Homework 4 – Part II – 40 points

Exercise J-2.4 (adopted from Jennrich 1995): The data below give the incidence of liver tumors in rats after exposure to various amounts of DDT:

Dose (ppm)	Animals Tested	Tumor Incidence
0	111	4
2	105	4
10	124	11
50	104	13
250	90	60

Assume that the number of tumor incidences follow a binomial distribution with incidence probability that satisfies an Armitage-Doll model

$$\pi = 1 - e^{-\beta_0 - \beta_1 x - \beta_2 x^2}$$

where x denotes the dose in ppm.

(a) [20 points] Write an R function to implement the iteratively reweighted least squares algorithm. Use the modified relative error as your convergence criteria with tolerance value of 10^{-6}. Use your function to obtain the maximum likelihood estimates for $\beta_0, \beta_1, \beta_2$.

(b) [5 points] Plot the probability of positive response for doses ranging from 0 to 250, based on the model that you fit in Part (a). Superimpose this plot by the observed proportions, that is the ratio of number of incidents and the number tested.

Exercise J-2.5: A scientist studied the effects of a very strong storm in trees in a forest in Minnesota as a function of the tree diameters and the severity of winds in the areas where the trees were located. The data for $n = 659$ trees is given in the attached file `blowBF.txt`. This dataset consists of the following variables:

D = diameter of the tree
S = severity of the storm at the location of the tree (in the scale of 0 to 1, with 1 being most severe).
y = 0 if the tree survived and 1 if the tree died (blown down)

The goal is to use logistic regression to model the probability that a tree is blown down as a function of $\log(D)$, the logarithm of the diameter of the tree, and the severity of storm S, including an intercept. Note that here $y_1, \ldots, y_n$ is a set of n observations where $y_i \sim Bernouli(\pi_i)$, where $y_i = 1$ indicates a tree is blown down and $y_i = 0$ indicates a tree is standing. We model the probability of success π_i (the probability that a tree is blown own) by the logit function

$$\pi_i(\boldsymbol{\beta}) = \frac{\exp\,(\boldsymbol{x}_i^T\boldsymbol{\beta})}{1 + \exp\,(\boldsymbol{x}_i^T\boldsymbol{\beta})} \quad for\ i = 1, \ldots, n \qquad (1)$$

where $\boldsymbol{x}_i^T = (x_{i1}, x_{i2}, \dots, x_{ip})$ is a vector of p observed covariates for case i (in this problem 1 for the intercept, log (D), and S), and $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^T$ is a set of p variables to be estimated.

(a) [7 points] Use the iteratively reweighted least squares program that you wrote in the previous problem to obtain the maximum likelihood estimates for the parameters.
(b) [5 points] Plot the three-dimensional graph of $\pi(\boldsymbol{\beta})$, using your maximum likelihood estimates and as a function of $X_1 = S$ and $X_2 = \log(D)$, For your plot, X_1 should range in the interval (0,1) and X_2 should range in the range of observed values for log (D).
(c) [3 points] What does your model predict for the probability that a tree with a diameter of 10 located in an area with wind severity of 0.3 is blown down?