

Practical 4: Logistic Regression Models

Comparison of Several Groups – Manual Implementation

Summer 2019: Question 2

A random sample of adults was classified according to their risk of contracting a certain disease. These adults were monitored for 10 years, after which, it was determined whether each adult had contracted the disease. The results are as follows:

Disease	Low Risk	Medium Risk	High Risk
Yes	20	15	15
No	180	85	35
Total	200	100	50

- (i) Formally specify logistic regression models that can be used to test the hypothesis that the rates of contracting the disease are the same in the three risk categories. Define the linear predictor for the appropriate full and reduced models. Use High Risk as the reference category. (15 marks)
- (ii) Calculate the maximum likelihood estimates of the probability of contracting the disease in each of the risk categories and overall. (5 marks)
- (iii) Given that the deviance for a Generalised Linear Model fitted to Binomial response data using the canonical link can be expressed as:

$$2 \sum_{i=1}^n y_i \log\left(\frac{y_i}{\hat{y}_i}\right) + (n_i - y_i) \log\left(\frac{n_i - y_i}{n_i - \hat{y}_i}\right),$$

calculate the deviances of the full and reduced logistic regression models. Use these to test whether the rates of contracting the disease are the same in the three risk categories. (18 marks)

- (iv) Calculate the Deviance residual for the Low Risk category under the reduced model. (12 marks) **ST6018 students can ignore this part.**

Implementation in R: