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.

Tray	Genotype 1 Pot		Genotype 2 Pot	
	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 ($j = 1, \ldots, 8$), 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, $t_j \sim N(0, \sigma_t^2) \;\forall\, j$, $p_{ij} \sim N(0, \sigma_p^2) \;\forall\, i, j$, $e_{ijk} \sim N(0, \sigma_e^2) \;\forall\, i, j, k$, 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 $\bar{y}_{ij.} = \frac{1}{2} \sum_{k=1}^{2} y_{ijk} \;\forall\, i, j$. Determine the distribution of $\bar{y}_{11.} - \bar{y}_{21.}$.

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

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

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

2. Now suppose the experiment described in problem 1 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), $t_j \sim N(0, \sigma_t^2)\ \forall\, j$, $p_{ij} \sim N(0, \sigma_p^2)\ \forall\, i, j$, $e_{ijk} \sim N(0, \sigma_e^2)\ \forall\, i, j, k$, 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.

(n) Compute an t-statistic that can be used to test for interaction between the factors *genotype* and *bacterial infection*.

(o) Find the p-value associated with the t-statistic in part (n).