Biost/Stat 533
Emerson, Spr 99
Homework #3 Key
April 28, 1999

1. Consider a linear regression model $\vec{Y} = \mathbf{X}\vec{\beta} + \vec{\epsilon}$ where the first column of the design matrix is filled with 1's (so we are fitting an intercept). Assume that $var(\vec{\epsilon}) = \sigma^2 \mathbf{I}_n$.

   a. Suppose $E[\epsilon_i] = \theta$ for $i = 1, \ldots, n$. What is the distribution for the OLS estimator $\widehat{\vec{\beta}}$? In particular, how does the assumption of nonzero mean for the errors alter the interpretation and distribution of the slope parameters $\beta_1, \ldots, \beta_{p-1}$?

   Ans: Under the above assumptions, we have that $E[\vec{Y}|\mathbf{X}] = \mathbf{X}\vec{\beta} + \theta\vec{1}_n$, where $\vec{1}_n$ is an n-vector containing all 1's, and $var(\vec{Y}|\mathbf{X}) = \sigma^2 \mathbf{I}_n$. To make this whole problem easier, we note that

$$\theta\vec{1}_n = \mathbf{X}\vec{\Delta} = \mathbf{X}\begin{pmatrix} \theta \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

   The OLSE $\widehat{\vec{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\vec{Y}$ thus has expectation

$$\begin{aligned} E[\widehat{\vec{\beta}}] &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T E[\vec{Y}] \\ &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{X}\vec{\beta} + \mathbf{X}\vec{\Delta}) \\ &= \vec{\beta} + \vec{\Delta} \end{aligned}$$

   Hence, we find that errors that do not have zero mean do not affect the expectation of $\hat{\beta}_1, \ldots, \hat{\beta}_{p-1}$ as the 2nd - $p$th elements of $\vec{\Delta}$ are 0. Furthermore, because the variance of a random variable is unaffected by adding a constant, the variance of the estimates is similarly unchanged.

   The interpretation of the slope parameter esimates is unchanged by such nonzero means for the error distribution.

   b. Let $\mathbf{X}^*$ be a design matrix derived from $\mathbf{X}$ by subtracting the corresponding column means from the elements in columns 2 through $p$. That is $X^*_{ij} = X_{ij} - \sum_{i=1}^{n} X_{ij}/n$. If we fit the regression model $\vec{Y} = \mathbf{X}^*\vec{\beta}^* + \vec{\epsilon}$, how does the OLS estimator $\widehat{\vec{\beta}}^*$ relate to $\widehat{\vec{\beta}}$ from the original problem. In particular, how does the interpretation and distribution of of each of the regression parameters change?

   Ans: This problem is easiest in matrix notation using partitioned matrices, though it can also be done by brute force. Consider the partitioning of $\mathbf{X}$ and $\mathbf{X}^*$ each into $n$ by 1 and $n$ by $p-1$ matrices

$$\mathbf{X} = (\,\vec{1}_n \quad \mathbf{W}\,) \qquad \mathbf{X}^* = (\,\vec{1}_n \quad \mathbf{W}^*\,)$$

   where $\mathbf{W}^* = \mathbf{W} - \frac{1}{n}\vec{1}_n\vec{1}_n^T\mathbf{W}$ has subtracted the means of the column of $\mathbf{W}$ from the elements in each corresponding column. We thus find

$$\mathbf{X}^T\mathbf{X} = \begin{pmatrix} \vec{1}_n^T \\ \mathbf{W}^T \end{pmatrix} (\,\vec{1}_n \quad \mathbf{W}\,) = \begin{pmatrix} n & \vec{1}_n^T\mathbf{W} \\ \mathbf{W}^T\vec{1}_n & \mathbf{W}^t\mathbf{W} \end{pmatrix}$$

   and

$$\mathbf{X}^{*T}\mathbf{X}^* = \begin{pmatrix} \vec{1}_n^T \\ ((\mathbf{I}_n - \frac{1}{n}\vec{1}_n\vec{1}_n^T)\mathbf{W})^T \end{pmatrix} (\,\vec{1}_n \quad (\mathbf{I}_n - \frac{1}{n}\vec{1}_n\vec{1}_n^T)\mathbf{W}\,) = \begin{pmatrix} n & \vec{0}_n^T \\ \vec{0}_n & \mathbf{W}^{*T}\mathbf{W}^* \end{pmatrix}$$

Now the inverse of a partitioned symmetric matrix can be given by

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

where $\mathbf{E} = \mathbf{D} - \mathbf{B}^T\mathbf{A}^{-1}\mathbf{B}$ and $\mathbf{F} = \mathbf{A}^{-1}\mathbf{B}$ (this is the form given by Seber, page 390). Now we merely need to compare $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\vec{Y}$ to $(\mathbf{X}^{*T}\mathbf{X}^*)^{-1}\mathbf{X}^{*T}\vec{Y}$. By straightforward matrix application of the above formula for the inverse and then matrix multiplication we find

$$(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T = \begin{pmatrix} \frac{1}{n}\vec{1}_n^T - \vec{1}_n^T\mathbf{W}(\mathbf{W}^T\mathbf{W} - \frac{1}{n}\mathbf{W}^T\vec{1}_n\vec{1}_n^T\mathbf{W})^{-1}\mathbf{W}^T(\mathbf{I}_n - \frac{1}{n}\vec{1}_n\vec{1}_n^T) \\ (\mathbf{W}^T\mathbf{W} - \frac{1}{n}\mathbf{W}^T\vec{1}_n\vec{1}_n^T\mathbf{W})^{-1}\mathbf{W}^T(\mathbf{I}_n - \frac{1}{n}\vec{1}_n\vec{1}_n^T) \end{pmatrix}$$

$$(\mathbf{X}^{*T}\mathbf{X}^*)^{-1}\mathbf{X}^{*T} = \begin{pmatrix} \frac{1}{n}\vec{1}_n^T \\ (\mathbf{W}^T\mathbf{W} - \frac{1}{n}\mathbf{W}^T\vec{1}_n\vec{1}_n^T\mathbf{W})^{-1}\mathbf{W}^T(\mathbf{I}_n - \frac{1}{n}\vec{1}_n\vec{1}_n^T) \end{pmatrix}$$

Note that in the above, the only difference is in the 1 by $n$ upper matrix in the partition. Thus, while $\hat{\beta}_0$ will not in general equal $\hat{\beta}_0^*$, we will have $\hat{\beta}_j = \hat{\beta}_j^*$ for $j = 1, \ldots, p-1$.

Now $var(\widehat{\vec{\beta}}) = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}$ and $var(\widehat{\vec{\beta}^*}) = \sigma^2(\mathbf{X}^{*T}\mathbf{X}^*)^{-1}$. Note that because the lower right matrix in the partition of $(\mathbf{X}^T\mathbf{X})^{-1}$

$$(\mathbf{W}^T\mathbf{W} - \frac{1}{n}\vec{1}_n^T\mathbf{W}\mathbf{W}^T\vec{1}_n)^{-1}$$

equals the lower right matrix in the partition of $(\mathbf{X}^{*T}\mathbf{X}^*)^{-1}$

$$(\mathbf{W}^{*T}\mathbf{W}^*)^{-1},$$

the covariance matrix for $(\hat{\beta}_1, \ldots, \hat{\beta}_{p-1})$ is equal to the covariance matrix for $(\hat{\beta}_1^*, \ldots, \hat{\beta}_{p-1}^*)$. So the estimated variances will also be the same for the slope estimates for the uncentered and the centered models.

The interpretation of the slope estimates is unchanged by centering, because the $j$th slope parameter continues to model the difference in means between two subjects who differ by one unit in their values of $X_j$ but are alike with respect to all other modelled covariates.

2. Consider a linear regression model relating response $\vec{Y}$ to an intercept and two predictor vectors $\vec{W}$ and $\vec{Z}$ (so design matrix $\mathbf{X} = (\vec{1}_n \quad \vec{W} \quad \vec{Z})$ has $X_{i1} \equiv 1$ for $i = 1, \ldots, n$ and $X_{i2} = W_i$ and $X_{i3} = Z_i$ and $\vec{\beta} = (\beta_0, \beta_1, \beta_2)^T$). Assume $E[\vec{\epsilon}] = \vec{0}$ and $var(\vec{\epsilon}) = \sigma^2\mathbf{I}_n$.

   a. Show that the correlation between OLS estimates $\hat{\beta}_1$ and $\hat{\beta}_2$ is opposite in sign to the sample correlation between $\vec{W}$ and $\vec{Z}$ and that the two slope estimates are uncorrelated if the sample correlation between $\vec{W}$ and $\vec{Z}$ is zero.

   Ans: I will work the first part of this problem in more generality, assuming $(p-1)$ covariates. Partition $\mathbf{X} = (\vec{1}_n \quad \mathbf{U})$ similar to problem 1. By problem 1, we can without loss of generality center the covariates to obtain $\mathbf{U}^* = (\mathbf{I}_n - \frac{1}{n}\vec{1}_n\vec{1}_n^T))\mathbf{U}$. Hence, the covariance between the $j$th and $k$th parameter estimates will the $(j,k)$th element of $\sigma^2(\mathbf{U}^{*T}\mathbf{U}^*)^{-1}$. Without loss of generality, assume $\sigma^2 = 1$. Due to the symmetry of the problem, it will be sufficient to consider the covariance between $\hat{\beta}_1$ and $\hat{\beta}_j$ for $j = 2, \ldots, p-1$. We thus further partition $\mathbf{U}^* = (\vec{W}^* \quad \mathbf{V}^*)$ into an $n$ by 1 matrix and an $n$ by $(p-2)$ matrix. Hence

$$\mathbf{U}^{*T}\mathbf{U}^* = \begin{pmatrix} \vec{W}^{*T}\vec{W} & \vec{W}^{*T}\mathbf{V}^* \\ \mathbf{V}^{*T}\vec{W}^* & \mathbf{V}^{*T}\mathbf{V}^* \end{pmatrix}$$

   Using the formula for the inverse of a partitioned matrix we find that the upper left matrix in the partition of $(\mathbf{U}^{*T}\mathbf{U}^*)^{-1}$ is

$$var(\hat{\beta}_1) = (\vec{W}^{*T}\vec{W}^*)^{-1} + (\vec{W}^{*T}\vec{W}^*)^{-1}\vec{W}^{*T}\mathbf{V}^*(\mathbf{V}^{*T}\mathbf{V}^* - \mathbf{V}^{*T}\vec{W}^*(\vec{W}^{*T}\vec{W}^*)^{-1}\vec{W}^{*T}\mathbf{V}^*)^{-1}\mathbf{V}^{*T}\vec{W}^*(\vec{W}^{*T}\vec{W}^*)^{-1}$$

and the upper right matrix in the partition of $(\mathbf{U}^{*T}\mathbf{U}^*)^{-1}$ is

$$cov(\hat{\beta}_1, (\hat{\beta}_2, \ldots, \hat{\beta}_{p-1})) = -(\vec{W}^{*T}\vec{W}^*)^{-1}\vec{W}^{*T}\mathbf{V}^*(\mathbf{V}^{*T}\mathbf{V}^* - \mathbf{V}^{*T}\vec{W}^*(\vec{W}^{*T}\vec{W}^*)^{-1}\vec{W}^{*T}\mathbf{V}^*)^{-1}$$

Now suppose $p = 3$ and $\mathbf{V}^* = \vec{Z}^*$. Then

$$\begin{aligned} \vec{W}^{*T}\vec{W}^* &= S_{WW} \\ \vec{W}^{*T}\mathbf{V}^* &= \vec{W}^{*T}\vec{Z}^* = S_{WZ} \\ \mathbf{V}^{*T}\mathbf{V}^* &= S_{ZZ} \end{aligned}$$

Letting $r_{WZ} = S_{WZ}/\sqrt{S_{WW}S_{ZZ}}$ be the sample correlation between $\vec{W}$ and $\vec{Z}$, we thus have

$$\begin{aligned} var(\hat{\beta}_1) &= \frac{1}{S_{WW}}\left(\frac{1}{1 - r_{WZ}^2}\right) \\ cov(\hat{\beta}_1, \hat{\beta}_2) &= \frac{-r_{WZ}}{(1 - r_{WZ}^2)\sqrt{S_{WW}S_{ZZ}}} \end{aligned}$$

By inspection, the covariance of $\hat{\beta}_1$ and $\hat{\beta}_2$ is opposite in sign to the sample correlation $r_{WZ}$, and it will only be zero if $\vec{W}$ and $\vec{Z}$ are uncorrelated.

b. Suppose we hold $S_{WW} = (\vec{W} - E[\vec{W}])^T(\vec{W} - E[\vec{W}])$, $S_{ZZ}$, and $\sigma^2$ constant, but we may freely vary $S_{WZ} = (\vec{W} - E[\vec{W}])^T(\vec{Z} - E[\vec{Z}])$. For what value of $S_{WZ}$ do we minimize the variance of $\hat{\beta}_1$ and $\hat{\beta}_2$? What does this suggest about our ability to test for an association between $Y$ and $W$ adjusting for $Z$ when $W$ and $Z$ are correlated?

Ans: From the results given above, it can be seen that the variance of $\hat{\beta}_1$ increases as the absolute value of the sample correlation between $\vec{W}$ and $\vec{Z}$ increases. Hence, we will have the greatest power to detect an association between $Y$ and $W$ when the sample correlation between $W$ and $Z$ is 0. This would be true on average if we sample in such a way that $W$ and $Z$ are independent (e.g., a completely randomized design), but we can obtain more efficient studies if we guarantee that $W$ and $Z$ are uncorrelated in our sample through experimental design.

This tendency for the standard error of a slope estimate to be increased by the modelling of a correlated variable is termed 'variance inflation'. Note that this effect exists even when the correlated variable does not predict the response. This in turn argues that adjusting for truly unimportant variables decreases the statistical power to detect associations between the response and other predictors.

3. Consider again the linear regression model in Problem 2 in which we will assume the true model is

$$\vec{Y} = \beta_0 + \vec{W}\beta_1 + \vec{Z}\beta_2 + \vec{\epsilon}$$

but we want to also consider fitting a model

$$\vec{Y} = \gamma_0 + \vec{W}\gamma_1 + \vec{\epsilon}^*$$

.

a. Under what conditions is the OLS estimate $\hat{\beta}_1$ equal to the OLS estimate $\hat{\gamma}_1$?

Ans: Suppose that $\vec{1}_n^T \vec{W} = \vec{1}_n^T \vec{Z} = 0$. (This can be achieved by centering the covariate vectors, and by problem 1, this does not affect the slope parameter estimates or distributions. Later we will consider the general case.) Define $\mathbf{W} = (\vec{1}_n \quad \vec{W})$ and $\mathbf{X} = (\mathbf{W} \quad \vec{Z})$. Then

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

and we want to find when $(0 \quad 1)\widehat{\vec{\gamma}} = (0 \quad 1 \quad 0)\widehat{\vec{\beta}}$. Following the approaches used above with $r = r_{WZ} = S_{WZ}/\sqrt{S_{WW}S_{ZZ}}$ we find

$$(\mathbf{W}^T\mathbf{W})^{-1} = \begin{pmatrix} \frac{1}{n} & 0 \\ 0 & \frac{1}{S_{WW}} \end{pmatrix}$$
$$(\mathbf{X}^T\mathbf{X})^{-1} = \begin{pmatrix} \frac{1}{n} & 0 & 0 \\ 0 & \frac{1}{S_{WW}(1-r^2)} & -\frac{r}{(1-r^2)\sqrt{S_{WW}S_{ZZ}}} \\ 0 & -\frac{r}{(1-r^2)\sqrt{S_{WW}S_{ZZ}}} & \frac{1}{S_{ZZ}(1-r^2)} \end{pmatrix}$$

Thus $\hat{\beta}_1 = \hat{\gamma}_1$ when

$$\frac{\vec{W}^T\vec{Y}}{(1-r^2)S_{WW}} - \frac{r\vec{Z}^T\vec{Y}}{(1-r^2)\sqrt{S_{WW}S_{ZZ}}} = \frac{\vec{W}^T\vec{Y}}{S_{WW}}$$

which in turn is satisfied if $r = 0$ or if

$$r = \sqrt{\frac{S_{WW}}{S_{ZZ}}}\frac{\vec{Z}^T\vec{Y}}{\vec{W}^T\vec{Y}}$$

Obviously, $\vec{Y}$ is random, and thus the second condition cannot be set by experimental design. We can set $r_{WZ} = 0$ by experimental design.

For arbitrary $\vec{W}$ and $\vec{Z}$, the above results obtain so long as the centered vectors have correlation 0. Of course, adding constants to vectors does not change their correlation, so for arbitrary $\vec{W}$ and $\vec{Z}$, $\hat{\gamma}_1 = \hat{\beta}_1$ so long as $S_{WZ}/\sqrt{S_{WW}S_{ZZ}} = 0$.

b. Under what conditions is the standard error of $\hat{\beta}_1$ equal to the standard error of $\hat{\gamma}_1$.

Ans: Now

$$var(\widehat{\vec{\beta}}) = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}, \text{ and}$$
$$var(\widehat{\vec{\gamma}}) = \tau^2(\mathbf{W}^T\mathbf{W})^{-1}$$

where $\sigma^2 = var(Y|W,Z)$ and $\tau^2 = var(Y|W) = \sigma^2 + \beta_2^2 var(Z|W)$. For the standard errors of $\hat{\beta}_2$ and $\hat{\gamma}_2$ to be equal, we must have

$$\sigma^2\frac{1}{(1-r^2)S_{WW}} = (\sigma^2 + \beta_2^2 var(Z|W))\frac{1}{S_{WW}}$$

This will be satisfied if $r = 0$ and $\beta_3 = 0$ or if $r = 0$ and $var(Z|W) = 0$. The above equation can also be satisfied by putting suitable restrictions on $var(Z|W) = \frac{r^2\sigma^2}{(\beta_2^2(1-r^2))}$ for nonzero $\beta_2$, but this is difficult to do by experimental design when $\beta_2$ is unknown.

c. Under what conditions is $\hat{\gamma}_1$ unbiased for $\beta_1$?

Ans:

$$\begin{aligned} E[\widehat{\vec{\gamma}}] &= (\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T E[\vec{Y}] \\ &= (\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\mathbf{X}\vec{\beta} \\ &= (\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T(\mathbf{W}(\beta_0 \quad \beta_1)^T + \vec{Z}\beta_2) \\ &= (\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\mathbf{W}(\beta_0 \quad \beta_1)^T + (\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\vec{Z}\beta_2) \\ &= (\beta_0 \quad \beta_1)^T + (\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\vec{Z}\beta_2) \end{aligned}$$

Hence, using our above results for the structure of $(\mathbf{W}^T\mathbf{W})^{-1}$, $\hat{\gamma}_1$ is unbiased for $\beta_1$ if only if $r_{WZ} = 0$ or $\beta_2 = 0$.

d. Under what conditions is $\hat{\gamma}_1$ BLUE for $\hat{\beta}_1$?

Ans: By Gauss-Markov theorem, $\widehat{\vec{\beta}}$ is BLUE for $\vec{\beta}$. Hence the only time that $\hat{\gamma}_1$ will be BLUE is when $\hat{\gamma}_1 = \hat{\beta}_1$ under the conditions of part (a.).

e. Suppose in particular that $\beta_1 = 0$ and $\beta_2 \neq 0$. What is the impact of this situation on the distribution of $\hat{\gamma}_1$, and how would $\hat{\gamma}_1$ compare to $\hat{\beta}_1$ from the full model? Compare this situation to the setting in which $\beta_2 = 0$ and $\beta_1 \neq 0$.

Ans: If $\beta_1 = 0$, $\beta_2 \neq 0$, and $r_{WZ} \neq 0$, the estimate $\hat{\gamma}_1$ will be biased towards finding an association between $Y$ and $W$ when there is truly none after conditioning on $Z$. $\hat{\beta}_1$ will tend to be close to zero, but $\hat{\gamma}_1$ will tend to be too large or too small depending upon the sign of $\beta_2$ and the sign of the correlation between $W$ and $Z$.

If $\beta_1 = 0$, $\beta_2 \neq 0$, and $r_{WZ} = 0$, the estimate $\hat{\gamma}_1$ will be unbiased for $\beta_1$. If $r_{WZ} = 0$ by design, the estimated standard error of $\hat{\gamma}_1$ will tend to be too large leading to confidence intervals that are too wide.

On the other hand, if $\beta_1 \neq 0$ and $\beta_2 = 0$, this is the situation where the smaller model provides regression estimates that are BLUE.

4. Consider a linear regression model $\vec{Y} = \mathbf{X}\vec{\beta} + \vec{\epsilon}$ where the first column of the design matrix is filled with 1's (so we are fitting an intercept). Consider adding an additional predictor $\vec{Z}$ to the model where, for some fixed $j$, $Z_i = 1$ if $i = j$ and $Z_i = 0$ otherwise. Let $\mathbf{X}^*$ be the augmented matrix in which the $(p+1)$th column is $\vec{Z}$, and consider fitting the regression model $\vec{Y} = \mathbf{X}^*\vec{\gamma} + \vec{\epsilon}^*$

a. How do the parameter estimates $\hat{\gamma}_0, \ldots, \hat{\gamma}_{p-1}$ differ from $\widehat{\vec{\beta}}$?

Ans: Without loss of generality, I consider the case of deleting the first case. To find $\widehat{\vec{\gamma}}$ I consider the partitioned matrix $\mathbf{X}^* = (\mathbf{X} \quad \vec{Z})$. Then letting $\vec{x}_1 = (\vec{Z}^T\mathbf{X})$ be the covariate vector for the first case, we have

$$\mathbf{X}^{*T}\mathbf{X}^* = \begin{pmatrix} \mathbf{X}^T\mathbf{X} & \vec{x}_1 \\ \vec{x}_1^T & 1 \end{pmatrix}$$

and using the formula for the inverse of a partitioned matrix given above, we find

$$(\mathbf{X}^{*T}\mathbf{X}^*)^{-1} = \begin{pmatrix} (\mathbf{X}^T\mathbf{X})^{-1} + \frac{(\mathbf{X}^T\mathbf{X})^{-1}\vec{x}_1\vec{x}_1^T(\mathbf{X}^T\mathbf{X})^{-1}}{(1-h_{11})} & -\frac{(\mathbf{X}^T\mathbf{X})^{-1}\vec{x}_1}{(1-h_{11})} \\ -\frac{\vec{x}_1^T(\mathbf{X}^T\mathbf{X})^{-1}}{(1-h_{11})} & \frac{1}{(1-h_{11})} \end{pmatrix}$$

where $h_{11} = \vec{x}_1^T(\mathbf{X}^T\mathbf{X})^{-1}\vec{x}_1$ is the first element on the diagonal of the hat matrix. Thus

$$\begin{aligned} \widehat{\vec{\gamma}} &= (\mathbf{X}^{*T}\mathbf{X}^*)^{-1}\mathbf{X}^{*T}\vec{Y} \\ &= \begin{pmatrix} \widehat{\vec{\beta}} - \frac{(\mathbf{X}^T\mathbf{X})^{-1}\vec{x}_1(\vec{Y}-\vec{x}_1^T\widehat{\vec{\beta}})}{(1-h_{11})} \\ \frac{(\vec{Y}-\vec{x}_1^T\widehat{\vec{\beta}})}{(1-h_{11})} \end{pmatrix} \end{aligned}$$

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

By inspection, then, adding the covariate $\vec{Z}$ to the model has $\hat{\beta}_j - \hat{\gamma}_j$ equal to the $j$th element of $(\mathbf{X}^T\mathbf{X})^{-1}\vec{x}_1(\vec{Y} - \vec{x}_1^T\widehat{\vec{\beta}})/(1 - h_{11})$.

It should be noted that $(\vec{x}_1^T \quad 1)\widehat{\vec{\gamma}} = Y_1$. Thus adding the covariate indicating a single case results in a model which predicts that case exactly.

b. How do the parameter estimates $\hat{\gamma}_0, \ldots, \hat{\gamma}_{p-1}$ differ from the estimates obtained by fitting the first model with the $j$th case deleted?

Ans: Note first that if we partition

$$\mathbf{X} = \begin{pmatrix} \vec{x}_1^T \\ \mathbf{W} \end{pmatrix} \qquad \vec{Y} = \begin{pmatrix} Y_1 \\ \vec{U} \end{pmatrix}$$

(so $\mathbf{W}$ contains rows 2 through $n$ of $\mathbf{X}$ and $\vec{U} = (Y_2, \ldots, Y_n)^T)$), then

$$\begin{aligned} \mathbf{X}^T\mathbf{X} &= \mathbf{W}^T\mathbf{W} + \vec{x}_1\vec{x}_1^T, \text{ so} \\ \mathbf{W}^T\mathbf{W} &= \mathbf{X}^T\mathbf{X} - \vec{x}_1\vec{x}_1^T, \text{ and} \\ \mathbf{X}^T\vec{Y} &= \vec{x}_1 Y_1 + \mathbf{W}^T\vec{U}. \end{aligned}$$

Now an alternative formula for the inverse of a partitioned symmetric matrix is given by

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

where $\mathbf{G} = \mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{B}^T$ and $\mathbf{F} = \mathbf{B}\mathbf{D}^{-1}$. (I derived this result when I did not have Seber handy and only vaguely remembered Seber's form. It is just an interchange of the rows and columns.) Using this formula and the above relations between $\mathbf{X}^T\mathbf{X}$ and $\mathbf{W}^T\mathbf{W}$, we find that $(\mathbf{X}^{*T}\mathbf{X}^*)^{-1}$, $\mathbf{X}^{*T}\vec{Y}$, and $\widehat{\vec{\gamma}}$ from part (a) can be written as

$$(\mathbf{X}^{*T}\mathbf{X}^*)^{-1} = \begin{pmatrix} (\mathbf{W}^T\mathbf{W})^{-1} & -(\mathbf{W}^T\mathbf{W})^{-1}\vec{x}_1 \\ -\vec{x}_1^T(\mathbf{W}^T\mathbf{W})^{-1} & 1 + \vec{x}_1^T(\mathbf{W}^T\mathbf{W})^{-1}\vec{x}_1 \end{pmatrix} \qquad \mathbf{X}^{*T}\vec{Y} = \begin{pmatrix} \mathbf{W}^T\vec{U} + \vec{x}_1 Y_1 \\ Y_1 \end{pmatrix}$$

and

$$\widehat{\vec{\gamma}} = \begin{pmatrix} (\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\vec{U} \\ Y_1 - \vec{x}_1^T(\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\vec{U} \end{pmatrix}$$

Thus we see that $(\hat{\gamma}_0, \ldots, \hat{\gamma}_{p-1})$ are exactly the OLS estimates that we would have obtained if the first case had been deleted from the dataset.

This result gives us computationally useful ways to compute the influence of individual cases: We can compute the change in the parameter estimates using the estimates from the full data case and the design matrix. We do not really have to fit separate regressions for every case deletion. I note, however, that we will not have such a result for other forms of regression. Furthermore, computing the difference in the P values is a little more difficult without actually fitting all the regressions.