EXERCISE 2

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

a) the model is

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

$$\text{logit}(\pi_i) = \beta_1 + \beta_2 \underbrace{\mathbb{1}(\text{class}_i = \text{first})}_{\text{dummy} = \begin{cases} 1 & \text{if class = first} \\ 0 & \text{if class} \neq \text{first} \end{cases}} + \beta_3 \underbrace{\mathbb{1}(\text{gender}_i = \text{man})}_{\text{dummy} = \begin{cases} 1 & \text{if gender}_i = \text{man} \\ 0 & \text{if gender}_i = \text{woman} \end{cases}} + \beta_4 \, \text{age}_i$$

The estimated model is

$$Y_i \sim \text{Bernoulli}(\hat{\pi}_i) \qquad \text{logit}(\hat{\pi}_i) = \log \frac{\hat{\pi}_i}{1-\hat{\pi}_i} = 1.50 + 2.01 \, \mathbb{1}(\text{class}_i = \text{first}) - 2.54 \, \mathbb{1}(\text{gender}_i = \text{man}) - 0.029 \cdot \text{age}_i$$

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

$p(y_1, \dots, y_n ; \underline{\pi}) = \prod_{i=1}^{n} \{ \pi_i^{y_i} (1-\pi_i)^{1-y_i} \}$

likelihood function

$$L(\underline{\pi}) = \prod_{i=1}^{n} \{ \pi_i^{y_i} (1-\pi_i)^{1-y_i} \}$$

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

where I denote with $\tilde{x}_i^T$ the row vector of the observed values of the covariates for individual i

the loglikelihood is

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

c) 1. z value of "Gender"

this is the observed value of the test statistic for testing

$H_0: \beta_3 = 0$ vs $H_1: \beta_3 \neq 0$

we use the statistic

$$z = \frac{\hat{\beta}_3 - 0}{\sqrt{[j(\hat{\beta})]^{-1}_{(3,3)}}} \overset{H_0}{\sim} N(0,1)$$

hence the needed quantity is $z^{obs} = \frac{-2.5473}{0.2017} = -12.629$

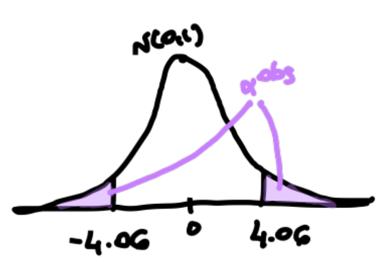
2. $\Pr(>|z|)$ for "Age"

this is the pvalue of the test $H_0: \beta_4 = 0$ vs $H_1: \beta_4 \neq 0$

Similar to the previous point, it is based on the quantity

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

Hence the pvalue is
$$\begin{aligned} \alpha^{obs} &= \mathbb{P}_{H_0}(|z| > |z^{obs}|) \\ &= 2 \, \mathbb{P}_{H_0}(z > |z^{obs}|) \\ &= 2 \, (1 - \Phi(|z^{obs}|)) \\ &= 2 \, (1 - \Phi(4.06)) \end{aligned}$$



$\Phi(4.06) \approx 1 \Rightarrow \alpha^{obs} \approx 0$

3. All of them.

d) $\hat{\pi}_A = (\hat{\pi}_A \mid \text{gender}_A = 0, \text{class}_A = 1, \text{age}_A = 30) = \frac{e^{\tilde{x}_A^T \hat{\beta}}}{1+e^{\tilde{x}_A^T \hat{\beta}}} = 0.9317$

since $\tilde{x}_A^T \hat{\beta} = 1.5003 + 2.0103 \cdot 1 - \cancel{2.5473} \cdot 0 - 0.0299 \cdot 30 = 2.6136$

$\Rightarrow \frac{e^{\tilde{x}_A^T \beta}}{1+e^{\tilde{x}_A^T \beta}} = \frac{e^{2.6136}}{1+e^{2.6136}} = 0.9317$

the odds for individual A are $\frac{\hat{\pi}_A}{1-\hat{\pi}_A} = 13.648$

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

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

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

The odds of individual A change of a multiplicative factor $e^{\hat{\beta}_4}$ to obtain the odds of individual B.

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

e) "class" is a dummy variable

Hence

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

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) $\begin{cases} H_0: \beta_2 = 0 \\ H_1: \beta_2 < 0 \end{cases}$

we use the test statistic

$$z = \frac{\hat{\beta}_2 - 0}{\hat{se}(\hat{\beta}_2)} \overset{H_0}{\sim} N(0,1)$$

the reject region is for large negative values

using a significance level α: reject if $z^{obs} < z_\alpha$

$z^{obs} = 8.11$ (in the table)

Hence I do not reject H_0 for all usual α

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

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

where $\tilde{\ell}(\text{saturated})$ 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

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

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

h) 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 loglikelihood is $\tilde{\ell}(\text{null})$

under H_1, I have the complete model

the maximum of the loglikelihood is $\hat{\ell}(\text{model})$

The LR test for testing the model is

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

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

Hence the observed value of the test is

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

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

I reject H_0