

Coursework 5: STAT 570

Philip Pham

October 28, 2018

1. Consider the data given in Table 1, which are a simplified version of those reported in Breslow and Day (1980). These data arose from a case-control study that was carried out to investigate the relationship between esophageal cancer and various risk factors. Disease status is denoted Y with $Y = 0$ and $Y = 1$ corresponding to without/with disease and alcohol consumption is represented by X with $X = 0$ and $X = 1$ denoting less than 80g and greater than or equal to 80g on average per day. Let the probabilities of high alcohol consumption in the cases and controls be denoted

$$p_1 = \mathbb{P}(X = 1 \mid Y = 1) \text{ and } p_2 = \mathbb{P}(X = 1 \mid Y = 0), \quad (1)$$

respectively.

$X = 0$	$X = 1$
---------	---------

Table 1: Case-control data: $Y = 1$ corresponds to the event of esophageal cancer, and $X = 1$ exposure to greater than 80g of alcohol per day. There are 200 cases and 775 controls.