EXERCISE 2 exam 22/02/2024

Y_i = survival of individual i $\quad i = 1, \dots, 714$

a) the model is

- $Y_i \sim \text{Bernoulli}(\pi_i)$ independent $\quad i = 1, \dots, n \quad n = 714$
- linear predictor

$$\eta_i = \underline{x}_i^T \underline{\beta} = \beta_1 + \beta_2 \underbrace{\mathbb{1}(\text{class}_i = \text{first})}_{\in\{0,1\}} + \beta_3 \underbrace{\mathbb{1}(\text{gender}_i = \text{man})}_{\in\{0,1\}} + \beta_4 \text{age}_i$$

$\begin{cases} 1 & \text{if class} = \text{first} \\ 0 & \text{if class} \neq \text{first} \end{cases}$ $\quad$ $\begin{cases} 1 & \text{if gender}_i = \text{man} \\ 0 & \text{if gender}_i = \text{woman} \end{cases}$

- logit link function: $g(\pi_i) = \text{logit}(\pi_i) = \log \frac{\pi_i}{1-\pi_i} = \eta_i$

b) $p(y_i; \pi_i) = \pi_i^{y_i} (1-\pi_i)^{1-y_i}$

$p(y_1, \dots, y_n; \underline{\pi}) = \prod_{i=1}^n \left\{ \pi_i^{y_i} (1-\pi_i)^{1-y_i} \right\}$ where $\pi_i = \frac{e^{\underline{x}_i^T \underline{\beta}}}{1 + e^{\underline{x}_i^T \underline{\beta}}}$

likelihood function

$$L(\underline{\beta}) = \prod_{i=1}^n \left\{ \left(\frac{e^{\underline{x}_i^T \underline{\beta}}}{1+e^{\underline{x}_i^T \underline{\beta}}} \right)^{y_i} \left(\frac{1}{1+e^{\underline{x}_i^T \underline{\beta}}} \right)^{1-y_i} \right\}$$

the log-likelihood is

$$\begin{aligned} \ell(\underline{\beta}) &= \log L(\underline{\beta}) \\ &= \sum_{i=1}^n \left\{ \underline{x}_i^T \underline{\beta}\, y_i - y_i \log(1+e^{\underline{x}_i^T \underline{\beta}}) - (1-y_i) \log(1+e^{\underline{x}_i^T \underline{\beta}}) \right\} \\ &= \sum_{i=1}^n \left\{ \underline{x}_i^T \underline{\beta}\, y_i - \cancel{y_i \log(1+e^{\underline{x}_i^T \underline{\beta}})} - \log(1+e^{\underline{x}_i^T \underline{\beta}}) + \cancel{y_i \log(1+e^{\underline{x}_i^T \underline{\beta}})} \right\} \\ &= \sum_{i=1}^n \left\{ \underline{x}_i^T \underline{\beta}\, y_i - \log(1+e^{\underline{x}_i^T \underline{\beta}}) \right\} \end{aligned}$$

c) 1. $\hat{z}$ value of "Age"

this is the observed value of the test statistic for testing

$H_0: \beta_4 = 0$ vs $H_1: \beta_4 \neq 0$

we use the statistic

$$\hat{z} = \frac{\hat{\beta}_4}{\sqrt{[I(\hat{\beta})^{-1}]_{(4,4)}}} \overset{H_0}{\sim} N(0,1)$$

hence the needed quantity is $z^{obs} = \frac{-0.0299}{0.00714} = -4.04$

2. $Pr(>|z|)$

this is the p-value of the test

$$\begin{aligned} \alpha^{obs} &= P_{H_0}(|Z| > |z^{obs}|) \\ &= 2\, P_{H_0}(Z > |z^{obs}|) \\ &= 2\,(1 - \Phi(|z^{obs}|)) \\ &= 2\,(1 - \Phi(4.04)) \end{aligned}$$

$\Phi(4.04) \approx 1 \Rightarrow \alpha^{obs} \approx 0$

3. All of them are significant

d) $\hat{\pi}_A = \left(\hat{\pi}_A \mid \text{gender}_A = 0,\ \text{class}_A = 1,\ \text{age}_A = 30\right) = \frac{e^{\underline{x}_A^T \hat{\beta}}}{1 + e^{\underline{x}_A^T \hat{\beta}}}$

the odds are $\frac{\hat{\pi}_A}{1-\hat{\pi}_A} = e^{\underline{x}_A^T \hat{\beta}}$

$\underline{x}_A^T \hat{\beta} = 1 \cdot 2.5003 + 2.0103 \cdot 1 - 2.5278 \cdot 0 - 0.0299 \cdot 30 = 2.6136$

odds $= e^{\underline{x}_A^T \hat{\beta}} = e^{2.6136} = 13.648$

we know that

$\text{logit}\,\hat{\pi}_A = \hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_4 \cdot 30$

$\text{logit}\,\hat{\pi}_B = \hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_4 \cdot 31 = \hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_4 \cdot 30 + \hat{\beta}_4$

Hence $\text{logit}\,\hat{\pi}_B - \text{logit}\,\hat{\pi}_A = \hat{\beta}_4$

$$\log \frac{\hat{\pi}_B}{1-\hat{\pi}_B} - \log \frac{\hat{\pi}_A}{1-\hat{\pi}_A} = \log \frac{\frac{\hat{\pi}_B}{1-\hat{\pi}_B}}{\frac{\hat{\pi}_A}{1-\hat{\pi}_A}} = \hat{\beta}_4$$

$$\Rightarrow \frac{\frac{\hat{\pi}_B}{1-\hat{\pi}_B}}{\frac{\hat{\pi}_A}{1-\hat{\pi}_A}} = e^{\hat{\beta}_4} \qquad \text{odds}_B = 0.97\ \text{odds}_A$$

The odds of individual A are multiplied by $e^{\hat{\beta}_4} = 0.97$ to obtain the odds of individual B.

That is, $\text{odds}_B = \frac{\hat{\pi}_B}{1-\hat{\pi}_B} = e^{-0.0299} \cdot 13.648 = 13.246$

In this case, $e^{\hat{\beta}_4} = 0.97 < 1$, hence the odds decrease for individual B w.r.t. individual A.

e) "class" is a dummy variable

Hence

$$\hat{\beta}_2 = \log \frac{\frac{\hat{P}(Y_i=1 \mid \text{class}_i = 1)}{\hat{P}(Y_i=0 \mid \text{class}_i = 1)}}{\frac{\hat{P}(Y_i=1 \mid \text{class}_i = 0)}{\hat{P}(Y_i=0 \mid \text{class}_i = 0)}} \Rightarrow e^{\hat{\beta}_2} = \frac{(\text{odds} \mid \text{class}_i = 1)}{(\text{odds} \mid \text{class}_i = 0)}$$

The odds of a person in third class are multiplied by $e^{\hat{\beta}_2} = e^{2.0103} = 7.4657$ to obtain the odds of a person in first class (keeping the other covariates fixed).

Since $\hat{\beta}_2$ is positive, a person with a first-class ticket has a higher probability of surviving, compared to a person with a second or third-class ticket.

f) the residual deviance is the lik.-ratio test between the saturated model and the proposed model:

$$D(\text{model}) = 2\left\{ \hat{\ell}(\text{saturated}) - \hat{\ell}(\text{model}) \right\}$$

where $\hat{\ell}(\text{saturated})$ is the maximum of the log-likelihood under a model with n parameters,

and $\hat{\ell}(\text{model})$ is the maximum of the log-likelihood under the current model

The null deviance is

$$D(\text{null}) = 2\left\{ \hat{\ell}(\text{saturated}) - \tilde{\ell}(\text{null}) \right\}$$

where $\tilde{\ell}(\text{null})$ is the maximum of the log-likelihood under a model with a single parameter π

g) we want to test

$\begin{cases} H_0: \beta_2 = \beta_3 = \beta_4 = 0 \\ H_1: \text{at least one is} \neq 0 \end{cases}$

under H_0, the model is $Y_i \sim \text{Ber}(\pi)$ $\quad \text{logit}(\pi) = \beta_1$ "null model"

the maximum of the log-likelihood is $\tilde{\ell}(\text{null})$

under H_1, I have the complete model

the maximum of the log-likelihood is $\hat{\ell}(\text{model})$

The LR test for testing the model is

$$W = 2(\hat{\ell}(\text{model}) - \tilde{\ell}(\text{null})) \overset{\cdot}{\sim} \chi^2_{p-1} = \chi^2_3 \quad \text{under } H_0$$

$$\begin{aligned} D(\text{null}) - D(\text{model}) &= 2\left\{ \cancel{\hat{\ell}(\text{saturated})} - \tilde{\ell}(\text{null}) - \cancel{\hat{\ell}(\text{saturated})} + \hat{\ell}(\text{model}) \right\} \\ &= 2\left\{ \hat{\ell}(\text{model}) - \tilde{\ell}(\text{null}) \right\} \end{aligned}$$

Hence the observed value of the test is

$w^{obs} = 964.52 - 675.14 = 289.38$

Reject region: $R = (\chi^2_{3;1-\alpha}; +\infty)$

if $\alpha = 0.05$, $\chi^2_{3;0.95} = 7.81$

I reject H_0

