STAT 510

FINAL EXAM

SPRING 2016

1. a) 5	2. 8	3. a) 6	4. a) 6
b) 5		b) 6	b) 6
c) 5		c) 6	c) 8
d) 8		d) 6	d) 6
		e) 6	e) 8
			f) 5

1a) $\underset{\sim}{c}'\beta$ IS ESTIMABLE IF AND ONLY IF THERE EXISTS A VECTOR $\underset{\sim}{a}$ SUCH THAT $\underset{\sim}{c}' = \underset{\sim}{a}'X$. THIS IS EQUIVALENT TO SAYING THAT $\underset{\sim}{c}'\underset{\sim}{\beta}$ IS A LINEAR COMBINATION OF THE ELEMENTS OF $E(\underset{\sim}{y}) = X\underset{\sim}{\beta}$ FOR ALL $\underset{\sim}{\beta} \in \mathbb{R}^p$.

b) $\underset{\sim}{c}(X'X)^- X'\underset{\sim}{y}$

c) $$\hat{\sigma}^2 = \frac{\underset{\sim}{y}'(I - X(X'X)^- X')\underset{\sim}{y}}{n-r}$$

1d) FROM OUR COURSE NOTES, WE KNOW

i) $\underline{c}'\hat{\beta} \sim N(\underline{c}'\beta, \sigma^2 \underline{c}'(X'X)^{-}\underline{c})$,

WHICH IS EQUIVALENT TO $Z = \dfrac{\underline{c}'\hat{\beta} - \underline{c}'\beta}{\sqrt{\sigma^2 \underline{c}'(X'X)^{-}\underline{c}}} \sim N(0,1)$

ii) $W = \dfrac{(n-r)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-r}$

iii) Z AND W ARE INDEPENDENT.

Now, $\dfrac{\underline{c}'\hat{\beta} - \underline{c}'\beta}{\sqrt{\hat{\sigma}^2 \underline{c}'(X'X)^{-}\underline{c}}} = \dfrac{\underline{c}'\hat{\beta} - \underline{c}'\beta}{\sqrt{\sigma^2 \underline{c}'(X'X)^{-}\underline{c}}} \dfrac{\sqrt{\sigma^2 \underline{c}'(X'X)^{-}\underline{c}}}{\sqrt{\hat{\sigma}^2 \underline{c}'(X'X)^{-}\underline{c}}} = Z \Big/ \sqrt{\hat{\sigma}^2/\sigma^2}$

$= Z \Big/ \sqrt{W/n-r} \sim t_{n-r}$

BECAUSE WE KNOW THAT t_{df} IS A STANDARD NORMAL DIVIDED BY THE SQUARE ROOT OF AN INDEPENDENT χ^2 OVER ITS DEGREES OF FREEDOM.

2. LET Y_{ijk} BE THE MEASUREMENT OF PROTEIN QUANTITY FOR STRAIN i, TIME j, AND PIG k.

$$\mathrm{Cov}(Y_{i1k}, Y_{i3k}) = \mathrm{Cov}(\mu_{i1} + P_k + e_{i1k}, \mu_{i3} + P_k + e_{i3k})$$

$$= \mathrm{Cov}(P_k, P_k) + \mathrm{Cov}(e_{i1k}, e_{i3k})$$

$$= \mathrm{VAR}(P_k) + \mathrm{CORR}(e_{i1k}, e_{i3k})\sqrt{\mathrm{VAR}(e_{i1k})\,\mathrm{VAR}(e_{i3k})}$$

$$= \sigma_p^2 + \rho^2\sqrt{\sigma_e^2\,\sigma_e^2}$$

$$= \sigma_p^2 + \rho^2\sigma_e^2,$$ WHICH IS ESTIMATED BY

$$6.5047 + (0.6772)^2\,8.3603.$$

3a) LET $a_{ik} = \dfrac{y_{i1k} + y_{i2k}}{2} = \bar{\mu}_{i\cdot} + P_k + \bar{e}_{i\cdot k}$

$= \bar{\mu}_{i\cdot} + \varepsilon_{ik}$, WHERE $\varepsilon_{ik} = P_k + \bar{e}_{i\cdot k}$

NOTE THAT THE ε_{ik} TERMS ARE iid $N(0, \sigma^2)$, WHERE $\sigma^2 = \sigma_p^2 + \frac{\sigma_e^2}{2}$.

THUS, A TWO SAMPLE t-TEST CAN BE USED TO TEST

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

FROM THE R OUTPUT OF THE ANANLYSIS OF AVERAGES,

WE HAVE

$$t = \frac{84.892 - 80.454}{\sqrt{2.169^2 + 1.534^2}}$$

3b) LET $d_{ik} = y_{i1k} - y_{i2k} = \mu_{i1} - \mu_{i2} + e_{i1k} - e_{i2k}$

$$\equiv \delta_i + \eta_{ik},$$

WHERE $\delta_i = \mu_{i1} - \mu_{i2}$ AND $\eta_{ik} = e_{i1k} - e_{i2k}$.

NOTE THAT THE η_{ik} TERMS ARE iid $N(0, \sigma_\eta^2)$,

WHERE $\sigma_\eta^2 = 2\sigma_e^2$.

THE TEST OF INFECTION MAIN EFFECT IS A TEST OF

$$H_0: \frac{\mu_{11} + \mu_{21}}{2} = \frac{\mu_{12} + \mu_{22}}{2}$$

$$\Longleftrightarrow$$

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

$$\Longleftrightarrow$$

$$H_0: \delta_1 + \delta_2 = 0.$$

3b) (CONTINUED)

FROM THE LAST ANALYSIS OF THE DIFFERENCES IN R,

WE CAN TEST $H_0: \delta_1 + \delta_2 = 0$ WITH

$$t = \frac{8.250 + 1.492}{\sqrt{2.439^2 + 1.724^2}}.$$

3c) IT IS STRAIGHTFORWARD TO SEE THAT A TEST FOR INTERACTION IS A TEST OF $H_0: \delta_1 = \delta_2 \iff H_0: \delta_1 - \delta_2 = 0.$

THUS,

$$t = \frac{8.250 - 1.492}{\sqrt{2.439^2 + 1.724^2}}$$

IS THE RELEVANT TEST STATISTIC.

3d) $\hat{\sigma}_m^2 = 2\hat{\sigma}_e^2 = 5.974^2$

$$\Rightarrow \hat{\sigma}_e^2 = \frac{5.974^2}{2}$$

3e) $\hat{\sigma}^2 = \hat{\sigma}_p^2 + \frac{\hat{\sigma}_e^2}{2} = 5.313^2$

$$\Rightarrow \hat{\sigma}_p^2 = 5.313^2 - \frac{5.974^2}{4}$$

THE ANSWERS TO PARTS a) THROUGH e) ABOVE MATCH TESTS AND ESTIMATES OBTAINED BY FITTING THE FULL LINEAR MIXED EFFECTS MODEL $\underline{y} = X\underline{\beta} + Z\underline{u} + \underline{e}$.

4a) $BIC = -2\ell(\hat{\underset{\sim}{\theta}}) + k\log(n)$

$$= -2\ell(\hat{\underset{\sim}{\theta}}) + 2k - 2k + k\log(n)$$

$$= AIC - 2k + k\log(n)$$

$$= 350.8 - 4 + 2\log(50)$$

$$= 346.8 + 2\log(50)$$

4 b) $0.25 = \dfrac{1}{1+\exp(-(-1.652+0.077\hat{x}))}$

$$\iff \ln\left(\frac{.25}{1-.25}\right) = -1.652 + 0.077\hat{x}$$

$$\iff \frac{\ln(1/3) + 1.652}{0.077} = \frac{1.652 - \ln(3)}{0.077} = \hat{x}$$

4 c) $\hat{x} = \dfrac{-\hat{\beta}_0 - \ln(3)}{\hat{\beta}_1}$

$$\frac{\partial \hat{x}}{\partial \hat{\beta}_0} = -\frac{1}{\hat{\beta}_1}$$

$$\frac{\partial \hat{x}}{\partial \hat{\beta}_1} = \frac{\hat{\beta}_0 + \ln(3)}{\hat{\beta}_1^2}$$

LET $\underline{d} = \begin{bmatrix} -1/\hat{\beta}_1 \\ \dfrac{\hat{\beta}_0 + \ln(3)}{\hat{\beta}_1^2} \end{bmatrix} = \begin{bmatrix} -1/.077 \\ \dfrac{-1.652 + \ln(3)}{(0.077)^2} \end{bmatrix} \equiv \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}$

4c) (CONTINUED)

BY THE DELTA METHOD,

$$\hat{VAR}(\hat{x}) = \underset{\sim}{d}' \, \hat{VAR}(\hat{\beta}) \underset{\sim}{d}$$

$$= d_1^2 (0.00392) - 2 d_1 d_2 (0.00032) + d_2^2 (0.00003)$$

$$\Longrightarrow \quad SE(\hat{x}) = \sqrt{d_1^2 (0.00392) - 2 d_1 d_2 (0.00032) + d_2^2 (0.00003)}$$

HOWEVER, NOTE THAT THERE IS EVIDENCE OF OVERDISPERSION.

A χ^2_{48} HAS MEAN 48 AND SD $= \sqrt{2 \times 48} < 10$.

THE RESIDUAL DEVIANCE FOR MODEL 1 IS MORE THAN 6 SDs ABOVE THE NULL MEAN: $110.30 > 48 + 6 \times 10$.

THUS, A MORE APPROPRIATE SE IS

$$\sqrt{\frac{110.30}{48} \left[d_1^2 (0.00392) - 2 d_1 d_2 (0.00032) + d_2^2 (0.00003) \right]}$$

4d) EACH ADDITIONAL HOUR OF TRAINING INCREASES THE ESTIMATED ODDS OF SUCCESS BY THE MULTIPLICATIVE FACTOR $\exp(0.077)$.

4e) THE FIT OF MODEL 2 SHOWS CLEAR EVIDENCE OF OVERDISPERSION. A χ^2_{45} RANDOM VARIABLE HAS A MEAN OF 45 AND A STANDARD DEVIATION OF $\sqrt{2 \times 45} < 10$. THUS, THE RESIDUAL DEVIANCE FOR MODEL 2 IS WELL OVER 5 STANDARD DEVIATIONS ABOVE THE NULL MEAN:

$$98.209 > 45 + 5\sqrt{2 \times 45}.$$

THUS, WE NEED TO ADJUST FOR OVERDISPERSION WHEN COMPARING MODELS 1 AND 2.

4 e) (CONTINUED)

THE RELEVANT TEST STATISTIC IS

$$F = \frac{(110.30 - 98.209)/(48-45)}{98.209/45}$$

4 f) $F_{3,45}$