EXERCISE 2 (exam practice)

$n = 30 \qquad (y_1, \dots, y_{10}, y_{11}, \dots, y_{30})$

$$x_i = \begin{cases} 1 & i = 1, \dots, 10 \\ 0 & i = 11, \dots, 30 \end{cases}$$

$Y_i \sim Ber(\pi_i) \qquad logit(\pi_i) = \beta_1 + \beta_2 x_i$

$$\Leftrightarrow \log \frac{\pi_i}{1-\pi_i} = \beta_1 + \beta_2 x_i \qquad \Leftrightarrow \pi_i = \frac{e^{\beta_1 + \beta_2 x_i}}{1 + e^{\beta_1 + \beta_2 x_i}}$$

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

$$p(y_1, \dots, y_{30} ; \underline{\pi}) = \prod_{i=1}^{n} \pi_i^{y_i} (1-\pi_i)^{1-y_i} \qquad \text{with } \pi_i = \frac{e^{\beta_1 + \beta_2 x_i}}{1 + e^{\beta_1 + \beta_2 x_i}}$$

likelihood function

$$L(\underline{\pi}) = \prod_{i=1}^{n} \pi_i^{y_i} (1-\pi_i)^{1-y_i} \rightarrow L(\underline{\beta}) = \prod_{i=1}^{n} \left(\frac{e^{\beta_1 + \beta_2 x_i}}{1 + e^{\beta_1 + \beta_2 x_i}} \right)^{y_i} \left(\frac{1}{1 + e^{\beta_1 + \beta_2 x_i}} \right)^{1-y_i}$$

log-likelihood function

$$\ell(\underline{\pi}) = \log L(\underline{\pi})$$

$$= \sum_{i=1}^{n} y_i \underbrace{\log \pi_i}_{\beta_1 + \beta_2 x_i - \log(1+e^{\beta_1+\beta_2 x_i})} + (1-y_i) \underbrace{\log(1-\pi_i)}_{\cancel{\log(1)} - \log(1+e^{\beta_1+\beta_2 x_i})}$$

hence

$$\ell(\underline{\beta}) = \sum_{i=1}^{n} y_i (\beta_1 + \beta_2 x_i) - \cancel{y_i \log(1+e^{\beta_1+\beta_2 x_i})} - \log(1+e^{\beta_1+\beta_2 x_i}) + \cancel{y_i \log(1+e^{\beta_1+\beta_2 x_i})}$$

$$= \sum_{i=1}^{n} \left\{ y_i (\beta_1 + \beta_2 x_i) - \log(1 + e^{\beta_1 + \beta_2 x_i}) \right\}$$

$$= \beta_1 \sum_{i=1}^{n} y_i + \beta_2 \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} \log(1 + e^{\beta_1 + \beta_2 x_i})$$

finally, the score function is

$$\ell_*(\beta) = \frac{\partial \ell(\underline{\beta})}{\partial \beta_j} \qquad j = 1, 2$$

$$= \begin{cases} \dfrac{\partial \ell(\underline{\beta})}{\partial \beta_1} = \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \dfrac{1}{1+e^{\beta_1+\beta_2 x_i}} \cdot e^{\beta_1+\beta_2 x_i} = 30\bar{y} - \sum_{i=1}^{10} \dfrac{e^{\beta_1+\beta_2}}{1+e^{\beta_1+\beta_2}} + \sum_{i=11}^{30} \dfrac{e^{\beta_1}}{1+e^{\beta_1}} \\ \dfrac{\partial \ell(\underline{\beta})}{\partial \beta_2} = \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} \dfrac{1}{1+e^{\beta_1+\beta_2 x_i}} \cdot e^{\beta_1+\beta_2 x_i} \cdot \underbrace{x_i}_{\text{is } = 0 \text{ for } i = 11, \dots, 30} = \sum_{i=1}^{10} y_i - \sum_{i=1}^{10} \dfrac{e^{\beta_1+\beta_2}}{1+e^{\beta_1+\beta_2}} \end{cases}$$

If I denote with $\bar{y}_1$ the mean of the first 10 observations, and $\bar{y}_2$ the mean of the last 20, I get:

$$30\bar{y} = \sum_{i=1}^{30} y_i = 10\bar{y}_1 + 20\bar{y}_2$$

$$\sum_{i=1}^{10} y_i = 10\bar{y}_1$$

$$= \begin{cases} 10\bar{y}_1 + 20\bar{y}_2 - 10\dfrac{e^{\beta_1+\beta_2}}{1+e^{\beta_1+\beta_2}} - 20\dfrac{e^{\beta_1}}{1+e^{\beta_1}} \\ 10\bar{y}_1 - 10\dfrac{e^{\beta_1+\beta_2}}{1+e^{\beta_1+\beta_2}} \end{cases}$$

likelihood equations: $\ell_*(\underline{\beta}) = 0$

we can solve them analytically and find the MLE in this case

$$\begin{cases} 10\bar{y}_1 + 20\bar{y}_2 - 10\dfrac{e^{\beta_1+\beta_2}}{1+e^{\beta_1+\beta_2}} - 20\dfrac{e^{\beta_1}}{1+e^{\beta_1}} = 0 \\ 10\bar{y}_1 - 10\dfrac{e^{\beta_1+\beta_2}}{1+e^{\beta_1+\beta_2}} = 0 \end{cases}$$

proportion of successes in the first 10 obs.

$$(\text{eq. } 2): \quad \bar{y}_1 = \frac{e^{\beta_1+\beta_2}}{1+e^{\beta_1+\beta_2}} \rightarrow \hat{\beta}_1 + \hat{\beta}_2 = \log \frac{\bar{y}_1}{1-\bar{y}_1} = \log \frac{\hat{\pi}_1}{1-\hat{\pi}_1}$$

with $\hat{\pi}_1 = \bar{y}_1$ estimate of the probability for observations in the first group $(i = 1, \dots, 10)$

$$(\text{eq } 1): \quad 10\bar{y}_1 + 20\bar{y}_2 - 10\underbrace{\frac{e^{\hat{\beta}_1+\hat{\beta}_2}}{1+e^{\hat{\beta}_1+\hat{\beta}_2}}}_{\bar{y}_1} - 20\frac{e^{\beta_1}}{1+e^{\beta_1}} = 0$$

$$\cancel{10\bar{y}_1} + 20\bar{y}_2 - \cancel{10\bar{y}_1} - 20\frac{e^{\hat{\beta}_1}}{1+e^{\hat{\beta}_1}} = 0 \Longrightarrow \frac{e^{\hat{\beta}_1}}{1+e^{\hat{\beta}_1}} = \bar{y}_2 \Rightarrow \hat{\beta}_1 = \log \frac{\bar{y}_2}{1-\bar{y}_2} = \log \frac{\hat{\pi}_2}{1-\hat{\pi}_2}$$

Hence

$$\hat{\beta}_1 = logit(\bar{y}_2)$$

$$\hat{\beta}_1 + \hat{\beta}_2 = logit(\bar{y}_1) \Rightarrow \hat{\beta}_2 = logit(\bar{y}_1) - logit(\bar{y}_2)$$

The fitted model is

$$Y_i \sim Ber(\hat{\pi}_i) \qquad logit(\hat{\pi}_i) = 1.3863 - 2.0794\, x_i$$

b) $\hat{\pi}_i$ when $x_i = 0$ is

$$P(Y_i = 1 | x_i = 0) = \frac{e^{\hat{\beta}_1}}{1+e^{\hat{\beta}_1}} = 0.800$$

$\hat{\pi}_i$ when $x_i = 1$ is

$$P(Y_i = 1 | x_i = 1) = \frac{e^{\hat{\beta}_1+\hat{\beta}_2}}{1+e^{\hat{\beta}_1+\hat{\beta}_2}} = 0.333$$

when $x_i = 0$ the odds are

$$\frac{\text{prob. success} | x_i = 0}{\text{prob. failure} | x_i = 0} = \frac{P(Y_i = 1 | x_i = 0)}{P(Y_i = 0 | x_i = 0)} = \frac{\left(\frac{e^{\hat{\beta}_1}}{1+e^{\hat{\beta}_1}}\right)}{\left(\frac{1}{1+e^{\hat{\beta}_1}}\right)} = \frac{0.800}{0.200} = 4.00 \qquad (= e^{\hat{\beta}_1})$$

odds $\cdot$ 100 = 400 = number of expected successes every 100 failures

→ when $x = 0$, I expect 400 successes every 100 failures

when $x_i = 1$ the odds are

$$\frac{\text{prob. success} | x_i = 1}{\text{prob. failure} | x_i = 1} = \frac{P(Y_i = 1 | x_i = 1)}{P(Y_i = 0 | x_i = 1)} = \frac{\left(\frac{e^{\hat{\beta}_1+\hat{\beta}_2}}{1+e^{\hat{\beta}_1+\hat{\beta}_2}}\right)}{\left(\frac{1}{1+e^{\hat{\beta}_1+\hat{\beta}_2}}\right)} = \frac{0.333}{0.666} = 0.500 \qquad (= e^{\hat{\beta}_1+\hat{\beta}_2})$$

→ when $x = 1$, I expect 50 successes every 100 failures

Finally, the odds ratio is

$$\frac{\left(\frac{\pi_i}{1-\pi_i} | x_i = 1\right)}{\left(\frac{\pi_i}{1-\pi_i} | x_i = 0\right)} = \frac{e^{\hat{\beta}_1+\hat{\beta}_2}}{e^{\hat{\beta}_1}} = e^{\hat{\beta}_2} = 0.1250$$

The odds for the group $x_i = 0$ are multiplied by 0.1250 to obtain the odds at $x_i = 1$

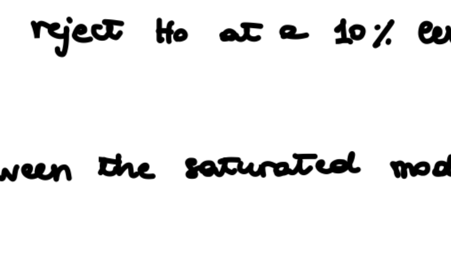
c) $$\begin{cases} H_0: \beta_2 = -1 \\ H_1: \beta_2 < -1 \end{cases}$$

The test statistic $\quad z = \dfrac{\hat{\beta}_2 - (-1)}{\sqrt{[j(\hat{\beta})]^{-1}_{2,2}}} \overset{H_0}{\sim} N(0,1)$

From the summary

$$\sqrt{[j(\hat{\beta})]^{-1}_{2,2}} = 0.7826 \qquad \hat{\beta}_2 = -2.0794$$

$$z^{obs} = \frac{-2.0794 + 1}{0.7826} = -1.3792$$



The reject region here is for negative values

Using a significance level α, I reject H_0 if $z^{obs} < z_\alpha$

$\alpha = 5\% \qquad z_\alpha = z_{0.05} = -z_{0.95} = -1.64 \qquad$ I do not reject H_0 at 5% level

$\alpha = 10\% \qquad z_\alpha = z_{0.10} = -z_{0.90} = -1.28 \qquad$ I reject H_0 at a 10% level

d) 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