## 6 Data Science (djw1005)

A social scientist is studying social mobility. For each individual $i$ in the dataset, she has a record of their education level $e_i$, their income $q_i$, and their parents' income $p_i$. For simplicity she takes education level to be a binary variable, $e_i \in \{0, 1\}$. Treat these observations as independent samples of random variables $(E, Q, P)$.

(*a*) Propose a model for $E$ as a function of $P$, and explain how to fit it. Your model must treat $E$ as binary, not as real-valued. [6 marks]

(*b*) Propose a model for $Q$ as a function of $E$, and explain how to fit it. [4 marks]

(*c*) Propose a model for $Q$ as a function of both $P$ and $E$, and explain how to fit it. [4 marks]

(*d*) The social scientist believes that in an ideal world one's income should depend only on one's educational level. She wishes to test the hypothesis that this dataset is consistent with her ideal world. Explain how to test this hypothesis. Give pseudocode. [6 marks]