

# WYBÓR MODELU STATYSTYCZNEGO

Piotr Pokarowski

29 kwietnia 2010

# Model liniowy – przykład

Modelowanie zależności między zawartością tłuszczu w ciele (Bodyfat) a 13 predyktorami (Age, Weight, Abdomen itp.)

$$\text{Bodyfat}_i \approx \beta_0 + \beta_1 * \text{Age}_i + \beta_2 * \text{Weight}_i + \dots + \beta_{13} * \text{Abdomen}_i,$$

gdzie  $i = 1, \dots, 252$  – 'niezależne' obserwacje.

- Które predyktory można pominąć ?
- Czy dodać interakcję  $\text{Age}_i * \text{Weight}_i$  ?

# Model liniowy – problem wyboru modelu

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i,$$

gdzie  $i = 1, \dots, n$ ;  $\varepsilon_i$  – niezal. zm. los. o rozkładzie  $N(0, \sigma^2)$ .

Równoważnie

$$y = X\beta + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2 I_n), \quad X - \text{macierz } n \times (p + 1).$$

Wiersze  $X$  – obserwacje, kolumny  $X$  – predyktory.

Chcemy wybrać model  $j \subseteq \{1, \dots, p\} =: \text{full}$ .

$X_j$  – podmacierz  $X$  o kolumnach indeksowanych elementami  $j$ ,  
 $\beta_j$  – podwektor  $\beta$  o elementach indeksowanych elementami  $j$ .

# Idealne kryterium Akaike

$y = \mu + \varepsilon$ ,  $y_* = \mu + \varepsilon_*$  – nowa replikacja modelu,

$\varepsilon, \varepsilon_* \sim N(0, \sigma^2 I_n)$ ,  $\varepsilon, \varepsilon_*$  są niezależne.

$\hat{y}_j := H_j y$  – predyktor liniowy  $y_*$  (estymator  $\mu$ ).

Najczęściej  $H_j = X_j(X_j^T X_j)^{-1} X_j^T$ .

Idealne AIC (Akaike Information Criterion)

$$\text{idAIC} := \operatorname{argmin}_j \mathbf{E} |y_* - \hat{y}_j|^2 = \operatorname{argmin}_j \mathbf{E} |\mu - \hat{y}_j|^2$$

# Kryterium Akaike (Akaike 1973)

Stw. 1.

$$\mathbf{E}|y_* - \hat{y}_j|^2 = \mathbf{E}|\mu - \hat{y}_j|^2 + n\sigma^2$$

Stw. 2.

$$\mathbf{E}|y_* - \hat{y}_j|^2 = \mathbf{E}|y - \hat{y}_j|^2 + 2\text{tr}(H_j)\sigma^2 \approx |y - \hat{y}_j|^2 + 2|j|\hat{\sigma}_{full}^2,$$

gdzie  $\hat{\sigma}_{full}^2 = |y - \hat{y}_{full}|^2 / (n - p)$  – nieobciążony estymator  $\sigma^2$ .

Wn.

$$\text{idAIC} \approx \text{argmin}_j (|y - \hat{y}_j|^2 + 2|j|\hat{\sigma}_{full}^2) =: \text{AIC}$$

# Idealne kryterium bayesowskie

Czynnik Bayesa

$$B_{j0}(y) := \frac{p(y|j)}{p(y|0)} = \frac{\int f(y|\beta_j, \beta_0)\pi_j(\beta_j, \beta_0) d(\beta_j, \beta_0)}{\int f(y|\beta_0)\pi_0(\beta_0) d(\beta_0)}$$

Jeśli każdy model jest tak samo prawdopodobny, to

$$B_{j0}(y) = \frac{p(j|y)}{p(0|y)}$$

Idealne BIC (Bayesian Information Criterion)

$$\text{idBIC} := \operatorname{argmax}_j B_{j0}(y) = \operatorname{argmax}_j p(j|y)$$

# Regularny bayesowski model liniowy

Zakładamy między innymi:

- $X^T X/n \rightarrow C > 0$  przy  $n \rightarrow \infty$ ,
- $\pi_j(\beta_j, \beta_0) = \nu_j(\beta_j|\beta_0)\pi_0(\beta_0)$ ,
- $\nu_j(\beta_j|\beta_0) \sim N(0, n\sigma^2(X_j^T X_j)^{-1})$ ,  
 $\nu_j$  – normal unit information prior.

# Kryterium bayesowskie (Schwarz 1978)

Tw. (Pauker 1998)

Dla regularnych bayesowskich modeli liniowych

$$\log B_{j_0}(y) = S_{j_0}(y) + O_P(n^{-1/2}),$$

gdzie

$$S_{j_0}(y) = \frac{|y - \hat{y}_0|^2 - |y - \hat{y}_j|^2}{2\sigma^2} - \frac{|j|}{2} \log(n).$$

Wn.

$$\text{idBIC} \approx \operatorname{argmin}_j (|y - \hat{y}_j|^2 + |j| \log(n) \hat{\sigma}_{full}^2) =: \text{BIC}$$

# Porównanie AIC z BIC

$$\text{AIC} = \operatorname{argmin}_j (|y - \hat{y}_j|^2 + 2|j|\hat{\sigma}_{full}^2)$$

$$\text{BIC} = \operatorname{argmin}_j (|y - \hat{y}_j|^2 + |j| \log(n)\hat{\sigma}_{full}^2)$$

Inne wyprowadzenie przy nieznanym  $\sigma^2$ :

$$\text{AIC} = \operatorname{argmin}_j (n \log |y - \hat{y}_j|^2 + 2|j|)$$

$$\text{BIC} = \operatorname{argmin}_j (n \log |y - \hat{y}_j|^2 + |j| \log(n))$$

Ogólne kryterium wyboru modelu:

$$\operatorname{argmin}_j (\text{GOODNESS} - \text{OF} - \text{FIT}_j + \text{PENALTY}_j)$$

# Zgodność selekcji

Zgodność to asymptotyczna ( $n \rightarrow \infty$ ) poprawność wyboru modelu.

Niech  $t$  – minimalny w sensie inkluzji model prawdziwy, czyli taki, że  $y = X_t\beta_t + \varepsilon$ .

Tw. Dla regularnych modeli liniowych

$\mathbf{P}(AIC = t) \rightarrow 1$  przy  $n \rightarrow \infty$ ,

$\mathbf{P}(BIC = t) \rightarrow 1$  przy  $n \rightarrow \infty$ .

# Zachłanna selekcja

Metoda dwuetapowa Zhenga i Loha (1995).

(1) Znajdź takie posortowanie  $(k_1, \dots, k_p)$  predyktorów, że

$$|y - \hat{y}_{\bar{k}_1}|^2 \geq \dots \geq |y - \hat{y}_{\bar{k}_p}|^2,$$

gdzie  $\bar{k} = full - \{k\}$ .

(2)  $\Lambda := \{\{k_1\}, \{k_1, k_2\}, \dots, full\}$

$$\text{greedyBIC} = \operatorname{argmin}_{j \in \Lambda} (|y - \hat{y}_j|^2 + |j| \log(n) \hat{\sigma}_{full}^2)$$

Tw Dla regularnych modeli liniowych  $\mathbf{P}(\text{greedyBIC} = t) \rightarrow 1$ .