BIOS 736: HW2

Thomas Hsiao

Problem 1

Table 1: Naive model fit on main study data

term	estimate	std.error	statistic	p.value
(Intercept)	0.210	0.117	1.794	0.076
w	0.156	0.093	1.678	0.097

The slope parameter in the naive regression is estimated to be 1.56 (95% CI: 1.37, 1.74) with a SE of 0.0927.

This is subjected to measurement error in the covariate W. We can implement a correction using the provided external validation data. The correction is as follows.

$$\hat{\beta} = \frac{\hat{\beta}^*}{\hat{\lambda}}, \hat{\text{Var}}\{\hat{\beta}\} = \frac{\hat{\text{Var}}\{\hat{\beta}^*\} + \hat{\beta}^2 \hat{\text{Var}}(\hat{\lambda})/\hat{\lambda}^2}{\hat{\lambda}^2}$$

Table 2: MEM fit on validation study data

term	estimate	std.error	statistic	p.value
(Intercept)	-0.010	0.064	-0.157	0.875
w	0.301	0.045	6.731	0.000

The corrected slope estimate is 0.517 (95% CI: -0.268, 1.30) with a SE of 0.401.

As is usually the case, the slope estimate for the surrogate results is attenuated - and the standard error underestimated. Once corrected, we get a much lower slope estimate with much more uncertainty - so much so that the CI crosses 0. This is a clear indication that measurement error can have a substantial impact on the results of any analysis, and the conclusions drawn can change drastically once accounted for.

Problem 2

See attached handwritten section.

Problem 3

The main contribution of Tang et al. (2015) is the extension of misclassification methods under relaxed assumptions to models with a binary response, a key binary exposure variable, and covariates influencing the measurement error using maximum likelihood (ML) and an internal validation design. The motivation stems from the HIV Epidemiology Research Study (HERS), where the main research question of interest is quantifying the association between bacterial vaginosis (BV) and Trichomoniasis (Trich). Non-differential error and simple differential error mechanisms of the dataset in the exposure and response is deemed to be an untenable assumption due to the complex misclassification process of BV and Trich. By using an internal validation set, the authors show large differences in estimates of the target parameter (odds ratio) between gold standard data, and the error-prone data, demonstrating the need for measurement error methodology. Barron (1977) first introduced the "matrix method" as a way to handle misclassification error in the exposure variable in a 2x2 table, and also extended it to error in the outcome assuming independent and nondifferential error. This paper extends the original method to inclusion of covariates associating exposure and outcome, as well as relaxing the independent and nondifferential error assumption to other scenarios (independent and differential, dependent and differential) using likelihood methods.