8 Data Science (djw1005)

(a) In a COVID vaccine trial, $n_0$ subjects were given a placebo and $n_1$ were given the vaccine; $x_0$ of the placebo subjects developed the disease and $x_1$ of the vaccinated subjects. Considering the probability model $X_k \sim \text{Binomial}(n_k, p_k)$, the vaccine efficacy is defined to be $e = 1 - p_1/p_0$.

(i) State the maximum likelihood estimators for $p_0$ and $p_1$. Give a formula for the maximum likelihood estimator for $e$. [2 marks]

(ii) Explain how to compute a 95% confidence interval for $e$. Also explain how to test whether $e > 0.5$. [7 marks]

(b) Further data about the trial has been made available, and we learn that subjects weren’t all enrolled for the same length of time. We are given a full dataset consisting of three features, the predictor variable $d_i$ and the response variables $(t_i, c_i)$ for subject $i$. Here $d_i = 1$ if the subject received the vaccine and $d_i = 0$ otherwise; $c_i = 1$ if the subject developed the disease and $c_i = 0$ otherwise; and $t_i$ is the day on which the subject developed the disease if $c_i = 1$, and the number of days enrolled in the trial otherwise.

Consider the following probability model. Among vaccinated subjects, the vaccine is effective with probability $f$ and ineffective otherwise. Effectively vaccinated subjects never get the disease. For ineffectively vaccinated subjects, and for subjects on placebo, each day there is a probability $q$ of developing the disease. The parameters $f$ and $q$ are unknown.

(i) For a subject $i$ who received placebo, give an expression for the likelihood of the pair $(t_i, c_i)$. [4 marks]

(ii) For a subject $i$ who received the vaccine, give an expression for the likelihood of the pair $(t_i, c_i)$. [4 marks]

(iii) Give an expression for the log likelihood of the entire dataset. [3 marks]

[Note: Your answers to this question should be symbolic. But you may like to know that in the low-dose part of the Oxford–AstraZeneca trial, $n_0 = 1374$, $n_1 = 1367$, $x_0 = 30$, and $x_1 = 3$.]