Box-Cox transformation is defined by the following equation:

$$y^*(\lambda) = \begin{cases} \frac{y^\lambda - 1}{\lambda}, & \text{if } \lambda \neq 0 \\ \ln(y), & \text{if } \lambda = 0 \end{cases}$$

- prove that it is continuous with respect to λ
- write down log-likelihood function and its partial derivative with respect to λ for regression line model driven random variable Y.

Důkaz spojitosti

$$\lim_{\lambda \to 0} \frac{y^\lambda - 1}{\lambda} = \ln(y)$$

dosadit $\lambda = 0$

$$\lim_{\lambda \to 0} \frac{y^\lambda - 1}{\lambda} = \left[\frac{0}{0}\right] = \lim_{\lambda \to 0} \frac{y^\lambda \cdot \ln(y)}{1} = \ln(y)$$

$$L(\sigma, \lambda, \beta_1, \beta_2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} e^{\frac{\left(\frac{y_i^\lambda - 1}{\lambda} - (\beta_1 + \beta_2 y_i)\right)^2}{2\sigma^2}} \cdot y_i^{\lambda - 1} =$$

$$y^* = \frac{y^\lambda - 1}{\lambda}$$

Transformace nezachovává PDF $= 1$ $\int f(x) = 1$

$\Rightarrow$ substituce a oprava

$$dy^* = \frac{\lambda y^{\lambda-1}}{\lambda} dy$$

$$dy^* = y^{\lambda-1} dy$$

$$= -n \ln\left(\sqrt{2\pi}\,\sigma\right) + \frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(\frac{y_i^\lambda - 1}{\lambda} - (\beta_1 + \beta_2 y_i)\right)^2 + (\lambda - 1) \sum_{i=1}^{n} \ln(y_i)$$

$$\frac{d\ell}{d\lambda} = \frac{1}{2\sigma^2} \sum_{i=1}^{n} 2\left(\frac{y_i^\lambda - 1}{\lambda} - (\beta_1 + \beta_2 y_i)\right) \cdot \frac{y_i^\lambda \cdot \ln(y_i) \cdot \lambda - (y_i^\lambda - 1) \cdot 1}{\lambda^2} + \sum_{i=1}^{n} \ln(y_i)$$

You successfully fitted a quadratic regression model through a dataset with $n = 100$ observations for predictor values $\in (-1; 1)$. Assume model $y = \beta_1 + \beta_2 x + \beta_3 x^2$, you obtained 95% CI for all coefficients $\beta_1 \in (9{,}762;\ 10{,}327)$, $\beta_2 \in (1{,}636;\ 2{,}289)$, $\beta_3 \in (4{,}007;\ 5{,}272)$ and residual standard deviation $s_{res} = 0,950093$. $\sum_{i=1}^{n}(y_i - \bar{y})^2 = 407{,}1588$

$$(\mathbf{X}^T\mathbf{X})^{-1} = \begin{bmatrix} 0{,}022504 & 0 & -0{,}03752 \\ 0 & 0{,}030003 & 0 \\ -0{,}03752 & 0 & 0{,}112556 \end{bmatrix}$$

a

- Compute R^2 and test whether all terms in regression model are statistically significant. b
- For observations
c $x = [-0{,}65; 0{,}53; 0{,}89]$; $y = [-8{,}532; 14{,}431; 13{,}265]$ compute Leverages, Standardized and Studentized residuals and Cook's distance.

d
- Compute 95% confidence intervals for prediction and expected value of a model for $x_{00} = 0$ and $x_0 = 0{,}53$

ⓐ $R^2 = 1 - \frac{S_e}{\sum_{i=1}^{n}(y-\bar{y})^2} = 1 - \frac{}{407{,}1588} = 0{,}7849$

$S_{res}^2 = \frac{S_e}{n-k} \Rightarrow S_e = S_{res}^2 \cdot (100-3) = 87{,}5596$

ⓑ $H_0: \beta_i = 0 \quad H_A: \beta_i \neq 0$

$t = \frac{\beta_i}{S_{res} \cdot \sqrt{g^{ii}}}$

$\beta_1: b_1 = 10{,}0445$ $\qquad t = \frac{\frac{9{,}762 + 10{,}327}{2}}{S_{res}\sqrt{g^{11}}} = 70{,}475$

$\beta_2: b_2 = 1{,}9625$ $\qquad t = 11{,}925$

$\beta_3: b_3 = 4{,}6395$ $\qquad t = 14{,}555$

$t = \frac{(0,1,0)\cdot\begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} - (0,1,0)\begin{pmatrix} b_1 \\ 0 \\ b_3 \end{pmatrix}}{S_{res}\sqrt{(0,1,0)(X^TX)^{-1}\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}}}$

$\bar{W}_{0,05} = \langle -t_{0,975}(97); t_{0,975}(97) \rangle$

$\langle -1{,}985 \; ; \; 1{,}985 \rangle$

$\Rightarrow \beta_i \notin \bar{W}$

$\Rightarrow$ vše se významně liší od 0

= všechny β_i jsou významné

ⓒ $x = -0{,}65 \qquad y = -8{,}532$

$H = X(X^TX)^{-1}X^T \qquad X = \begin{pmatrix} 1 & x_1 & x_1^2 \\ & \vdots & \\ 1 & x_n & x_n^2 \end{pmatrix}$

$h_{ii} = ?$ $\qquad$ Leverage

$h_{ii} = (1; -0{,}65; 0{,}65^2)(X^TX)^{-1}\begin{pmatrix} 1 \\ -0{,}65 \\ -0{,}65^2 \end{pmatrix} =$

$= (0{,}0066; -0{,}0195; 0{,}01)\begin{pmatrix} 1 \\ -0{,}65 \\ -0{,}65^2 \end{pmatrix} = 0{,}0235 = h_{ii}$

$D_i = \frac{e_i^2}{kS_{res}^2} \cdot \frac{h_{ii}}{(1-h_{ii})^2} = 3{,}376$

$e_i = y_i - b_1 - b_2 x_i - b_3 x_i^2 = -19{,}2629 \qquad k = 3$

$$r_i = \frac{e_i}{\sqrt{s_{res}^2(1-h_{ii})}} = -20{,}52$$

$$r_{(i)} = e_i\sqrt{\frac{(n-k)\cdot e_i}{S_e(1-h_{ii})-e_i^2}} = -47{,}28$$!

↳ vychází < 0 => nejde ze zadaných hodnot

(d) $CI = \left(\hat{y}(x_0) - s_{rest}\, t_{1-\frac{\alpha}{2}}(n-k)\sqrt{d^2(x_0)} \,;\, \hat{y}(x_0) + \dots \right)$

$PI = \left(\hat{y}(x_0) - s_{rest}\, t_{1-\frac{\alpha}{2}}(n-k)\sqrt{d^2(x_0)+1} \,;\, \hat{y}(x_0) + \dots \right)$

$x_0 = 0$

$\hat{y}(0) = 10{,}0445$ $\quad t_{0{,}975}(97) = 1{,}985$

$$d^2(x_0) = (1,0,0)(X^TX)^{-1}\begin{pmatrix}1\\0\\0\end{pmatrix} = 0{,}022504$$

$CI = \langle 9{,}76 \;;\; 10{,}327 \rangle$

$PI = \langle 8{,}137 \;;\; 11{,}95 \rangle$

$x_0 = 0{,}53$

$\hat{y}(0{,}53) = 12{,}3879$

$$d^2(0{,}53) = (1,\,0{,}53,\,0{,}53^2)(X^TX)^{-1}\begin{pmatrix}1\\0{,}53\\0{,}53^2\end{pmatrix} = \quad 0{,}019$$

$$= (0{,}0119 \;;\; 0{,}0159 \;;\; -0{,}0059)\begin{pmatrix}1\\0{,}53\\0{,}53^2\end{pmatrix} = \underline{\underline{0{,}0186}}$$