STAT3015/4030/7030 Generalised Linear Modelling

Tutorial 4

1. A study was conducted to examine the relationship between age and blood pressure. The data for 54 healthy adult women are contained in the data file `BP.txt`, the first column containing ages and the second containing diastolic blood pressures.

   (a) Fit a simple linear regression to this data and examine the residuals versus predictor plot. What do you notice?

   (b) To account for heteroscedasticity, it is sometimes suggested to fit the model:

   $$\begin{aligned} Y_i =& \beta_0 + \beta_1 x_i + \epsilon_i \\ \epsilon_i \sim& \mathcal{N}(0, \sigma_i^2) \end{aligned}$$

   where $\sigma_i^2 = \sigma^2 / w_i^2$ and the w_i^2's being weights used to indicate how the variance of the data changes from data point to data point. For this data, we might try using $w_i = 1/\sqrt{x_i}$ or $w_i = 1/x_i$. Why might these be appropriate choices? Fit weighted regressions using both of these suggestions. How do the parameter estimates (including the estimate of σ^2) change? Plot the weighted residuals, $w_i e_i$, versus predictor values for each weighting scheme. Which set of weights seems more appropriate to this dataset?

   (c) Why must we use the weighted residuals rather than the ordinary residuals from the weighted regressions in our plots to assess which weights are the most appropriate? [HINT: Recall that the variance of the ordinary residuals, e_i, is approximately the same as the variance of the error random variable ϵ_i; namely σ_i^2.]

2. The potency of an anaesthetic agent is measured in terms of the minimum concentration at which at least 50% of patients exhibit no response to stimulation. Thirty patients are administered a particular anaesthetic at various predetermined concentrations for 15 minutes before a stimulus was applied. The response variable was simply an indicator as to whether the patient responded or not. A GLM model with binomial error structure and link function

$$p(x_i) = g^{-1}(\beta_0 + \beta_1 x_i)$$

was fit to predict the probability of response (p) given the level of anaesthetic (x_i) where g was either the probit, logistic, or complementary log-log function. The focus of the experiment which gathered this data was to find the concentration value x at which the probability of responding to the stimulus was 50%, i.e., to estimate the value of x which satisfies the equation $p(x_i) = 0.5$. The table below gives the coefficient estimates for the three different link functions:

	$\widehat{\beta}_0$	$\widehat{\beta}_1$
Probit	3.8579	−3.3245
Logit	6.4685	−5.5676
Comp log-log	3.7316	−3.6370

Estimate the 50%-response concentration level for each of the three different link functions. Does our choice of link function influence our estimate in this problem?