ΘΕΜΑ 2019

Δ1 $\eta_x = \ln \frac{\mu_x}{n_x - \mu_x} = \underline{X}'\vec{\beta}$, οπότε $\ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4}$,

όπου i: δείκτης δείγματος.

Η Deviance $D(\hat{\beta}) = 2\sum_{i=1}^{m}\left(y_i \ln\left(\frac{y_i}{\hat{\mu}_i}\right) + (n_i - y_i)\ln\left(\frac{n_i - y_i}{n_i - \hat{\mu}_i}\right)\right)$, $\hat{\mu}_i = n_i \hat{p}_i$,

μπορεί να γραφεί ως

$$D(\hat{\beta}) = \sum_{i=1}^{n} (r_i^D)^2$$, όπου r_i^D το υπόλοιπο deviance, δηλαδή

$$r_i^D = \text{sgn}(y_i - \hat{\mu}_i)\left\{2y_i \ln\left(\frac{y_i}{\hat{\mu}_i}\right) + 2(n_i - y_i)\ln\left(\frac{n_i - y_i}{n_i - \hat{\mu}_i}\right)\right\}^{1/2}$$

Ένα μοντέλο λογιστικής παλινδρόμησης μπορεί να αξιολογηθεί με τη βοήθεια index plots, καθώς και γραφήματα της κανονικής και ημι-κανονικής κατανομής. Αυτό, διότι η κατανομή των υπολοίπων Pearson και deviance δεν προσεγγίζεται από την κανονική κατανομή στην περίπτωση $n=1$ (binary data). (Η Pearson δεν είναι γενικά καλή)

Δ2 $Y=1$: επιβίωση X_1: ηλικία, X_2: πίεση, X_3: σφυγμός, $X_4 = 1$: εισαγωγή

(i) $z_1 = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)} = -3.211$, p-value $= 2 * \text{pnorm}(-3.211, \text{lower.tail} = \text{TRUE})$

$z_2 = \frac{\hat{\beta}_2}{se(\hat{\beta}_2)} = 2.266$, p-value $= 2 * \text{pnorm}(2.266, \text{lower.tail} = \text{FALSE})$

κ.ο.κ.

Για το AIC του μοντέλου 2: $AIC = -2\hat{\ell} + 2p = 175.82$, $p=4$

$D_2 - D_1 = 167.82 - 167.05 = 0.77$, p-val: Ακολουθεί την $\chi^2_{2-1} = \chi^2_1$

Άρα p-val = pchisq(0.77, 1, lower.tail = FALSE) = 0.380

$M_2 \subset M_1$, με μικρή διαφορά deviance και υψηλό p-value → απορρίπτουμε

$D_3 - D_2 = 173.08 - 167.82 = 5.26$, $p_3 - p_2 = 1$, άρα p-val = pchisq(5.26, 1, lower.tail=FALSE)
$= 0.0218$

Σημαντική διαφορά $D_3 - D_2$ και μικρό p-value → απορρίπτουμε το M_3
$M_3 \subset M_2$

Άρα βέλτιστο το M_2.

(ii) Για 95% διάστημα εμπιστοσύνης θέλουμε $\alpha = 0.05$, άρα είναι το

$$\left[\exp\left(\hat{\beta}_4 - \underbrace{z_{0.025}}_{|\text{qnorm}(0.025)|} \cdot se(\hat{\beta}_4)\right), \exp\left(\hat{\beta}_4 + z_{0.025} \cdot se(\hat{\beta}_4)\right)\right] = [0.02296, 0.4485915]$$