1. Suppose that you are in a context where the true linear regression model between an $n \times 1$ response $Y$ and predictors $X$ and $Z$ is

$$Y = X\beta + Z\gamma + v,$$

where $v$ is an $n \times 1$ error vector, with $E[v] = 0$ and $\text{Var}(v) = \sigma^2 I_{n \times n}$. However, you only measure $X$. Assuming incorrectly a model of the form

$$Y = X\beta + e,$$

you proceed to estimate $\beta$ via ordinary least squares (OLS), i.e., $\hat{\beta} = (X^TX)^{-1}X^TY$. Assume $\sigma^2$ is known.

(a) Provide conditions on $X$, $Z$, and $v$ such that, despite working with an incorrect model, the OLS estimate $\hat{\beta}$ defined above is unbiased under the true model, i.e., $E[\hat{\beta} | X] = \beta$.

(b) Suppose that $Z$ and $v$ are independent of each other. Provide an expression for $\text{Var}(\hat{\beta} | X)$ and discuss whether or not this condition is sufficient to make this variance equal to the usual OLS variance, i.e., equal to $\sigma^2(X^TX)^{-1}$.

(c) You decide to do a z-test for $H_0 : \beta_j = 0$, for one of the elements $\beta_j$ of your coefficient vector $\beta$. Specifically, you compare the statistic $z = \hat{\beta}_j / \text{se}(\hat{\beta}_j)$ to the 0.05 critical value of the standard normal distribution. Under the conditions that you established in (a), and the condition assumed in (b), and an assumption of a normal error distribution for $v$, do you expect the actual size of this test (i.e., the probability of Type I error) to be greater than, less than, or equal to 0.05.

(d) Randomization is sometimes used in designed studies in the following manner. A set of experimental conditions is established ahead of time, the various combinations of which are encoded in the rows of $X$. Then, subjects in the study are randomly assigned to the conditions. Effectively, this is equivalent to randomly assigning the values in $Y$, and their corresponding values in $Z$, to rows of $X$. Does this use of randomization satisfy the conditions of part (a)? Part (b)? Explain your answer.
2. Assume a standard regression model, for response $y$, given covariates $x$, i.e.,

$$y_i = x_i^T \beta + e_i,$$

where the $e_i$ are i.i.d. errors with mean zero and variance 1. A so-called $M$-estimator of $\beta$ is defined as

$$\hat{\beta}_M = \arg \min_{\beta} \sum_{i=1}^{n} \rho(y_i - x_i^T \beta),$$

(1)

where $\rho$ is a differentiable function of its argument. This type of estimator has been used in so-called robust regression, where the OLS choice of $\rho(u) = u^2$ is replaced by a function less sensitive to outliers. A classic example of such a function is Huber’s loss function,

$$\rho(u) = \begin{cases} u^2/2 & \text{if } u \leq c \\ cu - c^2/2 & \text{otherwise} \end{cases},$$

for some positive constant $c$.

(a) Differentiating the M-estimate criterion function in (1), and defining $u_i = y_i - x_i^T \beta$, show that the estimator $\hat{\beta}_M$ is a solution to the system of equations

$$\sum_{i=1}^{n} w(u_i) x_{ij} (y_i - x_i^T \beta) = 0, \quad \text{for } j = 1, \ldots, p,$$

(2)

where $w(u) = \frac{\rho'(u)}{u}$ and $p$ is the number of predictors in $x$.

(b) For the Huber loss function, show that

$$w(u) = \begin{cases} 1 & \text{if } u \leq c \\ c/u & \text{otherwise} \end{cases}.$$

(c) Note that the system of equations in (2) looks like the system of equations that must be solved when doing weighted least squares, if we ignore the fact that the values $w(u_i)$ depend on the value $\beta$ to be estimated. Exploiting this connection, provide an algorithm (written semi-formally, in pseudo-code) for how you might solve this system of equations in an iterative fashion. [Hint: The actual algorithm used in practice is called iteratively reweighted least squares (IRWLS).]

(d) Provide an expression for how you might think to estimate the variance of your estimate $\hat{\beta}_M$ produced by the algorithm you describe in part (c).