Random and mixed effects models

Fixed effect: Three fields were available for an agricultural yield experiment. The experiment is conducted on those fields. The mean yield of this particular strain of wheat is the main interest of the investigators, but if the fields have important effects on the yields, then the investigators would like to know that as well. A "repetition" of this experiment would use the same three fields. The AOV model is

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

were Y_{ij} is the yield of the jth plot ($j = 1, 2, ..., n_i$) on the ith field ($i = 1, 2, 3$), μ is the grand mean, and α_i is the effect of the ith field, with $\Sigma\alpha_i = 0$, and ϵ_{ij} are independent normal$(0, \sigma_\epsilon^2)$ random variables.

In this model, field is a "fixed effect". The statistical model describes the whole ensemble of possible repetitions of the experiment *on these three fields*. Along with μ, the fixed parameters α_1, α_2, and α_3 are the quantities of interest.

Random effect: Three fields are selected at random from all the fields in a region. The experiment is conducted as described above, using the three selected fields. A "repetition" of the experiment would involve selecting three new fields at random, and the chance that any of the first three fields are selected again is very small. The AOV model is

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

were Y_{ij} is the yield of the jth plot ($j = 1, 2, ..., n_i$) on the ith field ($i = 1, 2, 3$), μ is the grand mean, ϵ_{ij} are independent normal$(0, \sigma_\epsilon^2)$ random variables, and now the α_i's are assumed to be independent normal$(0, \sigma_\alpha^2)$ random variables (with the α_i's independent of the ϵ_{ij}'s).

In this model, field is a "random effect". The statistical model describes the whole ensemble of possible repetitions of the experiment i*n the region from which the fields were selected*. Along with μ, the interest focuses on the parameter σ_α^2 and its relative magnitude in relation to σ_ϵ^2.

One important consequence of random effects is that the responses (Y_{ij}'s) are no longer independent. The random α_i's induce correlations among the responses. The responses jointly have a multivariate normal distribution.

Fixed effects model: levels of factors used in the analysis would remain the same if the experiment were repeated. Inference in the analysis is to the population of observations that could be generated from those fixed factor levels.

Random effects model: levels of factors used in the experiment are randomly selected from a population of possible levels. Inference in the analysis is to the population from which the levels were selected.

Mixed effects model: some factors are fixed, some are random.

Estimation: along with the magnitudes of any fixed effects, the variances of the random effects become parameters of interest

μ: grand mean
(the mean productivity over all the fields in the region from which the fields were selected)
σ_ϵ^2 variance component from plots
σ_α^2 variance component from fields
$\frac{\sigma_\alpha^2}{\sigma_\alpha^2+\sigma_\epsilon^2}$ proportion of variance in response due to fields

In the 1980s, PROC GLM in SAS was altered to accomodate random effects (via the RANDOM statement). It used least squares and method of moment matching to estimate the parameters in the models, and F tests based on expected mean squares to do hypotheses testing, in order to utilize the existing algorithms programmed in GLM.

Contemporary approach is to use full maximum likelihood (ML) or restricted maximum likelihood (REML) estimation. These require computerized numerical maximization (of the multivariate normal likelihood). The hypothesis tests are based on asymptotic maximum likelihood theory. This newer approach was implemented by SAS in PROC MIXED, which has essentially supplanted the use of PROC GLM for random and mixed effects models.

Random blocks and random treatment effects

Text example 17.2: interested in effects of different chemical analysts in measuring DNA content of tooth plaque

subjects: 3 females chosen at random
treatments: 3 analysts chosen at random

	subject (block)			
analyst	1	2	3	$\cdots$
1	5.2	6.0	7.2	
2	4.8	6.1	6.9	$\cdots$
3	5.4	6.2	7.4	
$\vdots$		$\vdots$		

Model: $Y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$

μ: grand mean

α_i: random effect due to the ith analyst
$\alpha_i \sim \text{normal}(0, \sigma_\alpha^2)$, $i = 1, 2, ..., a$

β_j: random effect due to the jth subject
$\beta_j \sim \text{normal}\left(0, \sigma_\beta^2\right)$, $j = 1, 2, ..., b$

$\epsilon_{ij} \sim \text{normal}(0, \sigma_\epsilon^2)$

α_i's, β_j's, ϵ_{ij}'s are all independent

Tests of interest (for the effect of factor A)

	A fixed	A random
H_0:	$\alpha_i = 0$, all i	$\sigma_\alpha^2 = 0$
H_a:	$\alpha_i \neq 0$, all i	$\sigma_\alpha^2 > 0$

	B fixed	B random
H_0:	$\beta_j = 0$, all j	$\sigma_\beta^2 = 0$
H_a:	$\beta_j \neq 0$, all j	$\sigma_\beta^2 > 0$

Two factor factorial, both factors random

Model: $Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$

μ: grand mean

$\alpha_i \sim \text{normal}(0, \sigma_\alpha^2)$, $i = 1, 2, ..., a$

$\beta_j \sim \text{normal}\left(0, \sigma_\beta^2\right)$, $j = 1, 2, ..., b$

$(\alpha\beta)_{ij} \sim \text{normal}\left(0, \sigma_{\alpha\beta}^2\right)$

$\epsilon_{ijk} \sim \text{normal}(0, \sigma_\epsilon^2)$, $k = 1, 2, ..., n$

α_i's, β_j's, $(\alpha\beta)_{ij}$'s, ϵ_{ijk}'s are all independent

Tests of interest

Interaction:

	fixed	random
H_0:	$(\alpha\beta)_{ij} = 0$, all i, j	$\sigma_{\alpha\beta}^2 = 0$
H_a:	$(\alpha\beta)_{ij} \neq 0$, some i, j	$\sigma_{\alpha\beta}^2 > 0$

Effect of factor A (proceed similarly for factor B)

	fixed	random
H_0:	$\alpha_i = 0$, all i	$\sigma_\alpha^2 = 0$
H_a:	$\alpha_i \neq 0$, all i	$\sigma_\alpha^2 > 0$

In the random case, there is no real problem with interpretation of significant interaction and significant main effect. One is interested in identifying the components of variability (**variance components**) σ_α^2, σ_β^2, $\sigma_{\alpha\beta}^2$, and σ_ϵ^2.

Mixed effects models: one random, one fixed

$a \times b$ factorial design with:

A, fixed effect
B, random effect

Model: $Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$

μ: grand mean
α_i is fixed effect of level i of factor A, $i = 1, 2, ..., a$
$\beta_j \sim \text{normal}\left(0, \sigma_\beta^2\right)$, $j = 1, 2, ..., b$
$(\alpha\beta)_{ij} \sim \text{normal}\left(0, \sigma_{\alpha\beta}^2\right)$,
$\epsilon_{ijk} \sim \text{normal}(0, \sigma_\epsilon^2)$, $k = 1, 2, ..., n$

α_i's, β_j's, $(\alpha\beta)_{ij}$'s, ϵ_{ijk}'s are all independent

Factor A hypotheses

H_0: $\alpha_i = 0$, all i
H_a: $\alpha_i \neq 0$, some i

Factor B hypotheses

H_0: $\sigma_\beta^2 = 0$
H_a: $\sigma_\beta^2 > 0$

Interaction hypotheses

H_0: $\sigma_{\alpha\beta}^2 = 0$
H_a: $\sigma_{\alpha\beta}^2 > 0$