

Instructions: This is a closed-notes, closed-book exam. No calculator or electronic device of any kind may be used. Use nothing but a pen or pencil. Please write your name and answers on blank paper. Please do not write your answers on the pages with the questions. For questions that require extensive numerical calculations that cannot be done easily without a calculator, simply set up the calculation and leave it at that. For example, $(3.45 - 1.67)/\sqrt{2.34}$ would be an acceptable answer.

1. Consider the model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\mathbf{y}_{n \times 1}$, $\mathbf{X}_{n \times p}$, and $\boldsymbol{\beta}_{p \times 1}$ are as defined in our course notes and

$$\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{V})$$

for some unknown $\sigma^2 > 0$ and some known symmetric, positive definite matrix \mathbf{V} . Suppose $\mathbf{C}\boldsymbol{\beta}$ is an estimable function of $\boldsymbol{\beta}$. Complete the following parts. In each case, no derivation is required. You may simply write down an answer without proving that your answer is correct.

- Provide an expression for the BLUE of $\mathbf{C}\boldsymbol{\beta}$ in terms of \mathbf{C} , \mathbf{y} , \mathbf{X} , and \mathbf{V} .
 - State the distribution of the BLUE of $\mathbf{C}\boldsymbol{\beta}$.
 - Suppose $\mathbf{C} = \mathbf{c}'$ for some vector \mathbf{c} ; i.e., assume \mathbf{C} has just one row. Provide an expression for a 95% confidence interval for $\mathbf{C}\boldsymbol{\beta} = \mathbf{c}'\boldsymbol{\beta}$ in terms of \mathbf{c} , \mathbf{y} , \mathbf{X} , \mathbf{V} , and an appropriate distribution quantile.
2. Recall the classic split-plot experimental design discussed in class involving a field with four blocks, the whole-plot treatment factor genotype with three levels, and the split-plot treatment factor fertilizer with four levels. With i indexing genotypes, j indexing fertilizer amounts, and k indexing blocks, we proposed the model

$$y_{ijk} = \mu_{ij} + b_k + w_{ik} + e_{ijk},$$

where

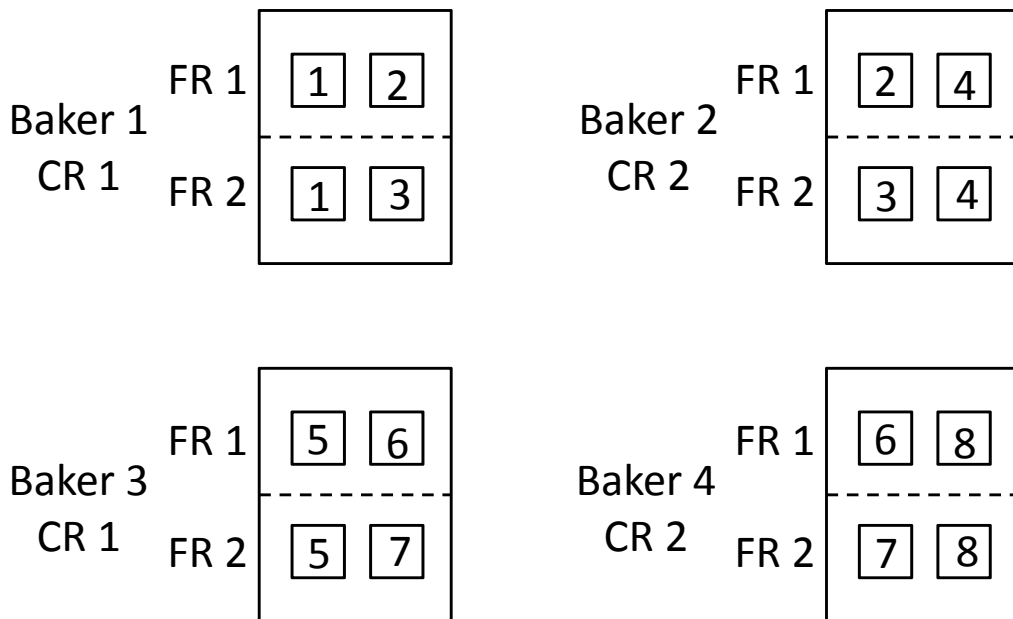
- y_{ijk} is the yield from the split-plot experimental unit for genotype i , fertilizer j , and block k ;
- μ_{ij} = mean for genotype i , fertilizer j ;
- b_k = random effect for block k with $b_k \sim N(0, \sigma_b^2)$;
- w_{ik} = random effect for the whole plot assigned genotype i in block k with $w_{ik} \sim N(0, \sigma_w^2)$;
- e_{ijk} = random error effect for genotype i , fertilizer j , block k with $e_{ijk} \sim N(0, \sigma_e^2)$;
- σ_b^2 , σ_w^2 , and σ_e^2 are positive variance components; and
- all random effects are independent.

Suppose we were to construct an ANOVA table that included one line for each of the factors block, genotype, and fertilizer, and all the interactions: block \times genotype, block \times fertilizer, genotype \times fertilizer, and block \times genotype \times fertilizer. The sum of squares for block \times fertilizer is

$$3 \sum_{j=1}^4 \sum_{k=1}^4 (\bar{y}_{.jk} - \bar{y}_{.j\cdot} - \bar{y}_{\cdot k} + \bar{y}_{\cdot\cdot})^2.$$

Derive the expected mean square for block \times fertilizer under the assumption that the model proposed above is the correct model.

3. Researchers were interested in determining which combination of cake recipe and frosting recipe would yield the best results. Two cake recipes (labeled CR1 and CR2) and two frosting recipes (labeled FR1 and FR2) were considered. The two cake recipes were randomly assigned to four bakers with two bakers for each cake recipe. Each baker prepared and baked a cake according to the recipe he or she was assigned. While cakes were baking and cooling, each baker prepared one batch of frosting using each of the frosting recipes. Half of each cake was randomly selected and covered with frosting prepared using FR1. The other half of each cake was covered with frosting prepared using FR2. Two pieces of frosted cake from each half of each cake were cut and delivered to eight trained judges for evaluation. Each judge tasted and separately assigned a score to two pieces of frosted cake. For the purposes of this problem, you may assume that each score follows a normal distribution. The diagram below shows the four cakes baked by the four bakers as large rectangles. The dashed line on each rectangle shows the dividing point that separates FR1 frosting from FR2 frosting. The numbered squares within each rectangle show the pieces of cake each judge scored. For example, the two squares marked with the number 1 show that judge number 1 scored two pieces of cake from a cake prepared with CR1 by baker 1. One of those pieces was frosted with FR1, and the other was frosted with FR2.



Let y_{ijk} denote the score assigned by judge k to a piece of cake prepared with CR_i and frosted with FR_j . Define the response vector by

$$\mathbf{y} = [y_{111}, y_{112}, y_{121}, y_{123}, y_{212}, y_{214}, y_{223}, y_{224}, y_{115}, y_{116}, y_{125}, y_{127}, y_{216}, y_{218}, y_{227}, y_{228}]'$$

Consider a linear mixed-effects model of the form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e},$$

where the components of the vector e will be assumed to look just like the components of y except with the letter e in place of the letter y .

- (a) Indicate the model you would fit to the data by providing expressions for X , β , Z , and u . You may use Kronecker product notation if you wish.
- (b) To complete your model specification, state the assumed distribution of

$$\begin{bmatrix} u \\ e \end{bmatrix}.$$

4. Suppose researchers wished to estimate the mean concentration of a chemical in the blood of mice living near a particular farm. Four mice were trapped near the farm and were assumed to be a simple random sample from the mouse population of interest. A blood sample was drawn from each mouse, and the concentration of the chemical was measured for each sample. To get some information about potential variation in the measurement process, the sample for the fourth mouse was measured twice. Let y_{ij} denote the j th measurement of the blood sample for the i th mouse. With this notation, our data consist of $y_{11}, y_{21}, y_{31}, y_{41}$, and y_{42} . Suppose

$$y_{ij} = \mu + u_i + e_{ij},$$

where $\mu \in \mathbb{R}$ is the parameter of interest and

$$u_1, u_2, u_3, u_4 \stackrel{i.i.d.}{\sim} N(0, \sigma_u^2) \text{ independent of } e_{11}, e_{21}, e_{31}, e_{41}, e_{42} \stackrel{i.i.d.}{\sim} N(0, \sigma_e^2)$$

for some positive variance components σ_u^2 and σ_e^2 .

- (a) Provide an expression for an unbiased estimator of σ_e^2 in terms of the data $y_{11}, y_{21}, y_{31}, y_{41}, y_{42}$.
 - (b) Provide an expression for an unbiased estimator of σ_u^2 in terms of the data $y_{11}, y_{21}, y_{31}, y_{41}, y_{42}$. Full credit will be awarded for any unbiased estimator. Thus, you may wish to report a simple unbiased estimator even if it does not make full use of all the available data.
5. Researchers are interested in determining ways to deter deer from eating ornamental plants. An experiment was conducted in eight fields. Within each field, four 15 meter \times 15 meter squares of land were studied. Within each field, researchers used the the following procedure:
 - Two of the four squares were randomly selected to be planted with plant type 1. The other two squares were planted with plant type 2.
 - One of the two squares planted with plant type 1 was randomly selected, and a fence was placed around the selected square. The other plant type 1 square was not surrounded by a fence.
 - One of the two squares planted with plant type 2 was randomly selected, and a fence was placed around the selected square. The other plant type 2 square was not surrounded by a fence.
 - Each of the four squares was divided into two rectangles of equal size. One rectangle within each square was randomly selected, and the plants growing in that rectangle were treated with a chemical. The other rectangle in each square was not treated with a chemical.
 - At the conclusion of the study, the amount of living plant biomass was measured for each rectangle.

This experiment involves four factors: field (levels 1 through 8), plant type (levels 1 and 2), fence (levels 1=yes and 2=no), and chemical (levels 1=yes and 2=no). Let y_{ijkl} be the amount of living plant biomass measured for plant type i , fence j , chemical k , and field l . Suppose

$$y_{ijkl} = \mu_{ijk} + f_l + s_{ijl} + e_{ijkl},$$

where $\mu_{ijk} \in \mathbb{R}$, $f_l \sim N(0, \sigma_f^2)$, $s_{ijl} \sim N(0, \sigma_s^2)$, $e_{ijkl} \sim N(0, \sigma_e^2)$, $\sigma_f^2 \in \mathbb{R}^+$, $\sigma_s^2 \in \mathbb{R}^+$, $\sigma_e^2 \in \mathbb{R}^+$, and all random effects are independent. The following table gives \bar{y}_{ijk} for all i, j, k .

	Chemically Treated		Not Chemically Treated	
	Fence	No Fence	Fence	No Fence
Plant Type 1	18	12	15	8
Plant Type 2	16	15	14	7

Suppose an ANOVA table was constructed that contained main effects for the factors field, plant type, fence, and chemical – and all possible interactions among the factors field, plant type, fence, and chemical. The following table gives information related to a subset of lines in that ANOVA table.

Source	Mean Square	Expected Mean Square
field	121.0	$8\sigma_f^2 + 2\sigma_s^2 + \sigma_e^2$
field×plant type	16.0	$2\sigma_s^2 + \sigma_e^2$
field×fence	16.5	$2\sigma_s^2 + \sigma_e^2$
field×chemical	2.3	σ_e^2
field×plant type×fence	15.5	$2\sigma_s^2 + \sigma_e^2$
field×plant type×chemical	1.7	σ_e^2
field×fence×chemical	2.1	σ_e^2
field×plant type×fence×chemical	1.9	σ_e^2

- Compute the BLUE of $\bar{\mu}_{.1} - \bar{\mu}_{.2}$. (the difference relevant for assessing the main effect of fence).
- Compute a standard error for your estimate in part (a).
- Compute the BLUE of $\bar{\mu}_{.11} - \bar{\mu}_{.21}$ (the difference relevant for assessing the effect of fence when plants are chemically treated).
- Compute a standard error for your estimate in part (c).
- Compute an F -statistic for testing the null hypothesis of no fence×chemical interactions.