EXAM 1 SOLUTIONS SPRING 2012

1. 6 POINTS

2. 4 POINTS EACH EXCEPT (c) AND (d) WHICH WERE WORTH 5 POINTS

3. a) 5 POINTS
   b) 6 POINTS

4. a) 4 POINTS
   b) 4 POINTS
   c) 5 POINTS
   d) 7 POINTS
   e) 5 POINTS
   f) 6 POINTS
   g) 5 POINTS
   h) 5 POINTS

1. $P_X = X(X'X)^- X' = XB$.

(WHERE $B = (X'X)^- X'$)

THUS, EVERY COLUMN OF P_X IS A LINEAR COMBINATION OF COLUMNS OF X.

$\therefore \mathcal{C}(P_X) \subseteq \mathcal{C}(X)$.

$X = P_X X$. THUS, EVERY COLUMN OF X IS A LINEAR COMBINATION OF COLUMNS OF P_X. $\therefore \mathcal{C}(X) \subseteq \mathcal{C}(P_X)$.

$\therefore \mathcal{C}(P_X) = \mathcal{C}(X)$ $\blacksquare$

1. $\underset{\sim}{a} \in \mathcal{C}(X) \Rightarrow \exists$ (THERE EXISTS) $\underset{\sim}{b} \ni$ (SUCH THAT)

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

$$\boxed{\text{AN ALTERNATIVE PROOF}}$$

$$\Rightarrow \exists\, \underset{\sim}{b} \ni \underset{\sim}{a} = P_X X \underset{\sim}{b}$$

(because $P_X X = X$)

$$\Rightarrow \exists\, \underset{\sim}{c}\ (\underset{\sim}{c} = X\underset{\sim}{b}) \ni \underset{\sim}{a} = P_X \underset{\sim}{c}$$

$$\Rightarrow \underset{\sim}{a} \in \mathcal{C}(P_X).$$

$\therefore\ \mathcal{C}(X) \subseteq \mathcal{C}(P_X).$

$$\underset{\sim}{a} \in \mathcal{C}(P_X) \Rightarrow \exists\, \underset{\sim}{b} \ni \underset{\sim}{a} = P_X \underset{\sim}{b} = X(X'X)^- X'\underset{\sim}{b}$$

$$\Rightarrow \exists\, \underset{\sim}{c}\ (\underset{\sim}{c} = (X'X)^- X'\underset{\sim}{b}) \ni \underset{\sim}{a} = X\underset{\sim}{c}$$

$$\Rightarrow \underset{\sim}{a} \in \mathcal{C}(X)$$

$\therefore\ \mathcal{C}(P_X) \subseteq \mathcal{C}(X).$ THUS, $\mathcal{C}(X) = \mathcal{C}(P_X)$ $\blacksquare$

2a) $$X = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 & -1 \end{bmatrix}$$

2 b) $E(Y_{12}) = E(\beta_1 - \beta_2 + \varepsilon_{12})$

$$= \beta_1 - \beta_2 + E(\varepsilon_{12})$$

$$= \beta_1 - \beta_2 + 0$$

$$= \beta_1 - \beta_2.$$

$\therefore$ Y_{12} IS A LINEAR UNBIASED ESTIMATOR OF $\beta_1 - \beta_2$.

$\therefore$ $\beta_1 - \beta_2$ IS ESTIMABLE

2b) THERE ARE MANY OTHER WAYS TO SEE THAT $\beta_1 - \beta_2$ IS ESTIMABLE. FOR EXAMPLE, WE KNOW LINEAR COMBINATIONS OF

$$E(\underline{y}) = X\underline{\beta} = \begin{bmatrix} \beta_1 - \beta_2 \\ \beta_3 - \beta_4 \\ \beta_2 - \beta_5 \\ \beta_1 - \beta_5 \end{bmatrix}$$

ARE ESTIMABLE.

WITH $A = [1, 0, 0, 0]$, $A\,E(\underline{y}) = \beta_1 - \beta_2$. $\boxtimes$

2 b) AN ALTERNATIVE PROOF IS AS FOLLOWS:

$\beta_1 - \beta_2 = C\underset{\sim}{\beta}$, WHERE $C = [1, -1, 0, 0, 0]$

BECAUSE $AX = C$, WHERE $A = [1, 0, 0, 0]$,

THE ESTIMABILITY OF $\beta_1 - \beta_2$ IS GUARANTEED.

2c) $\beta_1 - \beta_3$ IS NOT ESTIMABLE.

TO SEE THIS NOTE THAT

$$X\underset{\sim}{d} = \underset{\sim}{0} \Longrightarrow \begin{array}{l} d_1 - d_2 = 0 \\ d_3 - d_4 = 0 \\ d_2 - d_5 = 0 \\ d_1 - d_5 = 0 \end{array} \Rightarrow \underset{\sim}{d} = \begin{bmatrix} d_1 \\ d_1 \\ d_3 \\ d_3 \\ d_1 \end{bmatrix}$$

$\beta_1 - \beta_3 = C\underset{\sim}{\beta}$, WHERE $C = [1, 0, -1, 0, 0]$.

$C\underset{\sim}{d} = d_1 - d_3$, WHICH IS NOT NECESSARILY

0. $\therefore$ $\beta_1 - \beta_3$ IS NOT ESTIMABLE.

THIS FOLLOWS FROM RESULT 3.8 (ii)

FROM KEN KOEHLER'S NOTES ON

ESTIMABILITY THAT WE LEARNED

ABOUT IN HW 2, PROBLEM 11.

WE COULD ALSO USE THIS RESULT

TO PROVE THAT $\beta_1 - \beta_2$ IS ESTIMABLE

IN PART (b) BECAUSE

$$[1,-1,0,0,0]\begin{bmatrix} d_1 \\ d_1 \\ d_3 \\ d_3 \\ d_1 \end{bmatrix} = d_1 - d_1 = 0 \quad \forall\, d_1, d_3 \in \mathbb{R}.$$

HERE IS ANOTHER WAY TO SHOW THAT $\beta_1 - \beta_3$ IS NON ESTIMABLE. WE KNOW ONLY LINEAR COMBINATIONS OF $E(\underset{\sim}{y})$ ARE ESTIMABLE. A LINEAR COMBINATION OF $E(\underset{\sim}{y})$ LOOKS LIKE

$$\underset{\sim}{a}' E(\underset{\sim}{y}) = [a_1, a_2, a_3, a_4] \begin{bmatrix} \beta_1 - \beta_2 \\ \beta_3 - \beta_4 \\ \beta_2 - \beta_5 \\ \beta_1 - \beta_5 \end{bmatrix} =$$

$$a_1(\beta_1 - \beta_2) + a_2(\beta_3 - \beta_4) + a_3(\beta_2 - \beta_5)$$

$$+ a_4(\beta_1 - \beta_5) = (a_1 + a_4)\beta_1 + (a_3 - a_1)\beta_2$$

$$+ a_2\beta_3 - a_2\beta_4$$

$$+ (-a_3 - a_4)\beta_5.$$

IN ORDER FOR THIS LINEAR COMBINATION TO EQUAL $\beta_1 - \beta_3$, WE MUST HAVE THE COEFFICIENT OF β_3 AS -1 (i.e. $a_2 = -1$) AND THE COEFFICIENT OF β_4 AS 0 (i.e., $a_2 = 0$).

BECAUSE a_2 CANNOT BE -1 AND 0, THERE DOES NOT EXIST $\underline{a}$ SUCH THAT $\underline{a}'E(\underline{Y}) = \beta_1 - \beta_3$. THEREFORE, $\beta_1 - \beta_3$ IS NON ESTIMABLE.

EQUIVALENTLY,

$$(a_1 + a_4)\beta_1 + (a_3 - a_1)\beta_2 + a_2\beta_3 - a_2\beta_4 + (-a_3 - a_4)\beta_5 = \beta_1 - \beta_3 \iff a_1 + a_4 = 1,\ a_3 - a_1 = 0,\ a_2 = 1,\ -a_2 = 0,$$

AND $-(a_3 + a_4) = 0$.

HOWEVER, THIS SYSTEM OF EQUATIONS HAS NO SOLUTION BECAUSE $1 \neq 0$.

2d)

$$X'X = \begin{bmatrix} 2 & -1 & 0 & 0 & -1 \\ -1 & 2 & 0 & 0 & -1 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ -1 & -1 & 0 & 0 & 2 \end{bmatrix}$$

$$(X'X)^{-} = \begin{bmatrix} 2/3 & 1/3 & 0 & 0 & 0 \\ 1/3 & 2/3 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

TO GET 2d) RIGHT, IT IS IMPORTANT TO COMPUTE $X'X$ CORRECTLY, RECOGNIZE THAT MATRIX HAS RANK 3, AND IDENTIFY A 3×3 SUBMATRIX THAT IS EASY TO INVERT. RECALL THAT $\text{RANK}(X) = \text{RANK}(X'X)$.

$\text{RANK}(X)$ IS 3 BECAUSE THE FIRST 3 ROWS ARE CLEARLY LINEARLY INDEPENDENT. FURTHERMORE, THE REMAINING ROW (ROW 4) IS THE SUM OF THE FIRST AND THIRD ROWS. THUS, $\text{RANK}(X) = \text{RANK}(X'X) = 3$.

I WAS HOPING IT WOULD BE RELATIVELY EASY FOR YOU TO INVERT THE 3x3 MATRIX IN THE UPPER LEFT CORNER OF $X'X$.

THIS MATRIX IS

$$\begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} A^* & \underset{\sim}{0} \\ \underset{\sim}{0}' & 1 \end{bmatrix},$$

WHERE $A^* = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$.

WE KNOW $\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} \begin{bmatrix} C & 0 \\ 0 & D \end{bmatrix} = \begin{bmatrix} AC & 0 \\ 0 & BD \end{bmatrix}$

FOR ARBITRARY MATRICES A, B, C, D OF APPROPRIATE SIZE. THUS, WE KNOW THAT

$\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} & 0 \\ 0 & B^{-1} \end{bmatrix}$ IF A AND B ARE NONSINGULAR. THUS,

$$\begin{bmatrix} A^* & \underset{\sim}{0} \\ \underset{\sim}{0}' & 1 \end{bmatrix}^{-1} = \begin{bmatrix} A^{*-1} & \underset{\sim}{0} \\ \underset{\sim}{0}' & 1 \end{bmatrix}.$$

$$A^{*-1} = \frac{1}{(2)(2)-(-1)(-1)} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

$$= \begin{bmatrix} 2/3 & 1/3 \\ 1/3 & 2/3 \end{bmatrix}$$

$$\therefore \begin{bmatrix} A^* & \underset{\sim}{0} \\ \underset{\sim}{0}' & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 2/3 & 1/3 & 0 \\ 1/3 & 2/3 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

THE G.I. FOLLOWS FROM OUR ALGORITHM FOR FINDING GENERALIZED INVERSES.

2e) $X'X\underset{\sim}{b} = X'\underset{\sim}{y}$

2f) IF YOU WERE ABLE TO COMPUTE $(X'X)^-$, A SOLUTION IS OBTAINED BY COMPUTING

$$(X'X)^- X'\underset{\sim}{y} = \begin{bmatrix} \frac{1}{3}y_{12} + \frac{2}{3}y_{15} + \frac{1}{3}y_{15} \\ -\frac{1}{3}y_{12} + \frac{1}{3}y_{15} + \frac{2}{3}y_{25} \\ y_{34} \\ 0 \\ 0 \end{bmatrix}$$

IF YOU WERE NOT ABLE TO COMPUTE $(X'X)^{-}$, A SOLUTION TO THE NORMAL EQUATIONS CAN STILL BE OBTAINED USING ALGEBRA!

$$X'X\underline{b} = X'\underline{Y} \iff$$

$$\begin{aligned}
2b_1 - b_2 - b_5 &= Y_{12} + Y_{15} \\
-b_1 + 2b_2 - b_5 &= -Y_{12} + Y_{25} \\
b_3 - b_4 &= Y_{34} \\
-b_3 + b_4 &= -Y_{34} \\
-b_1 - b_2 + 2b_5 &= -Y_{25} - Y_{15}
\end{aligned}$$

THE LAST 2 EQUATIONS ARE REDUNDANT BECAUSE THE 4^{TH} IS NEGATIVE OF THE 3^{RD} AND THE 5^{TH} IS NEGATIVE OF THE SUM OF THE FIRST AND SECOND EQUATIONS. THUS, WE NEED TO FIND A SOLUTION TO

$$
\begin{aligned}
2b_1 - b_2 - b_5 &= y_{12} + y_{15} \\
-b_1 + 2b_2 - b_5 &= -y_{12} + y_{25} \\
b_3 - b_4 &= y_{34}
\end{aligned}
$$

THIS SYSTEM OF EQUATIONS HAS 5 UNKNOWNS $(b_1, \ldots, b_5)$ AND 3 EQUATIONS. THERE ARE AN

INFINITE NUMBER OF SOLUTIONS. IF WE TAKE $b_4 = b_5 = 0$, WE HAVE

$$2b_1 - b_2 = y_{12} + y_{15}$$
$$-b_1 + 2b_2 = -y_{12} + y_{25}$$
$$b_3 = y_{34}$$

SOLVING THIS SYSTEM YIELDS

$$b_1 = \frac{1}{3} y_{12} + \frac{2}{3} y_{15} + \frac{1}{3} y_{15}$$
$$b_2 = -\frac{1}{3} y_{12} + \frac{1}{3} y_{15} + \frac{2}{3} y_{25}$$
$$b_3 = y_{34}$$

WHICH IS THE SAME ANSWER AS $(X'X)^{-}X'y$ FOR

THE CHOICE OF $(X'X)^{-}$ PROVIDED PREVIOUSLY. AN INFINITE NUMBER OF OTHER SOLUTIONS ARE POSSIBLE.

2g) BASED ON OUR SOLUTION TO THE NORMAL EQUATIONS,

$$\underset{\sim}{c}'\hat{\underset{\sim}{\beta}} = [1, 0, 0, 0, -1](X'X)^{-}X'\underset{\sim}{y}$$

$$= \frac{1}{3} y_{12} + \frac{2}{3} y_{15} + \frac{1}{3} y_{25}.$$

BASED ON WHAT WE HAVE RECENTLY LEARNED IN CLASS, THIS MAKES SENSE.

NOTE THAT y_{15} AND $y_{12} + y_{25}$ ARE INDEPENDENT UNBIASED ESTIMATORS OF

$\beta_1 - \beta_5$. $VAR(Y_{15}) = \sigma^2$ AND

$VAR(Y_{12} + Y_{25}) = 2\sigma^2$. USING INVERSE

VARIANCE WEIGHTING THE BLUE IS

$$\frac{\left(\frac{1}{\sigma^2}\right)}{\left(\frac{1}{\sigma^2} + \frac{1}{2\sigma^2}\right)} Y_{15} + \frac{\left(\frac{1}{2\sigma^2}\right)}{\left(\frac{1}{\sigma^2} + \frac{1}{2\sigma^2}\right)} (Y_{12} + Y_{25})$$

$$= \frac{2}{3} Y_{15} + \frac{1}{3} (Y_{12} + Y_{25})$$

$$= \frac{1}{3} Y_{12} + \frac{2}{3} Y_{15} + \frac{1}{3} Y_{25}.$$

2h) $\widehat{\beta_1 - \beta_5} = \frac{1}{3} Y_{12} + \frac{2}{3} Y_{15} + \frac{1}{3} Y_{25}$

$$= \frac{1}{3}(\beta_1 - \beta_2 + \varepsilon_{12}) + \frac{2}{3}(\beta_1 - \beta_5 + \varepsilon_{15})$$

$$+ \frac{1}{3}(\beta_2 - \beta_5 + \varepsilon_{25})$$

$$= \beta_1 - \beta_5 + \frac{1}{3}\varepsilon_{12} + \frac{2}{3}\varepsilon_{15} + \frac{1}{3}\varepsilon_{25}$$

THUS, $E(\widehat{\beta_1 - \beta_5}) = \beta_1 - \beta_5 + E\left(\frac{1}{3}\varepsilon_{12} + \frac{2}{3}\varepsilon_{15} + \frac{1}{3}\varepsilon_{25}\right)$

AND $\widehat{\beta_1 - \beta_5}$ WILL BE UNBIASED AS LONG AS

$$E\left(\frac{1}{3}\varepsilon_{12} + \frac{2}{3}\varepsilon_{15} + \frac{1}{3}\varepsilon_{25}\right) = 0.$$

2h) (CONTINUED) IN GENERAL, THE OLS ESTIMATOR OF AN ESTIMABLE $\underset{\sim}{c}'\underset{\sim}{\beta}$ IS UNBIASED AS LONG AS $E(\underset{\sim}{\varepsilon}) = \underset{\sim}{0}$. NOTHING MORE IS REQUIRED.

2i) WE KNOW THE OLS IS THE BLUE WHEN THE GAUSS-MARKOV MODEL HOLDS, i.e., $E(\underset{\sim}{\varepsilon}) = \underset{\sim}{0}$ AND $Var(\underset{\sim}{\varepsilon}) = \sigma^2 I$

WE DON'T NEED NORMALITY OR INDEPENDENCE (JUST COVARIANCE 0) OF THE ε TERMS.

WE CAN RELAX ASSUMPTIONS FURTHER IN THIS SPECIAL CASE, BUT I DID NOT EXPECT ANYTHING WEAKER THAN THE GAUSS-MARKOV ASSUMPTIONS.

2 j) $E(Y_{15}) = \beta_1 - \beta_5$ $\therefore$ Y_{15} IS ANOTHER LINEAR UNBIASED ESTIMATOR:

$$3a)\ \begin{bmatrix} \hat{\underset{\sim}{y}} \\ \underset{\sim}{y} - \hat{\underset{\sim}{y}} \end{bmatrix} = \begin{bmatrix} P_X \\ I - P_X \end{bmatrix} \underset{\sim}{y} \sim N(\underset{\sim}{\mu}, \Sigma)$$

$$\underset{\sim}{\mu} = \begin{bmatrix} P_X \\ I - P_X \end{bmatrix} E(\underset{\sim}{y}) = \begin{bmatrix} P_X \\ I - P_X \end{bmatrix} X\underset{\sim}{\beta} = \begin{bmatrix} P_X X \underset{\sim}{\beta} \\ (I - P_X) X \underset{\sim}{\beta} \end{bmatrix}$$

$$= \begin{bmatrix} X\underset{\sim}{\beta} \\ \underset{\sim}{0} \end{bmatrix}$$

3 a) (CONTINUED)

$$\Sigma = \begin{bmatrix} P_X \\ I-P_X \end{bmatrix} \mathrm{VAR}(\underline{y}) \begin{bmatrix} P_X' & (I-P_X)' \end{bmatrix}$$

$$= \sigma^2 \begin{bmatrix} P_X P_X' & P_X (I-P_X)' \\ (I-P_X)P_X' & (I-P_X)(I-P_X)' \end{bmatrix}$$

$$= \sigma^2 \begin{bmatrix} P_X^2 & P_X (I-P_X) \\ (I-P_X)P_X & (I-P_X)^2 \end{bmatrix}$$

$$= \sigma^2 \begin{bmatrix} P_X & 0 \\ 0 & I-P_X \end{bmatrix}.$$

3a) (CONTINUED) THUS,

$$\begin{bmatrix} \hat{\underset{\sim}{y}} \\ \underset{\sim}{y} - \hat{\underset{\sim}{y}} \end{bmatrix} \sim N\left(\begin{bmatrix} X\underset{\sim}{\beta} \\ 0 \end{bmatrix}, \sigma^2 \begin{bmatrix} P_X & 0 \\ 0 & I - P_X \end{bmatrix} \right).$$

3b) $\hat{y}'\hat{y} = (P_X y)'(P_X y) = y' P_X' P_X y$

$$= y' P_X P_X y = y' P_X y.$$

NOTE THAT $P_X \, VAR(y) = P_X \sigma^2 I = \sigma^2 P_X$.

$\sigma^2 P_X$ IS NOT IDEMPOTENT BECAUSE

$(\sigma^2 P_X)(\sigma^2 P_X) = \sigma^4 P_X \neq \sigma^2 P_X$.

THUS, CONSIDER

$$\frac{\hat{y}'\hat{y}}{\sigma^2} = y' \frac{P_X}{\sigma^2} y.$$

$\frac{P_X}{\sigma^2} \text{VAR}(\underset{\sim}{y}) = \frac{P_X}{\sigma^2} \sigma^2 I = P_X$, WHICH IS IDEMPOTENT. THUS, $\frac{\hat{\underset{\sim}{y}}'\hat{\underset{\sim}{y}}}{\sigma^2}$ HAS A χ^2 DISTRIBUTION. THE

$$DF = \text{RANK}\left(\frac{P_X}{\sigma^2}\right) = \text{RANK}(P_X) = \text{RANK}(X) = r.$$

$$NCP = (X\underset{\sim}{\beta})' \frac{P_X}{\sigma^2} (X\underset{\sim}{\beta}) = \frac{\underset{\sim}{\beta}'X'X\underset{\sim}{\beta}}{\sigma^2}.$$

THUS, $\frac{\hat{\underline{Y}}'\hat{\underline{Y}}}{\sigma^2} \sim \chi^2_r\left(\underline{\beta}'X'X\underline{\beta}/\sigma^2\right)$

WHICH IMPLIES

$$\hat{\underline{Y}}'\hat{\underline{Y}} \sim \sigma^2\chi^2_r\left(\underline{\beta}'X'X\underline{\beta}/\sigma^2\right)$$

(A SCALED χ^2 DISTRIBUTION).

4 a) μ_1 IS THE MEAN FOR TREATMENT 1.

BECAUSE R USES THE DESIGN MATRIX

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

THE MEAN FOR TREATMENT 1 IS ESTIMATED BY R'S ESTIMATE OF THE INTERCEPT, i.e.,

351.

4b) μ_2 IS THE MEAN FOR TREATMENT 2.

THIS SHOULD BE CLEAR BECAUSE OUR MODEL FOR THE DATA WAS STATED AS $y_{ij} = \mu_i + \varepsilon_{ij}$, WHERE ε_{ij} TERMS iid $N(0, \sigma^2)$.

IN R, THE MEAN FOR TREATMENT 2 IS ESTIMATED BY $\widehat{\text{INTERCEPT}} + \widehat{\text{Dose2}}$ (SEE ROWS 3 & 4 OF DESIGN MATRIX). THUS,

$$\hat{\mu}_2 = 351 - 10 = 341.$$

4c) $\hat{\mu}_2 = \dfrac{Y_{21} + Y_{22}}{2} = \overline{Y}_{2\cdot}$

$\mathrm{VAR}(\hat{\mu}_2) = \dfrac{\sigma^2}{2}$

$\hat{\sigma}^2 = \dfrac{SSE}{n-r} = \dfrac{432.5}{10-5} = \dfrac{432.5}{5}$

THUS, $SE(\hat{\mu}_2) = \sqrt{\widehat{\mathrm{VAR}}(\hat{\mu}_2)}$

$= \sqrt{\hat{\sigma}^2/2}$

$= \sqrt{\dfrac{432.5}{(2)(5)}}$

4c) (CONTINUED)

I WAS HOPING THAT MANY OF YOU WOULD RECOGNIZE THAT $SE(\hat{\mu}_1) = \cdots = SE(\hat{\mu}_5)$ DUE TO THE BALANCED DESIGN. THUS,

$$SE(\hat{\mu}_2) = SE(\hat{\mu}_1) = SE(\widehat{\text{INTERCEPT}})$$

$$= 6.576.$$

THIS IS THE SAME AS

$\sqrt{\dfrac{432.5}{(2)(5)}}$ DERIVED ON THE PREVIOUS PAGE.

4c) SOME OF YOU COMPUTED

$SE(\hat{M}_2)$ BY ASSUMING

$VAR(\widehat{INTERCEPT} + \widehat{DOSE2})$

$= VAR(\widehat{INTERCEPT}) + VAR(\widehat{DOSE2})$.

THE PROBLEM IS THAT $\widehat{INTERCEPT}$ AND

$\widehat{DOSE2}$ ($\overline{Y}_{1\cdot}$ AND $\overline{Y}_{2\cdot} - \overline{Y}_{1\cdot}$, RESPECTIVELY)

ARE NOT INDEPENDENT.

4d) $\widehat{\text{DOSE2}}$ IS AN ESTIMATE OF $\mu_2 - \mu_1$. THUS, A t-STATISTIC FOR TESTING $H_0: \mu_1 = \mu_2$ IS GIVEN IN THE R OUTPUT.

$t = -1.075$, P-VALUE $= 0.3314$

THIS INDICATES THAT THERE IS NO SIGNIFICANT EVIDENCE OF A DIFFERENCE BETWEEN μ_1 AND μ_2. THE DISTRIBUTION

OF THIS t-STATISTIC IS NONCENTRAL t WITH $n-r = 10-5 = 5$ DF AND

$$NCP = \frac{\mu_1 - \mu_2}{\sqrt{\sigma^2\left(\frac{1}{2}+\frac{1}{2}\right)}} = (\mu_1-\mu_2)/\sigma.$$

MANY OF YOU FAILED TO STATE THE DISTRIBUTION OF THE TEST STATISTIC OR FAILED TO PROVIDE THE DF OR NCP. I TRIED TO CLUE YOU IN ON THIS BY MY STATEMENT OF "(BE VERY PRECISE)" WHEN I ASKED FOR THE DISTRIBUTION OF THE TEST STATISTIC.

4 e) THE t-STATISTIC FOR TESTING

$H_0: \mu_3 = \mu_4$ IS $t = \dfrac{\bar{Y}_{3.} - \bar{Y}_{4.}}{\sqrt{MSE\left(\frac{1}{2}+\frac{1}{2}\right)}} = \dfrac{\bar{Y}_{3.} - \bar{Y}_{4.}}{\sqrt{MSE}}$

THIS CAN BE COMPUTED FROM THE R OUTPUT AS $t = \dfrac{(-6-(-17))}{\sqrt{432.5/5}} = \dfrac{11}{\sqrt{432.5/5}}$

THUS, THE F-STAT IS $\dfrac{11^2}{432.5/5}$.

4e) (CONTINUED)

YOU MIGHT INSTEAD HAVE NOTICED FROM THE R OUTPUT THAT THE SE OF A DIFFERENCE IN TREATMENT MEANS (WHICH IS THE SAME FOR ALL TREATMENT PAIRS DUE TO THE BALANCED DESIGN) IS 9.301. THE LEADS TO THE SAME F-STAT AS ON THE PREVIOUS PAGE $\left(\frac{11}{9.301}\right)^2$.

4e) (CONTINUED) SOME OF YOU GAVE THE GENERAL FORMULA FOR AN F-STAT. THIS PROBLEM ASKED YOU TO "USE THE R CODE AND PARTIAL OUTPUT PROVIDED ... TO ANSWER THE FOLLOWING QUESTIONS." THUS, THE POINT WAS TO SEE IF YOU UNDERSTOOD ENOUGH ABOUT THE RELATIONSHIPS OF VARIOUS QUANTITIES TO COME UP WITH A NUMERICAL ANSWER.

4f) WE NEED TO CONDUCT A TEST FOR "LACK OF FIT." THIS CAN BE DONE BY COMPARING FULL AND REDUCED MODELS

$$F = \frac{(SSE_{REDUCED} - SSE_{FULL}) / (DFE_{RED} - DFE_{FULL})}{SSE_{FULL} / DFE_{FULL}}$$

$$= \frac{(1038.5 - 432.5) / [(10-2) - (10-5)]}{432.5/(10-5)}$$

$$= \frac{(1038.5 - 432.5)/3}{432.5/5}$$

THIS IS AN F-STATISTIC WITH DF 3 AND 5. THE p-VALUE CAN BE OBTAINED FROM THE LAST ANOVA TABLE IN THE OUTPUT, WHICH SHOWS THAT $p = 0.1907591$. THUS, THERE IS NO SIGNIFICANT EVIDENCE OF LACK OF LINEAR FIT.

4g) THE KEY TO THIS PROBLEM IS TO NOTICE THAT THE DOSES ARE NOT EQUALLY SPACED. IF THE MEANS FALL ALONG A LINE, THEN THERE EXIST β_0 AND β_1 SUCH THAT

$$\left.\begin{array}{l} \mu_1 = \beta_0 + \beta_1(0) \\ \mu_2 = \beta_0 + \beta_1(2) \\ \mu_3 = \beta_0 + \beta_1(4) \\ \mu_4 = \beta_0 + \beta_1(8) \\ \mu_5 = \beta_0 + \beta_1(16) \end{array}\right\} \iff \left\{\begin{array}{l} \mu_2 - \mu_1 = 2\beta_1 \\ \mu_3 - \mu_2 = 2\beta_1 \\ \mu_4 - \mu_3 = 4\beta_1 \\ \mu_5 - \mu_4 = 8\beta_1 \end{array}\right.$$

$$M_2 - M_1 = M_3 - M_2 \iff -1\,M_1 + 2\,M_2 - M_3 = 0$$

AND

$$M_4 - M_3 = 2(M_3 - M_2) \iff 2M_2 - 3M_3 + M_4 = 0$$

AND

$$M_5 - M_4 = 2(M_4 - M_3) \iff 2M_3 - 3M_4 + M_5 = 0$$

$$C = \begin{bmatrix} -1 & 2 & -1 & 0 & 0 \\ 0 & 2 & -3 & 1 & 0 \\ 0 & 0 & 2 & -3 & 1 \end{bmatrix} \qquad \underline{d} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

4h)

DF	SS	MS	F
1	5899.6	5899.6	5899.6/(432.5/5)
3	1038.5-432.5	202	202/(432.5/5)
5	432.5	432.5/5	

THESE ENTRIES ALL FOLLOW FROM THE SUMS OF SQUARES

$$\underset{\sim}{Y}'(P_2-P_1)\underset{\sim}{Y},\quad \underset{\sim}{Y}'(P_3-P_2)\underset{\sim}{Y},\quad \underset{\sim}{Y}'(P_4-P_3)\underset{\sim}{Y}$$

WHERE $X_1 = \underset{10\times1}{\tilde{1}}$,

$$X_2 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 2 \\ 1 & 2 \\ 1 & 4 \\ 1 & 4 \\ 1 & 8 \\ 1 & 8 \\ 1 & 16 \\ 1 & 16 \end{bmatrix},$$

$X_3 =$ DESIGN MATRIX X FROM PART (a), OR $\underset{5\times5}{I} \otimes \underset{2\times1}{\tilde{1}}$

AND $X_4 = \underset{10\times10}{I}$.

$\boxed{\text{THE DF ARE DIFFERENCES IN RANKS OF THESE MATRICES, i.e., } 2-1,\ 5-2,\ 10-5.}$

RECALL THAT THE SS CAN BE SEEN AS REDUCTIONS IN SSE WHEN PROJECTING ONTO A LARGER COLUMN SPACE COMPARED TO PROJECTING ONTO A SMALLER COLUMN SPACE. E.G.

$$\underset{\sim}{y}'(P_3 - P_2)\underset{\sim}{y} = \underset{\sim}{y}'\left[(I-P_2)-(I-P_3)\right]\underset{\sim}{y}$$

$$= \underset{\sim}{y}'(I-P_2)\underset{\sim}{y} - \underset{\sim}{y}'(I-P_3)\underset{\sim}{y}$$

$$= 1038.5 - 432.5 = 606$$

IS REDUCTION IN SSE WHEN PROJECTING ON X_3 INSTEAD OF X_2.