The One-way Random Effects Model

- **Model** (Balanced design)

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad i = 1, \dots, g; \quad j = 1, \dots, N,$$
$$\alpha_i \text{ iid } \sim N(0, \sigma_A^2), \quad \epsilon_{ij} \text{ iid } \sim N(0, \sigma^2), \quad \alpha_i \text{ and } \epsilon_{ij} \text{ independent.}$$

The model implies

$$\mathbb{E}(y_{ij}) = \mu, \quad \text{Var}(y_{ij}) = \sigma_A^2 + \sigma^2, \quad \text{Cov}(y_{ij}, y_{i'j'}) = \begin{cases} \sigma_A^2, & \text{if } i = i', j \neq j' \\ 0, & \text{if } i \neq i' \end{cases}$$

The *intraclass correlation* (ICC), the correlation among $\{y_{ij}\}_{j=1}^N$, is

$$\text{Corr}(y_{ij}, y_{ij'}) = \frac{\sigma_A^2}{\sigma^2 + \sigma_A^2}.$$

An alternative description of the model

$$y = \mu + e, \quad e \sim \mathsf{N}_n(\mathbf{0}, V),$$

where

$$V = N\sigma_A^2(\mathbf{I}_g \otimes \mathbf{J}_N) + \sigma^2 \mathbf{I}_n = N\sigma_A^2 M_G + \sigma^2 \mathbf{I}_n, \quad \mathbf{J}_N = \frac{1}{N}\mathbf{1}_N \mathbf{1}_N^t.$$

- **ANOVA table**

Source	SS	df	MS	$\mathbb{E}$[MS]	F-stat
$\mathcal{M}_{G\cdot 0}$	$\text{SS}_{G\cdot 0} = \|\hat{y}_G - \hat{y}_0\|^2$	$g-1$	$\text{MS}_{G\cdot 0}$	$N\sigma_A^2 + \sigma^2$	$\text{MS}_{G\cdot 0}/\text{MSE}$
$\mathcal{M}_G^\perp$	$\text{SSE} = \|y - \hat{y}_G\|^2$	$n-g$	MSE	σ^2	—
$\mathcal{M}_0^\perp$	$\text{SST} = \|y - \hat{y}_0\|^2$	$n-1$	—	—	—

  - Note that $y \sim \mathsf{N}_n(\mu\mathbf{1}, V)$, where

$$\begin{aligned} V &= N\sigma_A^2 M_G + \sigma^2 \mathbf{I}_n \\ &= N\sigma_A^2 M_G + \sigma^2 M_G + \sigma^2(\mathbf{I}_n - M_G) \\ &= (N\sigma_A^2 + \sigma^2)M_0 + (N\sigma_A^2 + \sigma^2)M_{G\cdot 0} + \sigma^2(\mathbf{I}_n - M_G) \end{aligned}$$

  where at the last line, we decompose the variance of y into three components from three orthogonal subspaces: $\mathcal{M}_0$, $\mathcal{M}_{G\cdot 0}$ and $\mathcal{M}_G^\perp$.

- The projection of y toward $\mathcal{M}_{G\cdot 0}$ is distributed as

$$M_{G\cdot 0}y \sim N(0, M_{G\cdot 0}VM_{G\cdot 0}),$$

where

$$M_{G\cdot 0}VM_{G\cdot 0} = (N\sigma_A^2 + \sigma^2)M_{G\cdot 0}.$$

Therefore $\text{SS}_{G\cdot 0} \sim (N\sigma_A^2 + \sigma^2)\chi^2_{g-1}$.

- The projection of y toward $\mathcal{M}_G^{\perp}$ is distributed as

$$(\mathbf{I} - M_G)y \sim N(0, (\mathbf{I} - M_G)V(\mathbf{I} - M_G)),$$

where

$$(\mathbf{I} - M_G)V(\mathbf{I} - M_G) = \sigma^2(\mathbf{I} - M_G).$$

Therefore $\text{SSE} \sim \sigma^2\chi^2_{n-g}$.

- So the F-stat follows a scaled F-distribution,

$$\text{F-stat} = \left(\frac{N\sigma_A^2 + \sigma^2}{\sigma^2}\right)\frac{\chi^2_{g-1}/(g-1)}{\chi^2_{n-g}/(n-g)} \sim \frac{N\sigma_A^2 + \sigma^2}{\sigma^2}F_{g-1,n-g}.$$

And Under $H_0 : \sigma_A^2 = 0$, F-test $\sim F_{g-1,n-g}$.

- ANOVA estimators (Method of Moments) of σ_A^2 and σ^2,

$$\hat{\sigma}_e^2 = \text{MSE}, \quad \hat{\sigma}_A^2 = \frac{\text{MS}_{G\cdot 0} - \text{MSE}}{N}. \tag{1}$$

Note: It is possible that $\text{MS}_{G\cdot 0} < \text{MSE}$, that is, the ANOVA estimate $\hat{\sigma}_A^2$ is negative. This is a strong evidence against the alternative hypothesis $\sigma_A^2 > 0$; this would rarely happen in practice if we compute $\hat{\sigma}_A^2$ only if the ANOVA test is significant.

- **GLSE of μ**

For a linear regression model $y = X\beta + \epsilon$ with $e \sim N(0, V)$, the GLSE of β is given by

$$\begin{aligned} \text{GLSE}(\beta) &= \arg\min_{\beta}(y - X\beta)^t V^{-1}(y - X\beta) \\ &= (X^t V^{-1} X)^{-1} X^t V^{-1} y. \end{aligned} \tag{2}$$

For one-way random effects model, we have $X = \mathbf{1}_n$ and also suppose σ^2 and σ_A^2 are known. Next we derive the GLSE of μ for balanced data

Recall that

$$V = N\sigma_A^2(\mathbf{I}_g \otimes \mathbf{J}_N) + \sigma^2\mathbf{I}_n = \mathbf{I}_g \otimes (N\sigma_A^2\mathbf{J}_N + \sigma^2\mathbf{I}_N).$$

Using the fact that

$$(a\mathbf{J}_m + b\mathbf{I}_m)^{-1} = \frac{1}{b}(\mathbf{I}_m - \frac{a}{a+b}\mathbf{J}_m), \tag{3}$$

we have

$$V^{-1} = \frac{1}{\sigma^2}\mathbf{I}_g \otimes \Big(\mathbf{I}_N - \frac{N\sigma_A^2}{N\sigma_A^2+\sigma^2}\mathbf{J}_N\Big). \tag{4}$$

Plugging $X = \mathbf{1}_n$ and (4) into (2), we obtain

$$\text{GLSE}(\mu) = \bar{y}_{\cdot\cdot}.$$

Further we have

$$\text{Var}(\bar{y}_{\cdot\cdot}) = \frac{N\sigma_A^2+\sigma^2}{n} = \frac{\sigma_A^2+\sigma^2/N}{g},$$

$$95\% \text{ CI } : \bar{y}_{\cdot\cdot} \pm t_{g-1,2.5\%}\sqrt{\frac{\text{MS}_{G\cdot 0}}{n}}.$$

Why is the degree of freedom here $g-1$ instead of $n-g$ or $n-1$? View the overall average $\bar{y}_{\cdot\cdot}$ as the average of g group means, namely,

$$\bar{y}_{\cdot\cdot} = \frac{1}{g}\sum_{i=1}^{g}\bar{y}_{i\cdot},$$

where $\bar{y}_{i\cdot}$'s are independent and each has variance $\sigma_A^2 + \sigma^2/N$.

- **MLE of** $(\mu, \sigma^2, \sigma_A^2)$

  Previously we adopt a two-stage approach for the one-way random effects model: first estimate the variance components (σ^2, σ_A^2) using (1), and then estimate μ based on the GLS approach. Here we derive the (joint) MLE for all three parameters $(\mu, \sigma^2, \sigma_A^2)$.

  Recall that $y \sim N(\mu, V)$, so

$$\log L = -\frac{n}{2}\log 2\pi - \frac{1}{2}\log|V| - \frac{1}{2}(y-\mu)^t V^{-1}(y-\mu).$$

  For balanced design, using the following decomposition of $\sigma^2 V^{-1}$

$$\begin{aligned}
\sigma^2 V^{-1} &= \mathbf{I}_n - \frac{N\sigma_A^2}{N\sigma_A^2+\sigma^2}\mathbf{I}_g \otimes \mathbf{J}_N \\
&= \mathbf{I}_n - \frac{N\sigma_A^2}{N\sigma_A^2+\sigma^2}M_G \\
&= (\mathbf{I}_n - M_G) + M_{G\cdot 0} + M_0 - \frac{N\sigma_A^2}{N\sigma_A^2+\sigma^2}(M_{G\cdot 0} + M_0) \\
&= (\mathbf{I}_n - M_G) + \frac{\sigma^2}{N\sigma_A^2+\sigma^2}M_{G\cdot 0} + \frac{\sigma^2}{N\sigma_A^2+\sigma^2}M_0,
\end{aligned}$$

we can write the log-likelihood as

$$\log L = -\frac{n}{2}\log 2\pi - \frac{g}{2}\log(N\sigma_A^2+\sigma^2) - \frac{n-g}{2}\log\sigma^2$$
$$-\frac{1}{2\sigma^2}\text{SSE} - \frac{1}{2(N\sigma_A^2+\sigma^2)}\text{SS}_{G\cdot 0} - \frac{n}{2(N\sigma_A^2+\sigma^2)}(\bar{y}_{\cdot\cdot}-\mu)^2. \quad (5)$$

Setting to zero the partial derivatives of $\log L$ with respect to μ, σ^2, and σ_A^2, we obtain the following solutions,

$$\tilde{\mu} = \bar{y}_{\cdot\cdot}, \quad \tilde{\sigma}^2 = \text{MSE}, \quad \tilde{\sigma}_A^2 = \frac{\text{SS}_{G\cdot 0}/g - \text{MSE}}{N} = \frac{\frac{g-1}{g}\text{MS}_{G\cdot 0} - \text{MSE}}{N},$$

and the MLEs are

$$\text{when} \quad \tilde{\sigma}_A^2 \geq 0, \quad \text{MLE} = (\tilde{\mu}, \tilde{\sigma}^2, \tilde{\sigma}_A^2)$$
$$\text{when} \quad \tilde{\sigma}_A^2 < 0, \quad \text{MLE} = (\tilde{\mu}, \frac{\text{SST}}{n}, 0)$$

What's the difference between the MLE of (σ^2, σ_A^2) and the corresponding ANOVA estimates?

- **REML estimation of** (σ^2, σ_A^2)

  An adaption of ML is to maximize the marginal likelihood

  $$L(\sigma_A^2, \sigma^2; y) = \int L(\mu, \sigma_A^2, \sigma^2; y)d\mu,$$

  which gives rise to the restricted (or residual) MLE.

  Below is the RMLE for one-way random effects model:

  When ANOVA estimation $\hat{\sigma}_A^2 \geq 0$, REML estimators are ANOVA estimators;

  When ANOVA estimation $\hat{\sigma}_A^2 < 0$, REML$(\sigma^2) = \frac{\text{SST}}{n-1}$, REML (σ_A^2)=0.

- **Prediction of the random effects**

  We consider the problem of predicting random variables $\mu + \alpha$ or α given data $Y = (\bar{y}_{1\cdot}, \ldots, \bar{y}_{g\cdot})$. Here $\alpha = (\alpha_1, \ldots, \alpha_g)^t$, and Y denotes the group averages.

  - Consider a general setup where (U, Y) are two random vectors with a joint distribution $p(u, y)$. The goal is to predict U conditioning on observing $Y = y$. We measure the error between a predictor $f(y)$ that is a function of the observed y and the true value u by expected squared error (aka MSE)

    $$\mathbb{E}\|f(Y) - U\|^2 = \int \|f(y) - u\|^2 p(u, y)dudy. \quad (6)$$

    The best predictor (BP), that is, the one minimizing (6), is given by

    $$\text{BP} = \mathbb{E}(U \mid Y = y).$$

    However it is difficult to calculate BP in practice since the joint distribution function $p(u, y)$ is often unknown.

– Instead of looking for the best one among all predictors, let us focus our attention on linear predictors, that is, $f(y) = a + By$. It turns out that the best linear predictor (BLP), that is, the one minimizing (6) among all linear predictors, depends only on the first and second moments of (U, Y). Suppose

$$\mathbb{E}\begin{pmatrix} U \\ Y \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \text{Cov}\begin{pmatrix} U \\ Y \end{pmatrix} = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{12}^t & \Omega_{22} \end{pmatrix}. \tag{7}$$

Then

$$\text{BLP} = \mu_1 + \Omega_{12}\Omega_{22}^{-1}(y - \mu_2).$$

For example, let U denote the ith random effect α_i and Y denote the average of the ith group $\bar{y}_{i\cdot}$, we have

$$\mathbb{E}\begin{pmatrix} \alpha_i \\ \bar{y}_{i\cdot} \end{pmatrix} = \begin{pmatrix} 0 \\ \mu \end{pmatrix}, \quad \text{Cov}\begin{pmatrix} U \\ Y \end{pmatrix} = \begin{pmatrix} \sigma_A^2 & \sigma_A^2 \\ \sigma_A^2 & \sigma_A^2 + \sigma^2/N \end{pmatrix},$$

and

$$\text{BLP}(\alpha_i) = \frac{\sigma_A^2/N}{\sigma_A^2 + \sigma^2/N}(\bar{y}_{i\cdot} - \mu).$$

– In practice, the problem of estimating the unknown parameters (such as μ_1, μ_2, Ω's) in the expression of BLP still remains. To weaken the assumption that μ_1 and μ_2 are known, we impose some structure on the two mean vectors by assuming

$$\mu_1 = X_1\beta, \quad \mu_2 = X_2\beta.$$

That is, the two means are related via a common parameter β which could be estimated from Y. This gives rise to the best linear unbiased predictor (BLUP),

$$\text{BLUP} = X_1\hat{\beta} + \Omega_{12}\Omega_{22}^{-1}(y - X_2\hat{\beta}),$$

where $\hat{\beta}$ is a BLUE of β.

– For the one-way random effects model,

$$\text{BLUP}(\alpha_i) = \frac{\sigma_A^2}{\sigma_A^2 + \sigma^2/N}(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot}) \tag{8}$$

$$\text{BLUP}(\mu + \alpha_i) = \bar{y}_{\cdot\cdot} + \frac{\sigma_A^2}{\sigma_A^2 + \sigma^2/N}(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot}), \tag{9}$$

where in practice we need to plug in the estimates for σ^2 and σ_A^2.
Note that both estimates have the so-called "shrinkage" effect: the BLUP of $\hat{\alpha}_i$ is not equal to $(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})$, but a shrunk version of $(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})$ toward zero. Similarly, the BLUP of $\mu + \alpha_i$ is not equal to the group mean $\bar{y}_{i\cdot}$, but a shrunk version of $\bar{y}_{i\cdot}$ toward the overall average $\bar{y}_{\cdot\cdot}$.

- **Model** (Unbalanced design)

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad i = 1, \ldots, g; \quad j = 1, \ldots, n_i,$$
$$\alpha_i \text{ iid } \sim N(0, \sigma_A^2), \quad \epsilon_{ij} \text{ iid } \sim N(0, \sigma^2), \quad \alpha_i \text{ and } \epsilon_{ij} \text{ independent.}$$

An alternative description of the model

$$y = \mu + e, \quad e \sim \mathsf{N}_n(\mathbf{0}, V),$$

where

$$V = \text{diag}\Big\{\sigma^2 \mathbf{I}_{n_i} + n_i \sigma_A^2 \mathbf{J}_{n_i}\Big\}_{i=1}^{g}.$$

- **ANOVA table**

- GLSE of μ

Using (3), we have

$$V^{-1} = \text{diag}\Big\{\frac{1}{\sigma^2}\Big(\mathbf{I}_{n_i} + \frac{n_i \sigma_A^2}{\sigma^2 + n_i \sigma_A^2}\mathbf{J}_{n_i}\Big)\Big\}_{i=1}^{g}. \tag{10}$$

Then we can derive the GLSE estimate of μ using (2),

$$\text{GLSE}(\mu) = \frac{\sum_{i=1}^{g} \frac{n_i \bar{y}_{i\cdot}}{n_i \sigma_A^2 + \sigma^2}}{\sum_{i=1}^{g} \frac{n_i}{n_i \sigma_A^2 + \sigma^2}} = \frac{\sum_{i=1}^{g} \bar{y}_{i\cdot}/\text{Var}(\bar{y}_{i\cdot})}{\sum_{i=1}^{g} 1/\text{Var}(\bar{y}_{i\cdot})},$$

which is a weighted average of the group means, with weights reciprocally related to the variance of each group mean.

- **MLE of** $(\mu, \sigma^2, \sigma_A^2)$

Using (10), we can write the log-likelihood as

$$\begin{aligned}\log L \quad = \quad & -\frac{n}{2}\log 2\pi - \frac{1}{2}\sum_{i=1}^{g}(n_i - 1)\log \sigma^2 - \frac{1}{2}\sum_{i=1}^{g}\log\left(\sigma^2 + n_i \sigma_A^2\right) \\ & -\sum_{i=1}^{g}\frac{n_i}{n_i \sigma_A^2 + \sigma^2}(\bar{y}_{i\cdot} - \mu)^2.\end{aligned}$$