Biost/Stat 533
Emerson, Spr 09
Homework #2 Key
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 $\vec{a}^T\widehat{\vec{\beta}}$.

   Ans: For notational convenience define $\overline{Y}_0 = \sum_{i=1}^{n_0} Y_i/n_0$ and $\overline{Y}_1 = \sum_{i=n_0+1}^{n} Y_i/n_1$. By straightforward matrix operations we find

$$\mathbf{X}^T\mathbf{X} = \begin{pmatrix} \vec{1}_n^T\vec{1}_n & \vec{1}_n^T\vec{w} \\ \vec{w}^T\vec{1}_n & \vec{w}^T\vec{w} \end{pmatrix} = \begin{pmatrix} n_0 + n_1 & n_0 \\ n_0 & n_0 \end{pmatrix} \qquad (\mathbf{X}^T\mathbf{X})^{-1} = \begin{pmatrix} \frac{1}{n_1} & -\frac{1}{n_1} \\ -\frac{1}{n_1} & \frac{1}{n_0} + \frac{1}{n_1} \end{pmatrix}$$

$$\mathbf{X}^T\vec{Y} = \begin{pmatrix} n_0\overline{Y}_0 + n_1\overline{Y}_1 \\ n_0\overline{Y}_0 \end{pmatrix} \qquad \widehat{\vec{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\vec{Y} = \begin{pmatrix} \overline{Y}_1 \\ \overline{Y}_0 - \overline{Y}_1 \end{pmatrix}$$

   Taking the expectation of $\widehat{\vec{\beta}}$ we find

$$E[\widehat{\vec{\beta}}] = \begin{pmatrix} \mu_1 \\ \mu_0 - \mu_1 \end{pmatrix}$$

   Thus for $\vec{a} = (0 \quad -1)^T$, we have $\vec{a}^T\vec{\beta} = \mu_1 - \mu_0$. So

$$\begin{aligned} \vec{a}^T\widehat{\vec{\beta}} &= \overline{Y}_1 - \overline{Y}_0 \\ E[\vec{a}^T\widehat{\vec{\beta}}] &= \vec{a}^T E[\widehat{\vec{\beta}}] = \mu_1 - \mu_0 \\ Var[\vec{a}^T\widehat{\vec{\beta}}] &= \vec{a}^T Var[\widehat{\vec{\beta}}]\vec{a} = \sigma^2\vec{a}^T(\mathbf{X}^T\mathbf{X})^{-1}\vec{a} = \sigma^2(\frac{1}{n_0} + \frac{1}{n_1}) \end{aligned}$$

   (Note that this is the estimate and variance for the two sample Z test.)

   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 $\vec{a}^T\widehat{\vec{\beta}}$.

   Ans: By straightforward matrix operations we find

$$\mathbf{X}^T\mathbf{X} = \begin{pmatrix} \vec{1}_n^T\vec{1}_n & \vec{1}_n^T\vec{z} \\ \vec{z}^T\vec{1}_n & \vec{z}^T\vec{z} \end{pmatrix} = \begin{pmatrix} n_0 + n_1 & n_1 \\ n_1 & n_1 \end{pmatrix} \qquad (\mathbf{X}^T\mathbf{X})^{-1} = \begin{pmatrix} \frac{1}{n_0} & -\frac{1}{n_0} \\ -\frac{1}{n_0} & \frac{1}{n_0} + \frac{1}{n_1} \end{pmatrix}$$

$$\mathbf{X}^T\vec{Y} = \begin{pmatrix} n_0\overline{Y}_0 + n_1\overline{Y}_1 \\ n_1\overline{Y}_1 \end{pmatrix} \qquad \widehat{\vec{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\vec{Y} = \begin{pmatrix} \overline{Y}_0 \\ \overline{Y}_1 - \overline{Y}_0 \end{pmatrix}$$

Taking the expectation of $\widehat{\vec{\beta}}$ we find

$$E[\widehat{\vec{\beta}}] = \begin{pmatrix} \mu_0 \\ \mu_1 - \mu_0 \end{pmatrix}$$

Thus for $\vec{a} = (0 \quad 1)^T$, we have $\vec{a}^T \vec{\beta} = \mu_1 - \mu_0$. So

$$\begin{aligned} \vec{a}^T \widehat{\vec{\beta}} &= \overline{Y}_1 - \overline{Y}_0 \\ E[\vec{a}^T \widehat{\vec{\beta}}] &= \vec{a}^T E[\widehat{\vec{\beta}}] = \mu_1 - \mu_0 \\ Var[\vec{a}^T \widehat{\vec{\beta}}] &= \vec{a}^T Var[\widehat{\vec{\beta}}] \vec{a} = \sigma^2 \vec{a}^T (\mathbf{X}^T \mathbf{X})^{-1} \vec{a} = \sigma^2 (\frac{1}{n_0} + \frac{1}{n_1}) \end{aligned}$$

(Note again that this is the estimate and variance for the two sample Z test.)

c. Using design matrix $\mathbf{X} = (\vec{w} \quad \vec{z})$, 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 $\vec{a}^T \widehat{\vec{\beta}}$.

Ans: By straightforward matrix operations we find

$$\mathbf{X}^T \mathbf{X} = \begin{pmatrix} \vec{w}^T \vec{1}_n & \vec{w}^T \vec{z} \\ \vec{z}^T \vec{w} & \vec{z}^T \vec{z} \end{pmatrix} = \begin{pmatrix} n_0 & 0 \\ 0 & n_1 \end{pmatrix} \qquad (\mathbf{X}^T \mathbf{X})^{-1} = \begin{pmatrix} \frac{1}{n_0} & 0 \\ 0 & \frac{1}{n_1} \end{pmatrix}$$

$$\mathbf{X}^T \vec{Y} = \begin{pmatrix} n_0 \overline{Y}_0 \\ n_1 \overline{Y}_1 \end{pmatrix} \qquad \widehat{\vec{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \vec{Y} = \begin{pmatrix} \overline{Y}_0 \\ \overline{Y}_1 \end{pmatrix}$$

Taking the expectation of $\widehat{\vec{\beta}}$ we find

$$E[\widehat{\vec{\beta}}] = \begin{pmatrix} \mu_0 \\ \mu_1 \end{pmatrix}$$

Thus for $\vec{a} = (-1 \quad 1)^T$, we have $\vec{a}^T \vec{\beta} = \mu_1 - \mu_0$. So

$$\begin{aligned} \vec{a}^T \widehat{\vec{\beta}} &= \overline{Y}_1 - \overline{Y}_0 \\ E[\vec{a}^T \widehat{\vec{\beta}}] &= \vec{a}^T E[\widehat{\vec{\beta}}] = \mu_1 - \mu_0 \\ Var[\vec{a}^T \widehat{\vec{\beta}}] &= \vec{a}^T Var[\widehat{\vec{\beta}}] \vec{a} = \sigma^2 \vec{a}^T (\mathbf{X}^T \mathbf{X})^{-1} \vec{a} = \sigma^2 (\frac{1}{n_0} + \frac{1}{n_1}) \end{aligned}$$

(Note again that this is the estimate and variance for the two sample Z test.)

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 $\vec{a}^T \widehat{\vec{\beta}}$.

Ans: In this case, the design matrix is not of full rank, because $\vec{w} + \vec{z} = \vec{1}_n$. We thus have to find a LSE either by eliminating a linearly dependent column, augmenting the design matrix with a constraint, or using a generalized inverse. As the last method is commonly effected using one of the first two, I will only demonstrate those first two methods.

Since any two columns of $\mathbf{X}$ are linearly independent, I decide to eliminate the third column. Thus I can use the results of part (a) to find the estimates $\hat{\beta}_0 = \overline{Y}_1$ and $\hat{\beta}_1 = \overline{Y}_0 - \overline{Y}_1$, and I set $\hat{\beta}_2 = 0$. We can define the generalized inverse

$$(\mathbf{X}^T \mathbf{X})^- = \begin{pmatrix} \frac{1}{n_1} & -\frac{1}{n_1} & 0 \\ -\frac{1}{n_1} & \frac{1}{n_0} + \frac{1}{n_1} & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Taking the expectation of $\widehat{\vec{\beta}}$ we find

$$E[\widehat{\vec{\beta}}] = \begin{pmatrix} \mu_1 \\ \mu_0 - \mu_1 \end{pmatrix}$$

Thus for $\vec{a} = (0 \quad -1 \quad 0)^T$, we have $\vec{a}^T\vec{\beta} = \mu_1 - \mu_0$. So

$$\begin{aligned} \vec{a}^T\widehat{\vec{\beta}} &= \overline{Y}_1 - \overline{Y}_0 \\ E[\vec{a}^T\widehat{\vec{\beta}}] &= \vec{a}^T E[\widehat{\vec{\beta}}] = \mu_1 - \mu_0 \\ Var[\vec{a}^T\widehat{\vec{\beta}}] &= \vec{a}^T Var[\widehat{\vec{\beta}}]\vec{a} = \sigma^2\vec{a}^T(\mathbf{X}^T\mathbf{X})^-\vec{a} = \sigma^2(\frac{1}{n_0} + \frac{1}{n_1}) \end{aligned}$$

An alternative approach to this problem would have been to use an identifiability constraint. One such constraint that might typically used in this situation is that $\hat{\beta}_1 + \hat{\beta}_2 = 0$. We thus create augmented vectors and matrices

$$\vec{U} = \begin{pmatrix} \vec{Y} \\ 0 \end{pmatrix} \qquad \mathbf{V} = \begin{pmatrix} \vec{1}_n & \vec{w} & \vec{z} \\ 0 & 1 & 1 \end{pmatrix}$$

By straightforward matrix operations we find

$$\mathbf{V}^T\mathbf{V} = \begin{pmatrix} \vec{1}_n^T\vec{1}_n & \vec{1}_n^T\vec{w} & \vec{1}_n^T\vec{z} \\ \vec{w}^T\vec{1}_n & \vec{w}^T\vec{w}+1 & \vec{w}^T\vec{z}+1 \\ \vec{z}^T\vec{1}_n & \vec{z}^T\vec{w}+1 & \vec{z}^T\vec{z}+1 \end{pmatrix} = \begin{pmatrix} n_0+n_1 & n_0 & n_1 \\ n_0 & n_0+1 & 1 \\ n_1 & 1 & n_1+1 \end{pmatrix}$$

$$(\mathbf{V}^T\mathbf{V})^{-1} = \begin{pmatrix} \frac{n+n_0n_1}{4n_0n_1} & \frac{n_1-n_0-n_0n_1}{4n_0n_1} & \frac{n_0-n_1-n_0n_1}{4n_0n_1} \\ \frac{n_1-n_0-n_0n_1}{4n_0n_1} & \frac{n+n_0n_1}{4n_0n_1} & \frac{n_0n_1-n}{4n_0n_1} \\ \frac{n_0-n_1-n_0n_1}{4n_0n_1} & \frac{n_0n_1-n}{4n_0n_1} & \frac{n+n_0n_1}{4n_0n_1} \end{pmatrix}$$

$$\mathbf{V}^T\vec{U} = \begin{pmatrix} n_0\overline{Y}_0 + n_1\overline{Y}_1 \\ n_0\overline{Y}_0 \\ n_1\overline{Y}_1 \end{pmatrix} \qquad \widehat{\vec{\beta}} = (\mathbf{V}^T\mathbf{V})^{-1}\mathbf{V}^T\vec{U} = \begin{pmatrix} \frac{\overline{Y}_0+\overline{Y}_1}{2} \\ \frac{\overline{Y}_0-\overline{Y}_1}{2} \\ \frac{\overline{Y}_1-\overline{Y}_0}{2} \end{pmatrix}$$

where to find the inverse of the 3 by 3 matrix I used the following result for the inverse of a symmetric partitioned matrix

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{D} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{A}^{-1} + \mathbf{F}\mathbf{E}^{-1}\mathbf{F}^T & -\mathbf{F}\mathbf{E}^{-1} \\ -\mathbf{E}^{-1}\mathbf{F}^T & \mathbf{E}^{-1} \end{pmatrix}$$

where $\mathbf{E} = \mathbf{D} - \mathbf{B}^T\mathbf{A}^{-1}\mathbf{B}$ and $\mathbf{F} = \mathbf{A}^{-1}\mathbf{B}$ (see Appendix A of Seber and Lee). Taking the expectation of $\widehat{\vec{\beta}}$ we find

$$E[\widehat{\vec{\beta}}] = \begin{pmatrix} (\mu_0+\mu_1)/2 \\ (\mu_0-\mu_1)/2 \\ (\mu_1-\mu_0)/2 \end{pmatrix}$$

Thus for $\vec{a} = (0 \quad -1 \quad 1)^T$, we have $\vec{a}^T\vec{\beta} = \mu_1 - \mu_0$ (note that $\vec{a}$ must be in the range space of $\mathbf{X}^T$). So

$$\begin{aligned} \vec{a}^T\widehat{\vec{\beta}} &= \overline{Y}_1 - \overline{Y}_0 \\ E[\vec{a}^T\widehat{\vec{\beta}}] &= \vec{a}^T E[\widehat{\vec{\beta}}] = \mu_1 - \mu_0 \\ Var[\vec{a}^T\widehat{\vec{\beta}}] &= \vec{a}^T Var[\widehat{\vec{\beta}}]\vec{a} = \sigma^2\vec{a}^T(\mathbf{V}^T\mathbf{V})^{-1}\vec{a} = \sigma^2(\frac{1}{n_0} + \frac{1}{n_1}) \end{aligned}$$

(Note again that this is the estimate and variance for the two sample Z test.)

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 $\widehat{\vec{\gamma}}$.

Ans: $\widehat{\vec{\beta}}$ and $\widehat{\vec{\gamma}}$ are LSE, so we know

$$\mathbf{X}^T\mathbf{X}\widehat{\vec{\beta}} = \mathbf{X}^T\vec{Y}$$
$$\mathbf{W}^T\mathbf{W}\widehat{\vec{\gamma}} = \mathbf{W}^T\vec{Y}$$

Because $\mathbf{X}$ and $\mathbf{W}$ have the same range space, we know there exist matrices $\mathbf{A}$ and $\mathbf{B}$ (not necessarily unique if $\mathbf{X}$ and $\mathbf{W}$ are not full rank) such that $\mathbf{W} = \mathbf{XA}$ and $\mathbf{X} = \mathbf{WB}$. Hence we find

$$\mathbf{W}^T\mathbf{W}\widehat{\vec{\gamma}} = \mathbf{W}^T\vec{Y} = \mathbf{A}^T\mathbf{X}^T\vec{Y} = \mathbf{A}^T\mathbf{X}^T\mathbf{X}\widehat{\vec{\beta}} = \mathbf{W}^T\mathbf{X}\widehat{\vec{\beta}}$$
$$\mathbf{X}^T\mathbf{X}\widehat{\vec{\beta}} = \mathbf{X}^T\vec{Y} = \mathbf{B}^T\mathbf{W}^T\vec{Y} = \mathbf{B}^T\mathbf{W}^T\mathbf{W}\widehat{\vec{\gamma}} = \mathbf{X}^T\mathbf{W}\widehat{\vec{\gamma}}$$

and from this we find

$$\mathbf{W}^T(\mathbf{W}\widehat{\vec{\gamma}} - \mathbf{X}\widehat{\vec{\beta}}) = \vec{0}$$
$$\mathbf{X}^T(\mathbf{W}\widehat{\vec{\gamma}} - \mathbf{X}\widehat{\vec{\beta}}) = \vec{0}$$

Now from $\widehat{\vec{\gamma}} = (\mathbf{W}^T\mathbf{W})^-\mathbf{W}^T\vec{Y}$ it is clear that $\widehat{\vec{\gamma}}$ is in the range space of $\mathbf{W}^T$, and similary $\widehat{\vec{\beta}}$ is in the range space of $\mathbf{X}^T$, which by hypothesis is equal to the range space of $\mathbf{W}^T$. Thus we can conclude $\mathbf{W}\widehat{\vec{\gamma}} - \mathbf{X}\widehat{\vec{\beta}} = \vec{0}$ and $\mathbf{W}\widehat{\vec{\gamma}} = \mathbf{X}\widehat{\vec{\beta}}$, and the two regression models are alternative parameterizations of each other. Furthermore, because $\mathbf{W} = \mathbf{XA}$ and $\mathbf{X} = \mathbf{WB}$,

$$\mathbf{W}^T\mathbf{W}(\widehat{\vec{\gamma}} - \mathbf{B}\widehat{\vec{\beta}}) = \vec{0}$$
$$\mathbf{X}^T\mathbf{X}(\mathbf{A}\widehat{\vec{\gamma}} - \widehat{\vec{\beta}}) = \vec{0}$$

and $\widehat{\vec{\gamma}} = \mathbf{B}\widehat{\vec{\beta}}$ and $\widehat{\vec{\beta}} = \mathbf{A}\widehat{\vec{\gamma}}$ for some matrices $\mathbf{B}$ and $\mathbf{A}$. Hence for estimable function $\vec{a}^T\vec{\beta}$, $\vec{b}^T\vec{\gamma} = \vec{a}^T\mathbf{A}\vec{\gamma}$ is estimable.

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 $\vec{a}^T\beta$.

   Ans: Using the laws of expectation we have

$$E[\vec{a}^T\widehat{\vec{\beta}}] = \vec{a}^T(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^TE[\vec{Y}] = \vec{a}^T(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{X}\vec{\beta}$$

Now because $\vec{a}^T\beta$ is estimable, we know by Proposition II.A.10 that there exists a vector $\vec{b} \in \mathcal{R}^n$ such that $\vec{a}^T = \vec{b}^T\mathbf{X}$. Furthermore, from the definition of a generalized inverse we know $\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{X} = \mathbf{X}^T\mathbf{X}$, so $\mathbf{X}(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{X} = \mathbf{X}$. Thus $\vec{a}^T(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{X}\vec{\beta} = \vec{b}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{X}\vec{\beta} = \vec{b}^T\mathbf{X}\vec{\beta} = \vec{a}^T\vec{\beta}$.

Using the results for the covariance of a vector product, we have

$$Var(\vec{a}^T\widehat{\vec{\beta}}) = \vec{a}^T(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^TVar(\vec{Y})\mathbf{X}(\mathbf{X}^T\mathbf{X})^-\vec{a} = \sigma^2\vec{a}^T(\mathbf{X}^T\mathbf{X})^-\vec{a}$$

when $Var(\vec{Y}) = \sigma^2 \mathbf{I}_n$.

For the general case we have that $\widehat{\vec{\beta}}_G$ is the OLSE for transformed model $\vec{Z} = \mathbf{W}\vec{\beta} + \vec{\epsilon}^*$, where $\vec{Z} = \mathbf{V}^{-1/2}\vec{Y}$, $\mathbf{W} = \mathbf{V}^{-1/2}\mathbf{X}$, and $\vec{\epsilon}^* \sim (\vec{0}, \mathbf{I}_n)$. And under the results given above, thus in this transformed problem OLSE $\vec{a}^T\widehat{\vec{\beta}}_G$ of estimable function $\vec{a}^T\vec{\beta}$ has expectation $\vec{a}^T\vec{\beta}$ as given above. The variance is found to be

$$\begin{aligned}
Var(\vec{a}^T\widehat{\vec{\beta}}_G) &= \vec{a}^T(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}^T\mathbf{V}^{-1}Var(\vec{Y})\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-}\vec{a} \\
&= \vec{a}^T(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{V}\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-}\vec{a} \\
&= \sigma^2\vec{a}^T(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-}\vec{a}
\end{aligned}$$

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

Ans: We again consider the transformed problem in which $\vec{\epsilon}^* \sim (\vec{0}, \mathbf{I}_n)$. Then by Proposition II.C.13 in the class notes, $\vec{a}^T\widehat{\vec{\beta}}_G$ is unique for all $\vec{a} \in \mathcal{R}^p$. $\vec{a}^T\widehat{\vec{\beta}}_G$ is also unbiased as noted above. Let $\vec{b}^T\vec{Z}$ be any other unbiased estimator. So $E[\vec{b}^T\vec{Z}] = \vec{b}^T\mathbf{W}\vec{\beta} = \vec{a}^T\vec{\beta}$ and $b^T\mathbf{W} = a^T$. $Var(\vec{b}^T\vec{Z}) = \vec{b}^T\vec{b}$ and $Var(\vec{a}^T\widehat{\vec{\beta}}_G) = \vec{a}^T(\mathbf{W}^T\mathbf{W})^{-}\vec{a} = \vec{b}^T\mathbf{W}(\mathbf{W}^T\mathbf{W})^{-}\mathbf{W}^T\vec{b}$. So

$$Var(\vec{b}^T\vec{Z}) - Var(\vec{a}^T\widehat{\vec{\beta}}_G) = \vec{b}^T(\mathbf{I}_n - \mathbf{W}(\mathbf{W}^T\mathbf{W})^{-}\mathbf{W}^T)\vec{b} = \vec{b}^T(\mathbf{I}_n - \mathbf{P})\vec{b}$$

And $(\mathbf{I}_n - \mathbf{P})(\mathbf{I}_n - \mathbf{P}) = (\mathbf{I}_n - \mathbf{P})$ and symmetric, so

$$Var(\vec{b}^T\vec{Y}) - Var(\vec{a}^T\widehat{\vec{\beta}}_G) = \vec{d}^T\vec{d} \geq 0$$

with equality only if $\vec{d} = \vec{0}$, which corresponds to $\vec{b}^T\vec{Y} = \vec{a}^T\widehat{\vec{\beta}}_G$. (Note that this proof proceeds exactly like the case for a design matrix of full rank, and that we establish the BLUE optimality in the transformed setting.)

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?

Ans: The OLSE $\widehat{\vec{\beta}}$ is found from

$$\mathbf{X}^T\mathbf{X} = \begin{pmatrix} \vec{w}^T\vec{w} & \vec{w}^T\vec{z} \\ \vec{z}^T\vec{w} & \vec{z}^T\vec{z} \end{pmatrix} = \begin{pmatrix} n_0 & 0 \\ 0 & n_1 \end{pmatrix} \qquad (\mathbf{X}^T\mathbf{X})^{-1} = \begin{pmatrix} \frac{1}{n_0} & 0 \\ 0 & \frac{1}{n_1} \end{pmatrix}$$

$$\mathbf{X}^T\vec{Y} = \begin{pmatrix} n_0\overline{Y}_0 \\ n_1\overline{Y}_1 \end{pmatrix} \qquad \widehat{\vec{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\vec{Y} = \begin{pmatrix} \overline{Y}_0 \\ \overline{Y}_1 \end{pmatrix}$$

Taking the expectation of $\widehat{\vec{\beta}}$ we find

$$E[\widehat{\vec{\beta}}] = \begin{pmatrix} \mu_0 \\ \mu_1 \end{pmatrix}$$

The variance of $\widehat{\vec{\beta}}$ is found by

$$Var(\widehat{\vec{\beta}}) = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T Var(\vec{Y})\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}$$

$$= \begin{pmatrix} \frac{1}{n_0} & 0 \\ 0 & \frac{1}{n_1} \end{pmatrix} \sigma^2 \begin{pmatrix} n_0(1+(n_0-1)\rho) & 0 \\ 0 & n_1(1+(n_1-1)\rho) \end{pmatrix} \begin{pmatrix} \frac{1}{n_0} & 0 \\ 0 & \frac{1}{n_1} \end{pmatrix}$$

$$= \sigma^2 \begin{pmatrix} \frac{1+(n_0-1)\rho}{n_0} & 0 \\ 0 & \frac{1+(n_1-1)\rho}{n_1} \end{pmatrix}$$

To find the GLSE $\widehat{\vec{\beta}}_G$, we first consider the form of $\mathbf{V}^{-1}$. Let $\mathbf{R}_m$ be a $m \times m$ matrix with 1's on the diagonal ane ρ elsewhere, and $\mathbf{O}$ be a conformable matrix full of 0's. Then

$$\mathbf{V} = \sigma^2 \begin{pmatrix} \mathbf{R}_{n_0} & \mathbf{O} \\ \mathbf{O} & \mathbf{R}_{n_1} \end{pmatrix} \quad \text{and} \quad \mathbf{V}^{-1} = \frac{1}{\sigma^2} \begin{pmatrix} \mathbf{R}_{n_0}^{-1} & \mathbf{O} \\ \mathbf{O} & \mathbf{R}_{n_1}^{-1} \end{pmatrix}$$

where $\mathbf{R}_m^{-1}$ has the same symmetrical structure as $\mathbf{R}_m$. Let the diagonal elements of $\mathbf{R}_m^{-1}$ be equal to r and the off diagonal elements be equal to s. Then because $\mathbf{R}_m\mathbf{R}_m^{-1} = \mathbf{I}_m$ we have the simultaneous equations

$$1 = r + (m-1)s\rho$$
$$0 = r\rho + s + (m-2)s\rho$$

which can be solved to yield

$$r = \frac{1+(m-2)\rho}{1+(m-2)\rho-(m-1)\rho^2}$$
$$s = -\frac{\rho}{1+(m-2)\rho-(m-1)\rho^2}$$

Let r_0 and s_0 be the values of r and s when $m = n_0$, and r_1 and s_1 be the values of r and s when $m = n_1$. From this we can then find

$$\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X} = \frac{1}{\sigma^2} \begin{pmatrix} n_0(r_0+(n_0-1)s_0) & 0 \\ 0 & n_1(r_1+(n_1-1)s_1) \end{pmatrix}$$

$$(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1} = \sigma^2 \begin{pmatrix} \frac{1}{n_0(r_0+(n_0-1)s_0)} & 0 \\ 0 & \frac{1}{n_1(r_1+(n_1-1)s_1)} \end{pmatrix}$$

$$\mathbf{X}^T\mathbf{V}^{-1}\vec{Y} = \begin{pmatrix} n_0(r_0+(n_0-1)s_0)\overline{Y}_0 \\ n_1(r_1+(n_1-1)s_1)\overline{Y}_1 \end{pmatrix}$$

$$\widehat{\vec{\beta}}_G = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\vec{Y} = \begin{pmatrix} \overline{Y}_0 \\ \overline{Y}_1 \end{pmatrix}$$

which is the same as the OLSE, and thus has the same expectation and variance (you can check that $\sigma^2(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}$ gives the same answer as found above– it does).

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.

Ans: The variance of $\widehat{\vec{\beta}}_G$ is given above. In order to estimate $\mu_1 - \mu_0$, we are interested in estimating $\vec{a}^T\vec{\beta}$, where $\vec{a} = (-1\ 1)^T$. The variance of the GLSE for that estimable function is thus

$$Var(\vec{a}^T\widehat{\vec{\beta}}_G) = \vec{a}^T Var(\widehat{\vec{\beta}}_G)\vec{a} = \sigma^2 \left(\frac{1 + (n_0 - 1)\rho}{n_0} + \frac{1 + (n_1 - 1)\rho}{n_1} \right)$$

Now we know n_0, n_1, and (by assumption) ρ. Hence to estimate the variance, we only need estimate σ^2. It would seem logical to consider the residuals $\vec{e} = \vec{Y} - \mathbf{X}\widehat{\vec{\beta}}_G$, which owing to the unbiasedness of the GLSE would have distribution

$$\vec{e} \sim \left(\vec{0}, \mathbf{V} = \sigma^2 \begin{pmatrix} \mathbf{R}_{n_0} & \mathbf{O} \\ \mathbf{O} & \mathbf{R}_{n_1} \end{pmatrix} \right).$$

Owing to the correlation among the residuals, the sample variance of the residuals will not estimate σ^2 directly. But we can transform the residuals to independence, and then take the sample variance of those transformed residuals. To do this we find a transformation matrix $\mathbf{A}$ such that $\mathbf{AVA}^T = \sigma^2\mathbf{I}_n$. We can find such a matrix by considering the linear algebra result that says that every symmetric positive definite matrix $\mathbf{V}$ can be expressed as a product involving an invertible symmetric matrix $\mathbf{V} = \mathbf{V}^{1/2}\mathbf{V}^{1/2}$ (where the notation is obviously mnemonic). For our purposes, then, we would want to find matrices $\mathbf{R}_{n_0}^{1/2}$ and $\mathbf{R}_{n_1}^{1/2}$, and then define our "whitening" transformation (terminology out of signal process, where independent errors simulate white noise) as

$$\mathbf{A} = \begin{pmatrix} \mathbf{R}_{n_0}^{-1/2} & \mathbf{O} \\ \mathbf{O} & \mathbf{R}_{n_1}^{-1/2} \end{pmatrix},$$

where, again, we use the obvious notation that $\mathbf{R}_{n_0}^{-1/2}$ is the inverse of $\mathbf{R}_{n_0}^{1/2}$.

On approach to finding $\mathbf{R}_{n_0}^{1/2}$ is to guess that it will have a structure similar to $\mathbf{R}$ with some constant a on the diagonal and another constant b on the off-diagonals. Then we would have that

$$\begin{aligned} 1 &= a^2 + (n_0 - 1)b^2 \\ \rho &= 2ab + (n_0 - 1)b^2 \end{aligned}$$

which can be solved to find

$$\begin{aligned} a &= \frac{(n_0 - 1)\sqrt{1-\rho} \pm \sqrt{1 + (n_0 - 1)\rho}}{n} \\ b &= \frac{-\sqrt{1-\rho} \pm \sqrt{1 + (n_0 - 1)\rho}}{n}. \end{aligned}$$

(Note that either the plus or the minus will work.) We would then have

$$\mathbf{A}\vec{e} \sim \left(\vec{0}, \sigma^2\mathbf{I}_n \right),$$

and could use

$$\hat{\sigma}^2 = \frac{1}{n}\vec{e}^T\mathbf{A}^T\mathbf{A}\vec{e}$$

as a consistent estimate (where consistency comes from WLLN). (Note that our true usual practice would be to divide by $n-2$ in this problem, due to the 2 dimensional $\widehat{\vec{\beta}}_G$. This would give an unbiased estimate of σ^2.)

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?

Ans: When we assume $\rho = 0$, we obtain

$$Var(\widehat{\vec{\beta}}) = \sigma^2 \begin{pmatrix} \frac{1}{n_0} & 0 \\ 0 & \frac{1}{n_1} \end{pmatrix}$$

$$Var(\vec{a}^T\widehat{\vec{\beta}}) = \sigma^2 \left(\frac{1}{n_0} + \frac{1}{n_1} \right)$$

Note that for positive ρ, the true variance is greater than that which would be estimated when we assume $\rho = 0$. Thus in this case where the data within groups defined by predictors are positively correlated, inference based on the assumption of independence with the true value of σ^2 would be anti-conservative. Of course, if we presume independence of the observations, we would not transform the residuals to estimate σ^2. For the same vector of residuals, we can show that for $\rho > 0$ (and this is a limit to which the correlation can be negative within the "exchangeable" structure for the correlation within a group that we are considering here)

$$\vec{e}^T\vec{e} - \vec{e}^T\mathbf{A}^T\mathbf{A}\vec{e} < 0,$$

thus by incorrectly assuming independence, when having to estimate our nuisance parameter σ^2, we will also underestimate σ^2, thereby making our inference even more anti-conservative.

Note that the degree of error we make depends both on ρ and the sample size n_0 and n_1 within "clusters": The increase in variability over what might be obtained with independent observations depends on the product of sample size and correlation. Hence, even very small correlation in large clusters (e.g., hospitals, schools, cities) causes a problem and must be accounted for.

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 $1 \le i \le n_0$ 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?

   Ans: The OLSE $\widehat{\vec{\beta}}$ is the same as given in problem 4a, and the expectation is the same as was given in that answer. The variance of $\widehat{\vec{\beta}}$ is found from the results for $(\mathbf{X}^T\mathbf{X})^{-1}$ with $n_0 = n_1$

$$\mathbf{V} = Var(\vec{Y}) = \sigma^2 \begin{pmatrix} \mathbf{I}_{n_0} & \rho\mathbf{I}_{n_0} \\ \rho\mathbf{I}_{n_0} & \mathbf{I}_{n_0} \end{pmatrix}$$

$$\begin{aligned} Var(\widehat{\vec{\beta}}) &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T Var(\vec{Y})\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} \\ &= \begin{pmatrix} \frac{1}{n_0} & 0 \\ 0 & \frac{1}{n_0} \end{pmatrix} \sigma^2 \begin{pmatrix} n_0 & n_0\rho \\ n_0\rho & n_0 \end{pmatrix} \begin{pmatrix} \frac{1}{n_0} & 0 \\ 0 & \frac{1}{n_0} \end{pmatrix} \\ &= \sigma^2 \begin{pmatrix} \frac{1}{n_0} & \frac{\rho}{n_0} \\ \frac{\rho}{n_0} & \frac{1}{n_0} \end{pmatrix} \end{aligned}$$

To find the GLSE $\widehat{\vec{\beta}}_G$, we use the result for inverse of a symmetric partitioned matrix to find

$$\mathbf{V}^{-1} = \frac{1}{\sigma^2} \begin{pmatrix} \frac{1}{1-\rho^2}\mathbf{I}_{n_0} & \frac{-\rho}{1-\rho^2}\mathbf{I}_{n_0} \\ \frac{-\rho}{1-\rho^2}\mathbf{I}_{n_0} & \frac{1}{1-\rho^2}\mathbf{I}_{n_0} \end{pmatrix}$$

. From this we can then find

$$\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X} = \frac{1}{\sigma^2}\begin{pmatrix} \frac{n_0}{1-\rho^2} & -\frac{n_0\rho}{1-\rho^2} \\ -\frac{n_0\rho}{1-\rho^2} & \frac{n_0}{1-\rho^2} \end{pmatrix} \qquad (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1} = \sigma^2\begin{pmatrix} \frac{1}{n_0} & \frac{\rho}{n_0} \\ \frac{\rho}{n_0} & \frac{1}{n_0} \end{pmatrix}$$

$$\mathbf{X}^T\mathbf{V}^{-1}\vec{Y} = \frac{1}{\sigma^2}\begin{pmatrix} \frac{n_0}{1-\rho^2}\overline{Y}_0 - \frac{n_0\rho}{1-\rho^2}\overline{Y}_1 \\ \frac{n_0}{1-\rho^2}\overline{Y}_1 - \frac{n_0\rho}{1-\rho^2}\overline{Y}_0 \end{pmatrix} \qquad \widehat{\vec{\beta}}_G = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\vec{Y} = \begin{pmatrix} \overline{Y}_0 \\ \overline{Y}_1 \end{pmatrix}$$

which is the same as the OLSE, and thus has the same expectation and variance (you can check that $\sigma^2(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}$ gives the same answer as found above– it does).

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.

Ans: The variance of $\widehat{\vec{\beta}}_G$ is given above. In order to estimate $\mu_1 - \mu_0$, we are interested in estimating $\vec{a}^T\vec{\beta}$, where $\vec{a} = (-1\ 1)^T$. The variance of the GLSE for that estimable function is thus

$$Var(\vec{a}^T\widehat{\vec{\beta}}_G) = \vec{a}^T Var(\widehat{\vec{\beta}}_G)\vec{a} = \sigma^2\frac{2(1-\rho)}{n_0}$$

We again have to estimate σ^2, which we can effect by methods similar to those used for problem 4. We can also think about it quite simply in this case: The paired observations would allow us to note that $Y_i - Y_{n_0+i} \sim \left(\mu_0 - \mu_1, \sigma^2(2-2\rho)\right)$, so we could take the sample variance of the paired differences and obtain an unbiased estimate of $\sigma^2(2-2\rho)$, and then use the known value of ρ to solve for an unbiased estimate of σ^2.

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?

Ans: When we assume $\rho = 0$, we obtain

$$Var(\widehat{\vec{\beta}}) = \sigma^2\begin{pmatrix} \frac{1}{n_0} & 0 \\ 0 & \frac{1}{n_1} \end{pmatrix}$$

$$Var(\vec{a}^T\widehat{\vec{\beta}}) = \sigma^2\left(\frac{2}{n_0}\right)$$

Note that for positive ρ, the true variance is less than that which would be estimated when we assume $\rho = 0$. Furthermore, when making inference using an estimate of σ^2, incorrectly assuming independence rather than a true positive correlation would overestimate σ^2. Thus in this case when the correlated observations are sampled at different values of the covariate, inference based on the assumption of independence would be conservative, resulting in a substantial loss of statistical power.

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?

Ans: As noted above, when we consider a cluster of correlated observations of response, if the correlation among the predictors is of the same sign as the correlation among the errors within that cluster, the true variance tends to be greater than the variance estimated under independence, and tests and confidence intervals will be anti-conservative. On the other hand, if the correlation among the predictors within a cluster is of opposite sign of the correlation among the errors, then the true variance tends to be smaller than the variance estimated under independence.

So, for instance, in problem 4 the predictors in a cluster were positively correlated in the sense that the cluster had all the same values for the predictor. In that problem, when $\rho > 0$, the estimated variance was too small. However, if $\rho < 0$ in that problem, the variance estimated under independence is too large. But as noted in problem 4, there is a lower bound on how negative a common correlation may be for a specific sample size within clusters: For a cluster size of 2, any correlation is possible, for a cluster size of n, we must have $\rho > -1/(n-1)$.

In problem 5, the predictors in a cluster were negatively correlated in the sense that repeated observations within a cluster were for different values of the predictor. In that problem, when $\rho > 0$, the variance estimated under independence was too large. On the other hand, if $\rho < 0$ the variance estimated under independence was too small, thereby leading to anti-conservative testing.