240A Solutions to Midterm Exam Winter 2006

1.(a) We have $\widehat{y} = 1.0 + 0.5x$ giving $\widehat{y}_1 = 1$, $\widehat{y}_2 = 2$ and $\widehat{y}_3 = 3$.
$s^2 = \frac{1}{n-2}\sum_i (y_i - \widehat{y}_i)^2 = (1-0)^2 + (2-4)^2 + (3-2)^2 = 1+4+1 = 6$. So $s = \sqrt{6} = 2.4495$.

(b) $s^2_{\widehat{\beta}_2} = s^2/\sum_i (x_i - \bar{x})^2 = 6/[(0-2)^2 + (2-2)^2 + (4-2)^2] = 6/8 = 0.75$. So $s_{\widehat{\beta}_2} = \sqrt{0.75} = 0.8660$.

(c) A 95% CI is $\widehat{\beta}_2 \pm t_{n-2;0.025} \times s_{\widehat{\beta}_2} = 0.75 \pm 12.706 \times 0.8660 = 0.50 \pm 11.00 = (-10.5, 11.5)$.

(d) $R^2 = 1 - ResSS/TotSS = 1 - 6/8 = 0.25$,
since $ResSS = \sum_i (y_i - \widehat{y}_i)^2 = 6$ from part (a)
and $TotSS = \sum_i (y_i - \bar{y})^2 = [(0-2)^2 + (4-2)^2 + (2-2)^2] = 8$.

2.(a) Here

$$\begin{aligned}
\mathbf{b}'\mathbf{A}\mathbf{b} &= \begin{bmatrix} b_1 & b_2 \end{bmatrix}' \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} b_1 & b_2 \end{bmatrix}' \begin{bmatrix} a_{11}b_1 + a_{12}b_2 \\ a_{21}b_1 + a_{22}b_2 \end{bmatrix} \\
&= b_1 a_{11} b_1 + b_1 a_{12} b_2 + b_2 a_{21} b_1 + b_2 a_{22} b_2.
\end{aligned}$$

(b) By direct differentiation $\partial \mathbf{b}'\mathbf{A}\mathbf{b}/\partial b_1 = 2a_{11}b_1 + a_{12}b_2 + a_{21}b_1$
and $\partial \mathbf{b}'\mathbf{A}\mathbf{b}/\partial b_2 = a_{12}b_1 + a_{21}b_1 + 2a_{22}b_2$.

(c) Now $a_{12} = a_{21}$ so $\partial \mathbf{b}'\mathbf{A}\mathbf{b}/\partial b_1 = 2a_{11}b_1 + 2a_{12}b_2$ and $\partial \mathbf{b}'\mathbf{A}\mathbf{b}/\partial b_2 = 2a_{21}b_1 + 2a_{22}b_2$. Stacking

$$\frac{\partial \mathbf{b}'\mathbf{A}\mathbf{b}}{\partial \mathbf{b}} = \begin{bmatrix} 2a_{11}b_1 + 2a_{12}b_2 \\ 2a_{21}b_1 + 2a_{22}b_2 \end{bmatrix} = 2\mathbf{A}\mathbf{b},$$

given $\mathbf{Ab}$ which is the last matrix in the second line of the part (a) answer.

3. Aside: derivation of the transformed model (not necessary)

$$\begin{aligned}
\mathbf{y} &= \beta_1 \mathbf{l} + \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{u} = \beta_1 \mathbf{l} + \mathbf{l}\bar{\mathbf{x}}_2'\boldsymbol{\beta}_2 + \mathbf{X}_2\boldsymbol{\beta}_2 - \mathbf{l}\bar{\mathbf{x}}_2'\boldsymbol{\beta}_2 + \mathbf{u} \\
&= (\beta_1 + \bar{\mathbf{x}}_2'\boldsymbol{\beta}_2)\mathbf{l} + (\mathbf{X}_2 - \mathbf{l}\bar{\mathbf{x}}_2')\boldsymbol{\beta}_2 + \mathbf{u} \\
&= \alpha_1 \mathbf{l} + \mathbf{X}_2^*\boldsymbol{\beta}_2 + \mathbf{u}, \quad \text{where } \alpha_1 = \beta_1 + \bar{\mathbf{x}}_2'\boldsymbol{\beta}_2
\end{aligned}$$

(a) We have

$$\mathbf{y} = \begin{bmatrix} \mathbf{l} & \mathbf{X}_2^* \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \boldsymbol{\beta}_2 \end{bmatrix} + \mathbf{u} = \mathbf{Z}\boldsymbol{\gamma} + \mathbf{u}.$$

(b) Intuitively the sum of deviations from a mean is zero. Formally:

$$\mathbf{l}'\mathbf{X}_2^* = \mathbf{l}'(\mathbf{X}_2 - \mathbf{l}\bar{\mathbf{x}}_2') = \mathbf{l}'\mathbf{X}_2 - \mathbf{l}'\mathbf{l}\bar{\mathbf{x}}_2' = N\bar{\mathbf{x}}_2' - N\bar{\mathbf{x}}_2' = \mathbf{0},$$

where we used $\mathbf{l}'\mathbf{X}_2 = N\bar{\mathbf{x}}_2'$, since $\bar{\mathbf{x}}_2' = \frac{1}{N}\mathbf{l}'\mathbf{X}_2$, and $\mathbf{l}'\mathbf{l} = N$.

(c) The usual $(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}$ becomes

$$\begin{aligned}
\begin{bmatrix} \widehat{\alpha}_1 \\ \widehat{\boldsymbol{\beta}}_2 \end{bmatrix} &= \begin{bmatrix} \mathbf{l}'\mathbf{l} & \mathbf{X}_2^{*\prime}\mathbf{l} \\ \mathbf{l}'\mathbf{X}_2^{*\prime} & \mathbf{X}_2^{*\prime}\mathbf{X}_2^* \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{l}'\mathbf{y} \\ \mathbf{X}_2^{*\prime}\mathbf{y} \end{bmatrix} = \begin{bmatrix} N & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2^{*\prime}\mathbf{X}_2^* \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{l}'\mathbf{y} \\ \mathbf{X}_2^{*\prime}\mathbf{y} \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{N} & \mathbf{0} \\ \mathbf{0} & (\mathbf{X}_2^{*\prime}\mathbf{X}_2^*)^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{l}'\mathbf{y} \\ \mathbf{X}_2^{*\prime}\mathbf{y} \end{bmatrix} = \begin{bmatrix} \frac{1}{N}\mathbf{l}'\mathbf{y} \\ (\mathbf{X}_2^{*\prime}\mathbf{X}_2^*)^{-1}\mathbf{X}_2^{*\prime}\mathbf{y} \end{bmatrix} = \begin{bmatrix} \bar{y} \\ (\mathbf{X}_2^{*\prime}\mathbf{X}_2^*)^{-1}\mathbf{X}_2^{*\prime}\mathbf{y} \end{bmatrix}
\end{aligned}$$

(d) And the usual $s^2(\mathbf{Z}'\mathbf{Z})^{-1}$ becomes

$$\widehat{\mathrm{V}}\begin{bmatrix} \widehat{\alpha}_1 \\ \widehat{\boldsymbol{\beta}}_2 \end{bmatrix} = s^2 \begin{bmatrix} \mathbf{l}'\mathbf{l} & \mathbf{X}_2^{*\prime}\mathbf{l} \\ \mathbf{l}'\mathbf{X}_2^{*\prime} & \mathbf{X}_2^{*\prime}\mathbf{X}_2^* \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{l}'\mathbf{l} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2^{*\prime}\mathbf{X}_2^* \end{bmatrix}^{-1} = s^2 \begin{bmatrix} \frac{1}{N} & \mathbf{0} \\ \mathbf{0} & (\mathbf{X}_2^{*\prime}\mathbf{X}_2^*)^{-1} \end{bmatrix}.$$

4.(a) Now

$$\begin{aligned}
\widehat{y}_f &= \widehat{\alpha}_1 + \mathbf{x}^{*\prime}_{2f}\widehat{\boldsymbol{\beta}}_2 = \left[\begin{array}{cc} 1 & \mathbf{x}^{*\prime}_{2f} \end{array}\right] \left[\begin{array}{c} \widehat{\alpha}_1 \\ \widehat{\boldsymbol{\beta}}_2 \end{array}\right] = \mathbf{c}'\widehat{\boldsymbol{\gamma}} \\
\widehat{\mathrm{V}}[\widehat{y}_f] &= \left[\begin{array}{cc} 1 & \mathbf{x}^{*\prime}_{2f} \end{array}\right] \widehat{\mathrm{V}} \left[\begin{array}{c} \widehat{\alpha}_1 \\ \widehat{\boldsymbol{\beta}}_2 \end{array}\right] \left[\begin{array}{c} 1 \\ \mathbf{x}^{*}_{2f} \end{array}\right] \\
&= \left[\begin{array}{cc} 1 & \mathbf{x}^{*\prime}_{2f} \end{array}\right] s^2 \left[\begin{array}{cc} \frac{1}{N} & \mathbf{0} \\ \mathbf{0} & (\mathbf{X}^{*\prime}_2\mathbf{X}^{*}_2)^{-1} \end{array}\right] \left[\begin{array}{c} 1 \\ \mathbf{x}^{*}_{2f} \end{array}\right] = s^2\left[\frac{1}{N} + \mathbf{x}^{*\prime}_{2f}(\mathbf{X}^{*\prime}_2\mathbf{X}^{*}_2)^{-1}\mathbf{x}^{*}_{2f}\right].
\end{aligned}$$

(b) A 95 percent confidence interval for $\widehat{y}_f$ as a prediction of $\mathrm{E}[y|\mathbf{x}_2 = \mathbf{x}_{2f}]$ is

$$\widehat{y}_f \pm t_{.025;N-k} \times s\sqrt{\frac{1}{N} + \mathbf{x}^{*\prime}_{2f}(\mathbf{X}^{*\prime}_2\mathbf{X}^{*}_2)^{-1}\mathbf{x}^{*}_{2f}}.$$

(c) This is minimized when $\mathbf{x}_{2f} = \bar{\mathbf{x}}_2$, so that $\mathbf{x}^{*}_{2f} = \mathbf{0}$.
[Then the 95% confidence interval is simply $\widehat{y}_f \pm t_{.025;N-k} \times s/\sqrt{N}$.]

5. We have $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{u}) = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}$.

(a) So

$$\mathrm{E}[\widehat{\boldsymbol{\beta}}] = \boldsymbol{\beta} + \mathrm{E}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}] = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathrm{E}[\mathbf{u}] = \boldsymbol{\beta} \text{ as } \mathrm{E}[\mathbf{u}] = \mathbf{0}.$$

(b) And

$$\begin{aligned}
\mathrm{V}[\widehat{\boldsymbol{\beta}}] &= \mathrm{E}[(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'] \\
&= \mathrm{E}[((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u})((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u})'] \\
&= \mathrm{E}[((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\mathbf{u}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathrm{E}[\mathbf{u}\mathbf{u}']\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}
\end{aligned}$$

(c) $\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}$ is a linear combination of normals, so is normal distributed.

6.(a) Statistically significant at 5% are exper, educ, and nonwhite, as $p < 0.05$.

(b) Jointly statistically significant at 5% as overall $F = 13.71$ has $p = 0.000 < 0.05$.

(c) Test $H_0 : \beta_3 = 0$ and $\beta_4 = 0$ (where I call the intercept β_1). This is $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$ where

$$\mathbf{R} = \left[\begin{array}{cccccc} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{array}\right] \text{ and } \mathbf{r} = \left[\begin{array}{c} 0 \\ 0 \end{array}\right].$$

Form the Wald test statistic W and reject H_0 if $\mathrm{W} > F_{0.05}(2, 175)$ where

$$\mathrm{W} = (\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{r})\left[\mathbf{R}\widehat{\mathrm{V}}[\widehat{\boldsymbol{\beta}}]\mathbf{R}'\right]^{-1}(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{r})$$

(d) Test $H_0 : \beta_{educ} = 0.10$ against $\beta_{educ} \neq 0.10$.

$$t = \frac{\widehat{\beta} - \beta^*}{s_{\widehat{\beta}}} = \frac{0.08369 - 0.10}{0.01244} = \frac{-0.01631}{0.01244} = -1.311. \ |t| < t_{.025}(175) \simeq z_{.025} = 1.96.$$

Do not reject H_0. Conclude that the claim is supported.

	Exam / 50			Exam / 50		
75*th* percentile	46.5	(93%)			B+	28 and above
Median	39	(78%)	A	43 and above	B	21 and above
25*th* percentile	33	(66%)	A-	35.5 and above		

Weakest answers (in order): 6(d), 6(c), 3 and 4, 5(c), 1 (details).