

Statistika

Regresní a korelační analýza
– Jednoduchá regresní a korelační analýza

Roman Biskup

Jihočeská univerzita v Českých Budějovicích
Ekonomická fakulta (Zemědělská fakulta)
Katedra aplikované matematiky a informatiky

2008/2009



1/8

Jednoduchá lineární regresní analýza

Model jednostranné závislosti s pevnými efekty

- ▶ Závislost dvou numerických statistických znaků – dále proměnné
 - ▶ závislá proměnná y (vysvětlovaná, endogenní)
 - ▶ nezávislá proměnná x (vysvětlující, exogenní)
- ▶ Situace reprezentovaná n dvojicemi hodnot $[x_i; y_i]$, pro $i = 1, \dots, n$
- ▶ Model lineární závislosti

$$y = \beta_0 + \beta_1 x,$$

- ▶ kde β_0 je tzv. **absolutní člen** a
- ▶ β_1 je tzv. **koefficient u lineárního členu**
- ▶ Předpoklad $y_i = \underbrace{\beta_0 + \beta_1 x_i}_{\text{model}} + \underbrace{\varepsilon_i}_{\text{chyba}}$, kde $\varepsilon_i \sim N(0; \sigma^2)$
- ▶ Odhady $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$, respektive $\hat{y} = b_0 + b_1 x$



3/8

Lineární regresní analýza

Nelineární regresní analýza



2/8

Odhad regresních koeficientů prostřednictvím MNČ I

- ▶ Minimalizovat funkci dvou proměnných β_0 a β_1
 - ▶ $S(\beta_0; \beta_1) = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2$
- ▶ Parciální derivace podle β_0 a β_1
 - ▶ $\frac{\partial S(\beta_0; \beta_1)}{\partial \beta_0} = 2 \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)(-1)$
 - ▶ $\frac{\partial S(\beta_0; \beta_1)}{\partial \beta_1} = 2 \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)(-x_i)$
- ▶ Řešení soustavy dvou rovnic o dvou neznámých β_0 a β_1
 - ▶ $\sum_{i=1}^n y_i - \sum_{i=1}^n \beta_0 - \sum_{i=1}^n \beta_1 x_i = 0$
 - ▶ $\sum_{i=1}^n x_i y_i - \sum_{i=1}^n \beta_0 x_i - \sum_{i=1}^n \beta_1 x_i^2 = 0$
- ▶ Odhady $\hat{\beta}_0 = b_0$ a $\hat{\beta}_1 = b_1$



4/8

Odhad regresních koeficientů prostřednictvím MNČ II

$$b_0 = \hat{\beta}_0 = \frac{\sum_{i=1}^n x_i \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i}{\left(\sum_{i=1}^n x_i\right)^2 - n \sum_{i=1}^n x_i^2}$$

$$b_1 = \hat{\beta}_1 = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}$$

5/8



Sdružené regresní přímky

Model oboustranné jednoduché lineární závislosti s pevnými efekty

► Volná volba závislosti

$$yx \quad y = \beta_0 + \beta_1 x$$

$$xy \quad x = \beta_0 + \beta_1 y$$

► Odhady regresních koeficientů

$$yx \quad \hat{\beta}_0 = a_{yx} \quad a \quad \hat{\beta}_1 = b_{yx}$$

$$xy \quad \hat{\beta}_0 = a_{xy} \quad a \quad \hat{\beta}_1 = b_{xy}$$

$$yx \quad a_{yx} = \frac{\sum x \sum xy - \sum x^2 \sum y}{(\sum x)^2 - n \sum x^2} \quad b_{yx} = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2}$$

$$xy \quad a_{xy} = \frac{\sum y \sum xy - \sum y^2 \sum x}{(\sum y)^2 - n \sum y^2} \quad b_{xy} = \frac{n \sum xy - \sum y \sum x}{n \sum y^2 - (\sum y)^2}$$

► Regresní rovnice

$$yx \quad \hat{y} = a_{yx} + b_{yx} x$$

$$xy \quad \hat{x} = a_{xy} + b_{xy} y$$

► Poznámky:

- Neplatí $b_{yx} = 1/b_{xy}$, což by platilo v případě funkční závislosti.
- Funkce popisující sdružené regresní přímky nejsou inverzní.
- ...

7/8



Interpretace výsledků lineární regrese

► Interpretace regresních koeficientů

β_0 Odhad absolutního členu – b_0

- b_0 udává jaká je průměrná úroveň závislé proměnné pokud je nezávislá proměnná nulová $\hat{y}(0)$.

β_1 Odhad koeficientu u lineárního členu – b_1

- b_1 tangenta směrového úhlu regresní přímky,
- b_1 udává o kolik se v průměru zvedne hodnota závislé proměnné, když se hodnota nezávisle proměnné zvedne o jednotku,
- $b_1 > 0$ – pozitivní (lineární) korelační závislost,
- $b_1 < 0$ – negativní (lineární) korelační závislost,
- $b_1 = 0$ – (lineární) nekorelovanost

► Interpretace odhadu $\bar{y}$

- Přesněji: Interpretace bodového odhadu závislé proměnné $\hat{y}(x_0)$ na dané úrovni nezávisle proměnné x_0 (pevný efekt)
 - $\hat{y}(x_0)$ udává jaká je průměrná hodnota závislé proměnné pro hodnotu nezávisle proměnné x_0

► Vyjádření regresní přímky s využitím $\bar{x}$ a $\bar{y}$

- $(y - \bar{y}) = b_1(x - \bar{x})$
- regresní přímka prochází bodem $[\bar{x}; \bar{y}]$

6/8



...

8/8

