# STAT 510

# FINAL EXAM SOLUTIONS

# SPRING 2019

| | |
|---|---|
| 1. | 10 |
| 2. | 15 |
| 3.a) | 12 |
| 3.b) | 6 |
| 3.c) | 10 |
| 3.d) | 9 |
| 3.e) | 6 |
| 4.a) | 8 |
| 4.b) | 8 |
| 4.c) | 8 |
| 4.d) | 8 |

1. THE PEARSON CHI-SQUARE STATISTIC IS

$$\sum_{i=1}^{100}\left(\frac{y_i - \hat{E}(y_i)}{\sqrt{\hat{VAR}(y_i)}}\right)^2 = \sum_{i=1}^{100}\left(\frac{y_i - \bar{y}.}{\sqrt{\bar{y}.}}\right)^2$$

$$= \frac{1}{\bar{y}.}\sum_{i=1}^{100}(y_i - \bar{y}.)^2$$

$$= \frac{1}{\bar{y}.}\, 99 \sum_{i=1}^{100}(y_i - \bar{y}.)^2 / (100-1)$$

$$= \frac{1}{62}\, 99 \cdot 93$$

$$= 99 \cdot 1.5 = 148.5$$

UNDER THE NULL, THE STAT IS APPROXIMATELY $\chi^2_{99}$. $E(\chi^2_{99}) = 99$, $VAR(\chi^2_{99}) = 198$

$SD(\chi^2_{99}) \approx 14$. THUS, THE STAT IS MORE THAN 3 STANDARD DEVIATIONS ABOVE THE NULL EXPECTATION. THERE IS EVIDENCE AGAINST THE ASSUMPTION OF iid POISSON OBSERVATIONS.

2. $y_{ijk} \mid \lambda_{ijk} \overset{ind}{\sim} \text{Poisson}(\lambda_{ijk})$

$$\ln(\lambda_{ijk}) = \mu_{jk} + f_i + s_{ij} + e_{ijk}$$

$f_i \overset{iid}{\sim} N(0, \sigma_f^2)$ (FIELD RANDOM EFFECTS)

$s_{ij} \overset{iid}{\sim} N(0, \sigma_s^2)$ (STRIP RANDOM EFFECTS)

$e_{ijk} \overset{iid}{\sim} N(0, \sigma_e^2)$ (TRAP RANDOM EFFECTS)

ALL $f_i$, $s_{ij}$, AND $e_{ijk}$ EFFECTS ARE INDEPENDENT OF EACH OTHER.

THE $\mu_{jk}$ TERMS ALLOW THE MEAN RESPONSE TO DEPEND ON BOTH TYPE AND DISTANCE ALONG EACH STRIP. THIS CELL-MEANS STRUCTURE ALLOWS FOR TYPE MAIN EFFECTS (WHICH ARE OF PRIMARY INTEREST), DISTANCE MAIN EFFECTS, AND TYPE-BY-DISTANCE INTERACTIONS.

THE PREVIOUS IS A NATURAL EXTENSION OF THE SPLIT-PLOT MODEL TO POISSON DATA.

A BETTER CHOICE MAY BE TO DROP THE $S_{ij}$ TERMS AND ASSUME

$$\underset{\sim}{e}_{ij} \equiv \begin{bmatrix} e_{ij1} \\ e_{ij2} \\ e_{ij3} \\ e_{ij4} \end{bmatrix} \overset{iid}{\sim} N\left(\underset{\sim}{0}, \sigma_e^2 \begin{bmatrix} 1 & \phi & \phi^2 & \phi^3 \\ \phi & 1 & \phi & \phi^2 \\ \phi^2 & \phi & 1 & \phi \\ \phi^3 & \phi^2 & \phi & 1 \end{bmatrix}\right)$$

THIS RECOGNIZES THE REPEATED MEASURES STRUCTURE OF THE EXPERIMENTAL DESIGN.

ANOTHER OPTION IS TO KEEP THE $S_{ij}$ TERMS AND STILL ADD THE CORRELATED ERRORS.

3. THIS IS A SPLIT-SPLIT-PLOT EXPERIMENT.

THE WHOLE-PLOT TREATMENT FACTOR CONSISTS OF THE COMBINATIONS OF TEMPERATURE, HUMIDITY, AND $CO_2$ LEVEL. THIS HAS $8-1=7$ DF THAT CAN BE BROKEN DOWN INTO 7 SINGLE-DF PIECES: T, H, C, T×H, T×C, H×C, T×H×C. TO REDUCE LINES, THIS FACTOR WILL BE WRITTEN AS THC IN ANOVA TABLE BELOW. WE WILL AS ABBREVIATE MONTH AS M, G = GENOTYPE, A = AGE, P = PLANT.

| SOURCE | DF |
|---|---|
| M | 3 |
| THC | 7 |
| M×THC = $WP_{ERROR}$ | 21 |
| G | 1 |
| G×THC | 7 |
| M×G + M×THC×G + P(M, THC, G) = $SP_{ERROR}$ | 280 |
| A | 1 |
| A×G | 1 |
| A×THC | 7 |
| A×G×THC | 7 |
| M×A + A×$WP_{ERROR}$ + A×$SP_{ERROR}$ = $SSP_{ERROR}$ | 304 |
| C. TOTAL | 639 |

$3 + 21 + (5-1)\times 4 \times 8 \times 2 = 280$

3. (CONTINUED)

IF YOU DON'T SEE HOW TO FIND THE WHOLE-PLOT ERROR, SPLIT-PLOT ERROR, AND SPLIT-SPLIT-PLOT ERROR DF IN THE PREVIOUS ANOVA TABLE, YOU CAN USE THE FOLLOWING ALTERNATIVE REASONING.

WHOLE-PLOT: THIS IS A RCBD WITH MONTHS AS BLOCKS AND GROWTH CHAMBERS AS EXPERIMENTAL UNITS. THERE ARE $4 \times 8 = 32$ WHOLE-PLOT EXERIMENTAL UNITS, SO THE WHOLE-PLOT ANOVA IS

| | |
|---|---|
| MONTH | $4-1$ |
| THC | $8-1$ |
| W.P. ERROR | $21 = (32-1)-(4-1)-(8-1)$ |
| C. TOTAL | $32-1$ |

SPLIT-PLOT: THE SPLIT-PLOT TREATMENT FACTOR IS GENOTYPE, AND THE SPLIT-PLOT EXPERIMONTAL UNITS ARE PLANTS. WE HAVE $4 \times 8 \times 5 \times 2 = 320$ PLANTS TOTAL. THUS, THE ANOVA IS

| | |
|---|---|
| MONTH | 3 |
| THC | 7 |
| MONTH x THC | 21 |
| G | 1 |
| G x THC | 7 |
| S.P. ERROR | $319-(3+7+21+1+7) = 280$ |
| C. TOTAL | 319 |

THE SPLIT-SPLIT-PLOT ERROR DF CAN BE OBTAINED BY SUBTRACTION ALSO BY FOLLOWING THE SAME STRATEGY.

3. (CONTINUED)

THE MODEL SPECIFIED IN THE PROBLEM ASSUMES THE CHAMBER EFFECTS ARE COMPLETELY NEW EACH MONTH AND INDEPENDENT OF THE CHAMBER EFFECTS IN ANY OTHER MONTH. AN ALTERNATIVE MODEL WOULD ASSUME A TOTAL OF 8 RATHER THAN 32 GROWTH CHAMBER EFFECTS. YET ANOTHER APPROACH WOULD CONSIDER A REPEATED-MEASURES CORRELATION STRUCTURE (LIKE AR(1)) WITHIN EACH GROWTH CHAMBER ACROSS MONTHS. BECAUSE NEW PLANTS ARE USED IN EACH GROWTH CHAMBER EACH MONTH, NEITHER OF THESE MODELS MAY BE BETTER THAN THE MODEL SPECIFIED IN THE PROBLEM STATEMENT. THERE ARE REPEATED-MEASURES ON PLANTS (TWO AGES). THE MODEL ASSUMES COMPOUND SYMMETRY STRUCTURE, WHICH IS THE SAME AS AR(1) FOR TWO OBSERVATIONS.

3. a) NO THREE-WAY INTERACTION MEANS THAT THE TWO-WAY INTERACTIONS ARE THE SAME FOR ALL LEVELS OF THE THIRD FACTOR. (WE HAD A HOMEWORK PROBLEM ABOUT THREE-WAY INTERACTION.) THUS,

$$H_0: \bar{\mu}_{111..} - \bar{\mu}_{121..} - \bar{\mu}_{211..} + \bar{\mu}_{221..} = \bar{\mu}_{112..} - \bar{\mu}_{122..} - \bar{\mu}_{212..} + \bar{\mu}_{222..}$$

$$\Longleftrightarrow H_0: \bar{\mu}_{111..} - \bar{\mu}_{121..} - \bar{\mu}_{211..} + \bar{\mu}_{221..} - \bar{\mu}_{112..} + \bar{\mu}_{122..} + \bar{\mu}_{212..} - \bar{\mu}_{222..} = 0$$

THUS, WE NEED TO ESTIMATE

BY REPLACING EACH $\bar{\mu}_{thc..}$ WITH ITS BLUE.

WE KNOW BLUE OF $\bar{\mu}_{thc..}$ IS

$$\bar{y}_{.thc...} = \bar{\mu}_{thc..} + \bar{u}_{.} + \bar{v}_{.th c} + \bar{w}_{.thc..} + \bar{e}_{.thc...}$$

THE $\bar{u}_{.}$ WILL CANCEL IN OUR CONTRAST. THUS,

VAR OF ESTIMATOR OF

IS $8\left[\sigma_v^2/4 + \sigma_w^2/40 + \sigma_e^2/80\right]$

$$= \frac{1}{10}\left[20\sigma_v^2 + 2\sigma_w^2 + \sigma_e^2\right]$$

3 a) (CONTINUED)

THUS,

$$t = \frac{8.6 - 7.5 - 10.8 + 13.3 - 9.6 + 8.6 + 11.6 - 14.0}{\sqrt{\frac{1}{10}\left[20 \times 0.1 + 2 \times 0.4 + 0.2\right]}} = \frac{.2}{\sqrt{3/10}}$$

3b) 21 (SEE ANOVA TABLE)

3c) 280 (SEE ANOVA TABLE)

3d) $\bar{Y}_{.11111.} - \bar{Y}_{.11112.} = \mu_{11111} - \mu_{11112} + \bar{e}_{.11111.} - \bar{e}_{.11112.}$

$$VAR(\bar{Y}_{.11111.} - \bar{Y}_{.11112.}) = 2\sigma_e^2/20 = \sigma_e^2/10$$

$$SE = \sqrt{\hat{\sigma}_e^2/10} = \sqrt{0.02}$$

3e) 304 (SEE ANOVA TABLE)

4 a) $F = \dfrac{0.1}{10.2/(40-8)} = \dfrac{3.2}{10.2}$

b) $F = \dfrac{(0.2+7.9+.1)/3}{10.2/(40-8)}$

c) $5.5 \pm 2.026\sqrt{1.2}$

$\uparrow t_{.975,\,40-3}$

d) $f(x) = \gamma_0 + \gamma_1 x + \gamma_2 x^2$

$$\frac{d f(x)}{d x} = \gamma_1 + 2\gamma_2 x$$

$$\frac{df(x)}{dx} = 0 \implies x = \frac{-\gamma_1}{2\gamma_2}$$

$$\implies \hat{x} = \frac{-5.5}{2(-.11)} = 25.$$

PART (e) WAS TO FIND A CONFIDENCE INTERVAL FOR THE TEMPERATURE AT WHICH EXPECTED TOTAL LEAF AREA IS MAXIMIZED, BUT THIS PART WAS REMOVED TO KEEP EXAM LENGTH DOWN.