Stat 5100 **Assignment 9**

Due: Sunday, April 20th 11:59PM in gradescope.

Problem 1 *A plant scientist was interested in comparing two plant genotypes (1 and 2). An experiment was conducted in a greenhouse with one table, eight trays, and sixteen pots. The table in the greenhouse held the eight trays with two pots on each tray. For each of the eight trays, two genotype 1 seeds were planted in one pot, and two genotype 2 seeds were planted in the other pot. The assignment of genotypes 1 and 2 to the two pots within each tray was determined by flipping a fair coin. The response of interest is a quantitative measurement of overall plant health that was calculated for each plant 42 days after planting. These quantitative measurements of overall plant health are presented as integers in Table 1 to make calculations easier, but please answer all questions as if each measurement is a realization from a normal distribution.*

Table 1. Measurements of overall plant health for each plant.

	Genotype 1 Pot		*Genotype 2 Pot*	
Tray	*Plant 1*	*Plant 2*	*Plant 1*	*Plant 2*
1	8	7	6	7
2	8	9	4	5
3	8	8	7	7
4	5	7	4	2
5	5	6	4	3
6	9	10	7	9
7	5	7	1	4
8	4	6	5	5

Let i index genotypes ($i = 1, 2$), j index trays ($j = 1, \dots, 8$), and k index plants within pots ($k = 1, 2$). Let y_{ijk} denote the response corresponding to genotype i, tray j, and plant k. Suppose

$$y_{ijk} = \mu_i + t_j + p_{ij} + e_{ijk} \quad \forall\, i, j, k, \tag{1}$$

where μ_1 and μ_2 are unknown real-valued parameters, $t_j \sim \mathcal{N}(0, \sigma_t^2)\ \forall\, j$, $p_{ij} \sim \mathcal{N}(0, \sigma_p^2)\ \forall\, i, j$, $e_{ijk} \sim \mathcal{N}(0, \sigma_e^2)\ \forall\, i, j, k$, and all t_j, p_{ij}, and e_{ijk} terms are mutually independent.

a) *Explain what the p_{ij} terms represent and provide one reason for including them in model (1).*

b) *Let $\bar{y}_{ij\cdot} = \frac{1}{2}\sum_{k=1}^{2} y_{ijk}\ \forall\, i, j$. Determine the distribution of $\bar{y}_{11\cdot} - \bar{y}_{21\cdot}$.*

c) *Compute the value of an unbiased estimator of the variance of $\bar{y}_{11\cdot} - \bar{y}_{21\cdot}$.*

d) *Provide a 95% confidence interval for $\mu_1 - \mu_2$.*

e) *Model (1) can be written in the form $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{e}$. Provide $\boldsymbol{X}$, $\boldsymbol{\beta}$, $\boldsymbol{Z}$, and $\boldsymbol{u}$.*

f) *Suppose the researchers would like to repeat their experiment again, using the same basic resources: eight trays, two pots per tray, sixteen seeds of genotype 1, and sixteen seeds of genotype 2. Would you recommend any changes to their experimental design? Explain why or why not.*

Problem 2 *This is a continuation of the previous question, Problem 1. Suppose the experiment actually involved a second factor –* bacterial infection *with levels 1=present and 2=absent – in addition to the factor* genotype*, randomly assigned to pots as discussed previously. Within each pot, one of the two plants was randomly selected for infection with a bacteria, which was applied by rubbing a gel containing the bacteria on the top leaf of the plant. The other plant in each pot was rubbed with the same gel but with the bacteria absent.*

Let y_{ijk} bet the response for the plant of genotype i on tray j that received bacterial infection k ($i = 1, 2; j = 1, \ldots, 8; k = 1, 2$). Suppose the data are the same as in Table 1 and arranged so that in each pot, Plant 1 corresponds to the plant infected with the bacteria and Plant 2 corresponds to the plant not infected with the bacteria. Suppose

$$y_{ijk} = \mu_{ik} + t_j + p_{ij} + e_{ijk} \quad \forall\, i, j, k, \tag{2}$$

where as in model (1), $t_j \sim \mathcal{N}(0, \sigma_t^2)\ \forall\, j$, $p_{ij} \sim \mathcal{N}(0, \sigma_p^2)\ \forall\, i, j$, $e_{ijk} \sim \mathcal{N}(0, \sigma_e^2)\ \forall\, i, j, k$, and all t_j, p_{ij}, and e_{ijk} terms are mutually independent.

a) This is a split-plot experiment. What are the whole-plot experimental units?

b) What are the split-plot experimental units?

c) What is the whole-plot treatment factor?

d) What is the split-plot treatment factor?

e) Create an ANOVA table with columns Source *and* Degrees of Freedom.

f) Give formulas for each of the Sums of Squares *of the ANOVA table. (Shortcut formulas for degrees of freedom and sums of squares work in this case because of the balanced experimental design.)*

g) Derive the expected mean square for the second to last line of the ANOVA table (the line right before corrected total). This line is typically called error or split-plot error.

h) Compute the value of the best linear unbiased estimator of $\mu_{11} - \mu_{12}$.

i) Derive an expression for the variance of the best linear unbiased estimator of $\mu_{11} - \mu_{12}$ in terms of model (2) parameters.

j) Compute a 95% confidence interval for $\mu_{11} - \mu_{12}$.

k) Determine the distribution of $y_{111} - y_{112} - y_{211} + y_{212}$.

l) Compute the value of the best linear unbiased estimator of $\mu_{11} - \mu_{12} - \mu_{21} + \mu_{22}$.

m) Derive an expression for the variance of the best linear unbiased estimator of $\mu_{11} - \mu_{12} - \mu_{21} + \mu_{22}$ in terms of model (2) parameters.

Problem 3 *Researchers created a device to test the effectiveness of helmets at reducing the stress caused by head impacts. The device includes a head-shaped sensor on which a helmet can be placed, as well as a striking weight that can produce impacts to the front or side of a helmet placed on the sensor. The intensity of each impact can be controlled by the researchers. When an impact is delivered, a measurement of the amount of stress experienced by the head-shaped sensor is recorded. A measurement of 0 indicates no stress, while a measurement of 100 indicates stress high enough to cause serious brain injury.*

The researchers used the device to test a total of 10 helmets consisting of 5 helmets of type 1 and 5 helmets of type 2. The 10 helmets were tested in random order. When each helmet was tested, it was struck a total of 4 times: once with low impact to the front, once with high impact to the front, once with low impact to the side, and once with high impact to the side. The order of the 4 impacts was determined separately for each helmet using the following procedure. A fair coin was flipped. If the result of the flip was heads, the first two impacts were front impacts and the last two impacts were side impacts. If the result of the flip was tails, the first two impacts were side impacts and the last two impacts were front impacts. For the first two impacts, the coin was flipped again. If the result of the flip was heads, the first impact was at low intensity and the second was at high intensity. If the result of the flip was tails, the first impact was at high intensity and the second at low intensity. A coin was flipped a third time to determine the order of the impact intensities for the third and fourth impacts so that each order (low and then high vs. high and then low) was equally likely.

Let $i = 1, 2$ index helmet types 1 and 2. Let $j = 1, \ldots, 5$ index helmets nested within helmet types. Let $k = 1, 2$ index the direction of impact, with $k = 1$ for front and $k = 2$ for side. Let $\ell = 1, 2$ index the intensity of the impact, with $\ell = 1$ for low and $\ell = 2$ for high. Let $y_{ijk\ell}$ be the stress measurement for the corresponding values of i, j, k, and ℓ. For $i = 1, 2$, $j = 1, \ldots, 5$, $k = 1, 2$, and $\ell = 1, 2$, consider the model

$$y_{ijk\ell} = \mu_{ik\ell} + a_{ij} + b_{ijk} + e_{ijk\ell}, \tag{3}$$

where the $\mu_{ik\ell}$ values are unknown parameters, $a_{ij} \sim \mathcal{N}(0, \sigma_a^2)$, $b_{ijk} \sim \mathcal{N}(0, \sigma_b^2)$, $e_{ijk\ell} \sim \mathcal{N}(0, \sigma_e^2)$, and all random terms are independent. Model (3) was fit to the dataset, and the following ANOVA table was obtained. Because we have a balanced experimental design, the type I and type III sums of squares are the same, and the lines of the ANOVA table can be reordered in a variety of ways without changing the results.

Source	*Sum of Squares*	*Expected Mean Square*
Type	*226*	
Direction	*255*	
Intensity	*8910*	
Type×Direction	*207*	
Type×Intensity	*2*	
Direction× Intensity	*7*	
Type×Direction×Intensity	*9*	
Helmet(Type)	*254*	$4\sigma_a^2 + 2\sigma_b^2 + \sigma_e^2$
Direction×Helmet(Type)	*114*	$2\sigma_b^2 + \sigma_e^2$
Error	*59*	σ_e^2
C. Total	*10043*	

a) *We learned a shortcut for expressing sums of squares in summation notation that works for balanced designs like the one considered here. Use that shortcut to express the sum of squares for Direction × Intensity using summation notation.*

b) *Compute a t statistic that can be used to test $H_0 : \bar{\mu}_{1\cdot\cdot} = \bar{\mu}_{2\cdot\cdot}$.*

c) *The statistic in part (b) has a noncentral t distribution. Provide an expression for the noncentrality parameter in terms of model (3) parameters.*

d) *Compute the value of an unbiased estimator for σ_a^2.*

a) *The best linear unbiased estimator of $\bar{\mu}_{12\cdot} - \bar{\mu}_{11\cdot}$ is equal to 0.5 for this dataset. Provide a 95% confidence interval for $\bar{\mu}_{12\cdot} - \bar{\mu}_{11\cdot}$.*

e) *Compute a standard error for the best linear unbiased estimator of $\mu_{121} - \mu_{111}$.*