EXERCISE 2

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

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

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

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

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

$p(y_1, \dots, y_{30} ; \underline{\pi}) = \prod_{i=1}^{n} \pi_i^{y_i} (1-\pi_i)^{1-y_i}$

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

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

$$= \begin{cases} \dfrac{\partial \ell(\beta)}{\partial \beta_1} = \displaystyle\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \frac{1}{1+e^{\beta_1+\beta_2 x_i}} \cdot e^{\beta_1+\beta_2 x_i} \\ \dfrac{\partial \ell(\beta)}{\partial \beta_2} = \displaystyle\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} \frac{1}{1+e^{\beta_1+\beta_2 x_i}} \cdot e^{\beta_1+\beta_2 x_i} \cdot x_i \end{cases}$$

The fitted model is

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

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

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

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

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

when $x_i = 0$ the odds are

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

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

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

when $x_i = 1$ the odds are

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

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

Finally, the odds ratio is

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

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

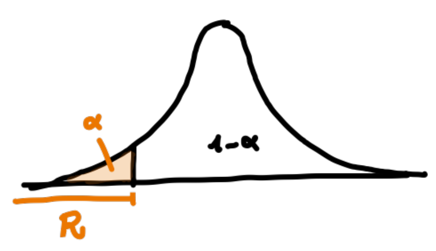
c) $\begin{cases} H_0 : \beta_2 = -1 \\ H_1 : \beta_2 < -1 \end{cases}$

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 $\alpha$, 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:

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

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

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

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