STAT 510
FINAL EXAM
SPRING 2018

1a) 5	2a) 5	3a) 7	4a) 4
b) 4	b) 6	b) 4	b) 4
c) 5	c) 6	c) 5	c) 4
d) 5	d) 7	d) 6	d) 5
	e) 7	e) 5	
		f) 6	

1a) THE COLUMN SPACE OF X IS THE SET OF ALL VECTORS THAT CAN BE EXPRESSED AS A LINEAR COMBINATION OF THE COLUMNS OF X.

$\{ \underset{\sim}{a} \in \mathbb{R}^n : \underset{\sim}{a} = X\underline{b} \text{ FOR SOME } \underline{b} \in \mathbb{R}^p \}$

OR

$\{ X\underline{b} : \underline{b} \in \mathbb{R}^p \}$

1b) C MUST BE EQUAL TO AX FOR SOME $q \times n$ MATRIX OF CONSTANTS A.

1c) $\sigma^2 C(X'X)^- C'$ SEE DERIVATION BELOW.

$$\mathrm{VAR}(C(X'X)^- X'\underset{\sim}{y}) = \mathrm{VAR}(AX(X'X)^- X'\underset{\sim}{y})$$

$$= \mathrm{VAR}(AP_x \underset{\sim}{y}) = AP_x \mathrm{VAR}(\underset{\sim}{y})(AP_x)'$$

$$= AP_x \sigma^2 I P_x' A' = \sigma^2 AP_x P_x' A'$$

$$= \sigma^2 AP_x P_x A' = \sigma^2 AP_x A'$$

$$= \sigma^2 AX(X'X)^- X'A' = \sigma^2 C(X'X)^- C'$$

1d) LET $\underset{\sim}{Z} = \frac{1}{\sigma}\underline{\varepsilon} = \frac{1}{\sigma} I \underline{\varepsilon}$.

CLEARLY $\underline{Z}$ IS A LINEAR TRANSFORMATION OF A MULTIVARIATE NORMAL VECTOR AND IS, THUS, MULTIVARIATE NORMAL.

$$E(\underline{Z}) = E\left(\frac{1}{\sigma} I \underline{\varepsilon}\right) = \frac{1}{\sigma} I\, E(\underline{\varepsilon}) = \underset{\sim}{0}$$

$$Var(\underline{Z}) = Var\left(\frac{1}{\sigma} I \underline{\varepsilon}\right) = \frac{1}{\sigma} I\, \sigma^2 I\, \frac{1}{\sigma} I = I.$$

THUS, $\underline{Z} \sim N(\underline{0}, I)$. THEREFORE,

$$\underline{\varepsilon}'\underline{\varepsilon}/\sigma^2 = \left(\frac{1}{\sigma}\underline{\varepsilon}\right)'\left(\frac{1}{\sigma}\underline{\varepsilon}\right) = \underline{Z}'\underset{\sim}{Z} \sim \chi^2_n.$$

2a) $(k-1)(\ell-1) = k\ell - k - \ell + 1$

$$\sum_{i=1}^{2} \sum_{j=1}^{5} \sum_{k=1}^{2} \sum_{\ell=1}^{2} \left(\bar{y}_{\cdot\cdot k\ell} - \bar{y}_{\cdot\cdot k\cdot} - \bar{y}_{\cdot\cdot\cdot\ell} + \bar{y}_{\cdot\cdot\cdot\cdot}\right)^2$$

$$= 10 \sum_{k=1}^{2} \sum_{\ell=1}^{2} \left(\bar{y}_{\cdot\cdot k\ell} - \bar{y}_{\cdot\cdot k\cdot} - \bar{y}_{\cdot\cdot\cdot\ell} + \bar{y}_{\cdot\cdot\cdot\cdot}\right)^2$$

2b) THIS IS A SPLIT-SPLIT-PLOT EXPERIMENT. $H_0: \bar{\mu}_{1\cdot\cdot} = \bar{\mu}_{2\cdot\cdot}$ IS THE NULL HYPOTHESIS THAT SAYS THERE IS NOW WHOLE PLOT-FACTOR (i.e., TYPE) MAIN EFFECT. THE WHOLE-PLOT-EXPERIMENTAL UNITS CORRESPOND TO HELMET (TYPE), SO THE F STATISTIC FOR TESTING H_0 IS

$$F = \frac{MS_{TYPE}}{MS_{HELMET(TYPE)}} = \frac{226/1}{254/8}.$$

THUS,

$$t = \sqrt{\frac{MS_{TYPE}}{MS_{HELMET(TYPE)}}} = \sqrt{\frac{226/1}{254/8}}$$

2 c) FROM THE PROVIDED EXPECTED MEAN SQUARES, IT IS STRAIGHTFORWARD TO SEE THAT

$$\frac{MS_{HELMET(TYPE)} - MS_{DIRECTION \times HELMET(TYPE)}}{4}$$

HAS EXPECTATION σ_a^2.

THUS, AN UNBIASED ESTIMATOR OF σ_a^2 TAKES THE VALUE

$$\frac{254/8 - 114/8}{4} = \frac{140}{32} = 4.375$$

2d) BECAUSE WE HAVE A BALANCED DESIGN, WE KNOW THE BLUE OF $\overline{\mu}_{12.} - \overline{\mu}_{11.}$ IS $\overline{y}_{1.2.} - \overline{y}_{1.1.}$, WHICH HAS VARIANCE

$$VAR\left(\overline{a}_{1.} + \overline{b}_{1.2} + \overline{e}_{1.2.} - \overline{a}_{1.} - \overline{b}_{1.1} - \overline{e}_{1.1.}\right)$$

$$= VAR\left(\overline{b}_{1.2.} - \overline{b}_{1.1} + \overline{e}_{1.2.} - \overline{e}_{1.1.}\right)$$

$$= \frac{2\sigma_b^2}{5} + \frac{2\sigma_e^2}{5 \times 2}$$

$$= \frac{1}{5}\left(2\sigma_b^2 + \sigma_e^2\right)$$

$$= \frac{1}{5} E\left(MS_{DIRECTION \times HELMET(TYPE)}\right)$$

THUS, $\widehat{VAR}\left(\overline{y}_{1.2.} - \overline{y}_{1.1.}\right) = \frac{1}{5}\left(\frac{114}{8}\right)$

$$= \frac{57}{20}$$

$$= 2.85$$

2d) (CONTINUED)

THUS, THE CONFIDENCE INTERVAL IS

$$0.5 \pm 2.306\sqrt{2.85}$$

$t_{.975,8}$

DF FOR DIRECTION x HELMET(TYPE)

2e) $\mathrm{VAR}(\bar{y}_{1.21} - \bar{y}_{1.11})$

$$= \mathrm{VAR}(\bar{a}_{1.} + \bar{b}_{1.2} + \bar{e}_{1.21} - \bar{a}_{1.} - \bar{b}_{1.1} - \bar{e}_{1.11})$$

$$= \mathrm{VAR}(\bar{b}_{1.2} - \bar{b}_{1.1} + \bar{e}_{1.21} - \bar{e}_{1.11})$$

$$= \frac{2\sigma_b^2}{5} + \frac{2\sigma_e^2}{5} = \frac{1}{5}(2\sigma_b^2 + 2\sigma_e^2)$$

$$= \frac{1}{5}\left[E(MS_{\text{DIRECTION x HELMET(TYPE)}}) + E(MS_{\text{ERROR}})\right]$$

2e) (CONTINUED)

THUS, SE IS

$$\sqrt{\frac{1}{5}\left(114/8 + 59/16\right)}$$

TO SEE THAT DF FOR ERROR IS 16,

NOTE THAT

ERROR = INTENSITY x HELMET (TYPE) +
INTENSITY x DIRECTION x HELMET (TYPE)

SO DF IS $(2-1) \times (5-1) \times 2 +$
$(2-1) \times (2-1) \times (5-1) \times 2 = 16$

ALTERNATIVELY, BY SUBTRACTION, WE HAVE

$(40-1) - (1+1+1+1+1+1+1+8+8) = 16$

3a) THE LIKELIHOOD RATIO STATISTIC IS

$$2\hat{\ell} - 2\hat{\ell}_0 = (2\hat{\ell}_s - 2\hat{\ell}_0) - (2\hat{\ell}_s - 2\hat{\ell})$$

$$= 88.105 - 37.674$$

HOWEVER, THERE IS EVIDENCE OF OVERDISPERSION: THE RESIDUAL DEVIANCE IS $2\hat{\ell}_s - 2\hat{\ell} = 37.67$. UNDER THE NULL OF NO OVERDISPERSION, THIS STATISTIC HAS APPROXIMATE DISTRIBUTION $\chi^2_{20-4} = \chi^2_{16}$. EVEN THOUGH THE APPROXIMATION MIGHT NOT BE PARTICULARLY GOOD IN THIS CASE, NOTE THAT $E(\chi^2_{16}) = 16$, $\mathrm{VAR}(\chi^2_{16}) = 2 \times 16 = 32$, $SD(\chi^2_{16}) = \sqrt{32} \approx 5.5$, AND

$$\frac{37.67 - 16}{5.5} = \frac{21.67}{5.5} \approx 4.$$

THUS, THE RESIDUAL DEVIANCE IS NEARLY 4 STANDARD DEVIATIONS ABOVE ITS EXPECTATION UNDER THE NULL HYPOTHESIS OF NO OVERDISPERSION. THUS, A BETTER STATISTIC FOR THE TEST OF INTEREST IS

$$\frac{(2\hat{\ell} - 2\hat{\ell}_0)/(4-1)}{\hat{\phi}} = \frac{(88.105 - 37.674)/3}{37.674/16}$$

3b) $F_{3,16}$

3c) ACCORDING TO THE MODEL, WE HAVE THE FOLLOWING TABLE:

GENOTYPE	$\log\left(\frac{\pi}{1-\pi}\right)$
1	$\beta_1 + \beta_3$ CHEM
2	$\beta_1 + \beta_2 + \beta_3$ CHEM $+ \beta_4$ CHEM

FOR GENOTYPE 1 AND CHEM = 0, WE HAVE

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_1 \Rightarrow \pi = \frac{1}{1+\exp(-\beta_1)}$$

$$\therefore \hat{\pi} = \frac{1}{1+\exp(1.5)}$$

3d) A confidence interval for β_1 with confidence level approximately equal to 95% is

$$\hat{\beta}_1 \pm t_{.975,16} \sqrt{\hat{\phi} \left[\widehat{I^{-1}(\beta)} \right]_{11}}$$

$$-1.50 \pm 2.12 \sqrt{(37.674/16)\, 0.151}$$

Let L and U be the lower and upper endpoints of this interval. Then the interval for π is

$$\left(\frac{1}{1+\exp(-L)} , \frac{1}{1+\exp(-U)} \right).$$

3e) FROM THE TABLE IN THE SOLUTION TO PART 3c), WE NEED x SUCH THAT

$$\beta_1 + \beta_3 x = \beta_1 + \beta_2 + \beta_3 x + \beta_4 x$$

i.e., $x = -\beta_2 / \beta_4$.

$$\hat{x} = \frac{-\hat{\beta}_2}{\hat{\beta}_4} = \frac{1.80}{0.10} = 18$$

3f) BECAUSE $h(\underline{\beta}) = -\beta_2/\beta_4$

IS A NONLINEAR FUNCTION OF $\underline{\beta}$,

WE USE THE DELTA METHOD TO

FIND A STANDARD ERROR.

$$\frac{\partial h(\underline{\beta})}{\partial \underline{\beta}} = \begin{bmatrix} \dfrac{\partial h(\underline{\beta})}{\partial \beta_1} \\[2ex] \dfrac{\partial h(\underline{\beta})}{\partial \beta_2} \\[2ex] \dfrac{\partial h(\underline{\beta})}{\partial \beta_3} \\[2ex] \dfrac{\partial h(\underline{\beta})}{\partial \beta_4} \end{bmatrix} = \begin{bmatrix} 0 \\[2ex] -1/\beta_4 \\[2ex] 0 \\[2ex] \beta_2/\beta_4^2 \end{bmatrix}$$

3 f) (CONTINUED)

$$\hat{\underset{\sim}{d}} \equiv \left.\frac{dh(\underline{\beta})}{d\underline{\beta}}\right|_{\underline{\beta}=\hat{\underline{\beta}}} = \begin{bmatrix} 0 \\ -1/_{-.10} \\ 0 \\ 1.8/_{(-.10)^2} \end{bmatrix} = \begin{bmatrix} 0 \\ 10 \\ 0 \\ 180 \end{bmatrix}$$

$$\hat{VAR}\left(h(\hat{\underline{\beta}})\right) = \hat{\underset{\sim}{d}}' \, \hat{VAR}(\hat{\underline{\beta}}) \, \hat{\underline{d}}$$

$$= \hat{\underline{d}}' \, \hat{\phi} \, \hat{I}^{-1}(\underline{\beta}) \, \hat{\underline{d}}$$

$$= \frac{37.674}{16} [10, 180] \begin{bmatrix} .261 & -.02 \\ -.02 & .002 \end{bmatrix} \begin{bmatrix} 10 \\ 180 \end{bmatrix}$$

$$= \frac{37.674}{16} \left(26.1 + 180^2(.002) - 2(.02)1800 \right)$$

$$= \frac{37.674}{16} \, 18.9$$

$$SE = \sqrt{\frac{37.674}{16} \, 18.9}$$

4a) $\beta_0, \beta_1, \beta_2$ 3

σ_d^2 1

$\Sigma_{\underline{b}}$ $3+2+1$

$\Sigma_{\underline{e}}$ $6+5+4+3+2+1$

31

DIMENSION OF MODEL PARAMETER SPACE IS 31.

4b) $-2\ln(\Lambda)$

$= 2[(-1064) - (-1235)]$

$= 2(1235 - 1064)$

$= 2 \times 171 = 342$

4 c) THE MODELS ARE THE SAME EXCEPT FOR THE ASSUMPTION ABOUT $\Sigma_{\underline{e}}$.

$\Sigma_{\underline{e}}$ HAS $6+5+4+3+2+1 = 21$ PARAMETERS IN MODEL D AND JUST 2 IN MODEL B. THUS, THE DIFFERENCE IN DIMENSIONS OF THE MODEL PARAMETER SPACES IS $21-2 = 19$.

$DF = 19$.

4 d) R would calculate AIC as follows:

Model Version	$AIC = -2\hat{\ell} + 2k$
A	$2634 + 2(10+1) = 2656$
B	$2470 + 2(10+2) = 2494$
C	$2316 + 2(10+2) = 2340$
D	$2128 + 2(10+21) = 2190$

Model version D has the lowest AIC and is therefor preferred over the other models. Note that the value of 10 in the calculations above is $3+1+3+2+1$ from part (a).