EXERCISE 2

Logistic regression : I denote the observed binary indicator of heart disease of individual i as y_i $i = 1, \dots, n$ $(n = 10\,000)$

1) model specification and assumptions

- $Y_i \sim \text{Bernoulli}(\pi_i)$ independent $i = 1, \dots, n$
- $\eta_i = \underline{\tilde{x}}_i^T \underline{\beta} = \beta_1 + \beta_2 \, \text{coffee}_i + \beta_3 \, \text{fastfood}_i + \beta_4 \, \text{income}_i$

  where $\text{coffee}_i = \begin{cases} 1 & \text{if individual } i \text{ drinks coffee} \\ 0 & \text{otherwise} \end{cases}$

- $\log \frac{\pi_i}{1-\pi_i} = \eta_i \iff \pi_i = \frac{e^{\eta_i}}{1+e^{\eta_i}}$ logit link

2) Estimate of the intercept:

$$z^{obs} = \frac{\hat{\beta}_1}{\hat{se}(\hat{\beta}_1)} \implies \hat{\beta}_1 = z^{obs} \cdot \hat{se}(\hat{\beta}_1) = 0.4923 \cdot (-22.08) = -10.87$$

z value and pvalue of coffee$_i$

$$z^{obs} = \frac{\hat{\beta}_2}{\hat{se}(\hat{\beta}_2)} = \frac{-0.6468}{0.2363} = -2.7372$$

$$\text{pvalue} = \mathbb{P}_{H_0}(|Z| > |z^{obs}|) = 2 \cdot \mathbb{P}_{H_0}(Z > |z^{obs}|) = 2 \cdot (1 - \Phi(2.7372))$$

from the table

$$0.995 < \Phi(2.7372) < 0.9975$$

$$0.0025 < 1 - \Phi(2.7372) < 0.005$$

$$0.005 < \text{pvalue} < 0.010$$

3) We first compute the estimated linear predictor $\hat{\eta}_i$

$$\hat{\eta}_i = \underline{\tilde{x}}_i^T \underline{\hat{\beta}} =$$

$$= \hat{\beta}_1 + \hat{\beta}_2 \cdot \underset{\underset{1}{\shortparallel}}{\text{coffee}_i} + \hat{\beta}_3 \, \underset{\underset{1000}{\shortparallel}}{\text{fastfood}_i} + \hat{\beta}_4 \, \underset{\underset{20\,000}{\shortparallel}}{\text{income}_i}$$

$$= -10.87 - 0.6468 + 0.0023 \cdot 1000 + 0$$

$$= -8.63$$

then, $\hat{\pi}_i = \frac{e^{\eta_i}}{1+e^{\eta_i}} = 0.000178$

4) In this case

$Y_i \sim \text{Ber}(\pi_i)$ indep

with $\pi_i = \frac{e^{\eta_i}}{1+e^{\eta_i}}$ and $\eta_i = \beta_0$

hence $\pi_i = \frac{e^{\beta_0}}{1+e^{\beta_0}}$ constant for all individuals

The MLE is $\tilde{\pi}_i = \bar{y}$ proportion of heart disease in the population

$$\implies \tilde{\beta}_0 = \log \frac{\bar{y}}{1-\bar{y}}$$

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

I use the likelihood ratio test

$$W = 2(\hat{\ell}(\text{model A}) - \tilde{\ell}(\text{model B})) \sim \chi^2_3 \text{ under } H_0$$

w^{obs} can be computed as

$$w^{obs} = D(\text{null}) - D(\text{model}) = 2920.6 - 1571.5 = 1349.1$$

I reject H_0.