Biost/Stat 533
Emerson, Spr 09
Homework #2
April 8, 2009

Written problems to be handed in Wednesday, April 15.

1. Suppose $Y_i \sim (\mu_0, \sigma^2)$ for $i = 1, \ldots, n_0$ and $Y_i \sim (\mu_1, \sigma^2)$ for $i = n_0 + 1, \ldots, n = n_0 + n_1$, with $Cov(Y_i, Y_j) = 0$ for $i \neq j$. We are interested in estimating $\mu_1 - \mu_0$. For notational convenience, let $\vec{w}$ be an n-vector such that $w_i = 1$ for $1 \leq i \leq n_0$ and $w_i = 0$ otherwise, and let $\vec{z} = \vec{1}_n - \vec{w}$. (In all parts of this problem please provide formulas in terms of simple statistics, not matrix notation.)

   a. Using design matrix $\mathbf{X} = (\vec{1}_n \quad \vec{w})$, find the ordinary least squares estimator $\widehat{\vec{\beta}}$ for regression model $\vec{Y} = \mathbf{X}\vec{\beta} + \vec{\epsilon}$. Find vector $\vec{a}$ such that estimable function $\vec{a}^T\vec{\beta} = \mu_1 - \mu_0$, and provide the formula and mean and variance for $\vec{a}^T\widehat{\vec{\beta}}$).

   b. Using design matrix $\mathbf{X} = (\vec{1}_n \quad \vec{z})$, find the ordinary least squares estimator $\widehat{\vec{\beta}}$ for regression model $\vec{Y} = \mathbf{X}\vec{\beta} + \vec{\epsilon}$. Find vector $\vec{a}$ such that estimable function $\vec{a}^T\vec{\beta} = \mu_1 - \mu_0$, and provide the formula and mean and variance for $\vec{a}^T\widehat{\vec{\beta}}$.

   c. Using design matrix $\mathbf{X} = (\vec{w} \quad \vec{z})$, find the bordinary least squares estimator $\widehat{\vec{\beta}}$ for regression model $\vec{Y} = \mathbf{X}\vec{\beta} + \vec{\epsilon}$. Find vector $\vec{a}$ such that estimable function $\vec{a}^T\vec{\beta} = \mu_1 - \mu_0$, and provide the formula and mean and variance for $\vec{a}^T\widehat{\vec{\beta}}$.

   d. Using design matrix $\mathbf{X} = (\vec{1}_n \quad \vec{w} \quad \vec{z})$, find the ordinary least squares estimator $\widehat{\vec{\beta}}$ for regression model $\vec{Y} = \mathbf{X}\vec{\beta} + \vec{\epsilon}$. Find vector $\vec{a}$ such that estimable function $\vec{a}^T\vec{\beta} = \mu_1 - \mu_0$, and provide the formula and mean and variance for $\vec{a}^T\widehat{\vec{\beta}}$.

2. Let $\mathbf{X}$ (dimension $n \times p$) and $\mathbf{W}$ (dimension $n \times r$) be design matrices with the same range spaces (so $\mathcal{R}[\mathbf{X}] = \mathcal{R}[\mathbf{W}]$, where $\mathcal{R}[\mathbf{X}] = \{\vec{y} : \vec{y} = \mathbf{X}\vec{a},\ \vec{a} \in \mathcal{R}^p\}$ and $\mathcal{R}[\mathbf{W}] = \{\vec{y} : \vec{y} = \mathbf{W}\vec{a},\ \vec{a} \in \mathcal{R}^r\}$). Show that regression models $\vec{Y} = \mathbf{X}\vec{\beta} + \vec{\epsilon}$ and $\vec{Y} = \mathbf{W}\vec{\gamma} + \vec{\epsilon}$ are alternative parameterizations of each other. Furthermore show that if $\vec{a}^T\vec{\beta}$ is an estimable function, then there exists an estimable function $\vec{b}^T\vec{\gamma}$ such that estimates $\vec{a}^T\widehat{\vec{\beta}}$ and $\vec{b}^T\widehat{\vec{\gamma}}$ are equal for all $\vec{Y} \in \mathcal{R}^n$ and all least squares estimators $\widehat{\vec{\beta}}$ and $\widehat{\vec{\gamma}}$.

3. Suppose n-vector $\vec{\epsilon}$ has $E[\vec{\epsilon}] = \vec{0}$ and $Cov[\vec{\epsilon}] = \mathbf{V}$ with $rank(\mathbf{V}) = n$. Let $\widehat{\vec{\beta}} = (\mathbf{X}^T\mathbf{X})^{-}\mathbf{X}^T\vec{Y}$ be the ordinary least squares estimator of $\vec{\beta}$ and $\widehat{\vec{\beta}}_G = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}^T\mathbf{V}^{-1}\vec{Y}$ be the generalized least squares estimator of $\vec{\beta}$ in regression model $\vec{Y} = \mathbf{X}\vec{\beta} + \vec{\epsilon}$.

   a. Find the mean and variance of estimators $\vec{a}^T\widehat{\vec{\beta}}$ and $\vec{a}^T\widehat{\vec{\beta}}_G$ of estimable function $\vec{a}^T\beta$.

   b. Show that a best linear unbiased estimator of estimable function $\vec{a}^T\vec{\beta}$ is $\vec{a}^T\widehat{\vec{\beta}}_G$.

4. Consider again the setting of problem 1 in which $Y_i \sim (\mu_0, \sigma^2)$ for $i = 1, \ldots, n_0$ and $Y_i \sim (\mu_1, \sigma^2)$ for $i = n_0 + 1, \ldots, n = n_0 + n_1 = 2n_0$, except observations within each group are correlated. That is, we have $Cov(Y_i, Y_j) = \rho\sigma^2$ for $i, j = 1, \ldots, n_0; i \neq j$, $Cov(Y_i, Y_j) = \rho\sigma^2$ for $i, j = n_0 + 1, \ldots, n; i \neq j$, and $Cov(Y_i, Y_j) = 0$ for $i = 1, \ldots, n_0; j = n_0 + 1, \ldots, n$. For notational convenience, let $\vec{w}$ be an n-vector such that $w_i = 1$ for $1 \leq i \leq n_0$ and $w_i = 0$ otherwise, and let $\vec{z} = \vec{1}_n - \vec{w}$. Consider linear regression model $\vec{Y} = \mathbf{X}\vec{\beta} + \vec{\epsilon}$ with $\mathbf{X} = (\vec{w} \quad \vec{z})$. We are interested in estimating $\vec{a}^T\vec{\beta} = \mu_1 - \mu_0$.

   a. Show that the ordinary least squares estimator $\widehat{\vec{\beta}}$ is equal to the generalized least squares estimator $\widehat{\vec{\beta}}_G$. What is the mean and variance of these estimators?

b. Provide an estimate of the variance of $\widehat{\vec{\beta}}_G$ and $\vec{a}^T\widehat{\vec{\beta}}_G$ assuming that ρ is known.

c. Provide an estimate of the variance of $\widehat{\vec{\beta}}$ and $\vec{a}^T\widehat{\vec{\beta}}$ under the assumption that the observations are independent. How do they compare to the answers in b?

5. Now consider the setting of problem 4 in which $Y_i \sim (\mu_0, \sigma^2)$ for $i = 1, \ldots, n_0$ and $Y_i \sim (\mu_1, \sigma^2)$ for $i = n_0 + 1, \ldots, n = n_0 + n_1 = 2n_0$, except observations are paired across groups. That is, we have $Cov(Y_i, Y_i) = \sigma^2$ for $i = 1, \ldots, n$, $Cov(Y_i, Y_{n_0+i}) = \rho\sigma^2$ for $i = 1, \ldots, n_0$, and $Cov(Y_i, Y_j) = 0$ otherwise. For notational convenience, let $\vec{w}$ be an n-vector such that $w_i = 1$ for $1 \leq i \leq n_0$ and $w_i = 0$ otherwise, and let $\vec{z} = \vec{1}_n - \vec{w}$. Consider linear regression model $\vec{Y} = \mathbf{X}\vec{\beta} + \vec{\epsilon}$ with $\mathbf{X} = (\vec{w} \quad \vec{z})$. We are interested in estimating $\vec{a}^T\vec{\beta} = \mu_1 - \mu_0$.

a. Show that the ordinary least squares estimator $\widehat{\vec{\beta}}$ is equal to the generalized least squares estimator $\widehat{\vec{\beta}}_G$. What is the mean and variance of these estimators?

b. Provide an estimate of the variance of $\widehat{\vec{\beta}}_G$ and $\vec{a}^T\widehat{\vec{\beta}}_G$ assuming that ρ is known.

c. Provide an estimate of the variance of $\widehat{\vec{\beta}}$ and $\vec{a}^T\widehat{\vec{\beta}}$ under the assumption that the observations are independent. How do they compare to the answers in b?

d. How does the effect of correlated observations affect an ordinary least squares analysis differ when the correlated observations are within groups sharing the same predictor values versus when the correlated observations have different predictor values?