Nelineární regresní funkce

Nelineární regrese

Na rozdíl od lineárních regresních modelů je třeba u nelineárních modelů počítat s řadou komplikací:

· neodhadnutelností některých parametrů,

· existencí minima funkce jen pro některé regresní modely,

· výskytem lokálních minim a sedlových bodů,

· špatnou podmíněností parametrů v regresním modelu,

· malým rozmezím experimentálních dat (zejména u parametrů vyjadřujících limitní chování modelu).

Nelineární regrese

Funkce **lineární v parametrech**

Aditivní typy funkcí:

- Kvadratická
- Lomená
- Logaritmická
- Kubická

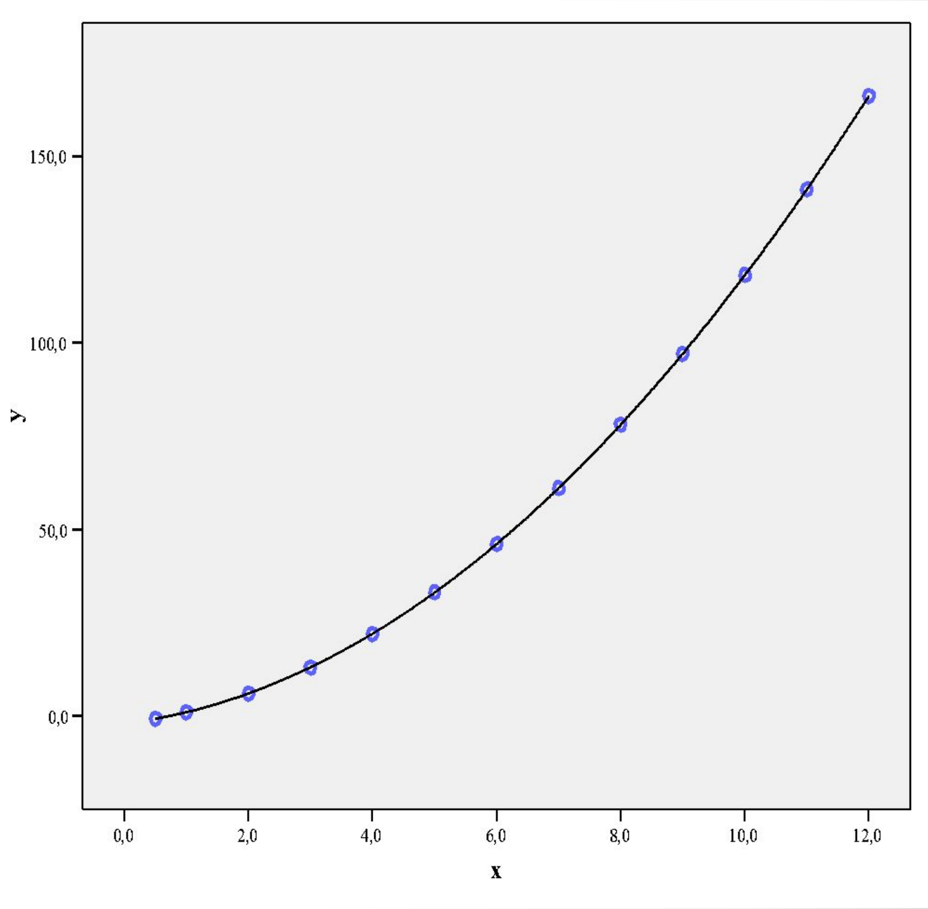
Funkce **nelineární v parametrech**

Multiplikativní typy funkcí:

- Exponenciální
- S - křivka
- Mocninná

Funkce kvadratická (parabola)

$$y'_i = a + bx_i + cx_i^2 + \varepsilon$$



Soustava normálních rovnic

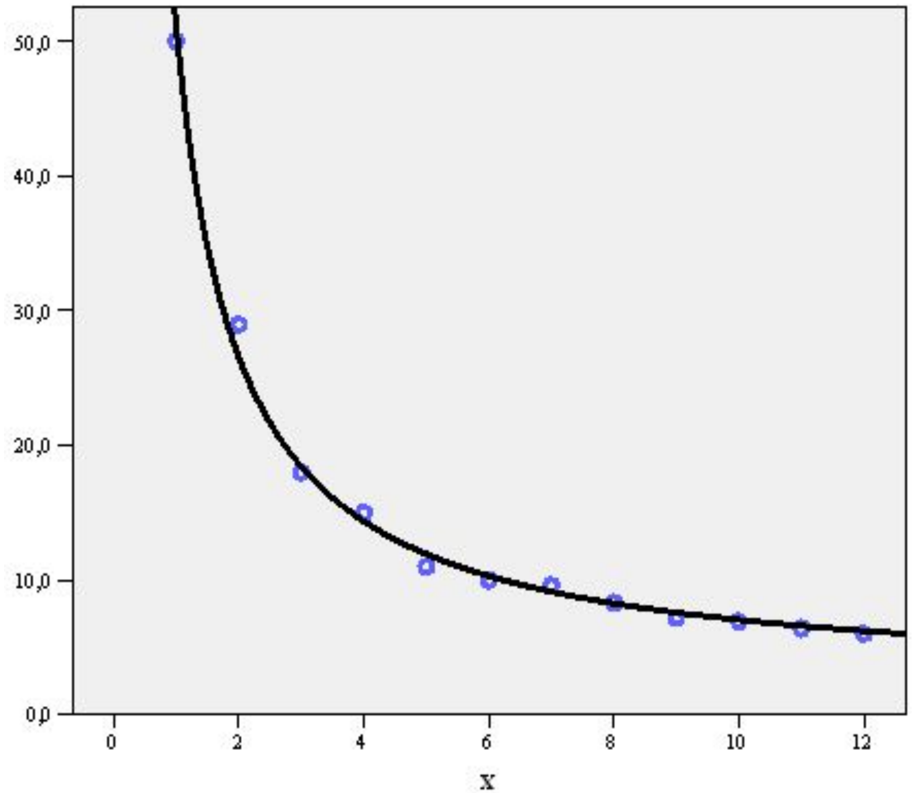
$$an + b\sum x_i + c\sum x_i^2 = \sum y_i$$

$$a\sum x_i + b\sum x_i^2 + c\sum x_i^3 = \sum y_i x_i$$

$$a\sum x_i^2 + b\sum x_i^3 + c\sum x_i^4 = \sum y_i x_i^2$$

Funkce lomená (hyperbola)

$$y'_i = a + b\frac{1}{x_i} + \varepsilon$$



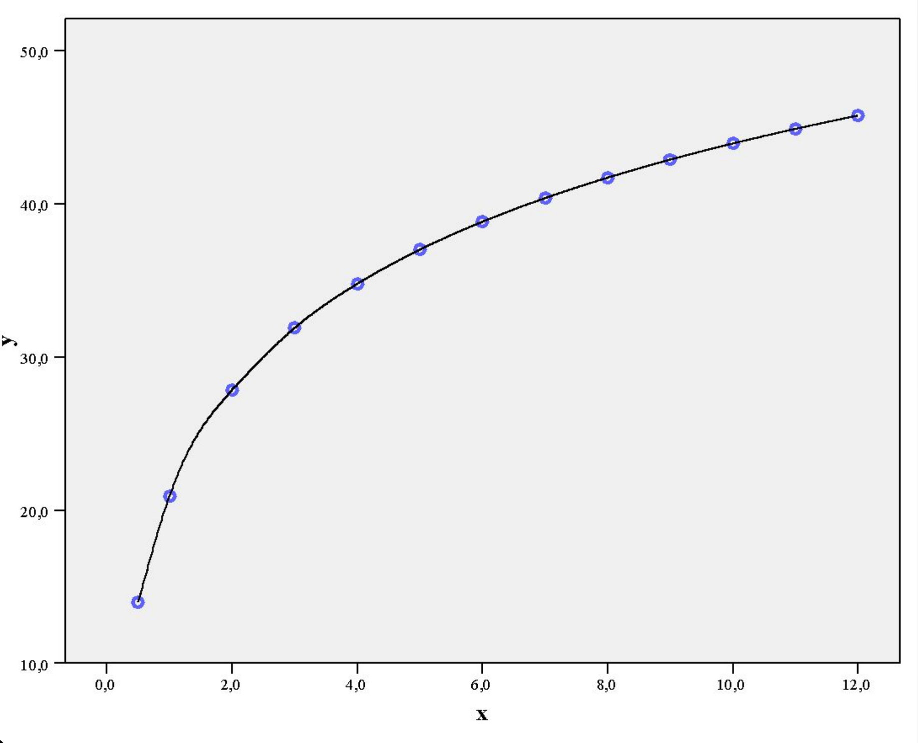
Soustava normálních rovnic

$$a.n + b\sum\frac{1}{x_i} = \sum y_i$$

$$a\sum\frac{1}{x_i} + b\sum\frac{1}{x_i^2} = \sum\frac{y_i}{x_i}$$

Funkce logaritmická (logaritmus)

$$y_i' = a + b \cdot ln\,(x_i\,) + \varepsilon$$



Soustava normálních rovnic

$$a\sum ln\,x_i + b\sum (ln\,x_i)^2 = \sum (ln\,x_i)\ y_i$$

$$a.n + b\sum ln\,x_i = \sum y_i$$

Funkce kubická

$$y'_i = a + bx_i + cx_i^2 + dx_i^3 + \varepsilon$$

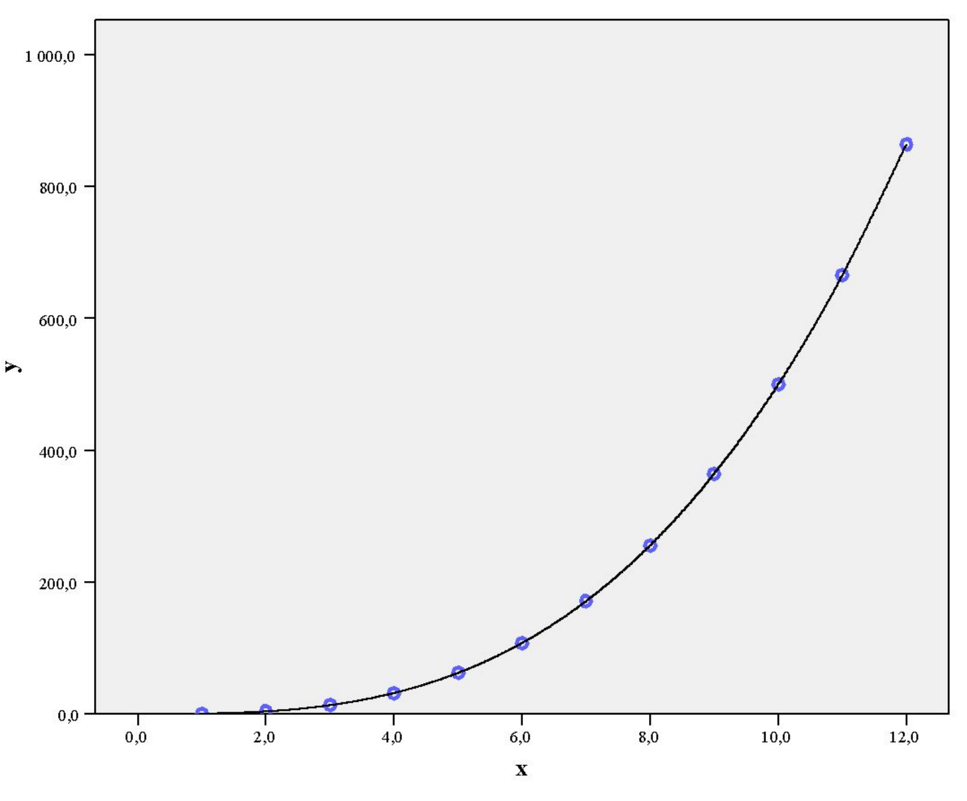
Nelineární regrese

Jak již bylo uvedeno odhad parametrů u regresních funkcí, které **nejsou lineární v parametrech**, **neprovádíme MNČ přímo**, protože její použití vede k soustavě nelineárních rovnic, z nichž zpravidla nedokážeme odhadnout přímo parametry ve formě vhodných výpočetních vzorců.

Používá se tedy způsob, kdy určitou regresní funkci, která je nelineární z hlediska parametrů, převedeme pomocí **linearizující transformace** na funkci **lineární v parametrech**.

Funkce mocninná

$$y'_i = a x_i^b \cdot (e^{\varepsilon})$$



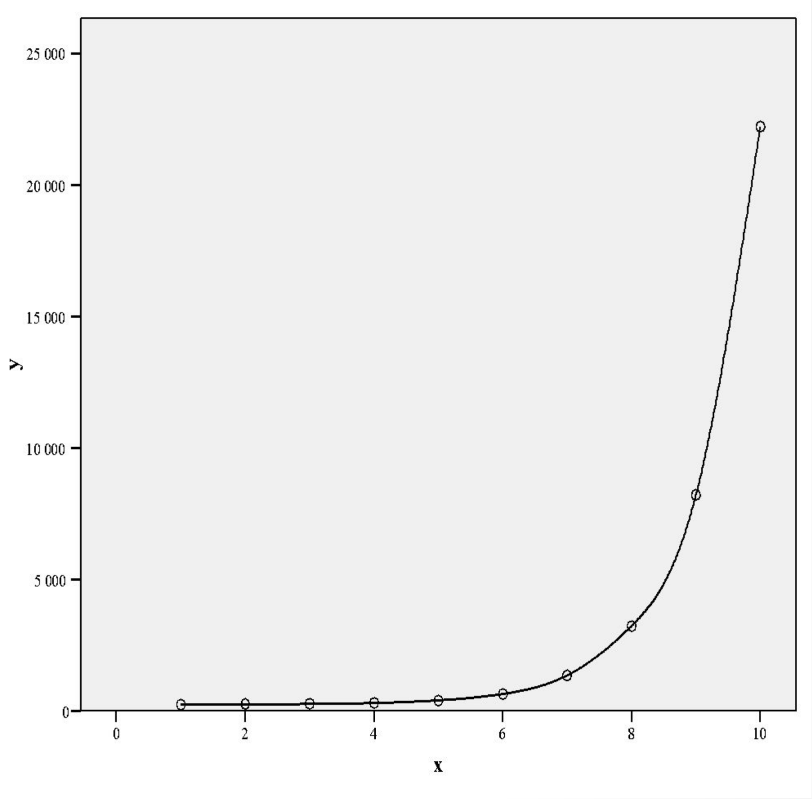
Soustava normálních rovnic

$$n.ln\,a + b\sum ln\,x_i = \sum ln\,y_i$$

$$lma\sum ln\,x_i + b\sum (ln\,x_i)^2 = \sum (ln\,y_i)(ln\,x_i)$$

Funkce exponenciální

$$y_i' = a\,exp(bx_i)\cdot(e^{\varepsilon})$$



Soustava normálních rovnic

$$n.ln\,a + ln\,b\sum x_i = \sum ln\,y_i$$

$$ln\,a\sum x_i + ln\,b\sum x_i^2 = \sum x_i\,ln\,y_i$$

Korelace při nelineární regresi

Při konstrukci míry ukazující na sílu závislosti vycházíme ze vztahu **empirických** a **vyrovnaných hodnot.**

$$\sum (y_i - \bar{y})^2 = \sum (\hat{y}_i - \bar{y})^2 + \sum (y_i - \hat{y}_i)^2$$

$$S_y \quad = \quad S_T \quad + \quad S_r$$

součet čtverců odchylek *y*	součet čtverců vyrovnaných hodnot	reziduální součet čtverců

Korelace při nelineární regresi

Sílu závislosti je možné měřit **index determinace.**

$$I_{yx}^2 = \frac{S_T}{S_y} = 1 - \frac{S_r}{S_y}$$

$$0 \leq I^2_{yx} \leq 1$$

Nízká hodnota indexu determinace nemusí ještě znamenat nízký stupeň závislosti mezi proměnnými, ale může signalizovat chybnou volbu regresní funkce.

Korelace při nelineární regresi

Sílu závislosti je možné měřit **index korelace.**

$$I_{yx} = \sqrt{I_{yx}^2} = \sqrt{1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}} = \sqrt{1 - \frac{S_r}{S_y}}$$

$$0 \leq I_{yx} \leq 1 \qquad I_{yx} \neq I_{xy}$$

Index korelace se používá k měření těsnosti závislosti pro libovolnou regresní funkci, jejíž parametry byly odhadnuty metodou nejmenších čtverců. Má menší vypovídací schopnost než index determinace.

Příklad

Ovlivňují investicích do reklamy výši tržeb ve firmě?

Konstrukce regresního modelu

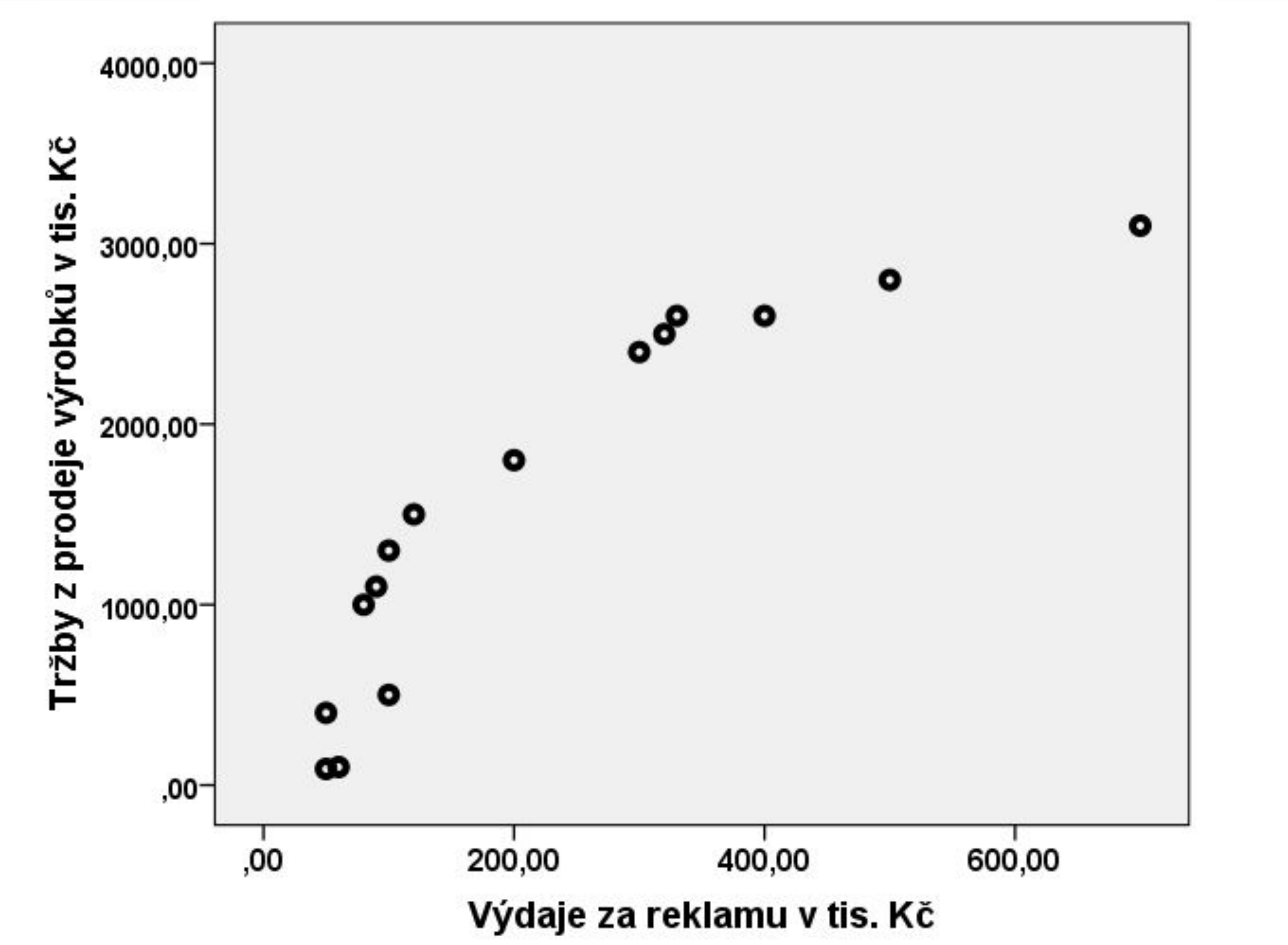
Proměnné:

x . . . investice do reklamy (v tis. Kč)

y . . . tržby z prodeje (v tis. Kč)

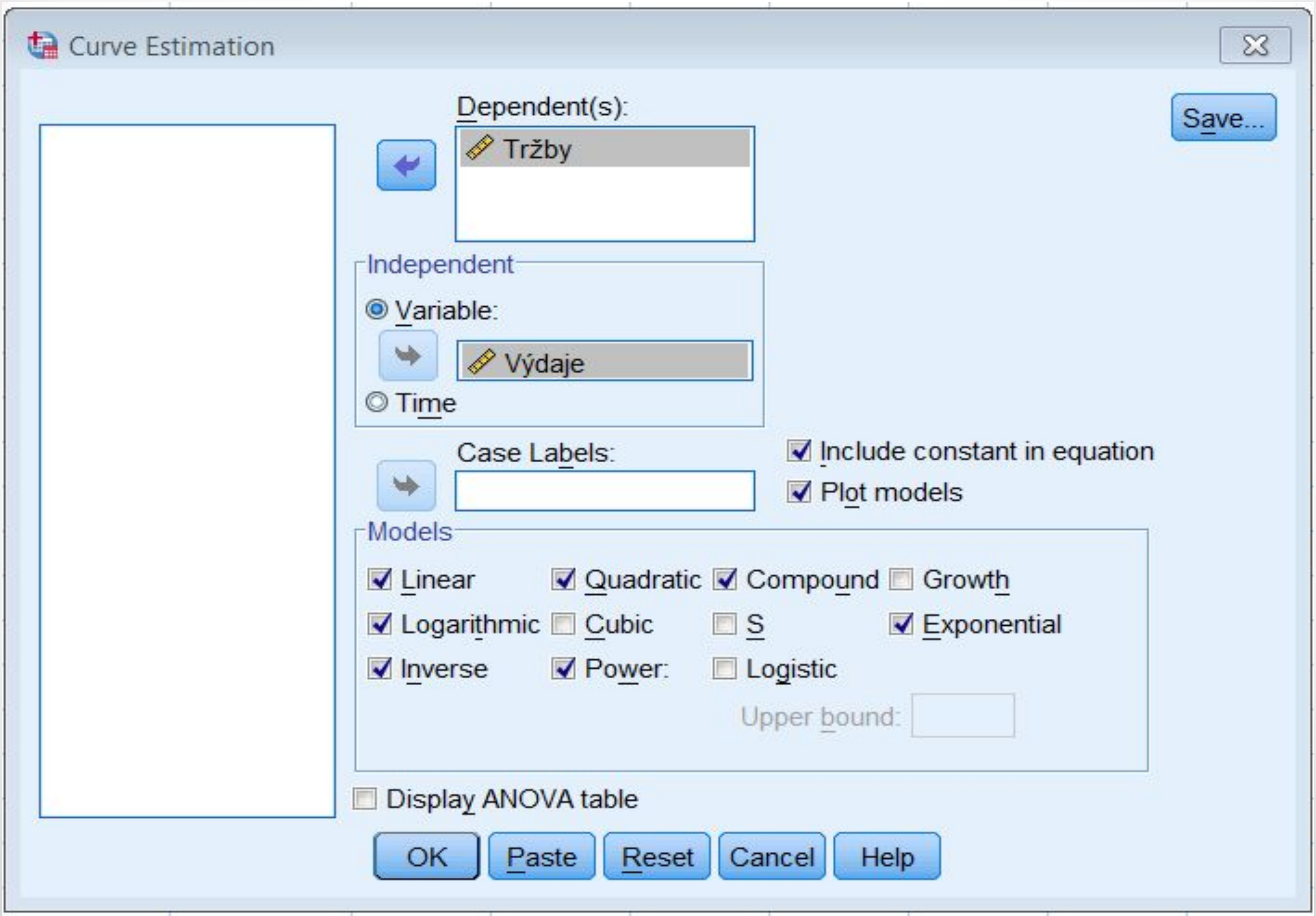
Příklad

Graf závislosti tržeb na výdajích (korelační pole)



Příklad

Volba regresní funkce v SPSS



Příklad
Hodnoty indexu determinace pro různé funkce

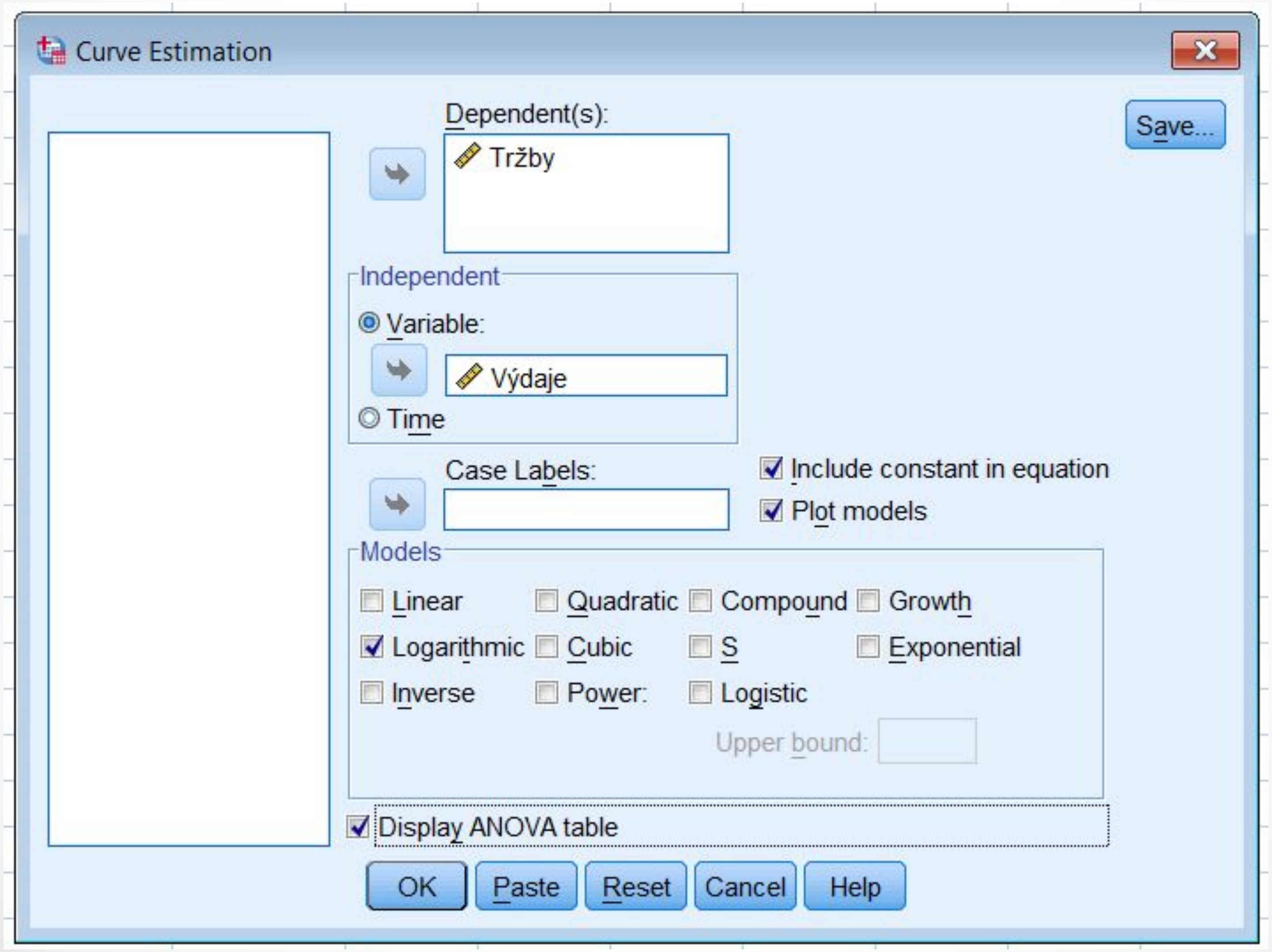
Model Summary and Parameter Estimates

Dependent Variable: Tržby z prodeje výrobků v tis. Kč

Equation	Model Summary					Parameter Estimates		
	R Square	F	df1	df2	Sig.	Constant	b1	b2
Linear	,801	52,417	1	13	,000	502,266	4,781	
Logarithmic	,949	239,702	1	13	,000	-4260,316	1151,988	
Inverse	,894	109,807	1	13	,000	2892,053	-151833,257	
Quadratic	,929	78,003	2	12	,000	-97,193	11,106	-,010
Compound	,483	12,125	1	13	,004	415,618	1,004	
Power	,700	30,356	1	13	,000	3,926	1,103	
Exponential	,483	12,125	1	13	,004	415,618	,004	

The independent variable is Výdaje za reklamu v tis. Kč.

Příklad
Volba logaritmického regresního modelu



Příklad

Regresní model pro logaritmickou funkci

Model Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate
,974	,949	,945	244,181

The independent variable is Výdaje za reklamu v tis. Kč.

Index determinace část rozptylu *y*, kterou lze vysvětlit regresním modelem

Opravená (adjustovaná) hodnota penalizuje regresní model za počet parametrů

Příklad

Regresní model pro logaritmickou funkci

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	14292046,10	1	14292046,10	239,702	,000
Residual	775113,897	13	59624,146		
Total	15067160,00	14			

The independent variable is Výdaje za reklamu v tis. Kč.

Test regresního modelu

Příklad

Regresní model pro logaritmickou funkci

Coefficients

	Unstandardized Coefficients		Standardized Coefficients		
	B	Std. Error	Beta	t	Sig.
ln(Výdaje za reklamu v tis. Kč)	1151,988	74,407	,974	15,482	,000
(Constant)	-4260,316	382,839		-11,128	,000

Parametry modelu

$$y_i' = a + b \cdot \ln x_i$$

$$y_i' = -4\,260{,}316 + 1151{,}988 \cdot \ln x_i$$

Příklad

Regresní model pro logaritmickou funkci