Lineární regrese a korelace

- Tvorba regresní funkce
- Posouzení její kvality

Základní pojmy

Závislosti

- Pevné
- Volné

- Regrese – jednostranné závislosti
- Korelace – vzájemné závislosti

Základní pojmy

Možnosti získání údajů pro regresi

- Pozorováním *n* statistických jednotek s časově prostorově a věcně vymezeným statistickým souborem.
- Pozorováním určité statistické jednotky v *n* různých časových okamžicích či intervalech.
- *N*-násobným opakováním určitého pokusu, jež je prováděn za stejných, či přibližně stejných podmínek.

Postupné kroky regresní analýzy

- Formulovat matematicky představy o charakteru regresní funkce.
- Posoudit vliv faktorů, které se rozhodneme do dané funkce nezahrnout.
- Odhadnout konkrétní podobu regresní funkce na základě zjištěných hodnot.
- Posoudit kvalitu této funkce

Postupné kroky regresní analýzy

Na základě regresní funkce lze provádět odhady:

- Interpolační – odhady prováděné uvnitř intervalu hodnot vysvětlující proměnné.
- Extrapolační – odhady prováděné mimo oblast měření.

Podle typu zvolené funkce poté rozlišujeme následující regrese

- Přímková (lineární)

$$y = b_o + b_1 \cdot x$$

- Parabolická (kvadratická)

$$y = b_o + b_1 \cdot x + b_2 \cdot x^2$$

- Polynomická *n*-tého stupně

$$y = b_o + b_1 \cdot x + b_2 \cdot x^2 + \ldots + b_n \cdot x^n$$

- Logaritmická

$$y = b_o + b_1 \ln x$$

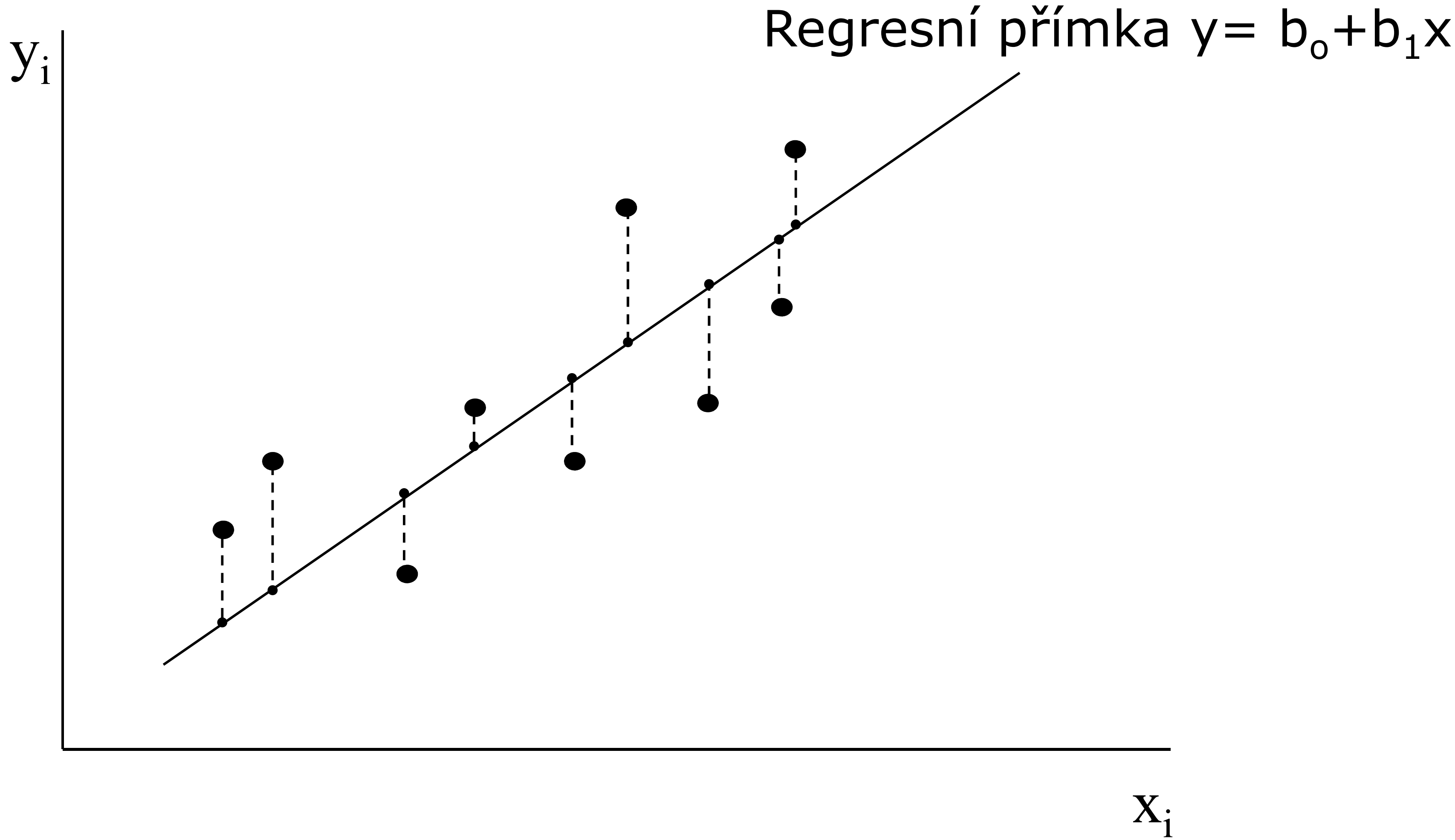
- Exponenciální

$$y = b_o \cdot b_1^x$$

- Hyperbolická

$$y = b_o + \frac{b_1}{x}$$

Odvození regresní funkce pomocí MNČ



Lineární regrese pomocí MNČ

- Absolutní člen (konstanta)

$$b_o = \frac{\sum y_i \sum x_i^{\ 2} - \sum x_i \sum y_i x_i}{n \sum x_i^{\ 2} - \left(\sum x_i\right)^2}$$

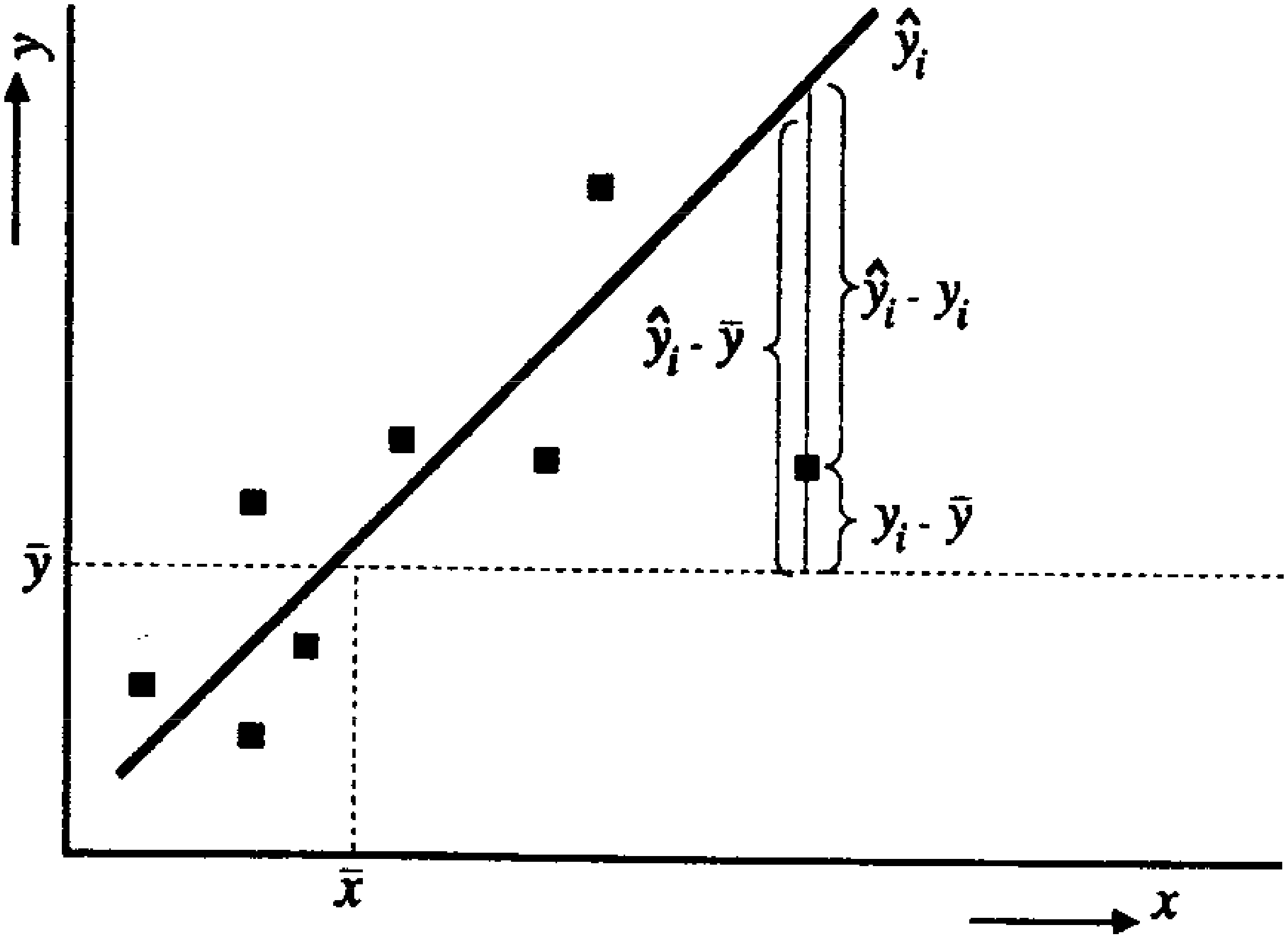
- Regresní koeficient

$$b_1 = \frac{n \sum y_i x_i - \sum x_i \sum y_i}{n \sum x_i^{\ 2} - \left(\sum x_i\right)^2} = \frac{s_{xy}}{s_x^2}$$

Prozkoumejte vztah mezi výdaji veřejných rozpočtů (VV) a HDP v ČR v letech 1993-2003. Na základě odhadnuté závislosti proveďte odhad výdajů v roce 2004.

	HDP b.c	**VV**
1993	**1020,3**	**420,00**
1994	**1182,8**	**494,60**
1995	**1466,7**	**573,60**
1996	**1660,6**	**638,71**
1997	**1785,1**	**685,13**
1998	**1962,5**	**737,48**
1999	**2041,4**	**755,05**
2000	**2150,1**	**842,67**
2001	**2315,3**	**903,32**
2002	**2414,7**	**918,58**
2003	**2532,4**	**1 115,60**

Kvalita regresní funkce



Rozptyly u regresní funkce

- Rozptyl empirických hodnot

$$s_y^2 = \frac{\sum_{i=1}^{n} \left(y_i - \overline{y}\right)^2}{n}$$

- Rozptyl teoretických (vyrovnaných) hodnot

$$s_{\widehat{y}}^2 = \frac{\sum_{i=1}^{n} \left(\widehat{y}_i - \overline{y}\right)^2}{n}$$

- Reziduální rozptyl

$$s_{(y-\widehat{y})}^2 = \frac{\sum_{i=1}^{n} \left(y_i - \widehat{y}_i\right)^2}{n}$$

Míry kvality vystižení

- Index determinace

$$R^2 = \frac{s_{\hat{y}}^2}{s_y^2}$$

- $R^2 = -1$ lineární závislost (nepřímá)

  $= 0$ lineární nezávislost

  $= 1$ lineární závislost (přímá)

Statistický test na nenulovost regresních parametrů bo a b1

$$T = \frac{b_i}{s_{bi}}$$

$t_{\alpha/2}$ - kvantil t-rozdělení pro hladinu významnosti α a počet stupňů volnosti = n-p

b_i - testovaný koeficient

s_{bi} - směrodatná odchylka tohoto koeficientu

Interval spolehlivosti pro regresní parametry

$$IS_{b_o} = \left(b_o - s_{b_o} t_{1-\frac{\alpha}{2}} ;\; b_o + s_{b_o} t_{1-\frac{\alpha}{2}} \right)$$

$$IS_{b_1} = \left(b_1 - s_{b_1} t_{1-\frac{\alpha}{2}} ;\; b_1 + s_{b_1} t_{1-\frac{\alpha}{2}} \right)$$

Funkce LINREGRESE

b_1	b_o
směrodatná odchylka koeficientu b_1	směrodatná odchylka koeficientu b_o
Index determinace	reziduální směrodatná odchylka
hodnota statistiky F	počet stupňů volnosti
regresní suma čtverců	reziduální suma čtverců