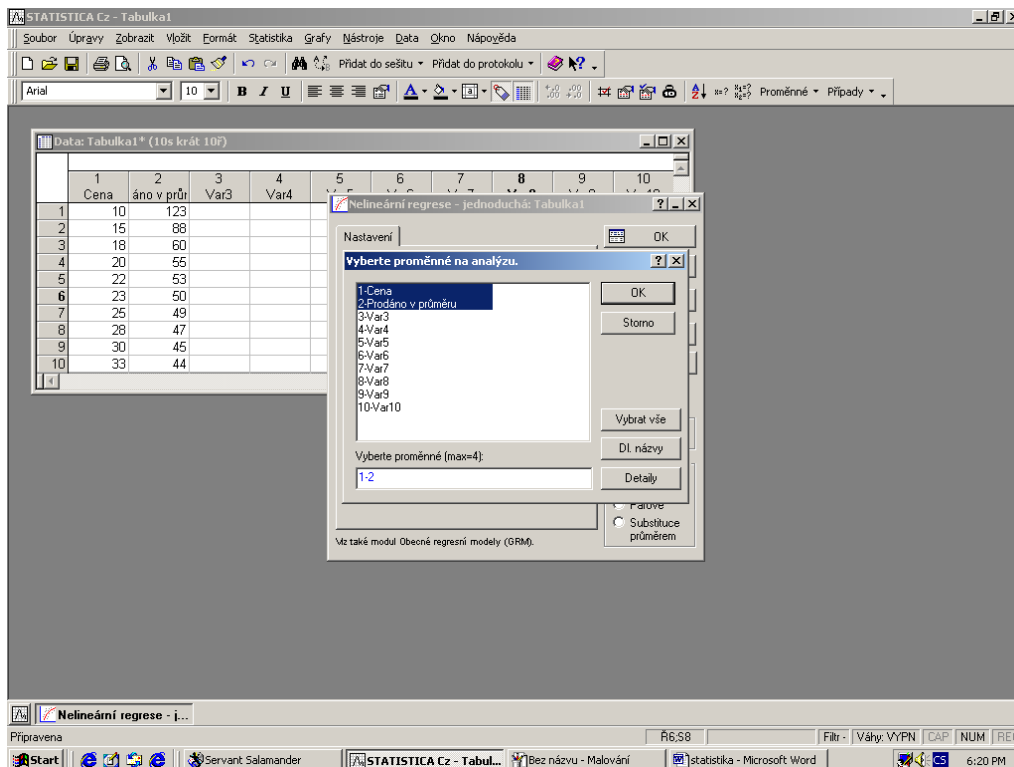


Objeví se okno *Nelineární regrese* –viz obr. 3. V něm nastavíme proměnné regresního modelu – viz obr. 4. Po stisknutí klávesy **OK** se vrátíme do okna *Nelineární regrese* (obr. 3) a stiskneme – li znovu tlačítko **OK**, otevře se okno *Nelineární komponenty regrese*- viz obr.5.

Obr. 3

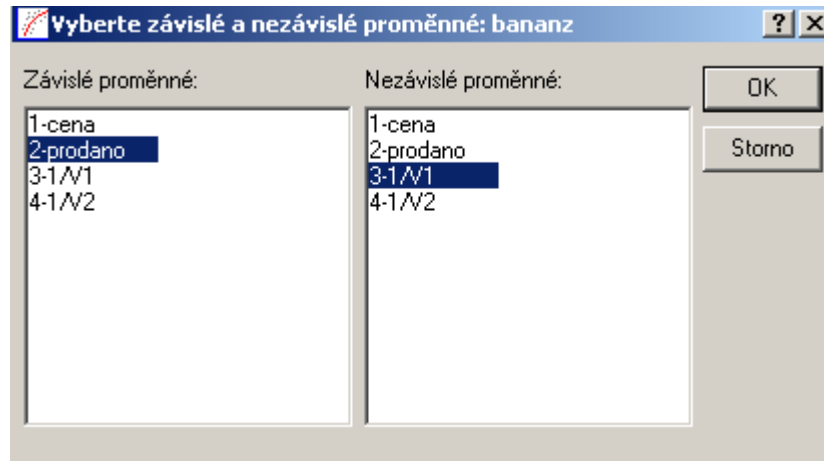


Obr. 4



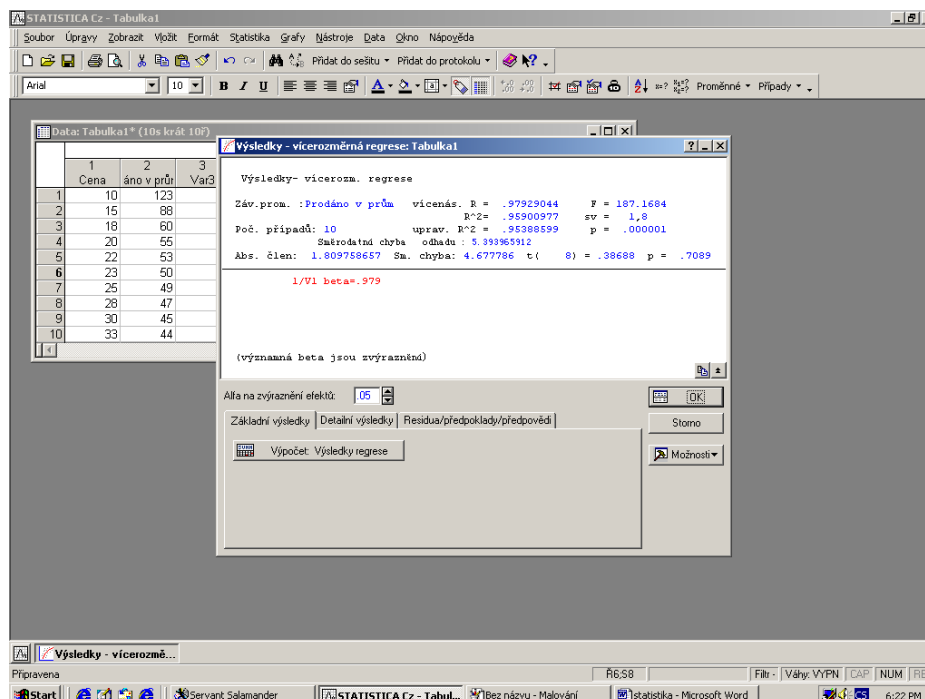
Nevýhodou je, že transformované proměnné jsou označeny pořadím (např. 1/V1), nikoli jménem (např. 1/cena). Musíme být tedy velmi opatrní. Na obr. 7 jsou zvoleny proměnné tak, aby byla modelována hyperbolická vazba mezi prodaným množstvím výrobku a jeho cenou. V tomto případě je závisle proměnnou „prodáno“ a nezávisle proměnnou „1/V1“. Stiskneme – li tlačítko **OK**, vrátíme se do okna *Definice modelu* (obr. 6).

Obr. 7



V něm zkontrolujeme, zda je nastavena *Standardní metoda* a zda je *absolutní člen zahrnut v modelu*. Poté provedeme výpočet stiskem **OK**. Otevře se okno výsledků –viz obr. 8. Má 3 karty. Nám zatím stačí karta *Základní výsledky*. V ní zvolíme tlačítko *Výpočet: Výsledky regrese*.

Obr. 8



Objeví se výstupní sestava - viz obr. 9. V její horní polovině je řada užitečných informací. Pro nás jsou důležité hodnoty R a R^2 , které interpretujeme jako *index či koeficient korelace* a *index či koeficient determinace*, podle toho, zda pracujeme s lineárním či nelineárním modelem.

Obr. 9

Výsledky regrese se závislou proměnnou : prodano (bananz)						
R= .97929044 R^2= .95900977 uprav. R^2= .95388599						
F(1,8)=187.17 p<.00000 Směrod. chyba odhadu : 5.3940						
N=10	Beta	Sm.chyba beta	B	Sm.chyba B	t(8)	Úroveň p
Abs.člen			1.810	4.67779	0.38688	0.708930
1/V1	0.979290	0.071581	1190.601	87.02616	13.68095	0.000001

Ve výstupní sestavě –viz obr. 9 - jsou pro vytvoření modelu důležité hodnoty, uvedené ve sloupci **B**. Jedná se o koeficienty u proměnných, jejichž název je uveden v prvním sloupci –tj. na obr. 9 se jedná o *Abs.člen* a *1/V1*. Ve výstupní sestavě obr. 9 se tedy jedná o model $y' = 1,81 + 1190,6/x$ s indexem korelace $I=0,979$ a indexem determinace $I^2=0,959$.

Čísla ve sloupci **Úroveň p** značí nejmenší hladiny významnosti, pro něž lze zamítnout hypotézu o nulových hodnotách regresních koeficientů. Čísla ve sloupci **t(B)** jsou pak hodnoty testovacích kritérií.

Konkrétně v prvním řádku jsou uvedeny výsledky testu $H_0: \alpha = 0$ proti $H_A : \text{non } H_0$, a v druhém řádku jsou uvedeny výsledky testu $H_0: \beta = 0$ proti $H_A : \text{non } H_0$.

Interpretujeme – li tedy hodnoty p z obr. 9, hypotézu o nulové hodnotě α zamítnout nelze a hypotézu o nulové hodnotě β zamítnout lze.

Pro nás z toho plyne, že model lze zjednodušit tím, že vypustíme absolutní člen. Stiskneme lištu *Výsledky-vícerozměrné*. Ta je zobrazena *v levé spodní části obrazovky*. Objeví se okno uvedené na obr. 8. (V něm můžeme mimo jiné v kartách *Detailní výsledky* či *Rezidua/předpoklady/předpovědi* zvolit provedení dalších výpočtů.) V okně na obr. 8 stiskneme tlačítko *Storno*. Vrátime se tak do okna *Definice modelu* (obr. 6). V něm v kartě *Detaily* rozbalíme položku *Abs. člen* a vybereme nabídku *Nastaven na 0*. Klikneme na *OK*. Provede se nový výpočet. Před tím však budeme upozorněni, že nelze srovnávat R^2 původní výstupní sestavy s hodnotou R^2 v sestavě zjednodušené. Je totiž počítán podle jiného vzorce. Nová výstupní sestava již nebude mít absolutní člen – viz obr.10.

Obr. 10

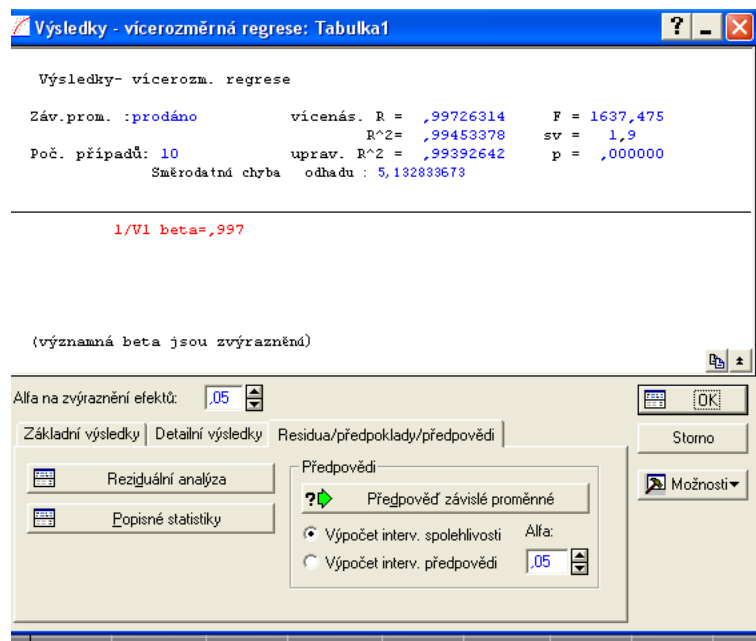
Výsledky regrese se závislou proměnnou : prodano (bananz)						
R= .99726314 R^2= .99453378 uprav. R^2= .99392642						
F(1,9)=1637.5 p<.00000 Směrod. chyba odhadu : 5.1328						
N=10	Beta	Sm.chyba beta	B	Sm.chyba B	t(9)	Úroveň p
1/V1	0.997263	0.024645	1221.952	30.19720	40.46573	0.000000

Na obr. 10 je uveden model $y' = 1221,952/x$ s indexem korelace $I=0,997$ a indexem determinace $I^2=0,994$. Tento model již zjednodušit nelze.

2) Predikce

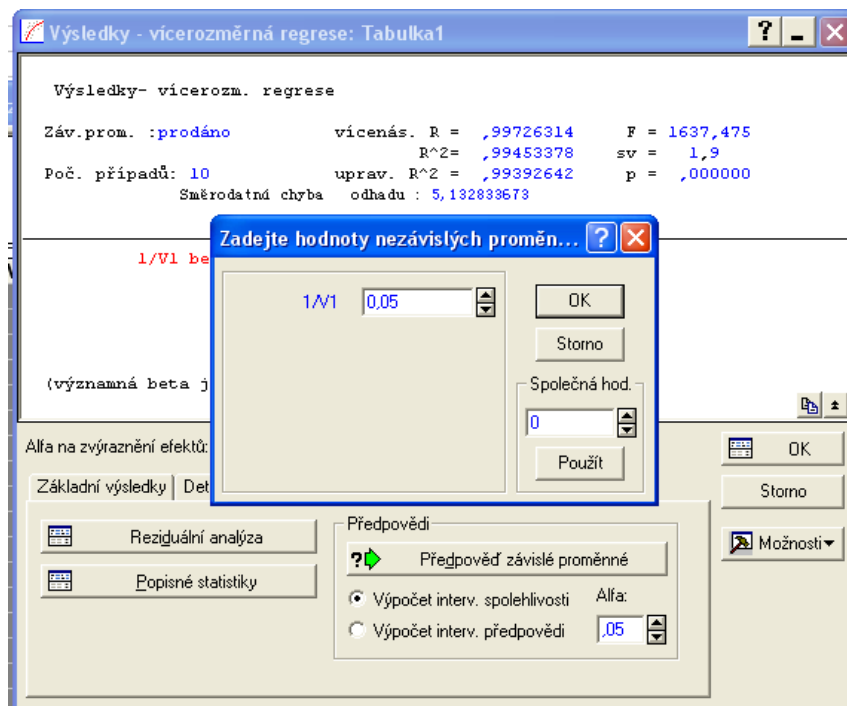
Predikci umožní provést nastavení karty *Rezidua/předpoklady/předpovědi* v okně *Výsledky – vícerozměrná regrese*- viz obr. 11. Do tohoto okna se nejrychleji vrátíme stiskem tlačítka *Výsledky –vícerozmě.*, umístěném ve spodní části obrazovky.

Obr. 11



Zvolíme-li nabídku **Výpočet interv. spolehlivosti**, aktualizujeme-li hladinu významnosti **Alfa** a zadáme-li hodnotu **nezávisle proměnné**, pro niž chceme predikci provést, pak vlastní predikci provedeme stiskem tlačítka **OK**. Výstupní sestava má tvar uvedený na obr. 13. Poznámka: nevýhodou je, že hodnoty **nezávisle proměnných** nelze zadat vzorcem, např. $1/20$, ale musíme uvést konkrétní hodnotu, např. $0,05$ – viz. obr. 12. Tu musíme vypočítat předem.

Obr. 12



Obr. 13

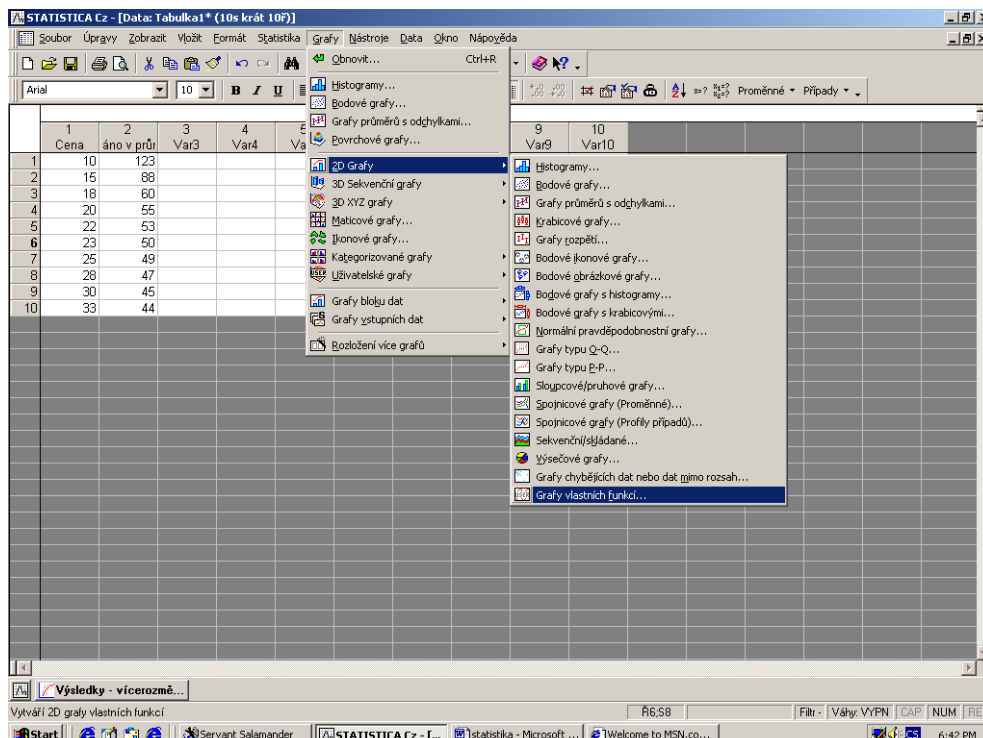
Předpovězené hodnoty (Tabulka1)			
proměnné: prodáno			
Proměnná	B-váž.	Hodnota	B-váž. * Hodnot
1/M1	1221,952	0,050000	61,09758
Předpověď			61,09758
-95,0%LS			57,68203
+95,0%LS			64,51312

Výstupní sestava obsahuje informace o příspěvcích nezávisle proměnných k odhadu závisle proměnné (*sloupec B-váž.*), bodový a intervalový odhad závisle proměnné (*sloupec B-váž. *Hodnot*).

3) Grafické znázornění známého regresního modelu

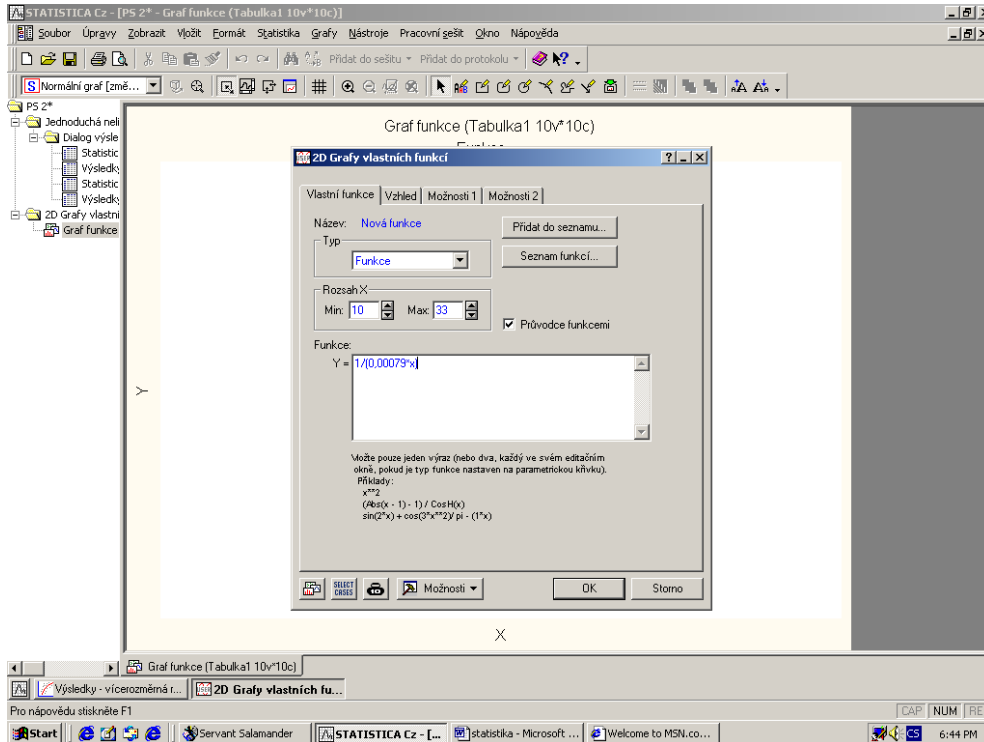
Známe-li předpis pro regresní model, můžeme si model graficky zobrazit. Zvolíme menu *Grafy*, submenu *2D grafy* a v něm nabídku *Grafy vlastních funkcí...*, viz obr. 14.

Obr. 14



V okně, které se otevře –viz. obr. 15, zvolíme rozsah závisle proměnné, tj. nastavíme políčka *Min.*, *Max.* a *předepíšeme tvar funkce*.

Obr. 15



Graf předepsané funkce se objeví po stisku tlačítka **OK**.

Obr.16

