STAT 510 FINAL SOLUTIONS

SPRING 2015

1a) THE F-TEST STATISTIC FOR LACK OF LINEAR FIT IS 1.1642. COMPARING THIS STATISTIC TO A F DISTRIBUTION WITH 3 AND 15 DEGREES OF FREEDOM YIELDS A P-VALUE OF 0.356. THERE IS NO SIGNIFICANT EVIDENCE OF LACK OF FIT. THE SIMPLE LINEAR REGRESSION MODEL FITS ADEQUATELY.

1b) $F = \dfrac{57.6}{(15.6+67.0)/(3+15)} \Longrightarrow |t| = \sqrt{\dfrac{57.6}{82.6/18}}$

(THE MOST COMMON WRONG ANSWER WAS $|t| = \sqrt{12.8455}$. THIS IS WRONG BECAUSE IT USES MSE FROM THE FIT OF THE CELL MEANS MODEL TO OBTAIN THE DENOMINATOR OF THE F STATISTIC. THE QUESTION ASKS FOR THE F STATISTIC THAT WOULD RESULT FROM FITTING THE SIMPLE LINEAR REGRESSION MODEL.)

2a) X IS THE MODEL MATRIX R WOULD USE FOR

lm(y ~ a + b + a:b) WHERE

a = factor(c(1,1,1,2,2,2)) AND

b = factor(c(1,1,2,1,2,2)).

THIS IS EQUIVALENT TO A CELL-MEANS MODEL WITH MODEL MATRIX

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

THUS, $P_X = P_{X^*} = X^*(X^{*\prime}X^*)^{-1}X^{*\prime}$

$$= \begin{bmatrix} 1/2 & 1/2 & 0 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \end{bmatrix}$$

2 b) THE FIRST COLUMN ON W IS THE FIRST COLUMN OF X MINUS THE SECOND COLUMN OF X. THE SECOND COLUMN OF W IS THE SECOND COLUMN OF X (i.e., $X\begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} = W$).

THUS, $\mathcal{C}(W) \subseteq \mathcal{C}(X)$. IT FOLLOWS THAT

$P_X W = W$.

3a) $X = \underset{2\times 2}{I} \otimes \underset{20\times 1}{\underset{\sim}{1}}$

$$\underset{\sim}{\beta} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

$$Z = \underset{10\times 10}{I} \otimes \underset{4\times 1}{\underset{\sim}{1}}$$

$$\underset{\sim}{u} = (P_{11}, P_{12}, P_{13}, P_{14}, P_{15}, P_{21}, P_{22}, P_{23}, P_{24}, P_{25})'$$

3b)

SOURCE	DF	SS
BREED	1	$20 \sum_{i=1}^{2} (\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot\cdot\cdot})^2$
PEN(BREED)	8	$4 \sum_{i=1}^{2} \sum_{j=1}^{5} (\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot})^2$
PIG(PEN, BREED)	30	$\sum_{i=1}^{2} \sum_{j=1}^{5} \sum_{k=1}^{4} (y_{ijk} - \bar{y}_{ij\cdot})^2$
C. TOTAL	39	$\sum_{i=1}^{2} \sum_{j=1}^{5} \sum_{k=1}^{4} (y_{ijk} - \bar{y}_{\cdot\cdot\cdot})^2$

$$\frac{MS_{BREED}}{MS_{PEN(BREED)}} = \frac{20 \sum_{i=1}^{2} (\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot\cdot\cdot})^2 / 1}{4 \sum_{i=1}^{2} \sum_{j=1}^{5} (\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot})^2 / 8}$$

THESE RESULTS FOLLOW DIRECTLY FROM SLIDE SET 12. THE SAME RESULT COULD BE OBTAINED BY SQUARING THE TWO-SAMPLE t-STATISTIC IN AN ANALYSIS OF PEN AVERAGES.

3c) $y_{ijkl} = \mu_{il} + u_{ij} + v_{ijk} + e_{ijkl}$

FOR $i = 1, 2$ AND $l = 1, \ldots, 7$, μ_{il} IS AN UNKNOWN PARAMETER IN $\mathbb{R}$. THE u_{ij} TERMS ARE $N(0, \sigma_u^2)$ FOR SOME UNKNOWN $\sigma_u^2 > 0$. THE v_{ijk} TERMS ARE $N(0, \sigma_v^2)$ FOR SOME UNKNOWN $\sigma_v^2 > 0$. THE e_{ijkl} TERMS ARE $N(0, \sigma^2)$. ALL u_{ij}, v_{ijk}, AND e_{ijkl} TERMS ARE MUTUALLY INDEPENDENT. THIS MODEL ALLOWS FOR SOME PIGS TO BE MORE ACTIVE THAN OTHERS AND FOR PIGS TO BE MORE ACTIVE ON SOME DAYS THAN OTHERS. ANOTHER REASONABLE CHOICE WOULD BE TO REPLACE $v_{ijk} + e_{ijkl}$ WITH ε_{ijkl} WHERE $\underset{\sim}{\varepsilon}_{ijk} = (\varepsilon_{ijk1}, \ldots, \varepsilon_{ijk7}) \overset{ind}{\sim} N(\underset{\sim}{0}, W)$ WITH

$$W = \eta^2 \begin{bmatrix} 1, \rho, \rho^2, \ldots, \rho^6 \\ \rho, 1, \rho, \ldots, \rho^5 \\ \vdots \quad \ddots \quad \vdots \\ \quad\quad \ddots \quad \rho \\ \rho^6, \rho^5, \ldots \rho, 1 \end{bmatrix}$$

FOR SOME $\eta^2 > 0$, $\rho \in (-1, 1)$.

4a) $\left\{1+\exp\left(-0.7485279+2(0.3237699)\right)\right\}^{-1}$

b) $-2\ell_1(\hat{\underset{\sim}{\theta}}_1)+2(2) \cong -9.36$

$-2\ell_2(\hat{\underset{\sim}{\theta}}_2)+2(3) \cong -23.98$

$-2\ell_3(\hat{\underset{\sim}{\theta}}_3)+2(7) \cong -17.69$

$\Rightarrow 2\ell_3(\hat{\underset{\sim}{\theta}}_3)-2\ell_2(\hat{\underset{\sim}{\theta}}_2) \cong (17.69+14)-(23.98+6)$

c) $2\ell_2(\hat{\underset{\sim}{\theta}}_2)-2\ell_1(\hat{\underset{\sim}{\theta}}_1) \cong (23.98+6)-(9.36+4)$

THIS STAT SHOULD BE APPROXIMATELY EQUAL TO

$Z^2=\left(\dfrac{.3237699}{SE}\right)^2$. THUS, $SE \approx \dfrac{.3237699}{\sqrt{(23.98+6)-(9.36+4)}}$

5a) $2\hat{l}_f - 2\hat{l}_0 = (2\hat{l}_s - 2\hat{l}_0) - (2\hat{l}_s - 2\hat{l}_f)$

$$= 17595.5 - 3109.6$$

b) (i) 2

(ii) $\exp(0.4352)$

(iii) 1

(iv) $\exp(0.4352 - 2\sqrt{\hat{\phi}}\, 0.02425)$

(v) $\exp(0.4352 + 2\sqrt{\hat{\phi}}\, 0.02425)$

$$\hat{\phi} = \frac{3109.6}{999}$$

WE SHOULD ADJUST FOR OVERDISPERSION BECAUSE 3109.6 IS FAR IN THE RIGHT TAIL OF THE χ^2_{999} DISTRIBUTION ($E(\chi^2_{999}) = 999$ $VAR(\chi^2_{999}) = 2(999)$ $SD(\chi^2_{999}) \simeq 45$)