STAT 510 SPRING 2014

EXAM 1 SOLUTIONS

1a) THE DATASET INCLUDES TWO MICE WHO RECEIVED DIET 2 AND WERE HOUSED IN CAGES WITHOUT RUNNING WHEELS. WE CAN SEE FROM THE MODEL THAT THE MEAN RESPONSE FOR THESE MICE IS ASSUMED TO BE THE SAME FOR BOTH MICE.

$$E(y_{211}) = E(y_{212}) = [1, -1, 1] \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}$$

$$= \beta_1 - \beta_2 + \beta_3$$

1a) (CONTINUED)

SOME OF YOU PROVIDED $\frac{Y_{211}+Y_{212}}{2}$, WHICH IS THE MEAN OF THE OBSERVED RESPONSES FOR THE TWO MICE IN THE EXPERIMENT. THIS IS NOT "THE MEAN RESPONSE OF A MOUSE WHO RECEIVED DIET 2 AND WAS HOUSED IN A CAGE WITHOUT A RUNNING WHEEL" ACCORDING TO THE MODEL.

$\frac{Y_{211}+Y_{212}}{2}$ INVOLVES THE DATA RATHER THAN OUR MODEL FOR THE DATA.

$\boxed{\text{5 POINTS}}$

1b) $X'X = \begin{bmatrix} 8 & 0 & 0 \\ 0 & 8 & 0 \\ 0 & 0 & 8 \end{bmatrix}$ $(X'X)^{-1} = \frac{1}{8} \underset{3\times3}{I}$

$$X'\underset{\sim}{y} = \begin{bmatrix} y_{\cdot\cdot\cdot} \\ y_{1\cdot\cdot} - y_{2\cdot\cdot} \\ y_{\cdot1\cdot} - y_{\cdot2\cdot} \end{bmatrix}$$

$$\hat{\underset{\sim}{\beta}} = (X'X)^{-1}X'\underset{\sim}{y} = \begin{bmatrix} \bar{y}_{\cdot\cdot\cdot} \\ \frac{\bar{y}_{1\cdot\cdot} - \bar{y}_{2\cdot\cdot}}{2} \\ \frac{\bar{y}_{\cdot1\cdot} - \bar{y}_{\cdot2\cdot}}{2} \end{bmatrix}$$

$\boxed{\text{10 POINTS}}$

1c) $[1\ -1\ 1]\,\hat{\underset{\sim}{\beta}} = \hat{\beta}_1 - \hat{\beta}_2 + \hat{\beta}_3$

$$= \bar{Y}_{\cdot\cdot\cdot} - \left(\frac{\bar{Y}_{1\cdot\cdot} - \bar{Y}_{2\cdot\cdot}}{2}\right) + \left(\frac{\bar{Y}_{\cdot1\cdot} - \bar{Y}_{\cdot2\cdot}}{2}\right)$$

$$= \frac{\bar{Y}_{1\cdot\cdot}}{2} + \frac{\bar{Y}_{2\cdot\cdot}}{2} - \frac{\bar{Y}_{1\cdot\cdot}}{2} + \frac{\bar{Y}_{2\cdot\cdot}}{2} + \frac{\bar{Y}_{\cdot1\cdot}}{2} - \frac{\bar{Y}_{\cdot2\cdot}}{2}$$

$$= \bar{Y}_{2\cdot\cdot} + \frac{\bar{Y}_{\cdot1\cdot}}{2} - \frac{\bar{Y}_{\cdot2\cdot}}{2} - \frac{\bar{Y}_{\cdot1\cdot}}{2} + \frac{\bar{Y}_{\cdot1\cdot}}{2}$$

$$= \bar{Y}_{2\cdot\cdot} + \bar{Y}_{\cdot1\cdot} - \bar{Y}_{\cdot\cdot\cdot}$$

$\boxed{\text{6 POINTS}}$

1d)

	NO EXERCISE	EXERCISE	
DIET 1	$\beta_1+\beta_2+\beta_3$	$\beta_1+\beta_2-\beta_3$	$\beta_1+\beta_2$
DIET 2	$\beta_1-\beta_2+\beta_3$	$\beta_1-\beta_2-\beta_3$	$\beta_1-\beta_2$

$H_0: \beta_1+\beta_2 = \beta_1-\beta_2 \iff \boxed{H_0: \beta_2 = 0}$ $\boxed{\text{6 POINTS}}$

1e) $\hat{\beta}_1 - \hat{\beta}_2 = \bar{Y}_{\cdot\cdot\cdot} - \dfrac{\bar{Y}_{1\cdot\cdot} - \bar{Y}_{2\cdot\cdot}}{2} = \boxed{\bar{Y}_{2\cdot\cdot}}$ $\boxed{\text{6 POINTS}}$

2 a) B1: $\dfrac{\left(\frac{7+9}{2}\right)+2}{2} = 5$

B2: $\dfrac{\left(\frac{8+10}{2}\right)+\left(\frac{0+2}{2}\right)}{2} = 5$

B3: $\dfrac{12+\left(\frac{0+4}{2}\right)}{2} = 7$

$\boxed{\text{6 POINTS}}$

THE MOST COMMON MISTAKE HERE WAS TO COMPUTE MEANS INSTEAD OF LSMEANS.

2 b) THE PROBLEM TESTS TO SEE IF YOU UNDERSTAND HOW TO TEST FOR THE MAIN EFFECT OF A FACTOR IN AN UNBALANCED TWO-FACTOR EXPERIMENT. WE ARE GIVEN ALL THE DATA HERE AND DON'T ACTUALLY NEED THE R CODE AND OUTPUT AT ALL TO DO THIS PROBLEM. YOU ARE ASKED TO ASSUME THE CELL MEANS MODEL WITH CONSTANT VARIANCE. YOU ONLY NEED THE FOLLOWING INFORMATION:

2b) (CONTINUED)

$\overline{Y}_{11\cdot} = \frac{7+9}{2} = 8$ $\quad$ $\overline{Y}_{12\cdot} = \frac{8+10}{2} = 9$ $\quad$ $\overline{Y}_{13\cdot} = 12$

$\overline{Y}_{21\cdot} = 2$ $\quad$ $\overline{Y}_{22\cdot} = \frac{0+2}{2} = 1$ $\quad$ $\overline{Y}_{23\cdot} = \frac{0+4}{2} = 2$

$n_{11} = 2$ $\quad$ $n_{12} = 2$ $\quad$ $n_{13} = 1$

$n_{21} = 1$ $\quad$ $n_{22} = 2$ $\quad$ $n_{23} = 2$

$$SSE = (7-8)^2 + (9-8)^2 + (8-9)^2 + (10-9)^2 + (12-12)^2$$
$$+ (2-2)^2 + (0-1)^2 + (2-1)^2 + (0-2)^2 + (4-2)^2$$

$$= 14$$

$$MSE = \frac{SSE}{10-6} = \frac{14}{4} = 7/2 = 3.5$$

2b) (CONTINUED)

NOW THE TEST FOR A FACTOR A MAIN EFFECT IS A TEST OF $H_0: \frac{\mu_{11}+\mu_{12}+\mu_{13}}{3} = \frac{\mu_{21}+\mu_{22}+\mu_{23}}{3}$

WHICH IS EQUIVALENT TO A TEST OF

$H_0: \mu_{11}+\mu_{12}+\mu_{13}-\mu_{21}-\mu_{22}-\mu_{23}=0.$

WE COULD TEST WHETHER THIS LINEAR COMBINATION IS ZERO WITH

$$t = \frac{\bar{Y}_{11.}+\bar{Y}_{12.}+\bar{Y}_{13.}-\bar{Y}_{21.}-\bar{Y}_{22.}-\bar{Y}_{23.}}{\sqrt{MSE\left(\frac{1}{n_{11}}+\frac{1}{n_{12}}+\frac{1}{n_{13}}+\frac{(-1)^2}{n_{21}}+\frac{(-1)^2}{n_{22}}+\frac{(-1)^2}{n_{23}}\right)}}.$$

2 b) (CONTINUED)

THE F STATISTIC IS t^2:

$$F = \frac{(8+9+12-2-1-2)^2}{3.5\left(\frac{1}{2}+\frac{1}{2}+1+1+\frac{1}{2}+\frac{1}{2}\right)} = \frac{24^2}{14} = \frac{144}{3.5}$$

I SPENT QUITE A BIT OF TIME IN CLASS TRYING TO WARN YOU NOT TO USE THE LINE FOR FACTOR A IN THE R ANOVA TABLE TO TEST FOR FACTOR A MAIN EFFECTS. IN THIS CASE, IT GIVES A NUMERICALLY SIMILAR ANSWER, BUT THIS IS NOT THE TEST FOR FACTOR A MAIN EFFECT.

2b) (CONTINUED)

THE TEST THAT R REPORTS FOR "A2" IS ALSO NOT THE TEST FOR FACTOR A MAIN EFFECT.

RATHER, THIS TESTS WHETHER $\mu_{11} = \mu_{21}$

BECAUSE R'S PARAMETERIZATION IS

	B1	B2	B3
A1	INT	INT + B2	INT + B3
A2	INT + A2	INT + A2 + B2 + A2:B2	INT + A2 + B3 + A2:B3

THE DIFFERENCE OF THESE CELLS IS A2.

I EXPECTED MOST OF YOU TO COMPLETE THE PROBLEM AS FOLLOWS:

2 b) $F = \dfrac{\hat{\beta}' C' \left[C (X'X)^{-1} C' \right]^{-1} C \hat{\beta} / 1}{MSE}$,

WHERE $C = [1, 1, 1, -1, -1, -1]$,

$$\hat{\beta} = \begin{bmatrix} (7+9)/2 \\ (8+10)/2 \\ 12 \\ 2 \\ (0+2)/2 \\ (0+4)/2 \end{bmatrix} = \begin{bmatrix} 8 \\ 9 \\ 12 \\ 2 \\ 1 \\ 2 \end{bmatrix}, \quad X = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

$$X'X = \begin{bmatrix} 2 & & & & & \\ & 2 & & & \text{\Large 0} & \\ & & 1 & & & \\ & & & 1 & & \\ & \text{\Large 0} & & & 2 & \\ & & & & & 2 \end{bmatrix}, \quad (X'X)^{-1} = \mathrm{DIAG}\left(\tfrac{1}{2}, \tfrac{1}{2}, 1, 1, \tfrac{1}{2}, \tfrac{1}{2}\right),$$

2 b) (CONTINUED)

$$C(X'X)^{-1}C' = \tfrac{1}{2} + \tfrac{1}{2} + 1 + 1 + \tfrac{1}{2} + \tfrac{1}{2} = 4,$$

$$\left[C(X'X)^{-1}C'\right]^{-1} = \frac{1}{4},$$

$$F = \frac{(8+9+12-2-1-2)^2}{4\,(3.5)} = \frac{(24)^2}{4\,(3.5)} = \frac{144}{3.5}. \quad \boxed{\text{12 POINTS}}$$

2c) $F = 0.7619$ (FROM THE LAST LINE OF THE ANOVA TABLE)

$\boxed{\text{8 POINTS}}$

3a) THIS IS THE DECREASE IN ERROR SUM OF SQUARES THAT RESULTS WHEN X_1 IS ADDED AS A LINEAR TERM TO A MODEL THAT CONTAINS ONLY AN INTERCEPT. $\boxed{\text{6 POINTS}}$

SOME OF YOU WROTE ABOUT CHANGES IN SUM OF SQUARES WITHOUT SPECIFYING WHICH SUM OF SQUARES YOU WERE REFERRING TO. OTHERS WERE NOT SPECIFIC ABOUT WHICH MODEL X_1 WAS ADDED TO.

3b) SOME OF YOU DIDN'T CATCH THAT "model (1)" WAS THE FULL MODEL IN THE FULL MODEL EQUATION THAT WAS LABELED WITH THE EQUATION LABEL "(1)". IF YOU CALCULATED MSE FOR SOME OTHER MODEL, I GAVE YOU FULL CREDIT.

ALL I WAS LOOKING FOR WAS

$\hat{\sigma}^2 = MSE = 83$ $\boxed{\text{6 POINTS}}$

3c) I EXPECTED ALL OF YOU TO KNOW FROM OUR ANOVA SLIDES THAT THE NCP IS $\frac{1}{2\sigma^2}\underset{\sim}{\beta}'X'(P_5 - P_4)X\underset{\sim}{\beta}$, WHERE

$P_4 = X_4(X_4'X_4)^{-}X_4'$, $X_4 = [\underset{\sim}{1}, \underset{\sim}{x}_1, \underset{\sim}{x}_2, \underset{\sim}{x}_3]$,

$P_5 = X_5(X_5'X_5)^{-}X_5'$, $X_5 = X = [\underset{\sim}{1}, \underset{\sim}{x}_1, \underset{\sim}{x}_2, \underset{\sim}{x}_3, \underset{\sim}{x}_4]$.

REPORTING THIS WAS WORTH $\boxed{\text{5 POINTS}}$

THE OTHER $\boxed{\text{5 POINTS}}$ COULD BE EARNED BY SIMPLIFYING THE EXPRESSION FOR THIS SPECIAL CASE. (SEE NEXT PAGE)

3 c) (CONTINUED)

$$\frac{1}{2\sigma^2}\underline{\beta}'X'(P_5-P_4)X\underline{\beta} = \frac{1}{2\sigma^2}\underline{\beta}'X'(P_5-P_4)'(P_5-P_4)X\underline{\beta}$$

$$= \frac{1}{2\sigma^2}\|(P_5-P_4)X\underline{\beta}\|^2 = \frac{1}{2\sigma^2}\|P_5X\underline{\beta} - P_4X\underline{\beta}\|^2$$

$$= \frac{1}{2\sigma^2}\|X\underline{\beta} - P_4[\underline{1}, \underline{x}_1, \underline{x}_2\, \underline{x}_3, \underline{x}_4]\underline{\beta}\|^2$$

$$= \frac{1}{2\sigma^2}\|X\underline{\beta} - [P_4\underline{1}, P_4\underline{x}_1, P_4\underline{x}_2, P_4\underline{x}_3, P_4\underline{x}4]\underline{\beta}\|^2$$

$$= \frac{1}{2\sigma^2}\|X\underline{\beta} - [\underline{1}, \underline{x}_1, \underline{x}_2, \underline{x}_3, P_4\underline{x}_4]\underline{\beta}\|^2$$

$$= \frac{1}{2\sigma^2}\|\beta_0\underline{1} + \beta_1\underline{x}_1 + \beta_2\underline{x}_2 + \beta_3\underline{x}_3 + \beta_4\underline{x}_4 - \beta_0\underline{1} - \beta_1\underline{x}_1 - \beta_2\underline{x}_2 - \beta_3\underline{x}_3 - \beta_4P_4\underline{x}_4\|^2$$

3c) (CONTINUED)

$$= \frac{1}{2\sigma^2} \| \beta_4 \underset{\sim}{x}_4 - \beta_4 P_4 \underset{\sim}{x}_4 \|^2$$

$$= \frac{\beta_4^2}{2\sigma^2} \| \underset{\sim}{x}_4 - P_4 \underset{\sim}{x}_4 \|^2$$

NOTE THAT WHEN $\underset{\sim}{x}_4$ IS A LINEAR COMBINATION OF $\underset{\sim}{1}, \underset{\sim}{x}_1, \underset{\sim}{x}_2, \underset{\sim}{x}_3$, THIS NCP WOULD BE ZERO,

HOWEVER, AS LONG AS $\underset{\sim}{x}_4$ IS NOT AN LC OF THE OTHER COLUMNS IN THE REGRESSION MODEL MATRIX,

THE NCP $= 0$ IF AND ONLY IF $\beta_4 = 0$.

3 d) $\dfrac{(172+14)/2}{83} = \dfrac{186/2}{83} = \dfrac{93}{83}$ $\boxed{\text{9 POINTS}}$

3 e) $\dfrac{2033}{(172+14+214109)/2581}$ $\boxed{\text{10 POINTS}}$