Fixed vs. Random Effects

The levels of a *fixed effect* are selected in a systematic fashion and inference is restricted to those levels.

The levels of a *random effect* can be thought of as a random sample from a larger population of possible levels (e.g., a random sample of technicians). Inference can be made about the entire population of levels.

Examples:

1. (Tomato varieties 1) A gardener is interested in the yields of four different varieties of tomato plant. She conducts an experiment with 24 tomato plants, 6 plants of each of the 4 varieties. There are 24 field plots available throughout her garden, each plot is large enough to accommodate one plant. She randomly allocates each plant to one of the 24 different plots.

2. (Tomato varieties 2) A gardener is interested in the yields of four different varieties of tomato plant. She conducts an experiment with 24 tomato plants, 6 plants of each of the 4 varieties. There are 12 field plots in her garden that she uses to grow tomatoes every year. Each plot is large enough to accommodate two plants.

3. (Tomato varieties 3) A gardener is interested in the yields of four different varieties of tomato plant. She conducts an experiment with 24 tomato plants, 6 plants of each of the 4 varieties. This year, there are 12 field plots available in her garden for the experiment. Each plot is large enough to accommodate two plants. However, she uses different parts of her garden for different vegetables every year.

Q: In examples 2 and 3, how should plants be assigned to plots? Randomly? Or should there be structure to it?

4. (Dairy bulls) A dairy purchases 50 Holstein bulls from a start-up artificial breeding corporation, and begins a study to determine if these bulls produce offspring that are high-yielding producers of milk. Enough semen from each bull is used so that 3 years later, approximately 60 daughter cows per bull will have milk producing records.

5. (Soils and fertilizers) Potted plant growth depends on the potting soil and the fertilizer it is grown in. Suppose 30 chrysanthemum plants, all of the same age and variety, are randomly allocated to 30 pots, one per pot, where each pot contains one combination of each of 6 soil mixtures and each of 5 fertilizers. Consider the appropriate statistical model when (a) these specific soils and fertilizers are to be evaluated, and (b) these soils represent samples from a topsoil distributor and these fertilizers represent samples from a garden products producer.

6. (Medications and clinics) A blood pressure study was undertaken to study the effect of some new medications. Fifteen clinics were randomly chosen throughout Washington State, with 5 patients on each of four treatments (placebo and three medications).

One-way Random Effects Model

$$Y_{ij} = \mu + a_i + \varepsilon_{ij}, \quad i = 1, \ldots, I; \; j = 1, \ldots, J,$$
$$a_i \sim N(0, \sigma_A^2), \quad \varepsilon_{ij} \sim N(0, \sigma_E^2).$$
$$\{a_1, \ldots, a_I, \varepsilon_{11}, \ldots, \varepsilon_{IJ}\} \text{ mutually independent.}$$

This model is appropriate when the levels of a factor are randomly sampled from a population. The random effect a_i represents the difference between the ith sampled level and the overall population mean. The variance σ_A^2 represents the variance in the population.

The a_i and the ε_{ij} are uncorrelated: $E[a_{i'}\varepsilon_{ij}] = 0 \; \forall \; i', i, j$.

What does the covariance matrix of the Y's look like?

$$cov[Y_{ij}, Y_{i'j'}] = \begin{cases} \sigma_A^2 + \sigma_E^2 & \text{if } \; i = i', j = j' \\ \sigma_A^2 & \text{if } \; i = i', j \neq j' \\ 0 & \text{if } \; i \neq i' \end{cases}$$

Observations within a factor level are correlated.

Inference is made on the *variance components* σ_E^2 and σ_A^2. In particular, we may wish to test $H : \sigma_A^2 = 0$. This replaces the test $H : \alpha_1 = \ldots = \alpha_I$ in the fixed effects model.

Correlation Structure

The model is

$$\mathbf{Y} = \mathbf{1}_n \mu + \mathbf{Z}\mathbf{a} + \boldsymbol{\varepsilon},$$

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \vdots \\ \mathbf{Y}_I \end{pmatrix}, \quad \mathbf{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_I \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \vdots \\ \boldsymbol{\varepsilon}_I \end{pmatrix},$$

$\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{iJ})'$, $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iJ})'$, $n = IJ$, and

$$\mathbf{Z} = \begin{pmatrix} \mathbf{1}_J & \mathbf{0}_J & \cdots & \mathbf{0}_J \\ \mathbf{0}_J & \mathbf{1}_J & \cdots & \mathbf{0}_J \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_J & \mathbf{0}_J & \cdots & \mathbf{1}_J \end{pmatrix}.$$

Therefore,

$$\begin{aligned} \operatorname{cov}(\mathbf{Y}) &= \operatorname{cov}(\mathbf{Z}\mathbf{a}) + \operatorname{cov}(\boldsymbol{\varepsilon}) \\ &= \mathbf{Z}\sigma_A^2 \mathbf{I}\mathbf{Z}' + \sigma_E^2 \mathbf{I} \\ &= \sigma_A^2 \mathbf{Z}\mathbf{Z}' + \sigma_E^2 \mathbf{I} \\ &= \begin{pmatrix} \sigma_A^2 \mathbf{1}_J \mathbf{1}_J' + \sigma_E^2 \mathbf{I}_{J\times J} & & \\ & \ddots & \\ & & \sigma_A^2 \mathbf{1}_J \mathbf{1}_J' + \sigma_E^2 \mathbf{I}_{J\times J} \end{pmatrix}, \end{aligned}$$

a block diagonal matrix, with the same covariance matrix for each i:

$$\begin{aligned}
\mathrm{cov}(\mathbf{Y}_i) &= \sigma_A^2 \mathbf{1}_J \mathbf{1}_J' + \sigma_E^2 \mathbf{I}_{J\times J} \\
&= \begin{pmatrix} \sigma_A^2 + \sigma_E^2 & \sigma_A^2 & \cdots & \sigma_A^2 \\ \sigma_A^2 & \sigma_A^2 + \sigma_E^2 & & \sigma_A^2 \\ \vdots & & \ddots & \\ \sigma_A^2 & \cdots & & \sigma_A^2 + \sigma_E^2 \end{pmatrix}.
\end{aligned}$$

Therefore, the correlation matrix of $\mathbf{Y}_i$ is

$$\mathrm{corr}(\mathbf{Y}_i) = \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & & \vdots \\ \vdots & & \ddots & \\ \rho & \cdots & & 1 \end{pmatrix}.$$

where

$$\rho = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_E^2}$$

is the *intraclass correlation coefficient.*

Distribution Theory

Recall the ANOVA decomposition:

$$SS_{\text{TOT}} = SS_A + SS_E$$

$$\sum_i \sum_j (Y_{ij} - \bar{Y}_{\cdot\cdot})^2 = J \sum_i (\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2 + \sum_i \sum_j (Y_{ij} - \bar{Y}_{i\cdot})^2.$$

We have

$$\begin{aligned} SS_A &= J \sum_i (\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2 \\ &= J \sum_i [a_i + \bar{\varepsilon}_{i\cdot} - (\bar{a}_\cdot + \bar{\varepsilon}_{\cdot\cdot})]^2 \\ &= J \sum_i (u_i - \bar{u})^2, \end{aligned}$$

with $u_i = a_i + \bar{\varepsilon}_{i\cdot} \sim N(0, \sigma_A^2 + \sigma_E^2/J)$, (independently), so

$$SS_A \sim (J\sigma_A^2 + \sigma_E^2)\, \chi^2_{I-1}.$$

Also,

$$SS_E = \sum_i \sum_j (\varepsilon_{ij} - \bar{\varepsilon}_{i\cdot})^2 \sim \sigma_E^2\, \chi^2_{I(J-1)},$$

because this is just the RSS from a one-way fixed-effects model for the ε_{ij}. It can further be proved that SS_A and SS_E are independent.

Therefore, we have

$$\frac{\sigma_E^2}{J\sigma_A^2 + \sigma_E^2} \frac{MS_A}{MS_E} \sim F_{I-1,I(J-1)},$$

where $MS_A = SS_A/(I-1)$ and $MS_E = SS_E/[I(J-1)]$. This gives a way of testing $H : \sigma_A^2 = 0$. If $\sigma_A^2 = 0$ then

$$\frac{MS_A}{MS_E} \sim F_{I-1,I(J-1)}.$$

Estimation of Variance Components

From the above distributional results, we have

$$E[MS_A] = J\sigma_A^2 + \sigma_E^2, \quad E[MS_E] = \sigma_E^2.$$

So we can obtain unbiased "method of moments" estimates of the variance components by

$$\hat{\sigma}_E^2 = MS_E, \quad \hat{\sigma}_A^2 = (MS_A - MS_E)/J.$$

Note there is no guarantee that $\hat{\sigma}_A^2$ will be positive. If $\hat{\sigma}_A^2 < 0$, it is common practice to set it to 0.

Alternatives to method-of-moments estimates are maximum likelihood (ML) estimates and REML estimates (BIOST/STAT 570).

Confidence Interval for σ_A^2/σ_E^2:

The probability statement

$$P\left(F_{I-1,I(J-1)}^{1-\alpha/2} < \frac{\sigma_E^2}{J\sigma_A^2 + \sigma_E^2}\,\frac{MS_A}{MS_E} < F_{I-1,I(J-1)}^{\alpha/2}\right) = 1 - \alpha$$

gives the following $100(1-\alpha)\%$ CI for σ_A^2/σ_E^2:

$$J^{-1}\left(\frac{MS_A}{MS_E}\,\frac{1}{F_{I-1,I(J-1)}^{\alpha/2}} - 1\right) \quad \text{to} \quad J^{-1}\left(\frac{MS_A}{MS_E}\,\frac{1}{F_{I-1,I(J-1)}^{1-\alpha/2}} - 1\right).$$

A CI for ρ can be obtained using $\rho^{-1} = 1 + (\sigma_A^2/\sigma_E^2)^{-1}$.

All of these distributional results derive from the assumption that all random effects are normally distributed.

Two-Way Random Effects Model

Model:

$$Y_{ijk} = \mu + a_i + b_j + c_{ij} + \varepsilon_{ijk},$$
$$i = 1, \ldots, I;\ j = 1, \ldots, J;\ k = 1, \ldots, K,$$

$$\left.\begin{array}{rcl} a_i & \sim & N(0, \sigma_A^2) \\ b_j & \sim & N(0, \sigma_B^2) \\ c_{ij} & \sim & N(0, \sigma_{AB}^2) \\ \varepsilon_{ijk} & \sim & N(0, \sigma_E^2) \end{array}\right\} \text{ mutually independent.}$$

Distribution Theory:

As before,

$$SS_E = \sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{ij.})^2 = \sum_i \sum_j \sum_k (\varepsilon_{ijk} - \bar{\varepsilon}_{ij.})^2 \sim \sigma_E^2\, \chi^2_{IJ(K-1)}.$$

Also

$$\begin{aligned} SS_{AB} &= \sum_i \sum_j \sum_k (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2 \\ &= K \sum_i \sum_j (u_{ij} - \bar{u}_{i.} - \bar{u}_{.j} + \bar{u}_{..})^2 \\ &\sim (K\sigma_{AB}^2 + \sigma_E^2)\, \chi^2_{(I-1)(J-1)}, \end{aligned}$$

where $u_{ij} = c_{ij} + \bar{\varepsilon}_{ij.} \sim N(0, \sigma_{AB}^2 + \sigma_E^2/K)$, (independently).

Similarly,

$$\begin{aligned} SS_A &= \sum_i \sum_j \sum_k (\bar{Y}_{i..} - \bar{Y}_{...})^2 \\ &= JK \sum_i (v_i - \bar{v})^2, \quad (v_i = a_i + \bar{c}_{i.} + \bar{\varepsilon}_{i..}) \\ &\sim (JK\sigma_A^2 + K\sigma_{AB}^2 + \sigma_E^2)\, \chi^2_{I-1}, \end{aligned}$$

and

$$SS_B \sim (IK\sigma_B^2 + K\sigma_{AB}^2 + \sigma_E^2)\,\chi^2_{J-1}.$$

It can again be shown that all SS's are independent:

$$\begin{aligned}
\frac{\sigma_E^2}{K\sigma_{AB}^2 + \sigma_E^2}\frac{MS_{AB}}{MS_E} &\sim F_{(I-1)(J-1),IJ(K-1)}, \\
\frac{K\sigma_{AB}^2 + \sigma_E^2}{JK\sigma_A^2 + K\sigma_{AB}^2 + \sigma_E^2}\frac{MS_A}{MS_{AB}} &\sim F_{I-1,(I-1)(J-1)}, \\
\frac{K\sigma_{AB}^2 + \sigma_E^2}{IK\sigma_B^2 + K\sigma_{AB}^2 + \sigma_E^2}\frac{MS_B}{MS_{AB}} &\sim F_{J-1,(I-1)(J-1)}.
\end{aligned}$$

Note: The interaction mean-square, not the error mean square, is used for testing the variance components for A and B. This is because

$$\frac{MS_A}{MS_{AB}} \sim F_{I-1,(I-1)(J-1)}, \quad \text{if } H : \sigma_A^2 = 0,$$

but the same cannot be said for $\frac{MS_A}{MS_E}$.

Fixed Effects ANOVA Table:

Source	df	SS	MS	$E[MS]$
A	$I-1$	$\sum_{ijk}(\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot})^2$	$MS_A = \frac{SS_A}{I-1}$	$\sigma_E^2 + \frac{JK\sum_i \alpha_i^2}{I-1}$
B	$J-1$	$\sum_{ijk}(\bar{Y}_{\cdot j\cdot} - \bar{Y}_{\cdot\cdot\cdot})^2$	$MS_B = \frac{SS_B}{J-1}$	$\sigma_E^2 + \frac{IK\sum_j \beta_j^2}{J-1}$
AB	$(I-1)(J-1)$	$\sum_{ijk}(\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdot\cdot\cdot})^2$	$MS_{AB} = \frac{SS_{AB}}{(I-1)(J-1)}$	$\sigma_E^2 + \frac{K\sum_{ij}(\alpha\beta)_{ij}^2}{(I-1)(J-1)}$
Error	$IJ(K-1)$	$\sum_{ijk}(Y_{ijk} - \bar{Y}_{ij\cdot})^2$	$MS_E = \frac{SS_E}{IJ(K-1)}$	σ_E^2
Total	$IJK-1$	$\sum_{ijk}(Y_{ijk} - \bar{Y}_{\cdot\cdot\cdot})^2$		

Random Effects ANOVA Table:

Source	df	SS	MS	$E[MS]$
A	$I-1$	$\sum_{ijk}(\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot})^2$	$MS_A = \frac{SS_A}{I-1}$	$\sigma_E^2 + K\sigma_{AB}^2 + JK\sigma_A^2$
B	$J-1$	$\sum_{ijk}(\bar{Y}_{\cdot j\cdot} - \bar{Y}_{\cdot\cdot\cdot})^2$	$MS_B = \frac{SS_B}{J-1}$	$\sigma_E^2 + K\sigma_{AB}^2 + IK\sigma_B^2$
AB	$(I-1)(J-1)$	$\sum_{ijk}(\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdot\cdot\cdot})^2$	$MS_{AB} = \frac{SS_{AB}}{(I-1)(J-1)}$	$\sigma_E^2 + K\sigma_{AB}^2$
Error	$IJ(K-1)$	$\sum_{ijk}(Y_{ijk} - \bar{Y}_{ij\cdot})^2$	$MS_E = \frac{SS_E}{IJ(K-1)}$	σ_E^2
Total	$IJK-1$	$\sum_{ijk}(Y_{ijk} - \bar{Y}_{\cdot\cdot\cdot})^2$		