

**STAT 511      Exam 2      Spring 2012**

1. If  $\mathbf{y} = \mathbf{1}\mu + \boldsymbol{\epsilon}$ , where  $\text{Var}(\boldsymbol{\epsilon}) = \sigma_1^2 \mathbf{I} + \sigma_2^2 \mathbf{1}\mathbf{1}'$  for some  $\sigma_1^2, \sigma_2^2 > 0$ , then the BLUE of  $\mu$  is  $\bar{y}$ , the average of the entries in the vector  $\mathbf{y}$ . Use this fact, along with whatever else you may know about best linear unbiased estimation, to compute the BLUE of  $\mu$  in the following situation.

Suppose

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} \sim N \left( \begin{bmatrix} \mu \\ \mu \\ \mu \\ \mu \\ \mu \end{bmatrix}, \begin{bmatrix} 5 & 1 & 1 & 1 & 0 \\ 1 & 5 & 1 & 1 & 0 \\ 1 & 1 & 5 & 1 & 0 \\ 1 & 1 & 1 & 5 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{bmatrix} \right), \text{ and we observe } \mathbf{y} = \begin{bmatrix} 5 \\ 3 \\ 8 \\ 8 \\ 9 \end{bmatrix}.$$

Note that finding the BLUE of  $\mu$  in this situation does not require inverting any matrices. All the computations needed to obtain the BLUE can be carried out by hand in a minute or two.

2. An experiment was conducted to study mean plant height of two genotypes exposed to three watering levels. The experiment was conducted in 4 greenhouses. Each greenhouse contained three tables. On each table, were 2 pots with 1 plant in each pot. The 2 plants on any given table consisted of 1 plant of one genotype and 1 plant of the other genotype, with genotypes randomly assigned to the pots. Within each greenhouse, the three watering levels were randomly assigned to the three tables, with one table per watering level. Thus, the two plants on any given table received the same amount of water throughout the experiment. At the conclusion of the experiment, the height of each plant was recorded.

For  $i = 1, 2, 3, 4; j = 1, 2, 3$ ; and  $k = 1, 2$ ; let  $y_{ijk}$  denote the height recorded for the plant associated with greenhouse  $i$ , watering level  $j$ , and genotype  $k$ . Consider the following model that will be referred to henceforth as MODEL 1.

$$y_{ijk} = \mu + g_i + \omega_j + t_{ij} + \gamma_k + \phi_{jk} + e_{ijk} \quad (i = 1, 2, 3, 4; j = 1, 2, 3; k = 1, 2),$$

where the  $g_i$  terms are  $N(0, \sigma_g^2)$ , the  $t_{ij}$  terms are  $N(0, \sigma_t^2)$ , the  $e_{ijk}$  terms are  $N(0, \sigma_e^2)$ , all these random terms are mutually independent, and the remaining terms in the model are unknown fixed parameters.

- (a) According to MODEL 1, what is the correlation between the heights of two plants growing together on the same table? (Note that this question is asking for the correlation rather than the covariance.)
- (b) In terms of the MODEL 1 parameters, write down the null hypothesis of no watering level main effects.

(c) MODEL 1 can be written as  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}$ , where

$$\mathbf{y} = (y_{111}, y_{112}, y_{121}, y_{122}, y_{131}, y_{132}, y_{211}, y_{212}, y_{221}, y_{222}, y_{231}, y_{232}, y_{311}, y_{312}, y_{321}, y_{322}, y_{331}, y_{332}, y_{411}, y_{412}, y_{421}, y_{422}, y_{431}, y_{432})'$$

Provide corresponding expressions for  $\mathbf{X}$ ,  $\boldsymbol{\beta}$ ,  $\mathbf{Z}$ , and  $\mathbf{u}$ . You may wish to express parts of your answer using Kronecker product notation.

(d) Now suppose data have been collected and stored in the vector  $\mathbf{y}$  in R. Suppose GH, WL, and GENO are factors in R corresponding to the experimental factors greenhouse, watering level, and genotype, respectively. Consider the follow R commands and output.

```
o=lm(y~GH*WL*GENO)
anova(o)
```

Analysis of Variance Table

```
Response: y
      Df Sum Sq Mean Sq
GH      3  113.3    37.8
WL      2  321.8   160.9
GENO    1    2.5     2.5
GH:WL   6  116.4    19.4
GH:GENO 3   11.7     3.9
WL:GENO 2   75.1    37.5
GH:WL:GENO 6   14.5     2.4
```

Suppose our goal is to analyze the data under the assumption that MODEL 1 is correct. Use information in the output to compute three  $F$ -statistics that can be used to test for

- i. watering level main effects,
- ii. genotype main effects, and
- iii. watering level by genotype interactions, respectively.

3. An experiment was conducted to compare the effects of several different management practices on the health of pigs. A total of 20 farms were involved in the experiment. Two housing types (HT1 and HT2) and two diets (D1 and D2) were considered. A completely randomized design was used to assign the four combinations of housing type and diet to the 20 farms with 5 farms for each combination of housing type and diet. Within each farm, 8 sows (mother pigs) were raised according to the assigned housing type and diet. Three piglets from one litter born to each sow were randomly assigned to 3 different vitamin levels (V1, V2, and V3), with one pig for each vitamin level per sow. When the piglets reached 9 weeks of age, a response of interest was measured for each piglet. The researchers were interested in understanding how the factors housing type, diet, and vitamin level and the possible interactions among these factors affect the mean response.

- (a) Consider the linear mixed-effects model that you would fit to data from this experiment and write down the corresponding ANOVA table with columns for *Source* and *Degrees of Freedom*. (You do not need to write down the model. Just provide the ANOVA table.) The last two lines of the table should have the entries *error* and *corrected total*. (It is not necessary to separately indicate the components of the *error* line.) You may abbreviate the names of other entries in the *Source* column, but please use intuitive abbreviations.
- (b) Circle each term in the *Source* column of your ANOVA table that corresponds to a random factor.

4. A total of 8 plants were assigned to two treatments using a completely randomized design with 4 plants per treatment group. The response variable of interest (the concentration of a chemical) was measured for each plant. Suppose the true chemical concentrations of plants are independent. Furthermore, for  $i = 1, 2$ , suppose the true chemical concentration of a plant receiving treatment  $i$  is normally distributed with a mean  $\mu_i$  and a variance  $\sigma_p^2$ .

A special measuring device is required to measure the response. When applied to a plant, that device returns a measurement that is equal to the true chemical concentration in the plant plus a normally distributed measurement error with mean zero and variance  $\sigma_e^2$ . Suppose all measurement errors are independent of each other and independent of the true chemical concentrations of the plants.

After measuring the response for 2 plants in each treatment group, the measuring device malfunctioned and needed to be replaced. The company provided a new measuring device that functions just like the old device except that the measurement error variance has been cut in half (i.e., reduced from  $\sigma_e^2$  to  $\sigma_e^2/2$ ). The remaining 2 plants in each treatment group were measured with the new device. Due to limited plant material, it was only possible to measure each plant one time. The following data were obtained.

Treatment	Plant ID	Measuring Device	Measurement of Chemical Concentration
1	1	old	65
1	2	old	59
1	3	new	64
1	4	new	63
2	5	old	82
2	6	old	76
2	7	new	77
2	8	new	70

- (a) Assume that  $\sigma_p^2/\sigma_e^2$  is known to be 3. Find the BLUE of  $\mu_2$ .
- (b) Again assume that  $\sigma_p^2/\sigma_e^2$  is known to be 3. Find the BLUP of the chemical concentration of the plant with ID number 8.
- (c) Let  $\mathbf{y}$  denote the random vector whose observed value is the last column of the data table. Find a matrix  $\mathbf{A}'$  such that  $\mathbf{A}'\mathbf{y}$  is a vector of uncorrelated error contrasts that could be used to find REML estimates of  $\sigma_p^2$  and  $\sigma_e^2$ . You do not need to actually find REML estimates. You need only to specify the matrix  $\mathbf{A}'$  with the desired properties.