

## 12 Regrese

### Teorie: Regresní model

Zjistíme, zda hodnoty  $Y$  (vysvětlovaná proměnná) závisí na hodnotě  $x$  (vysvětlující proměnná).

Uvažujeme model  $Y = f(\beta_0, \beta_1, \dots, \beta_k; x) + \varepsilon$ ,

kde  $\varepsilon \sim N(0, \sigma^2)$  reprezentuje nesystematické chyby.

Model odhadujeme na základě sady měření  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ .

Parametry modelu  $\beta_0, \beta_1, \dots, \beta_k$  lze odhadnout například metodou nejmenších čtverců, která

minimalizuje výraz  $S(\beta_0, \beta_1, \dots, \beta_k) = \sum_{i=1}^n (y_i - f(\beta_0, \beta_1, \dots, \beta_k; x_i))^2$

$$Y = \beta_0 + \beta_1 x + \varepsilon$$

lineární model

$$Y = \beta_0 \cdot e^{\beta_1 x} + \varepsilon$$

exponenciální model

$$Y = \beta_0 + \beta_1 \ln(x) + \varepsilon$$

logaritmický model

$$Y = \beta_0 \cdot x^{\beta_1} + \varepsilon$$

mocninná

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$$

polynomiální stupně 2

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$$

modely s více vysvětlujícími proměnnými

Pokud jsou splněny předpoklady modelu, lze též testovat

Koeficient determinace  $R^2 \in (0, 1)$  určuje podíl vysvětlených variabilit  $Y$  pomocí  $x$ .

F-test testuje hypotézu, zda model je významný.  $H_0 : \beta_1 = 0 \wedge \beta_2 = 0 \dots \wedge \beta_k = 0$

T-testy parametrů testují hypotézy, zda parametry jsou nulové.  $H_0 : \beta_j = 0, j = 0, 1, \dots, k$

V programu Excel existuje několik postupů

přidat spojnicí trendu do grafu

různé typy modelů

INTERCEPT(y;x)

odhadne  $\beta_0$  lineárního modelu

SLOPE(y;x)

odhadne  $\beta_1$  lineárního modelu

LINTREND(y;x;xNew)

vypočte  $y_{New} = \hat{\beta}_0 + \hat{\beta}_1 x_{New}$

LINREGRESE(y;x;koeficient  $\beta_0$  A/N;statistiky A/N)

odhadne  $\beta_0, \beta_1$

koeficient determinace  $R^2$

údaje pro testy o nulovosti parametrů

Analýza dat/Regrese

odhadne  $\beta_0, \beta_1$

koeficient determinace  $R^2$

p-hodnoty testů o významnosti modelu

intervalové odhady parametrů

p-hodnoty testů o nulovosti parametrů

(12.1) Byla provedena měření rychlosti zvuku ve vzduchu v závislosti na teplotě.

Teplota	-20	0	20	50	100
Rychlost	323	327	340	364	386

Odhadněte lineární model závislosti rychlosti zvuku na teplotě.

[  $rychlost = 331.159 + 0.561 \cdot teplota$ ,  $p$ -hodnota testu  $H_0 : \beta_1 = 0$  je 0.0012 ]

[ je prokázáno, že rychlost zvuku závisí na teplotě ]

- (12.2) V tabulce uvedeny hodnoty pružnosti materiálu v závislosti na čase. Navrhněte vhodný model a odhadněte jeho parametry.

Čas	0	2	4	6	8	10
Pružnost	100	79	70	68	51	49

[ Například model  $y = 95.742 \cdot e^{0.07x}$ , kdy koeficient determinace  $R^2 = 0.9515$  ]

- (12.3) Zajímá nás vliv dešťových srážek na výskyt plísně u brambor a na úrodu. Na pokusných polích v deseti oblastech byly zjištěny údaje o vydatnosti průměrných letních měsíčních srážek (v mm), podíl sazenic napadených plísní (v %) a výnos z jednoho hektaru (v t).

pole	srážky	plíseň	výnos
1	36	40	160
2	40	35	200
3	30	35	140
4	25	25	130
5	41	45	170
6	23	20	150
7	35	30	150
8	34	35	200
9	44	45	230
10	42	40	195

- (a) Ovlivňuje vydatnost srážek výskyt plísně ? Ovlivňuje vydatnost srážek výnosy ? Souvisí spolu výskyt plísně a výnosy ?  
 (b) Jaký je vliv srážek na výnos bez vlivu plísní ?  
 (c) Jaký je vliv plísní na výnos bez vlivu srážek ?  
 (d) Jaký je celkový vliv srážek a plísní na výnos ?

[ (a) Srážky - plísně a srážky - výnos -> souvisí spolu, plíseň - výnos -> nesouvisí spolu ]  
 [ použijeme test o nulovosti korelačního koeficientu ]

[ (b)  $vynos = 50.152 + 3.496 \cdot srazky$ , jeden milimetr srážek zvýší výnos o 3.496 tun, ]  
 [ koeficient determinace  $R^2 = 0.603$ , p-hodnota F-testu významnosti modelu = 0.008 ]

[ (c)  $vynos = 86.458 + 2.458 \cdot plisne$ , nezamítám ale  $H_0 : \beta_1 = 0$  tedy plísně výnos neovlivňují ]  
 [ koeficient determinace  $R^2 = 0.387$ , p-hodnota F-testu významnosti modelu = 0.055 ]

[ (d)  $vynos = 49.316 + 4.898 \cdot srazky - 1.378 \cdot plisne$  ]  
 [ jeden milimetr srážek zvýší výnos o 4.898 tun, jedno procento plísní sníží výnos o 1.378 tun. ]  
 [ koeficient determinace  $R^2 = 0.628$ , p-hodnota F-testu významnosti modelu = 0.031 ]