

1. a) The p_{ij} terms are random effects corresponding to pots. Pots are the experimental units in this experiment because the levels of the factor genotype were randomly assigned to pots. Because there are multiple observations per experimental unit, it is important to explicitly include a random effect for each experimental unit in the model. These random effects account for pot-to-pot variation in the response that may occur due differences among pots. The pot random effects allow for the correlation between the two observations from a given pot to be larger than the correlation between two observations from different pots.

- b) Note that $\bar{y}_{A1\cdot} - \bar{y}_{B1\cdot} = \mu_A - \mu_B + p_{A1} - p_{B1} + \bar{e}_{A1\cdot} - \bar{e}_{B1\cdot}$. Thus,

$$E(\bar{y}_{A1\cdot} - \bar{y}_{B1\cdot}) = \mu_A - \mu_B \text{ and } \text{Var}(\bar{y}_{A1\cdot} - \bar{y}_{B1\cdot}) = \sigma_p^2 + \sigma_p^2 + \sigma_e^2/2 + \sigma_e^2/2 = 2\sigma_p^2 + \sigma_e^2.$$

Because linear combinations of normal random variables are normal,

$$\bar{y}_{A1\cdot} - \bar{y}_{B1\cdot} \sim N(\mu_A - \mu_B, 2\sigma_p^2 + \sigma_e^2).$$

- c) Let $d_j = \bar{y}_{Aj\cdot} - \bar{y}_{Bj\cdot} \forall j$. From problem 1(b), it follows that

$$d_1, \dots, d_8 \stackrel{iid}{\sim} N(\mu_A - \mu_B, 2\sigma_p^2 + \sigma_e^2).$$

Thus,

$$s_d^2 = \sum_{j=1}^8 (d_j - \bar{d})^2 / (8 - 1)$$

is an unbiased estimator of $2\sigma_p^2 + \sigma_e^2$. Straightforward calculations yield the estimate $13.5/7 \approx 1.929$.

- d) $\bar{d} \pm t_{0.975,7} \sqrt{s_d^2/8} \iff 2 \pm 2.36 \sqrt{(13.5/7)/8} \iff 2 \pm 2.36 * 0.49 \iff 0.84 \text{ to } 3.16$
e) If we order the \mathbf{y} vector by working our way across trays with the genotype A pot first and the genotype B pot second within each tray, we have

$$\mathbf{X} = \mathbf{1}_{8 \times 1} \otimes \mathbf{I}_{2 \times 2} \otimes \mathbf{1}_{2 \times 1}, \quad \boldsymbol{\beta} = (\mu_1, \mu_2)', \quad \mathbf{Z} = [\mathbf{I}_{8 \times 8} \otimes \mathbf{1}_{4 \times 1}, \mathbf{I}_{16 \times 16} \otimes \mathbf{1}_{2 \times 1}]$$

$$\text{and } \mathbf{u} = [t_1, t_2, \dots, t_8, p_{A1}, p_{B1}, p_{A2}, p_{B2}, \dots, p_{A8}, p_{B8}]'.$$

- f) It would be better to put one plant of genotype A and one plant of genotype B in each of the 16 pots. Then the variance of the genotype A average minus the genotype B average would be $\sigma_e^2/8$ because both tray and pot random effects would cancel in the difference of averages. This variance ($\sigma_e^2/8$) is less than the variance for the estimated difference in genotype means for the original design: $(2\sigma_p^2 + \sigma_e^2)/8 = \sigma_p^2/4 + \sigma_e^2/8$. Thus, we have a more precise unbiased estimator of the difference in genotype means if we put one plant of each genotype in each pot.

-
2. a) The experiment has a split-plot design with soil moisture as the whole-plot treatment factor and tray as the whole-plot experimental unit. The F -statistic for testing for soil moisture main effects is

$$F = \text{MS}_{\text{soilmois}} / \text{MS}_{\text{tray}} = 36.125 / 7.062 \approx 5.12.$$

This statistic can also be computed by hand from the raw data. Find the average response for each tray. Use those averages as data to compute a two-sample t -statistic for comparing the mean of the trays assigned to the low soil moisture treatment to the mean of the trays assigned to the high soil moisture treatment. Square this t -statistic to get the F -statistic.

- b) The degrees of freedom are 1 and 6.
c) The null hypothesis can be stated in several equivalent ways. One correct answer is

$$H_0 : \mu_{A1} - \mu_{B1} = \mu_{A2} - \mu_{B2}.$$

- d) The experiment has a split-plot design with soil moisture as the whole-plot treatment factor and genotype as the split-plot treatment factor. The split-plot experimental units are pots, which correspond to combinations of tray and genotype. Thus, the $\text{tray}:\text{geno}$ term represents the error term for testing for two-factor interaction. The F -statistic for testing for interaction is

$$F = \text{MS}_{\text{soilmois:geno}} / \text{MS}_{\text{tray:geno}} = 10.125 / 0.562 \approx 18.0.$$

This F -statistic has 1 numerator degree of freedom. Thus, the corresponding t -statistic is $t \approx \sqrt{18.0} \approx \pm 4.24$.

This t -statistic can also be computed by hand from the raw data. Find the average response for each pot. Compute the difference between the genotype A pot average and the genotype B pot average for each tray. Use those differences as data to compute a two-sample t -statistic for comparing the mean of the differences for trays assigned to the low soil moisture treatment to the mean of the differences for trays assigned to the high soil moisture treatment.

- e) The t -statistic has 6 degrees of freedom.
f) The noncentrality parameter associated with the t -statistic is easy to obtain by replacing the estimates in the t -statistic formula with the parameters they estimate. From the description of the t -statistic calculation from the raw data, it follows that the noncentrality parameter is

$$\frac{(\mu_{A1} - \mu_{B1}) - (\mu_{A2} - \mu_{B2})}{\sqrt{(2\sigma_p^2 + \sigma_e^2)(1/4 + 1/4)}} = \frac{\mu_{A1} - \mu_{B1} - \mu_{A2} + \mu_{B2}}{\sqrt{\sigma_p^2 + \sigma_e^2/2}}.$$

-
3. a) The Best Linear Unbiased Estimator (BLUE) of $\bar{\mu}_A - \bar{\mu}_B$ is the average of the genotype A observations minus the average of the genotype B observations in both these balanced split-plot designs. In the first experiment, the variance of the differences in averages is

$$\text{Var}(\bar{p}_{A\cdot} - \bar{p}_{B\cdot} + \bar{e}_{A\cdot\cdot} - \bar{e}_{B\cdot\cdot}) = \sigma_p^2/4 + \sigma_e^2/8.$$

In the second experiment, the variance of the differences in averages is the same except for the addition of the term

$$\text{Var}[(t_2 + t_4 + t_5 + t_7)/4 - (t_1 + t_3 + t_6 + t_8)/4] = \sigma_t^2/2.$$

Because we would expect the estimate of $\sigma_t^2/2 + \sigma_p^2/4 + \sigma_e^2/8$ to be bigger than the estimate of $\sigma_p^2/4 + \sigma_e^2/8$, se_2 is expected to be bigger than se_1 .

- b) The BLUE uses inverse variance weighting to combine independent estimators. Because we do not know the variances, we use estimates of the variances to do the weighting. This yields the estimate

$$\frac{\frac{1}{se_1^2}2.0 + \frac{1}{se_2^2}1.5}{\frac{1}{se_1^2} + \frac{1}{se_2^2}} = \frac{(se_2^2)2.0 + (se_1^2)1.5}{se_1^2 + se_2^2}.$$