STAT 511 EXAM 1 SOLUTIONS SPRING 2010

1 a) See slide 1 of our course notes.

b) $X(X'X)^-X'\underset{\sim}{y}$

c) $X'X\underset{\sim}{b} = X'\underset{\sim}{y}$

d) YES. $X(X'X)^-X'\underset{\sim}{y} = X(X'X)^-X'X\underset{\sim}{b}$

$$= P_X X\underset{\sim}{b}$$

$$= X\underset{\sim}{b}.$$

2. Let A be any symmetric and idempotent matrix. Then

$$A = AA^-A = A(AA)^-A$$
$$= A(A'A)^-A' = P_A.$$

Thus, A is the orthogonal projection matrix that projects onto $\mathcal{C}(A)$.

3 a)
$$\begin{bmatrix} 1 & 1 & 0 & -1 \\ 1 & 1 & 0 & -1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & -1 \\ 1 & 0 & 1 & -1 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}$$

b) A quantity is estimable if and only if it can be written as a linear combination of $E(\underline{y}) = X\underline{\beta}$. or, equivalently, if there is a linear function of $\underline{y}$ whose expectation is equal to the quantity.

i) $\mu + s_1 = E(y_{121})$. Thus, $\mu + s_1$ is estimable.

ii) $\mu + s_1 + 10\beta = E(y_{121} + 10(y_{131} - y_{121})) = E(10y_{131} - 9y_{121})$

Thus, $\mu + s_1 + 10\beta$ is estimable.

3 b) (continued)

iii) $S_1 - S_2 = E(Y_{121} - Y_{221})$.

Thus, $S_1 - S_2$ is estimable.

iv) In order for μ to be estimable, we need to find $\underset{\sim}{a}$ so that $\underset{\sim}{a}' X \underset{\sim}{\beta} = \mu$, i.e.

$\underset{\sim}{a}' X = [1, 0, 0, 0]$.

$$\underset{\sim}{a}' X = [1, 0, 0, 0] \Longrightarrow \sum_{m=1}^{12} a_m = 1, \text{ and}$$

$$\sum_{m=1}^{6} a_m = 0, \text{ and}$$

$$\sum_{m=7}^{12} a_m = 0.$$

Because $\sum_{m=1}^{6} a_m = 0$ and $\sum_{m=7}^{12} a_m = 0 \Longrightarrow \sum_{m=1}^{12} a_m = 0$, there does not exist $\underset{\sim}{a}$ such that $\underset{\sim}{a}' X \underset{\sim}{\beta} = \mu$. Therefore, μ is not estimable.

There are many other ways to complete 3 b). Using the method of HW 2, Problem 7 would work like this.

Suppose $X\underset{\sim}{d} = \underset{\sim}{0}$. Then

$$\begin{aligned} d_1 + d_2 - d_4 &= 0 \\ d_1 + d_2 &= 0 \\ d_1 + d_2 + d_4 &= 0 \\ d_1 + d_3 - d_4 &= 0 \\ d_1 + d_3 &= 0 \\ d_1 + d_3 + d_4 &= 0 \end{aligned} \iff \begin{gathered} d_1 = -d_2 = -d_3 \\ \text{and} \\ d_4 = 0 \end{gathered}$$

which means that $X\underset{\sim}{d} = 0 \iff$ $\underset{\sim}{d}$ is of the form $[d, -d, -d, 0]'$, $d \in \mathbb{R}$.

Now note that

i) $\mu + s_1 = [1,1,0,0]\,\underset{\sim}{\beta}$

$$[1,1,0,0]\begin{bmatrix} d \\ -d \\ -d \\ 0 \end{bmatrix} = 0 \quad \forall\, d \in \mathbb{R}.$$

Thus, $\mu + s_1$ is estimable.

ii) $\mu + s_1 + 10\beta = [1,1,0,10]\,\underset{\sim}{\beta}$

$$[1,1,0,10]\begin{bmatrix} d \\ -d \\ -d \\ 0 \end{bmatrix} = 0 \quad \forall\, d \in \mathbb{R}.$$

Thus, $\mu + s_1 + 10\beta$ is estimable.

iii) $s_1 - s_2 = [0,1,-1,0]\,\underset{\sim}{\beta}$

$$[0,1,-1,0]\begin{bmatrix} d \\ -d \\ -d \\ 0 \end{bmatrix} = 0 \quad \forall\, d \in \mathbb{R}$$

Thus, $s_1 - s_2$ is estimable.

iv) $\mu = [1,0,0,0]\,\underset{\sim}{\beta}$ $\quad [1,0,0,0]\begin{bmatrix} d \\ -d \\ -d \\ 0 \end{bmatrix} = d \neq 0 \quad \forall\, d \in \mathbb{R}.$

Thus, μ is NOT ESTIMABLE.

3c) THE X MATRIX IN 3(a) HAS RANK 3. ANY ONE OF THE FIRST THREE COLUMNS CAN BE WRITTEN AS A LINEAR COMBINATION OF THE OTHER TWO. R WOULD DISCARD THE SECOND COLUMN OF THE MATRIX, BUT CALCULATIONS ARE MUCH EASIER IF COLUMNS OF THE MODEL MATRIX ARE ORTHOGONAL. THUS, I WILL REMOVE THE FIRST COLUMN AND USE

$$X = \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & 1 & -1 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

AS THE MODEL MATRIX.

OUR MODEL SPECIFICATION IN THE PROBLEM STATEMENT IMPLIES THAT $E(\underset{\sim}{y}) = \begin{bmatrix} \mu+s_1-\beta \\ \mu+s_1-\beta \\ \mu+s_1 \\ \mu+s_1 \\ \mu+s_1+\beta \\ \mu+s_1+\beta \\ \mu+s_2-\beta \\ \mu+s_2-\beta \\ \mu+s_2 \\ \mu+s_2 \\ \mu+s_2+\beta \\ \mu+s_2+\beta \end{bmatrix}$.

IT IS EASY TO SEE THAT MY X MATRIX MULTIPLIED ON THE RIGHT BY $\begin{bmatrix} \mu+s_1 \\ \mu+s_2 \\ \beta \end{bmatrix}$ WILL GIVE $E(\underset{\sim}{y})$.

THUS, $(X'X)^{-1}X'\underset{\sim}{y} = \hat{\underset{\sim}{\beta}} = \begin{bmatrix} \widehat{\mu+s_1} \\ \widehat{\mu+s_2} \\ \hat{\beta} \end{bmatrix}$.

3 d) The BLUE of $E(Y_{111})$ is

$$\underbrace{[1 \;\; 0 \;\; -1]}_{\text{First row of } X} \hat{\beta}, \text{ where } \hat{\beta} = (X'X)^{-1}X'y.$$

$$X'X = \begin{bmatrix} 6 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 8 \end{bmatrix} \qquad (X'X)^{-1} = \begin{bmatrix} 1/6 & 0 & 0 \\ 0 & 1/6 & 0 \\ 0 & 0 & 1/8 \end{bmatrix}$$

$$X'y = \begin{bmatrix} Y_{1\cdot\cdot} \\ Y_{2\cdot\cdot} \\ Y_{\cdot 3 \cdot} - Y_{\cdot 1 \cdot} \end{bmatrix} \qquad (X'X)^{-1}X'y = \begin{bmatrix} \bar{Y}_{1\cdot\cdot} \\ \bar{Y}_{2\cdot\cdot} \\ \frac{1}{2}(\bar{Y}_{\cdot 3 \cdot} - \bar{Y}_{\cdot 1 \cdot}) \end{bmatrix}$$

Therefore, $\widehat{E(Y_{111})} = \bar{Y}_{1\cdot\cdot} - \frac{1}{2}(\bar{Y}_{\cdot 3 \cdot} - \bar{Y}_{\cdot 1 \cdot})$.

3 e) $\widehat{\mu + S_1} = \bar{Y}_{1\cdot\cdot}$

$$\widehat{\mu + S_1 + 10\beta} = \bar{Y}_{1\cdot\cdot} + 5(\bar{Y}_{\cdot 3 \cdot} - \bar{Y}_{\cdot 1 \cdot})$$

$$\widehat{S_1 - S_2} = \bar{Y}_{1\cdot\cdot} - \bar{Y}_{2\cdot\cdot}$$

4a) The F-test at the very bottom of the output tests whether all model coefficients other than the intercept are simultaneously equal to Ø. Thus, this is a test of

one mean for all observations vs.

one mean for each combination of source and protein amount (the cell means model).

$F = 39.84$

$df = 5, 6$

p-value = 0.0001587

There is strong evidence of a difference among treatment means.

b) The design matrix used by R is

$$\begin{array}{cccccc}
\mu & s & x & x^2 & s{:}x & s{:}x^2 \\
\end{array}$$
$$\left[\begin{array}{cccccc}
1 & 0 & -1 & 1 & 0 & 0 \\
1 & 0 & -1 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 1 & 0 & 0 \\
1 & 0 & 1 & 1 & 0 & 0 \\
1 & 1 & -1 & 1 & -1 & 1 \\
1 & 1 & -1 & 1 & -1 & 1 \\
1 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1
\end{array}\right]$$

The mean for Source 2, Protein Amount 3 is given by either of the last two rows times β. Thus, the estimate is the sum of $\hat{\beta}$ $= 17.5 + 5 + 6.25 - .25 + 3.75 + .75$

$= 33$

4b) This can also be seen by noting that the model is

$$y_{ijk} = \beta_0 + \beta_1 S(i) + \beta_2 x_j + \beta_3 x_j^2 + \beta_4 S(i) * x_j + \beta_5 S(i) * x_j^2 + \varepsilon_{ijk}$$

where $S(i) = \begin{cases} 1 & \text{if } i=2 \\ 0 & \text{if } i=1 \end{cases}$ and $x_j = j-2$.

Thus, for $i=2$, $j=3$, the mean is

$$\beta_0 + \beta_1(1) + \beta_2(3-2) + \beta_3(3-2)^2 + \beta_4(1)(3-2) + \beta_5(1)(3-2)^2$$

$$= \beta_0 + \beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5.$$

c) The standard error of any cell mean

is $\sqrt{\hat{Var}(\bar{Y}_{ij\cdot})} = \sqrt{\hat{\sigma}^2/2}$

$$= \sqrt{MSE/2}$$

$$= \sqrt{3.25/2}.$$

Also, the estimated intercept is an estimate of μ_{11}, which has the same standard error as $\hat{\mu}_{23}$. R gives 1.2748 as SE for intercept

4d) From part (b), we can see that the average of the source 1 means is $[1, 0, 0, \frac{2}{3}, 0, 0]\,\underline{\beta}$

$\uparrow$

Average of 1^{st} 6 rows of X.

The average of the source 2 means is $[1, 1, 0, \frac{2}{3}, 0, \frac{2}{3}]\,\underline{\beta}$.

$\uparrow$

Average of the last 6 rows of X

Thus, $C = [0, 1, 0, 0, 0, \frac{2}{3}]$ will do.

$\uparrow$

Difference of vectors above.

4 e) $\left(19.5 + 0.37 + 28.13 + 0.04\right) / 9$

4 f) $$F = \frac{\left(SSE_{REDUCED} - SSE_{FULL}\right) / \left(dfE_{REDUCED} - dfE_{FULL}\right)}{MSE_{FULL}}$$

$$= \frac{\left(0.37 + 28.13 + .04\right) / (9-6)}{3.25}$$

$df = 3, 6$

5. $W \sim \chi^2_m(\delta^2) \Rightarrow W \overset{d}{=} (\underset{\sim}{z} + \underset{\sim}{\delta})'(\underset{\sim}{z} + \underset{\sim}{\delta})$

where $\underset{m\times 1}{\underset{\sim}{z}} \sim N(\underset{\sim}{0}, I)$ and $\underset{\sim}{\delta} = \underset{m\times 1}{\begin{bmatrix} \delta \\ 0 \\ \vdots \\ 0 \end{bmatrix}}$.

a) $E(W) = E(\underset{\sim}{z}'\underset{\sim}{z} + 2\underset{\sim}{\delta}'\underset{\sim}{z} + \underset{\sim}{\delta}'\underset{\sim}{\delta})$

$= E(\underset{\sim}{z}'\underset{\sim}{z}) + 2\underset{\sim}{\delta}'E(\underset{\sim}{z}) + \delta^2$

$= E(\chi^2_m) + 2\underset{\sim}{\delta}'\underset{\sim}{0} + \delta^2$

$= m + \delta^2$

b) $Var(W) = Var(\underset{\sim}{z}'\underset{\sim}{z} + 2\underset{\sim}{\delta}'\underset{\sim}{z} + \underset{\sim}{\delta}'\underset{\sim}{\delta})$

$= Var\left(\sum_{i=1}^{m} z_i^2 + 2\delta z_1\right)$

$= Var\left(\sum_{i=1}^{m} z_i^2\right) + 4\delta^2 Var(z_1) + 0$

(because $cov(z_1^2 + \cdots + z_m^2, z_1) = Cov(z_1^2, z_1) = 0$)

$= Var(\chi^2_m) + 4\delta^2 = 2m + 4\delta^2$.

Point values were set as follows:

1. a) 4
   b) 4
   c) 4
   d) 5

2. 6

3. a) 4
   b) 4 x 4
   c) 4
   d) 6
   e) 6

4. a) 5
   b) 5
   c) 5
   d) 5
   e) 5
   f) 6

5. a) 5
   b) 5