EXAM 1 SOLUTIONS

SPRING 2019

PROBLEM	POINTS POSSIBLE
1	10
2	8
3a)	12
3b)	6
3c)	5
3d)	5
3e)	5
3f)	12
4a)	6
4b)	10
4c)	11
4d)	10

1. $\mathcal{C}(X) = \mathcal{C}(W) \Longrightarrow XA = W$ FOR SOME A,

$WB = X$ FOR SOME B.

THUS,
$$\begin{aligned} P_X P_W &= P_X W(W'W)^- W' \\ &= P_X XA(W'W)^- W' \\ &= XA(W'W)^- W' \\ &= W(W'W)^- W' \\ &= P_W \end{aligned}$$

LIKEWISE $P_W P_X = P_W WB(X'X)^- X' = WB(X'X)^- X' = P_X$

NOW
$$\begin{aligned} (P_X - P_W)'(P_X - P_W) &= (P_X - P_W)(P_X - P_W) \\ &= P_X P_X - P_X P_W - P_W P_X + P_W P_W \\ &= P_X - P_W - P_X + P_W = 0. \end{aligned}$$

THEREFORE, $P_X - P_W = 0$, WHICH IMPLIES

$P_X = P_W$.

1. ALTERNATIVE SOLUTION.

$\forall \underset{\sim}{z} \in \mathbb{R}^n$, $P_X \underset{\sim}{z} = P_W \underset{\sim}{z}$ IS THE UNIQUE VECTOR IN $\mathcal{C}(X) = \mathcal{C}(W)$ THAT IS CLOSEST TO $\underset{\sim}{z}$ WITH RESPECT TO EUCLIDEAN DISTANCE.

(WE PROVED THERE EXISTS ONLY ONE VECTOR IN $\mathcal{C}(X) = \mathcal{C}(W)$ THAT IS CLOSEST TO $\underset{\sim}{z}$ IN HW2, PROBLEM 2.)

$$P_X \underset{\sim}{z} = P_W \underset{\sim}{z} \quad \forall \underset{\sim}{z} \in \mathbb{R}^n$$

$$\Rightarrow P_X \underset{\sim}{e}_i = P_W \underset{\sim}{e}_i \quad \forall i = 1, \dots, n$$

WHERE $\underset{\sim}{e}_i$ IS THE ith COLUMN OF $I_{n \times n}$.

$$\Rightarrow P_X I = P_W I$$

$$\Rightarrow P_X = P_W.$$

2. $E\left[(\hat{\theta}-\theta)^2\right]$

$$= E\left[\left(\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta\right)^2\right]$$

$$= E\left[(\hat{\theta} - E(\hat{\theta}))^2 + (E(\hat{\theta}) - \theta)^2 + 2(\hat{\theta} - E(\hat{\theta}))(E(\hat{\theta}) - \theta)\right]$$

$$= E(\hat{\theta} - E(\hat{\theta}))^2 + (E(\hat{\theta}) - \theta)^2 + 2(E(\hat{\theta}) - \theta)\, E(\hat{\theta} - E(\hat{\theta}))$$

$$= \text{VAR}(\hat{\theta}) + [\text{BIAS}(\hat{\theta})]^2 + 2(E(\hat{\theta}) - \theta)\underbrace{(E(\hat{\theta}) - E(\hat{\theta}))}_{0}$$

$$= \text{VAR}(\hat{\theta}) + [\text{BIAS}(\hat{\theta})]^2$$

3 a) $\theta = E(y_{3i}) = \beta_1 + \beta_2(4-4) = \beta_1$

$$X'X = \begin{bmatrix} 20 & 0 \\ 0 & 5[(-4)^2 + (-2)^2 + 0^2 + 6^2] \end{bmatrix}$$

$$= \begin{bmatrix} 20 & 0 \\ 0 & 280 \end{bmatrix} \Rightarrow (X'X)^{-1} = \begin{bmatrix} 1/20 & 0 \\ 0 & 1/280 \end{bmatrix}$$

$$X'y = \begin{bmatrix} 20\bar{y}_{..} \\ 5(-4\bar{y}_{1.} - 2\bar{y}_{2.} + 6\bar{y}_{4.}) \end{bmatrix}$$

$$(X'X)^{-1}X'y = \begin{bmatrix} \bar{y}_{..} \\ \dfrac{-2\bar{y}_{1.} - \bar{y}_{2.} + 3\bar{y}_{4.}}{28} \end{bmatrix}$$

$$\hat{\theta} = \hat{\beta}_1 = \bar{y}_{..}$$

3 b) WITH $\underline{c}' = [1, 0]$, WE HAVE

$$\mathrm{VAR}(\hat{\beta}_1) = \mathrm{VAR}(\underline{c}'\hat{\underline{\beta}}) = \sigma^2 \underline{c}'(X'X)^{-1}\underline{c}$$

$$= \sigma^2\left(\frac{1}{20}\right) = 36/20 = 9/5$$

c) $$E(\hat{\theta}) = E(\bar{Y}_{..}) = E\left[\frac{\bar{Y}_{1.} + \bar{Y}_{2.} + \bar{Y}_{3.} + \bar{Y}_{4.}}{4}\right]$$

$$= \frac{E(\bar{Y}_{1.}) + E(\bar{Y}_{2.}) + E(\bar{Y}_{3.}) + E(\bar{Y}_{4.})}{4}$$

$$= (\mu_1 + \mu_2 + \mu_3 + \mu_4)/4$$

$$= \frac{160 + 180 + 200 + 252}{4}$$

$$= 198$$

$$\mathrm{BIAS} = E(\hat{\theta}) - \theta = 198 - 200$$

$$= -2$$

d) $$9/5 + (-2)^2 = 5.8$$

$[\mathrm{VAR} + \mathrm{BIAS}^2]$

3e) $VAR(\hat{\mu}_3) = VAR(\bar{Y}_{3.}) = \frac{36}{5} = 7.2$

$$MSE(\hat{\mu}_3) = 7.2 + 0^2 = 7.2$$

$\left(E(\bar{Y}_{3.}) = \mu_3 = \theta \right.$ SO BIAS IS ZERO.$\left. \right)$

THUS, EVEN THOUGH

$\hat{\mu}_3 = \bar{Y}_{3.}$ IS THE BEST

LINEAR UNBIASED ESTIMATOR

OF θ, IT HAS HIGHER

MEAN SQUARED ERROR THAN THE

SLIGHTLY BIASED ESTIMATOR $\bar{Y}_{..}$

OBTAINED FROM THE FIT OF THE

INCORRECT MODEL.

3 f) $F = \underset{\sim}{y}'(P_X - P_{X_0})\underset{\sim}{y} / (r - r_0)$ over MSE:

$$F = \frac{\underset{\sim}{y}'(P_X - P_{X_0})\underset{\sim}{y} / (r - r_0)}{MSE}$$

$$\sim F_{r-r_0,\, n-r}\left(\frac{\underset{\sim}{\beta}' X' (P_X - P_{X_0}) X \underset{\sim}{\beta}}{2\sigma^2}\right)$$

$$\sim F_{r-r_0,\, n-r}\left(\frac{\|(P_X - P_{X_0}) X \underset{\sim}{\beta}\|^2}{2\sigma^2}\right)$$

$$\sim F_{r-r_0,\, n-r}\left(\frac{\| X\underset{\sim}{\beta} - P_{X_0} X \underset{\sim}{\beta}\|^2}{2\sigma^2}\right)$$

$n = 20$, $r = 4$, $r_0 = 2$

WITH $\underset{\sim}{1} = \underset{\sim}{1}_{5\times 1}$, WE HAVE

$$X\underset{\sim}{\beta} = \begin{bmatrix} \mu_1 \underset{\sim}{1} \\ \mu_2 \underset{\sim}{1} \\ \mu_3 \underset{\sim}{1} \\ \mu_4 \underset{\sim}{1} \end{bmatrix} = \begin{bmatrix} 160\, \underset{\sim}{1} \\ 180\, \underset{\sim}{1} \\ 200\, \underset{\sim}{1} \\ 252\, \underset{\sim}{1} \end{bmatrix}$$

AND

3f) (CONTINUED)

$$X_0 = \begin{bmatrix} \underset{\sim}{1} & -4\,\underline{1} \\ \underline{1} & -2\,\underline{1} \\ \underset{\sim}{1} & \underset{\sim}{0} \\ \underset{\sim}{1} & 6\,\underline{1} \end{bmatrix}.$$

BY THE SAME CALCULATIONS USED IN PART (a), WE HAVE

$$(X_0'X_0)^{-1}X_0'X\underline{\beta} = \begin{bmatrix} \bar{\mu}_{\cdot} \\ \dfrac{-2\mu_1 - \mu_2 + 3\mu_4}{28} \end{bmatrix}$$

$$= \begin{bmatrix} 198 \\ 9\tfrac{1}{7} \end{bmatrix}$$

THUS,

$$X_0(X_0'X_0)^{-1}X_0'X\underline{\beta} = \begin{bmatrix} (198 - 36\tfrac{4}{7})\,\underline{1} \\ (198 - 18\tfrac{2}{7})\,\underline{1} \\ 198\,\underline{1} \\ (198 + 54\tfrac{6}{7})\,\underline{1} \end{bmatrix} = \begin{bmatrix} 161\tfrac{3}{7}\,\underline{1} \\ 179\tfrac{5}{7}\,\underline{1} \\ 198\,\underline{1} \\ 252\tfrac{6}{7}\,\underline{1} \end{bmatrix}$$

3 f) (CONTINUED)

THUS,

$$X\underset{\sim}{\beta} - P_0 X\underset{\sim}{\beta} = \begin{bmatrix} -1\,3/7\,\underset{\sim}{1} \\ 2/7\,\underset{\sim}{1} \\ 2\,\underset{\sim}{1} \\ -6/7\,\underset{\sim}{1} \end{bmatrix}.$$

THUS, $\|X\underset{\sim}{\beta} - P_0 X\underset{\sim}{\beta}\|^2 = 5\left(10/7\right)^2 + 5\left(2/7\right)^2 + 5(2)^2 + 5\left(6/7\right)^2$

$$= 5\left[\frac{100+4+36}{49} + 4\right]$$

$$= 5\left[48/7\right] = \frac{240}{7}$$

THE NONCENTRALITY PARAMETER IS THEN $\frac{240}{2 \times 36 \times 7} = 10/21$. So $F_{2,16}(10/21)$

3f) (CONTINUED)

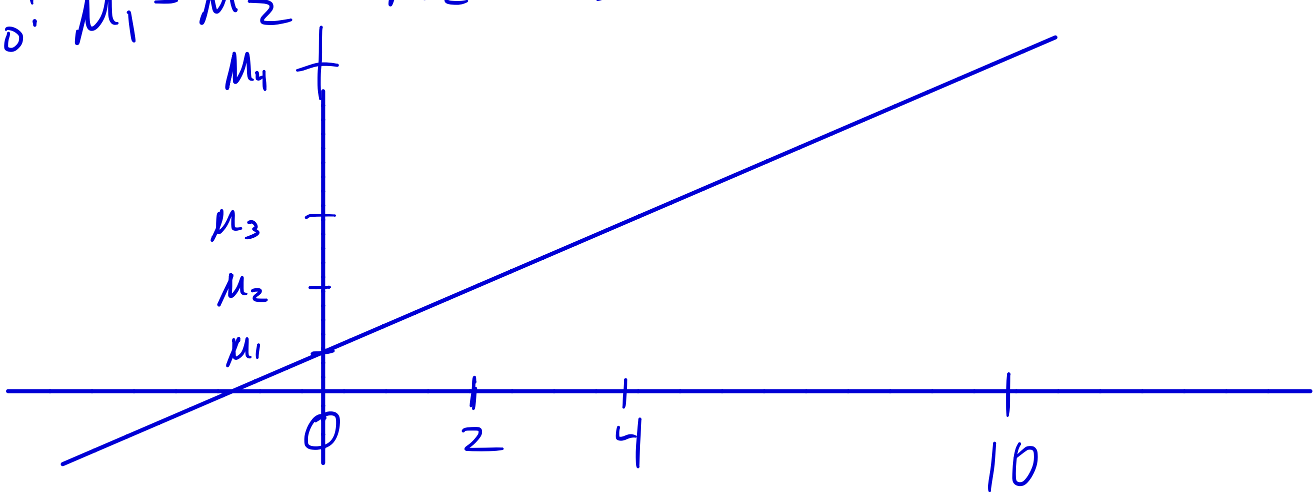
ALTERNATIVELY, THE NCP CAN BE COMPUTED AS $\dfrac{\beta' X' [C(X'X)^{-1} C']^{-1} X \beta}{2\sigma^2}$.

HOWEVER, WHEN SELECTING C, IT IS IMPORTANT TO NOTICE THAT X VALUES 0, 2, 4, AND 10 ARE NOT EQUALLY SPACED. ONE APPROPRIATE CHOICE FOR C IS

$$\begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 3 & -4 & 1 \end{bmatrix},$$

WHICH TESTS

$H_0: \mu_1 - \mu_2 = \mu_2 - \mu_3$ AND $3(\mu_3 - \mu_2) = \mu_4 - \mu_3$.



4 a) R PARAMETERIZATION IS

$\mu_{ij} = \mu + r_i + a_j + (ra)_{ij}$ WITH ALL PARAMETERS WITH ONE OR MORE 1 SUBSCRIPTS DISCARDED.

Region	Advertisement 1	Advertisement 2
1	μ	$\mu + a_2$
2	$\mu + r_2$	$\mu + r_2 + a_2 + (ra)_{22}$
3	$\mu + r_3$	$\mu + r_3 + a_2 + (ra)_{32}$
COLUMN AVERAGES:	$\mu + \frac{r_2 + r_3}{3}$	$\mu + \frac{r_2+r_3}{3} + a_2 + \frac{(ra)_{22} + (ra)_{32}}{3}$

ESTIMATED CELL MEANS

Region	Advertisement 1	Advertisement 2
1	123	$123 + (-28) = 95$
2	$123 + 56 = 179$	$123 + 56 + (-28) + (-11) = 140$
3	$123 + (-5) = 118$	$123 + (-5) + (-28) + 108 = 198$

4a) (CONTINUED)

SO LSMEAN FOR AD 1 IS

$$\hat{\mu} + \frac{\hat{r}_2 + \hat{r}_3}{3} = 123 + \frac{56-5}{3} = 140$$

OR $(123 + 179 + 118)/3 = 140$

LSMEAN FOR AD 2 IS

$$\hat{\mu} + \frac{\hat{r}_2 + \hat{r}_3}{3} + \hat{a}_2 + \frac{\widehat{(ra)}_{22} + \widehat{(ra)}_{23}}{3}$$

$$= 140 - 28 + \frac{-11 + 108}{3} = 112 + 32\tfrac{1}{3} = 144\tfrac{1}{3}$$

OR $(95 + 140 + 198)/3 = 144\tfrac{1}{3}$

4 b) $\mu_{11} - \mu_{12} = \mu - (\mu + a_2) = -a_2$

$\hat{\mu}_{11} - \hat{\mu}_{12} = -\hat{a}_2 = 28$

$$\text{VAR}(\hat{\mu}_{11} - \hat{\mu}_{12}) = \text{VAR}(\bar{y}_{11\cdot} - \bar{y}_{12\cdot})$$
$$= \text{VAR}(\bar{y}_{11\cdot}) + \text{VAR}(\bar{y}_{12\cdot})$$
$$= \sigma^2/5 + \sigma^2/5$$
$$= 2\sigma^2/5$$

$$\text{SE}(\hat{\mu}_{11} - \hat{\mu}_{12}) = \text{SE}(-\hat{a}_2) = \sqrt{2\hat{\sigma}^2/5}$$

WE ARE GIVEN $\text{SE}(\widehat{(ra)}_{32}) = 36$

NOTE $(ra)_{32} = \mu + r_3 + a_2 + (ra)_{32} - (\mu + r_3)$
$$- (\mu + a_2) + \mu$$
$$= \mu_{32} - \mu_{31} - \mu_{12} + \mu_{11}$$

THUS, $\text{SE}(\widehat{(ra)}_{32}) = \text{SE}(\hat{\mu}_{32} - \hat{\mu}_{31} - \hat{\mu}_{12} + \hat{\mu}_{11})$
$$= \sqrt{4\hat{\sigma}^2/5}$$

4b) (CONTINUED)

THUS, $SE(-\hat{a}_2) = \sqrt{36^2/2} = 36/\sqrt{2} = 18\sqrt{2}$.

ALSO, WE HAVE

$$\hat{\sigma}^2 = \frac{5 \times 36^2}{4} = 45 \times 36 = 5 \times 9 \times 6 \times 6 = 1620$$

$$n - r = 30 - 6 = 24$$

THUS, 95% CONFIDENCE INTERVAL

IS $28 \pm 2.064 \times 18\sqrt{2}$

4c) COMPARING THE ADDITIVE MODEL FOR LACK OF FIT RELATIVE TO THE CELL-MEANS MODEL IS THE SAME AS TESTING FOR INTERACTIONS BETWEEN REGION AND ADVERTISEMENT IN THE CELL-MEANS MODEL.

4c) (CONTINUED)

NO INTERACTIONS IS EQUIVALENT TO

$\mu_{11} - \mu_{12} = \mu_{21} - \mu_{22}$ AND

$$\frac{(\mu_{11} - \mu_{12}) + (\mu_{21} - \mu_{22})}{2} = \mu_{31} - \mu_{32},$$

WHICH IS EQUIVALENT TO

$$\begin{bmatrix} 1 & -1 & -1 & 1 & 0 & 0 \\ 1 & -1 & 1 & -1 & -2 & 2 \end{bmatrix} \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{21} \\ \mu_{22} \\ \mu_{31} \\ \mu_{32} \end{bmatrix} = \underset{\sim}{0}$$

OR $C\underset{\sim}{\beta} = \underset{\sim}{0}$, WHERE

$$C \equiv \begin{bmatrix} 1 & -1 & -1 & 1 & 0 & 0 \\ 1 & -1 & 1 & -1 & -2 & 2 \end{bmatrix}, \quad \underset{\sim}{\beta} \equiv \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{21} \\ \mu_{22} \\ \mu_{31} \\ \mu_{32} \end{bmatrix}.$$

4 c) (CONTINUED)

USING THE TABLE OF ESTIMATES FROM 4(a)

GIVES

$$C\hat{\underline{\beta}} = \begin{bmatrix} 123 - 95 - 179 + 140 \\ 123 - 95 + 179 - 140 - 236 + 396 \end{bmatrix}$$

$$= \begin{bmatrix} -11 \\ 227 \end{bmatrix}$$

$$C(X'X)^{-1}C' = C \tfrac{1}{5} I C' = \tfrac{1}{5} CC'$$

$$= \tfrac{1}{5} \begin{bmatrix} 4 & 0 \\ 0 & 12 \end{bmatrix}$$

4c) (CONTINUED)

$$\left[C(X'X)^{-1}C'\right]^{-1} = 5\begin{bmatrix} 1/4 & 0 \\ 0 & 1/12 \end{bmatrix}$$

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

$$= 5\left[\tfrac{1}{4}(-11)^2 + \tfrac{1}{12}(227)^2\right]$$

$$F = \frac{5\left[\tfrac{1}{4}(-11)^2 + \tfrac{1}{12}(227)^2\right]/2}{5 \times 36^2/4}$$

AN ALTERNATIVE SOLUTION IS TO COMPUTE THE TEST STATISTIC FOR TESTING $H_0: (r\alpha)_{22} = (r\alpha)_{32} = 0$ BECAUSE THIS NULL IS EQUIVALENT TO NO INTERACTION

4c) (ALTERNATIVE SOLUTION CONTINUED)

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

$\underset{\sim}{\beta} = [\mu, r_2, r_3, a_2, (ra)_{22}, (ra)_{32}]'$

$C\hat{\underset{\sim}{\beta}} = \begin{bmatrix} -11 \\ 108 \end{bmatrix}$. UNFORTUNATELY, THE MODEL MATRIX X THAT GOES WITH THE PARAMETER VECTOR

$\underset{\sim}{\beta} = [\mu, r_2, r_3, a_2, (ra)_{22}, (ra)_{32}]'$

DOES NOT LEAD TO $X'X$ THAT IS EASILY INVERTIBLE. THUS, IT IS BETTER TO RECOGNIZE THAT

$$\begin{bmatrix} -11 \\ 108 \end{bmatrix} = \begin{bmatrix} \widehat{(ra)}_{23} \\ \widehat{(ra)}_{33} \end{bmatrix} = \begin{bmatrix} \hat{\mu}_{22} - \hat{\mu}_{21} - \hat{\mu}_{12} + \hat{\mu}_{11} \\ \hat{\mu}_{32} - \hat{\mu}_{31} - \hat{\mu}_{12} + \hat{\mu}_{11} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & -1 & -1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} \hat{\mu}_{11} \\ \hat{\mu}_{12} \\ \hat{\mu}_{21} \\ \hat{\mu}_{22} \\ \hat{\mu}_{31} \\ \hat{\mu}_{32} \end{bmatrix} \equiv C^* \hat{\underset{\sim}{\mu}}.$$

Thus, $\text{Var}(C\hat{\beta}) = \text{Var}(C^*\hat{\underline{\mu}})$

$$= \sigma^2 C^* (X'X)^{-1} C^{*\prime}$$

$$= \sigma^2 C^* \left(\tfrac{1}{5} I\right) C^{*\prime}$$

$$= \frac{\sigma^2}{5} C^* C^{*\prime}$$

$$= \frac{\sigma^2}{5} \begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix}$$

$$= \frac{2\sigma^2}{5} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

$$\left[\widehat{\text{Var}}(C\hat{\beta})\right]^{-1} = \frac{5}{2\hat{\sigma}^2} \, \frac{1}{3} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

$$= \frac{5}{2\left(\frac{5 \times 36^2}{4}\right)} \, \frac{1}{3} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

$$= \frac{2}{3 \times 36^2} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

4c) (ALTERNATIVE SOLUTION CONTINUED)

$$F = (C\hat{\beta})' \left[\widehat{VAR}(C\hat{\beta})\right]^{-1} C\hat{\beta} / 2$$

$$= \frac{[-11, 108] \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} -11 \\ 108 \end{bmatrix}}{3 \times 36^2}$$

$$= [-11, 108] \begin{bmatrix} -130 \\ 227 \end{bmatrix} \Big/ (3 \times 36^2)$$

$$= \frac{(11)(130) + (108)(227)}{(3)\, 36^2}$$

BOTH VERSIONS OF THE F STATISTIC COMPUTE TO APPROXIMATELY 6.67.

4d) BASED ON ESTIMATED MEANS IN PART (a), WE HAVE THE FOLLOWING SIMPLE EFFECT ESTIMATES FOR AD 1 MEAN MINUS AD 2 MEAN:

REGION	
1	$\hat{\mu}_{11} - \hat{\mu}_{12} = 28$
2	$\hat{\mu}_{21} - \hat{\mu}_{22} = 39$
3	$\hat{\mu}_{31} - \hat{\mu}_{32} = -80$

THE MARGIN OF ERROR (t-QUANTILE x SE) FOR THESE ESTIMATES IS (FROM PART (b))

$$2.064 \times 18\sqrt{2} \approx 2 \times 18 \times 1.5 = 54$$

THUS, WE SEE A STATISTICALLY SIGNIFICANT DIFFERENCE FOR REGION 3 BECAUSE

$$0 \notin (-80-54, -80+54).$$

4d) NONE OF THE OTHER INTERVALS EXCLUDES ZERO. (ALTHOUGH IT IS NOT OBVIOUS FROM THE INFO GIVEN, A BONFERRONI CORRECTION WOULD USE $t_{1-0.025/3,\,24} = 2.57$ AS THE QUANTILE AND WOULD NOT CHANGE ANY CONCLUSIONS ABOUT THE STATISTICAL SIGNIFICANCE AT THE 0.05 LEVEL.) THUS, I MIGHT TELL THE COMPANY EXECUTIVE SOMETHING LIKE THE FOLLOWING:

4 d) (CONTINUED)

"IN REGION 3, POTENTIAL CUSTOMERS WHO RECEIVED ADVERTISEMENT 2 SPENT ON AVERAGE 80 CENTS MORE PER POTENTIAL CUSTOMER THAN POTENTIAL CUSTOMERS WHO RECEIVED ADVERTISEMENT 1. WE ESTIMATE THE MEAN SPENDING DIFFERENCE TO BE 80 ± 54 CENTS PER POTENTIAL CUSTOMER AND RECOMMEND ADVERTISEMENT 2 FOR POTENTIAL CUSTOMERS IN REGION 3. RESULTS ARE INCONCLUSIVE IN REGIONS 1 AND 2."

4 d) THE STATED PURPOSE OF THE EXPERIMENT IS TO COMPARE ADVERTISEMENTS, SO COMPARING REGIONS IS NOT LIKELY TO BE OF INTEREST FOR THIS STUDY. THE COMPANY WOULD PRESUMABLY HAVE MUCH MORE DATA THAT WOULD ADDRESS SALES IN EACH REGION.

MANY OF YOU WROTE DESCRIPTIONS THAT WOULD PROBABLY BE CLOSE TO IMPOSSIBLE FOR THE COMPANY EXECUTIVE TO UNDERSTAND (UNLESS SHE OR HE HAPPENED TO HAVE A DEGREE IN STATISTICS).