Random and mixed effects models

**ex 1. 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

\[ Y_{ij} = \mu + \alpha_i + \epsilon_{ij} \]

where \( Y_{ij} \) is the yield of the \( j \)th plot \( (j = 1, 2, \ldots, n_i) \) on the \( i \)th field \( (i = 1, 2, 3) \), \( \mu \) is the grand mean, and \( \alpha_i \) is the effect of the \( i \)th field, with \( \Sigma \alpha_i = 0 \), and \( \epsilon_{ij} \) are independent normal \( (0, \sigma^2) \) 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 \( \mu \), the fixed parameters \( \alpha_1, \alpha_2, \) and \( \alpha_3 \) are the quantities of interest.
**ex. 2 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 $j$th plot ($j = 1, 2, \ldots, n_i$) on the $i$th field ($i = 1, 2, 3$), $\mu$ is the grand mean, $\epsilon_{ij}$ are independent normal$(0, \sigma^2_\varepsilon)$ random variables, and now the $\alpha_i$'s are assumed to be independent normal$(0, \sigma^2_\alpha)$ random variables (with the $\alpha_i$'s independent of the $\epsilon_{ij}$'s).

In this model, field is a “random effect”. The statistical model describes the whole ensemble of possible repetitions of the experiment in the region from which the fields were selected. Along with $\mu$, the interest focuses on the parameter $\sigma^2_\alpha$ and its relative magnitude in relation to $\sigma^2_\varepsilon$. 
**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.

**Expected mean squares:** fixed effects (1-way AOV)

$$SS(\text{trt}) = \sum_{i=1}^{t} n_i (\bar{Y}_{i.} - \bar{Y}_{..})^2$$

$$SS(\text{error}) = \sum_{i=1}^{t} \sum_{j=1}^{n_i} (Y_{ij} - (\hat{\mu} + \hat{\alpha}_i))^2$$

$$MS(\text{trt}) = \frac{SS(\text{trt})}{t-1}$$

$$MS(\text{error}) = \frac{SS(\text{error})}{n_T-t}$$

$$H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_t = 0$$

$$H_a: \neq \text{ somewhere}$$
Principle of likelihood ratio leads to using

\[ F = \frac{\text{MS(trt)}}{\text{MS(error)}} \]

as a test statistic. Under \( H_0 \), \( F \) has an \( F(t - 1, n_T - t) \) distribution, and under \( H_a \), \( F \) has a noncentral \( F(t - 1, n_T - t, \lambda) \) distribution. Can show that

\[
\text{E(MS(trt))} = \sigma^2 + \frac{\sum n_i \alpha_i^2}{t - 1}
\]

\[
\text{E(MS(error))} = \sigma^2
\]

Under \( H_0 \), numerator and denominator has the same expected value, while under \( H_a \), numerator has an expected value enhanced by an amount related to the effect sizes.

**Expected mean squares: random effects**

\[
V(Y_{ij}) = \sigma^2 + \sigma^2 \text{ total variance}
\]

\[
\frac{\sigma^2}{\sigma^2 + \sigma^2}: \text{ proportion of total variance due to differences in factor levels}
\]

\( H_0: \sigma^2 = 0 \) (all levels same)
Hₐ: $\sigma_\alpha^2 > 0$ (levels are different)

SS(trt), MS(trt): same as fixed (in this 1-way AOV case)

SS(error), MS(error): same as fixed (ditto)

Can show that

$$E(\text{MS(trt)}) = \sigma_\varepsilon^2 + n'\sigma_\alpha^2$$

where

$$n' = \frac{1}{t - 1} \left( \sum n_i - \frac{\sum n_i^2}{\sum n_i} \right)$$

(if $n_1 = n_2 = \cdots = n_t = r$, then $n' = r$)

$$E(\text{MS(error)}) = \sigma_\varepsilon^2$$

Test statistic is again

$$F = \frac{\text{MS(trt)}}{\text{MS(error)}}$$

Under $H_0$, $F$ has an $F(t - 1, n_T - t)$ distribution. Reject $H_0$ if $F \geq f_\alpha$. 
Remarks:

- In the 1-way AOV situation, the analyses for fixed and random effects are the same, but the interpretation are different

- In more complex designs & AOV models, the analyses are different (one uses EMS to find out which test statistics to use)

Estimation: frequently more interested in the parameters themselves in random effects models

\( \mu: \) grand mean

ex. What is the mean productivity over all the fields in the region (from which the fields were selected)? Estimate is \( \hat{\mu} \).

\[ \hat{\mu} = \bar{Y}_.. \quad \text{(unbiased)} \]

(Note: random & mixed effects models become really complex when the design is unbalanced, even in the 1-way case. Hypothesis tests & CIs are frequently just approximate, based on asymptotic results. So, beyond here, formulas assume \( n_1 = n_2 = \cdots = n_t = r \) )

\[ \sigma^2_{Y..} = \frac{\sigma^2_e + r\sigma^2_\alpha}{n_T} \]
\[
\hat{\sigma}_{Y_{..}}^2 = \frac{\text{MS(trt)}}{n_T} \quad \text{(unbiased because)}
\]
\[
E(\text{MS(trt)}) = \sigma_{\epsilon}^2 + r\sigma_{\alpha}^2 
\]
Can show that \( \frac{Y_{..}-\mu}{\sqrt{\hat{\sigma}_{Y_{..}}^2}} \) has a Student's t\((t-1)\) distribution.

100(1 - \(\alpha\))% CI for \(\mu\):

\[
\bar{y}_{..} - t_{\alpha/2} \sqrt{\hat{\sigma}_{Y_{..}}^2}, \quad \bar{y}_{..} + t_{\alpha/2} \sqrt{\hat{\sigma}_{Y_{..}}^2}
\]

\(\sigma_{\epsilon}^2\): \(\hat{\sigma}_{\epsilon}^2 = \text{MS(error)}\)

\[
\frac{(n_T-t)\text{MS(error)}}{\sigma_{\epsilon}^2} \quad \text{has a chi-square}(n_T - t) \text{ distribution}
\]
(yields same confidence interval as for fixed effects)

\(\sigma_{\alpha}^2\): \(\hat{\sigma}_{\alpha}^2 = \frac{\text{MS(trt)} - \text{MS(error)}}{n'}\) \quad \text{(unbiased)}

Note: this can sometimes be less than zero!
Alternative estimates are available (REML, etc.); CIs only approximate.

\[
\frac{\sigma_{\alpha}^2}{\sigma_{\alpha}^2 + \sigma_{\epsilon}^2} : \quad \text{let}
\]
\[
L = \frac{1}{r} \left[ \frac{\text{MS(trt)}}{\text{MS(error)} f_{\alpha/2}} - 1 \right] \quad \text{(df = } t - 1, n_T - t) 
\]
\[
U = \frac{1}{r} \left[ \frac{\text{MS(trt)}}{\text{MS(error)} f_{1-\alpha/2}} - 1 \right]
\]
\( L, U \) is a 100(1 – \( \alpha \))\% CI for \( \frac{\sigma_a^2}{\sigma_e^2} \), and

\[
\frac{L}{1+L}, \frac{U}{1+U} \quad \text{is a 100(1 – \( \alpha \))\% CI for} \quad \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2}.
\]