OPTIMAL DESIGNS FOR LOGISTIC REGRESSION WITH A FIRST ORDER LINEAR PREDICTOR

K.G. Russell[†1]

[1] *University of Wollongong, Wollongong, Australia*

[†] E-mail: *kgr@uow.edu.au*

Little is known about how best to collect binary data. This paper examines an experimental situation where the resulting data are easy to analyse, but where the data collection is difficult to design.

Consider a set of Bernoulli trials in which the probability of success, π, is related to a single explanatory variable, x. Suppose that the logit of π is a first order function of x; i.e., $\ln[\pi/(1-\pi)] = \eta_x = \beta_0 + \beta_1 x$. The quantity η_x is called the *linear* (in the parameters) *predictor.* We wish to estimate the values of β_0 and β_1. Observations will be taken at n values of x ("support points"). Suppose that $m_i\,(>0)$ trials are conducted at $x = x_i$ $(i = 1, \ldots, n)$, where $x_1 < \cdots < x_n$ and $m_1 + \cdots + m_n = M$. For preliminary estimates of β_0 and β_1 and specified values of n and M, we wish to determine a design (the values of $x_1, \ldots, x_n$, $m_1, \ldots m_n$) that optimizes the information obtained from the data.

As the Maximum Likelihood estimators of β_0 and β_1 have some undesirable properties, we use the Maximum Penalized Likelihood (MPL) estimators (Firth, *Biometrika,* 1993, pp. 27–38). We examine various properties of the MPL estimators, and suggest the Integrated Mean Square Error as an appropriate measure of the optimality of a proposed design. Examples of optimal designs will be given for various values of β_0, β_1, n and M.