

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  $\bar{y}_{11\cdot} - \bar{y}_{21\cdot} = \mu_1 - \mu_2 + p_{11} - p_{21} + \bar{e}_{11\cdot} - \bar{e}_{21\cdot}$ . Thus,

$$E(\bar{y}_{11\cdot} - \bar{y}_{21\cdot}) = \mu_1 - \mu_2 \text{ and } \text{Var}(\bar{y}_{11\cdot} - \bar{y}_{21\cdot}) = \sigma_p^2 + \sigma_p^2 + \sigma_e^2/2 + \sigma_e^2/2 = 2\sigma_p^2 + \sigma_e^2.$$

Because linear combinations of normal random variables are normal,

$$\bar{y}_{11\cdot} - \bar{y}_{21\cdot} \sim N(\mu_1 - \mu_2, 2\sigma_p^2 + \sigma_e^2).$$

- c) Let  $d_j = \bar{y}_{1j\cdot} - \bar{y}_{2j\cdot} \forall j$ . From problem 1(b), it follows that

$$d_1, \dots, d_8 \stackrel{iid}{\sim} N(\mu_1 - \mu_2, 2\sigma_p^2 + \sigma_e^2).$$

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 1 pot first and the genotype 2 pot second within each tray, we have

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

$$\text{and } \mathbf{u} = [t_1, t_2, \dots, t_8, p_{11}, p_{21}, p_{12}, p_{22}, \dots, p_{18}, p_{28}]'$$

- f) It would be better to put one plant of genotype 1 and one plant of genotype 2 in each of the 16 pots. Then the variance of the genotype 1 average minus the genotype 2 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 whole-plot experimental units are pots.  
 b) The split-plot experimental units are plants.  
 c) The whole-plot experimental treatment factor is genotype.  
 d) The split-plot experimental treatment factor is bacterial infection.  
 e) The ANOVA table is

Source	DF	SS
trays	$8-1=7$	$\sum_{i=1}^2 \sum_{j=1}^8 \sum_{k=1}^2 (\bar{y}_{.j} - \bar{y}_{...})^2$
genotypes	$2-1=1$	$\sum_{i=1}^2 \sum_{j=1}^8 \sum_{k=1}^2 (\bar{y}_{i..} - \bar{y}_{...})^2$
trays $\times$ genotypes	$(8-1) \times (2-1) = 7$	$\sum_{i=1}^2 \sum_{j=1}^8 \sum_{k=1}^2 (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j} + \bar{y}_{...})^2$
infections	$2-1=1$	$\sum_{i=1}^2 \sum_{j=1}^8 \sum_{k=1}^2 (\bar{y}_{..k} - \bar{y}_{...})^2$
genotypes $\times$ infections	$(2-1) \times (2-1) = 1$	$\sum_{i=1}^2 \sum_{j=1}^8 \sum_{k=1}^2 (\bar{y}_{i.k} - \bar{y}_{i..} - \bar{y}_{..k} + \bar{y}_{...})^2$
error	$(8-1) \times (2-1)$ $+ (8-1) \times (2-1) \times (2-1) = 14$	$\sum_{i=1}^2 \sum_{j=1}^8 \sum_{k=1}^2 (y_{ijk} - \bar{y}_{ij.} - \bar{y}_{i.k} + \bar{y}_{i..})^2$
Corrected total	$32-1 = 31$	$\sum_{i=1}^2 \sum_{j=1}^8 \sum_{k=1}^2 (y_{ijk} - \bar{y}_{...})^2$

- f) See the above ANOVA table

g)

$$\begin{aligned}
E(MS_{error}) &= \frac{1}{14} E(SS_{error}) \\
&= \frac{1}{14} E\left\{ \sum_{i=1}^2 \sum_{j=1}^8 \sum_{k=1}^2 (y_{ijk} - \bar{y}_{ij.} + \bar{y}_{i.k} + \bar{y}_{i..})^2 \right\} \\
&= \frac{1}{14} E\left\{ \sum_{i=1}^2 \sum_{j=1}^8 \sum_{k=1}^2 \left[ (\mu_{ik} + t_j + p_{ij} + e_{ijk}) - (\bar{\mu}_i. + t_j + p_{ij} + \bar{e}_{ij.}) \right. \right. \\
&\quad \left. \left. - (\mu_{ik} + \bar{t}_. + \bar{p}_{i.} + \bar{e}_{i.k}) + (\bar{\mu}_i. + \bar{t}_. + \bar{p}_{i.} + \bar{e}_{i..}) \right]^2 \right\} \\
&= \frac{1}{14} E\left\{ \sum_{i=1}^2 \sum_{j=1}^8 \sum_{k=1}^2 (e_{ijk} - \bar{e}_{ij.} - \bar{e}_{i.k} + \bar{e}_{i..})^2 \right\} \\
&= \frac{1}{14} E\left\{ \sum_{i=1}^2 \sum_{j=1}^8 \sum_{k=1}^2 \left[ (e_{ijk} - \bar{e}_{ij.}) - (\bar{e}_{i.k} - \bar{e}_{i..}) \right]^2 \right\} \\
&= \frac{1}{14} \sum_{i=1}^2 \sum_{j=1}^8 E\left\{ \sum_{k=1}^2 \left[ (e_{ijk} - \bar{e}_{ij.}) - (\bar{e}_{i.k} - \bar{e}_{i..}) \right]^2 \right\} \\
&= \frac{1}{14} \sum_{i=1}^2 \sum_{j=1}^8 (2-1) \text{Var}(e_{ijk} - \bar{e}_{i.k}) \\
&= \frac{1}{14} \sum_{i=1}^2 \sum_{j=1}^8 (2-1) \text{Var}\left(\frac{7}{8}e_{ijk} - \frac{1}{8} \sum_{j' \neq j} e_{ij'k}\right) \\
&= \frac{1}{14} \times 2 \times 8 \times 1 \times \left[ \left(\frac{7}{8}\right)^2 + 7 \times \left(\frac{1}{8}\right)^2 \right] \sigma_e^2 \\
&= \sigma_e^2
\end{aligned}$$

h) The BLUE is  $\bar{y}_{1.1} - \bar{y}_{1.2} = -1$ .

i)

$$\begin{aligned}
\text{Var}(\widehat{\mu_{11}} - \widehat{\mu_{12}}) &= \text{Var}(\bar{y}_{1.1} - \bar{y}_{1.2}) \\
&= \text{Var}((\mu_{11} + \bar{t}_. + \bar{p}_{1.} + \bar{e}_{1.1}) - (\mu_{12} + \bar{t}_. + \bar{p}_{1.} + \bar{e}_{1.2})) \\
&= \text{Var}(\bar{e}_{1.1} - \bar{e}_{1.2}) \\
&= \frac{1}{4} \sigma_e^2
\end{aligned}$$

j) The estimate for the variance in (i) is obtained by dividing the  $MS_{error}$  in the second last line of the ANOVA table by 4, which is

---


$$\begin{aligned}\widehat{Var}(\widehat{\mu}_{11} - \widehat{\mu}_{12}) &= \frac{1}{4}\hat{\sigma}_e^2 \\ &= \frac{1}{4} \times 0.9286 \\ &= 0.2321\end{aligned}$$

with degrees of freedom 14 as indicated in the ANOVA table. So the 95% confidence interval for  $\mu_{11} - \mu_{12}$  is

$$\begin{aligned}\bar{y}_{1.1} - \bar{y}_{1.2} \pm t_{0.975,14} \sqrt{\widehat{Var}(\widehat{\mu}_{11} - \widehat{\mu}_{12})} \\ = -1 \pm 2.1448 \times 0.4818 \\ = (-2.0334, 0.0334)\end{aligned}$$

**k)**

$$\begin{aligned}y_{111} - y_{112} - y_{211} + y_{212} \\ = (\mu_{11} + t_1 + p_{11} + e_{111}) - (\mu_{12} + t_1 + p_{11} + e_{112}) \\ - (\mu_{21} + t_1 + p_{21} + e_{211}) + (\mu_{22} + t_1 + p_{21} + e_{212}) \\ = (\mu_{11} - \mu_{12} - \mu_{21} + \mu_{22}) + (e_{111} - e_{112} - e_{211} + e_{212}) \\ \sim N(\mu_{11} - \mu_{12} - \mu_{21} + \mu_{22}, 4\sigma_e^2)\end{aligned}$$

**l)** The BLUE is  $\bar{y}_{1.1} - \bar{y}_{1.2} - \bar{y}_{2.1} + \bar{y}_{2.2} = -0.5$ .

**m)**

$$\begin{aligned}Var(\bar{y}_{1.1} - \bar{y}_{1.2} - \bar{y}_{2.1} + \bar{y}_{2.2}) \\ = Var((\mu_{11} + \bar{t}_. + \bar{p}_{1.} + \bar{e}_{1.1}) - (\mu_{12} + \bar{t}_. + \bar{p}_{1.} + \bar{e}_{1.2}) \\ - (\mu_{21} + \bar{t}_. + \bar{p}_{2.} + \bar{e}_{2.1}) - (\mu_{22} + \bar{t}_. + \bar{p}_{2.} + \bar{e}_{2.2})) \\ = Var(\bar{e}_{1.1} - \bar{e}_{1.2} - \bar{e}_{2.1} + \bar{e}_{2.2}) \\ = \frac{4}{8}\sigma_e^2 \\ = \frac{1}{2}\sigma_e^2\end{aligned}$$

**n)** The null hypothesis is  $H_0: \mu_{11} - \mu_{12} - \mu_{21} + \mu_{22} = 0$ ,

$$\begin{aligned}\text{t-statistic} &= \frac{\bar{y}_{1.1} - \bar{y}_{1.2} - \bar{y}_{2.1} + \bar{y}_{2.2}}{\sqrt{\frac{1}{2}\hat{\sigma}_e^2}} \\ &= \frac{\bar{y}_{1.1} - \bar{y}_{1.2} - \bar{y}_{2.1} + \bar{y}_{2.2}}{\sqrt{\frac{1}{2}MS_{error}}} \\ &= -0.7338\end{aligned}$$

- 
- o) The degrees of freedom is 14 which is the degrees of freedom of the  $MS_{error}$ . Thus, the  $p$ -value =  $2P(t_{14} < -0.7338) = 0.475$ .