STAT 510
SPRING 2017
EXAM 1 SOLUTIONS

POINTS PER PROBLEM

1 a) 9	2 a) 9	3. 10
b) 9	b) 9	
c) 6	c) 5	
d) 10	d) 10	
e) 13	e) 10	

1. a) 95% Confidence Interval for $\mu_{W,4} - \mu_{B,4}$:

$$\underset{\sim}{c}' \hat{\underset{\sim}{\beta}} \pm 1.99 \sqrt{\hat{\sigma}^2 \underset{\sim}{c}' (X'X)^{-1} \underset{\sim}{c}}$$

$X'X = 10 \underset{8\times 8}{I}$ $\quad$ $\underset{\sim}{c}' = [0,0,0,-1,0,0,0,1]$

$(X'X)^{-1} = \frac{1}{10} I$ $\quad$ $\underset{\sim}{\beta}' = [\mu_{B,1}, \mu_{B,2}, \ldots, \mu_{W,4}]$

$$\hat{\mu}_{W,4} - \hat{\mu}_{B,4} \pm 1.99 \sqrt{18.3 \left(\frac{1}{10} + \frac{1}{10}\right)}$$

$$24.5 - 22.1 \pm 1.99 \sqrt{3.66}$$

1 b) $$t = \frac{\underset{\sim}{c}' \hat{\underset{\sim}{\beta}}}{\sqrt{\hat{\sigma}^2 \underset{\sim}{c}' (X'X)^{-1} \underset{\sim}{c}}}$$

$\underset{\sim}{c}' = (1, -1, 0, 0, -1, 1, 0, 0)$

$$t = \frac{23.0 - 27.2 - 27.8 + 23.2}{\sqrt{18.3 \; 4/10}}$$

1c) $$t = \frac{\underset{\sim}{c}'\hat{\underset{\sim}{\beta}}}{\sqrt{\hat{\sigma}^2 \underset{\sim}{c}'(X'X)^- \underset{\sim}{c}}} = \frac{\underset{\sim}{c}'\hat{\underset{\sim}{\beta}} - \underset{\sim}{c}'\underset{\sim}{\beta} + \underset{\sim}{c}'\underset{\sim}{\beta}}{\sqrt{\frac{\hat{\sigma}^2}{\sigma^2}\sigma^2 \underset{\sim}{c}'(X'X)^- \underset{\sim}{c}}}$$

$$= \frac{\dfrac{\underset{\sim}{c}'\hat{\underset{\sim}{\beta}} - \underset{\sim}{c}'\underset{\sim}{\beta}}{\sqrt{\sigma^2 \underset{\sim}{c}'(X'X)^- \underset{\sim}{c}}} + \dfrac{\underset{\sim}{c}'\underset{\sim}{\beta}}{\sqrt{\sigma^2 \underset{\sim}{c}'(X'X)^- \underset{\sim}{c}}}}{\sqrt{\hat{\sigma}^2/\sigma^2}}$$

$$= \frac{z + \delta}{\sqrt{w/(n-r)}}$$ WHERE

$$z \equiv \frac{\underset{\sim}{c}'\hat{\underset{\sim}{\beta}} - \underset{\sim}{c}'\underset{\sim}{\beta}}{\sqrt{\sigma^2 \underset{\sim}{c}'(X'X)^- \underset{\sim}{c}}} \sim N(0,1)$$ INDEPENDENT

OF $$W = \frac{(n-r)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-r}$$ AND

$$\delta = \frac{\underline{c}'\underline{\beta}}{\sqrt{\sigma^2 \underline{c}'(X'X)^{-}\underline{c}}}.$$

Thus, $$NCP = \frac{\underline{c}'\underline{\beta}}{\sqrt{\sigma^2 \underline{c}'(X'X)^{-}\underline{c}}}.$$

FOR A t DISTRIBUTION, IT IS EASY TO DIRECTLY WRITE DOWN THE NCP BY REPLACING ESTIMATORS IN THE STATISTIC WITH THE PARAMETERS THEY ESTIMATE. IN THIS PROBLEM, WE HAVE

$$NCP = \frac{\mu_{B,1} - \mu_{B,2} - \mu_{W,1} + \mu_{W,2}}{\sqrt{2\sigma^2/5}}$$

I WAS EXPECTING STUDENTS TO WRITE DOWN THIS ANSWER DIRECTLY WITHOUT ALL THE WORK ABOVE.

1 d) LET $X_1 = \underline{1}$ AND X_2 BE THE MODEL MATRIX THAT ALLOWS A DISTINCT MEAN FOR EACH GROUP 1, 2, 3, AND 4. $\text{RANK}(X_2) - \text{RANK}(X_1) = 3$.

$$\underset{\sim}{y}'(P_2 - P_1)\underset{\sim}{y} = \|P_2 \underset{\sim}{y} - P_1 \underset{\sim}{y}\|^2$$

$$= 20(\bar{y}_{\cdot 1} - \bar{y}_{\cdot\cdot})^2 + 20(\bar{y}_{\cdot 2} - \bar{y}_{\cdot\cdot})^2 + 20(\bar{y}_{\cdot 3} - \bar{y}_{\cdot\cdot})^2$$

$$+ 20(\bar{y}_{\cdot 4} - \bar{y}_{\cdot\cdot})^2$$

$$= 20[(25.4 - 24.75)^2 + (25.2 - 24.75)^2 + (25.1 - 24.75)^2 + (23.3 - 24.75)^2]$$

$$= 57$$

BECAUSE DATA ARE BALANCED WE KNOW ANOVA F-STAT FOR GROUP IS F STAT FOR TESTING GROUP MAIN EFFECTS:

$$F = \frac{\underset{\sim}{y}'(P_2 - P_1)\underset{\sim}{y} / 3}{MSE} = \frac{57/3}{18.3}$$

WHAT FOLLOWS IS AN ALTERNATIVE APPROACH TO FINDING F BASED ON TEST OF $H_0 : C\underset{\sim}{\beta} = 0$.

$$H_0: \bar{\mu}_{\cdot 1} = \bar{\mu}_{\cdot 2} = \bar{\mu}_{\cdot 3} = \bar{\mu}_{\cdot 4}$$

$$\iff \bar{\mu}_{\cdot 1} = \bar{\mu}_{\cdot 2} \text{ AND } \bar{\mu}_{\cdot 3} = \bar{\mu}_{\cdot 4} \text{ AND}$$

$$\frac{\bar{\mu}_{\cdot 1} + \bar{\mu}_{\cdot 2}}{2} = \frac{\bar{\mu}_{\cdot 3} + \bar{\mu}_{\cdot 4}}{2}$$

$$\iff C\underline{\beta} = \underset{\sim}{0} \quad \text{WHERE}$$

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

$$\underline{\beta}' = [\mu_{B,1}, \mu_{B,2}, \mu_{B,3}, \mu_{B,4}, \mu_{W,1}, \mu_{W,2}, \mu_{W,3}, \mu_{W,4}]$$

$$C(X'X)^{-1}C' = \frac{1}{10} CC' = \frac{1}{10}\begin{bmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 8 \end{bmatrix}$$

$$\left(C(X'X)^{-}C'\right)^{-1} = \frac{5}{4}\begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$C\hat{\beta} = \begin{bmatrix} 50.8 - 50.4 \\ 50.2 - 46.6 \\ 50.8 + 50.4 - 50.2 - 46.6 \end{bmatrix} = \begin{bmatrix} 0.4 \\ 3.6 \\ 4.4 \end{bmatrix}$$

$$(C\hat{\beta})'[C(X'X)^{-1}C']^{-1}C\hat{\beta}$$

$$= 5/4\,(2 * .4^2 + 2 * 3.6^2 + 4.4^2)$$

$$F = \frac{5/4\,(2 * .4^2 + 2 * 3.6^2 + 4.4^2)/3}{18.3} = \frac{57/3}{18.3}$$

1e)

Source	DF	SS
Race	1	
Group	3	57 ← From Part d)
Race x Group	3	
Error	72	18.3 * 72 ← Given MSE
C. Total	79	

The above is straightforward. Also, we know

$SS_{RACE} + SS_{GROUP} + SS_{RACE \times GROUP} = 2.95 \times 7 \times 18.3$

because the overall F statistic is

$$2.95 = \frac{\underline{y}'(P_4 - P_1)\underline{y}/7}{MSE}$$

GIVEN IN PROBLEM

$$= \frac{\left[\underline{y}'(P_4-P_3)\underline{y} + \underline{y}'(P_3-P_2)\underline{y} + \underline{y}'(P_2-P_1)\underline{y}\right]/7}{MSE}$$

$$= \frac{\left[SS_{RACE \times GROUP} + SS_{RACE} + SS_{GROUP}\right]/7}{18.3}.$$

IF WE CAN GET SS_{RACE}, THE REST WILL FOLLOW.

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

$$C\hat{\underline{\beta}} = 23 + 27.2 + 27.3 + 22.1 - 27.8 - 23.2 - 22.9 - 24.5$$
$$= 99.6 - 98.4 = 1.2$$

$$C(X'X)^{-1}C' = 8/10 \Rightarrow [C(X'X)^{-1}C']^{-1} = 5/4$$

$$SS_{RACE} = (C\hat{\underline{\beta}})'[C(X'X)^{-1}C']^{-1}C\hat{\underline{\beta}} = \frac{5 \times 1.2^2}{4} = 1.8$$

SO WE HAVE

SOURCE	DF	SS
RACE	1	1.8
GROUP	3	57
RACE X GROUP	3	$2.95 \times 7 \times 18.3 - SS_{RACE} - SS_{GROUP}$
ERROR	72	18.3×72
C. TOTAL	79	SUM OF ALL THE ABOVE OR $18.3 \times 72 + 2.95 \times 7 \times 18.3$

2 a) $SS_{C.TOTAL} = 9.63 + 6.37 = 16 = y'(I-P_1)y$

THIS IS SSE FOR INTERCEPT ONLY MODEL.

$$y'(P_2-P_1)y = 0.19$$

$$y'(I-P_2)y = y'(I-P_1)y - y'(P_2-P_1)y$$

$$= 16 - 0.19 = 15.81$$

$$\hat{\sigma}^2 = \frac{y'(I-P_2)y}{9-2} = \frac{15.81}{7}$$

2 b) $F = \dfrac{0.19/1}{15.81/7}$

2 c) 142.124

2 d)

SOURCE	DF	SS
TRT	2	9.63
x	1	?
ERROR	5	0.27
C. TOTAL	8	16.0

$$SS_x = 16 - (9.63 + 0.27) = 16 - 9.9 = 6.1$$

$$F = \frac{6.1}{0.27/5} \implies |t| = \sqrt{\frac{30.5}{0.27}}$$

INSPECTION OF DATA SHOWS THAT y DECREASES AS x INCREASES WITHIN EACH TREATMENT GROUP. THUS, $\hat{\alpha}_2 < 0$, AND $t = -\sqrt{30.5/0.27}$

2e) This problem asks for a brief report "for the researchers". Thus, I was hoping you would focus on conclusions relevant for the researchers, and explain those conclusions in a way that might make sense to a scientifically literate non-statistician.

Even without a calculator, you should be able to see that the t-stat computed in part (d) is very large in magnitude. Thus, both pre-treatment weight and treatment (see part c) are highly statistically significant in a model that includes both. We can also see that an additive model seems adequate compared to a more complex model that allows for interaction between pre-treatment weight and treatment. This additive model says the expected value of the response is a linear function of pretreatment weight within each treatment group. Thus, There are three regression lines, one for each treatment group. The slopes of the three lines are the same in the additive model, but the intercepts differ across treatment groups. We could say something like the following to the researchers.

"Both pre-treatment weight and treatment have a statistically significant association with the chemical level response. Within each treatment group, larger pre-treatment weights are associated with lower levels of the chemical. For any particular pretreatment weight held constant across the treatment groups, the mean level of the chemical differs in a statistically significant way across treatment groups, and the differences among treatment groups in mean response appear similar, regardless of which pre-treatment weight we consider."

3. A IS THE MODEL MATRIX FOR A QUADRATIC REGRESSION FUNCTION FOR DATA WITH TWO OBSERVATIONS AT EACH OF $x=1$, $x=2$, AND $x=3$.

BECAUSE A QUADRATIC FUNCTION CAN PASS THROUGH THE THREE POINTS $(1,\mu_1)$, $(2,\mu_2)$, $(3,\mu_3)$ FOR ANY μ_1, μ_2, μ_3, THIS IS JUST ANOTHER VERSION OF THE CELL MEANS MODEL MATRIX WITH THREE MEANS, i.e.,

$\mathcal{C}(A) = \mathcal{C}(X)$, WHERE

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

THUS, $P_A = P_X$

$$P_X = X(X'X)^{-1}X'$$

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

NOW NOTE THAT $\mathcal{C}(A) \subset \mathcal{C}(B)$ BECAUSE

$$B \begin{bmatrix} 1 & 2 & 4 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} = A.$$

THIS CAN ALSO BE SEEN BY NOTING THAT THE FIRST 3 COLUMNS OF B ARE LIKE A BUT WITH A CENTERED X VARIABLE $(x - \bar{x})$.

THUS, $P_A P_B = P_A = P_X$.