SPRING 2016

EXAM 1 SOLUTIONS

POINTS PER PROBLEM:

1a) 5	2 a) 5	3 a) 10	4 a) 5
b) 5	b) 20	b) 15	b) 10
c) 10			c) 10

1a)

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

NOTE THAT THE COLUMN SPACE OF X IS THE SAME AS THE COLUMN SPACE OF THIS MATRIX $\longrightarrow$

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \equiv W$$

SO THIS MODEL IS THE SAME AS THE CELL MEANS MODEL.

$$X = W \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} \Rightarrow \mathcal{C}(X) \subseteq \mathcal{C}(W)$$

$$W = X \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{bmatrix} \Rightarrow \mathcal{C}(W) \subseteq \mathcal{C}(X)$$

$$\therefore \mathcal{C}(X) = \mathcal{C}(W)$$

IN ADDITION TO THE MORE FORMAL ARGUMENT ON THE PREVIOUS PAGE, IT SHOULD BE CLEAR ALMOST IMMEDIATELY THAT THE MODEL IN QUESTION IS THE CELL MEANS MODEL.

BECAUSE β_1, β_2, β_3, AND β_4 CAN BE ANY REAL NUMBERS, SO ALSO, β_1, $\beta_1+\beta_2$, $\beta_1+\beta_2+\beta_3$, AND $\beta_1+\beta_2+\beta_3+\beta_4$ CAN BE ANY REAL NUMBERS WITH NO RESTRICTIONS. THUS, THIS MODEL IS THE SAME AS THE CELL MEANS MODEL.

1b) LET $\mu_1 = \beta_1$, $\mu_2 = \beta_1 + \beta_2$, $\mu_3 = \beta_1 + \beta_2 + \beta_3$, $\mu_4 = \beta_1 + \beta_2 + \beta_3 + \beta_4$.

$\beta_4 = \mu_4 - \mu_3$ WHOSE BLUE IS $\overline{Y}_{4.} - \overline{Y}_{3.}$ UNDER THE CELL MEANS MODEL.

THUS, BLUE OF β_4 IS $26.3 - 22.8 = 3.5$.

1c) $VAR(\hat{\beta}_4) = VAR(\hat{\mu}_4 - \hat{\mu}_3) = VAR(\overline{Y}_{4.} - \overline{Y}_{3.}) = \dfrac{\sigma^2}{4} + \dfrac{\sigma^2}{2}$

THUS, $SE(\hat{\beta}_4) = \sqrt{\hat{\sigma}^2\left[\frac{1}{4} + \frac{1}{2}\right]} = \sqrt{MSE\ 3/4}$

(COULD STOP HERE AND GET FULL CREDIT.)

$$MSE = \frac{SSE}{11-4} = \frac{(3-1)4.1 + (2-1)3.4 + (2-1)2.8 + (4-1)3.2}{7}$$

$$= \frac{8.2 + 3.4 + 2.8 + 9.6}{7} = \frac{24}{7}$$

$$SE(\hat{\beta}_4) = \sqrt{\frac{24}{7}\frac{3}{4}} = \sqrt{18/7}$$

2 a) THE DATA ARE

	B1	B2
A1	3, 5	10
A2	2	10, 12

THE BLUES OF THE CELL MEANS ARE

$$\hat{\mu}_{11} = \frac{3+5}{2} = 4 \qquad \hat{\mu}_{12} = 10$$

$$\hat{\mu}_{21} = 2 \qquad \hat{\mu}_{22} = \frac{10+12}{2} = 11.$$

THE LSMEANS FOR FACTOR A ARE

$$A1: \frac{1}{2}\hat{\mu}_{11} + \frac{1}{2}\hat{\mu}_{12} = \frac{4+10}{2} = 7$$

$$A2: \frac{1}{2}\hat{\mu}_{21} + \frac{1}{2}\hat{\mu}_{22} = \frac{2+11}{2} = 6.5.$$

2b) $\dfrac{3+5+10+2+10+12}{6} = 7$

THUS, C. TOTAL $= (3-7)^2 + (5-7)^2 + (10-7)^2 + (2-7)^2 + (10-7)^2 + (12-7)^2$

$= 16 + 4 + 9 + 25 + 9 + 25 = 88$

LET P_2 BE THE PROJECTION MATRIX FOR PROJECTION ONTO THE COLUMN SPACE OF THE MODEL MATRIX FOR THE MODEL THAT INCLUDES INTERCEPT AND FACTOR A. AS USUAL, LET P_1 BE THE PROJECTION MATRIX FOR THE MODEL MATRIX $\underset{\sim}{1}$.

$$\begin{aligned} SS(A|1) = \underset{\sim}{y}'(P_2-P_1)\underset{\sim}{y} &= \underset{\sim}{y}'(P_2-P_1)(P_2-P_1)\underset{\sim}{y} \\ &= \underset{\sim}{y}'(P_2-P_1)'(P_2-P_1)\underset{\sim}{y} \\ &= [(P_2-P_1)\underset{\sim}{y}]'[(P_2-P_1)\underset{\sim}{y}] \\ &= \|(P_2-P_1)\underset{\sim}{y}\|^2 = \|P_2\underset{\sim}{y} - P_1\underset{\sim}{y}\|^2 \end{aligned}$$

BECAUSE THE MODEL THAT INCLUDES AN INTERCEPT AND FACTOR A IS JUST A CELL MEANS MODEL WITH TWO CELLS (ONE FOR A1 AND ONE FOR A2), WE SHOULD KNOW IMMEDIATELY THAT

$$P_2 \underset{\sim}{y} = [\underbrace{6, 6, 6}_{\frac{3+5+10}{3}}, \underbrace{8, 8, 8}_{\frac{2+10+12}{3}}]'$$

WE SHOULD ALSO KNOW THAT $P_1 \underset{\sim}{y} = [7, 7, 7, 7, 7, 7]'$

THUS, $SS(A|1) = 3(6-7)^2 + 3(8-7)^2 = 6.$

ALTERNATIVELY, $SSE(1, A) = (3-6)^2 + (5-6)^2 + (10-6)^2 + (2-8)^2 + (10-8)^2 + (12-8)^2$

$= 82$

$SS(A|1) = SSE(1) - SSE(1, A) = SSTOTAL - 82 = 88 - 82 = 6.$

EITHER WAY, $SS(A|1) = 6.$

SS(B | 1, A) IS NOT AS EASY TO DEAL WITH.

WE CAN COMPUTE THAT LATER BY USING

$SS(A|1) + SS(B|1,A) + SS(AB|1,A,B) + SSE = SSTOTAL$

WE ALREADY HAVE $SS(A|1)$ AND SSTOTAL.

$$SSE = (3-4)^2 + (5-4)^2 + (10-10)^2 + (2-2)^2 + (10-11)^2 + (12-11)^2 = 4$$

$$SS(AB|1,A,B) = (C\hat{\underline{\beta}})'[C(X'X)^- C']^{-1} C\hat{\underline{\beta}},$$

WHERE $C = [1, -1, -1, 1]$, $X = \begin{bmatrix} 1&0&0&0 \\ 1&0&0&0 \\ 0&1&0&0 \\ 0&0&1&0 \\ 0&0&0&1 \\ 0&0&0&1 \end{bmatrix}$, $\hat{\underline{\beta}} = \begin{bmatrix} 4 \\ 10 \\ 2 \\ 11 \end{bmatrix}$, $(X'X)^- = \begin{bmatrix} \frac{1}{2}&0&0&0 \\ 0&1&0&0 \\ 0&0&1&0 \\ 0&0&0&\frac{1}{2} \end{bmatrix}$,

$C(X'X)^- C' = [\frac{1}{2}, -1, -1, \frac{1}{2}] \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix} = 3$, $C\hat{\underline{\beta}} = 4 - 10 - 2 + 11 = 3$,

$[C(X'X)^- C']^{-1} = \frac{1}{3}$. THUS, $SS(AB|1,A,B) = 3 \frac{1}{3} 3 = 3$.

IT FOLLOWS THAT $SS(B|1,A) = 88 - 6 - 3 - 4 = 75$.

WE HAVE

SOURCE	DF	SS
A	$1 = 2-1$	6
B	$1 = 3-2$	75
A x B	$1 = 4-3$	3
ERROR	$2 = 6-4$	4
C. TOTAL	$5 = 6-1$	88

3a) LET JANE BE INSTRUCTOR 1 AND JOHN BE INSTRUCTOR 2. FOR $i=1,2$ AND $j=1,2$, LET μ_{ij} BE THE EXPECTED VALUE OF A RATING GIVEN BY A STUDENT WHOSE ACTUAL INSTRUCTOR WAS i AND WHOSE PERCEIVED INSTRUCTOR WAS j. THE BEST ESTIMATES OF μ_{11}, μ_{12}, μ_{21}, AND μ_{22} ARE PROVIDED BY R AS

$$\hat{\mu}_{11} = 2.85$$

$$\hat{\mu}_{12} = 2.85 + 0.87$$

$$\hat{\mu}_{21} = 2.85 - 0.06$$

$$\hat{\mu}_{22} = 2.85 - 0.06 + 0.87 - 0.1831$$

THUS, 0.87 IS AN ESTIMATE OF $\mu_{12} - \mu_{11}$.

THIS MEANS THAT WHEN JANE IS THE ACTUAL INSTRUCTOR, THE EXPECTED VALUE OF HER RATING IS ESTIMATED TO BE 0.87 POINTS HIGHER WHEN STUDENTS BELEIVE SHE IS JOHN THAN WHEN STUDENTS ARE TOLD THE TRUTH ABOUT HER IDENTITY. THE STANDARD ERROR OF 0.3184 SAYS THAT THE "TYPICAL SIZE" OF THE ERROR MADE WHEN ESTIMATING $\mu_{12} - \mu_{11}$ (USING THE METHOD WE'VE USED HERE) IS APPROXIMATELY 0.3184 UNITS. THE p-VALUE OF 0.00941 INDICATES THAT AN ESTIMATED DIFFERENCE AS LARGE OR LARGER THAN 0.87 WOULD BE UNLIKELY TO OCCUR IF μ_{12} WERE EQUAL TO μ_{11}. THUS, JANE WAS RATED SIGNIFICANTLY HIGHER BY THE STUDENTS WHO BELIEVED SHE WAS JOHN THAN BY STUDENTS WHO KNEW SHE WAS JANE.

3 b) THE RELEVANT ESTIMATE IS

$$\frac{\hat{\mu}_{11}+\hat{\mu}_{21}}{2} - \frac{\hat{\mu}_{12}+\hat{\mu}_{22}}{2} = \frac{2.85+2.85-0.06}{2}$$

$$- \frac{2.85+0.87+2.85-0.06+0.87-0.1831}{2}$$

$$= -0.87 + \frac{0.1831}{2}$$

THE VARIANCE OF THE ESTIMATOR IS

$$\frac{1}{4}\left[\mathrm{VAR}(\hat{\mu}_{11}) + \mathrm{VAR}(\hat{\mu}_{21}) + \mathrm{VAR}(\hat{\mu}_{12}) + \mathrm{VAR}(\hat{\mu}_{22})\right]$$

$$= \frac{1}{4}\left[\frac{\sigma^2}{10} + \frac{\sigma^2}{10} + \frac{\sigma^2}{10} + \frac{\sigma^2}{13}\right] = \frac{\sigma^2}{4}\left(3/10 + 1/13\right)$$

WE WILL NEED $\hat{\sigma}^2$. NOTE $0.2252 = se(\hat{\mu}_{11}) = \sqrt{\hat{\sigma}^2/10}$.

THUS, $\hat{\sigma}^2 = 10 \times (0.2252)^2$. THE CONFIDENCE INTERVAL IS

$$-0.87 + \frac{0.1831}{2} \pm t_{43-4,\,0.975}\sqrt{\frac{10\times(0.2252)^2}{4}\left(3/10 + 1/13\right)}$$

4a) THE NORMAL EQUATIONS ARE

$$X'X\underline{b} = X'\underline{y}.$$

IF $\underline{b}_1$ IS A SOLUTION TO THE NORMAL EQUATIONS, THEN $\underline{b}_2$ WILL ALSO BE A SOLUTION TO THE NORMAL EQUATIONS IF $X\underline{b}_1 = X\underline{b}_2$. THIS SHOULD BE CLEAR BECAUSE $X\underline{b}_1 = X\underline{b}_2 \Rightarrow X'X\underline{b}_1 = X'X\underline{b}_2$.

LET $\underline{b}_1$ BE THE SOLUTION PROVIDED BY SAS.

EACH ELEMENT OF $X\underline{b}_1$ HAS THE FORM

$$\hat{\beta}_0 + \hat{\beta}_1 x + \hat{\phi}_i + \hat{\delta}_j + \hat{\gamma}_{ij},$$ WHERE

$$\underline{b}_1 = [\hat{\beta}_0, \hat{\beta}_1, \hat{\phi}_1, \hat{\phi}_2, \hat{\delta}_1, \hat{\delta}_2, \hat{\delta}_3, \hat{\gamma}_{11}, \hat{\gamma}_{12}, \hat{\gamma}_{13}, \hat{\gamma}_{21}, \hat{\gamma}_{22}, \hat{\gamma}_{23}]'.$$

WE CAN GET ANOTHER SOLUTION AS

$$\underline{b}_2 = [\hat{\beta}_0 - 1, \hat{\beta}_1, \hat{\phi}_1 + 1, \hat{\phi}_2 + 1, \hat{\delta}_1, \hat{\delta}_2, \hat{\delta}_3, \hat{\gamma}_{11}, \hat{\gamma}_{12}, \hat{\gamma}_{13}, \hat{\gamma}_{21}, \hat{\gamma}_{22}, \hat{\gamma}_{23}]'$$

BECAUSE $\hat{\beta}_0 + \hat{\beta}_1 x + \hat{\phi}_i + \hat{\delta}_j + \hat{\gamma}_{ij} = \hat{\beta}_0 - 1 + \hat{\beta}_1 x + \hat{\phi}_i + 1 + \hat{\delta}_j + \hat{\gamma}_{ij} \quad \forall\, i, j, x$

SO THAT $X\underline{b}_1 = X\underline{b}_2$ AND $X'X\underline{b}_1 = X'X\underline{b}_2$.

THUS,

$$\begin{bmatrix} 56.79-1 \\ 0.57 \\ -2.50+1 \\ 0.00+1 \\ 2.91 \\ -1.18 \\ 0.00 \\ 3.66 \\ 0.22 \\ 0.00 \\ 0.00 \\ 0.00 \\ 0.00 \end{bmatrix}$$

IS ONE OF INFINITELY MANY SOLUTIONS TO THE NORMAL EQUATIONS THAT IS DIFFERENT FROM THE SAS SOLUTION.

4 b) $$F = \frac{(11.25 + 191.60 + 20.75)/(1+2+2)}{551.68/23}$$

ALTERNATIVELY,

$$F = \frac{[(839.86 - 64.58) - 551.68]/(28-23)}{551.68/23}$$

4 c)

SOURCE	SS	DF
X	64.58	1
ERROR	839.86 − 64.58	30 − 2
C. TOTAL	839.86	30 − 1

$$F = \frac{64.58}{(839.86-64.58)/(30-2)} \Rightarrow |t| = \sqrt{\frac{64.58}{(839.86-64.58)/(30-2)}}$$