STAT612 Theory of Linear Model Last update: November 5, 2019

Notes 9: Variance Components and Mixed Models I

Book reference: Monahan (2008) *Instructor: Raymond Wong*

1 Introduction

In this notes, we introduce the mixed model. We begin with one-way model, followed by two-way model.

2 One-way model

2.1 One-way (fixed) ANOVA revisited

Recall the model for one-way ANOVA:

$$y_{ij} = \mu + \alpha_i + e_{ij}, \quad i = 1, \ldots, a, \quad j = 1, \ldots, n_i,$$

where $e_{ij} \overset{i.i.d.}{\sim} N(0, \sigma^2)$. In vector-matrix notations, we write $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$ where

$N \times (a+1)$

rank(X) = a

$$\mathbf{X} = \begin{bmatrix} \mathbf{1}_{n_1} & \mathbf{1}_{n_1} & \mathbf{0} & \cdots & \cdots & \mathbf{0} \\ \mathbf{1}_{n_2} & \mathbf{0} & \mathbf{1}_{n_2} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & & & \ddots & \\ \mathbf{1}_{n_c} & \mathbf{0} & \cdots & \cdots & \mathbf{0} & \mathbf{1}_{n_a} \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} \mu \\ \alpha_1 \\ \vdots \\ \alpha_a \end{bmatrix}, \quad \mathbf{e} \sim N_N(\mathbf{0}, \sigma^2 \mathbf{I}).$$

$\mu \quad \alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_a.$

It is shown in Example 2 of Notes 4 that

$$(\mathbf{X}^\intercal \mathbf{X})^- = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 1/n_1 & 0 & \cdots & 0 \\ 0 & 0 & 1/n_2 & & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & 1/n_a \end{bmatrix}, \quad \mathbf{P_1} = \frac{1}{N}\mathbf{J}_N, \quad \mathbf{P_X} = \begin{bmatrix} \frac{1}{n_1}\mathbf{J}_{n_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \frac{1}{n_2}\mathbf{J}_{n_2} & \cdots & \mathbf{0} \\ \vdots & & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \frac{1}{n_a}\mathbf{J}_{n_a} \end{bmatrix},$$

where $N = \sum_{i=1}^a n_i$ and $\mathbf{J}_m$ is a $m \times m$ matrix with all elements being 1. Recall that in Example 6 of Notes 6, we partition $\mathbf{X} = [\mathbf{1}, \mathbf{X}_1]$, and consider

$y^T y = y^T A_0 y + y^T A_1 y + y^T A_2 y.$

$$\mathbf{A}_0 = \mathbf{P_1}, \quad \mathbf{A}_1 = \mathbf{P_X} - \mathbf{P_1}, \quad \mathbf{A}_2 = \mathbf{I} - \mathbf{P_X}.$$

(When compared with Notes 6, a slightly different notations for $\mathbf{A}_i$ is used to match with the presentation below.) Here is the one-way ANOVA table (Example 6 of Notes 6):

Source	df	Projection	Sum of squares	Noncentrality
Mean	1	$\mathbf{P_1}$	$SSM = N\bar{y}^2$	$\phi_M = \frac{\boldsymbol{\eta}^\intercal \mathbf{P_1} \boldsymbol{\eta}}{2\sigma^2}$
Group	$a-1$	$\mathbf{P_X} - \mathbf{P_1}$	$SSA = \sum_{i=1}^a n_i \bar{y}_{i.}^2 - N\bar{y}^2$	$\phi_a = \frac{\boldsymbol{\eta}^\intercal (\mathbf{P_X} - \mathbf{P_1}) \boldsymbol{\eta}}{2\sigma^2}$
Error	$N-a$	$\mathbf{I} - \mathbf{P_X}$	$SSE = \sum_{i=1}^a \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})^2$	0

$\frac{SSA}{\sigma^2} = \frac{y^T A_1 y}{\sigma^2} = y^T \left(\frac{A_1}{\sigma^2}\right) y.$ $\quad \frac{A_1}{\sigma^2} V = \left(\frac{A_1}{\sigma^2}\right)(\sigma^2 I) = A_1$ is idempotent.

$\Rightarrow \frac{SSA}{\sigma^2} \sim \chi^2_{a-1}(\phi/2)$ $\quad \phi = \eta^T \left(\frac{A_1}{\sigma^2}\right) \eta = \eta^T \frac{(P_x - P_1)}{\sigma^2} \eta.$

$\eta = E(y)$

In particular, one can show that

$$\boldsymbol{\eta}^\intercal \mathbf{P_1} \boldsymbol{\eta} = \mathsf{E}(\mathbf{y})^\intercal \mathbf{P_1} \mathsf{E}(\mathbf{y}) = \frac{1}{N} \left[\sum_{i=1}^{a} \sum_{j=1}^{n_i} E(y_{ij}) \right]^2 = \frac{1}{N} \left[N\mu + \sum_{i=1}^{a} n_i \alpha_i \right]^2 = N(\mu + \bar{\alpha})^2$$

$$\boldsymbol{\eta}^\intercal (\mathbf{P_X} - \mathbf{P_1}) \boldsymbol{\eta} = \underbrace{\sum_{i=1}^{a} n_i(\mu + \alpha_i)^2}_{\eta^T P_x \eta} - \underbrace{N(\mu + \bar{\alpha})^2}_{\eta^T P_1 \eta} = \sum_{i=1}^{a} n_i(\alpha_i - \bar{\alpha})^2,$$

where $\bar{\alpha} = N^{-1} \sum_{i=1}^{a} n_i \alpha_i$.

2.2 Kronecker products

In below, we sometimes focus on the balanced setting ($n_i = n$ for all i). Corresponding notations can be much simplified by using the notations of Kronecker product. For any matrices $\mathbf{A} \in \mathbb{R}^{r \times s}$ and $\mathbf{B} \in \mathbb{R}^{u \times v}$, the Kronecker product $\mathbf{A} \otimes \mathbf{B}$ is defined as

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} A_{11}\mathbf{B} & A_{12}\mathbf{B} & \dots & A_{1s}\mathbf{B} \\ A_{21}\mathbf{B} & A_{22}\mathbf{B} & \dots & A_{2s}\mathbf{B} \\ \vdots & \vdots & & \vdots \\ A_{r1}\mathbf{B} & A_{r2}\mathbf{B} & \dots & A_{rs}\mathbf{B} \end{bmatrix} \in \mathbb{R}^{ru \times sv},$$

where A_{ij} is the (i, j)-th element of $\mathbf{A}$. For example,

$$\mathbf{1}_2 \otimes \mathbf{I}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{I}_2 \otimes \mathbf{J}_2 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}, \quad \mathbf{J}_2 \otimes \mathbf{I}_2 = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}.$$

Here are a few useful properties of Kronecker products in our following discussion. For matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}$ (which do not necessarily have the same dimensions), and scalars c and d,

$$\begin{aligned}
\mathbf{A} \otimes (\mathbf{B} + \mathbf{C}) &= \mathbf{A} \otimes \mathbf{B} + \mathbf{A} \otimes \mathbf{C} \\
\mathbf{A} \otimes (\mathbf{B} \otimes \mathbf{C}) &= (\mathbf{A} \otimes \mathbf{B}) \otimes \mathbf{C} \\
(c\mathbf{A}) \otimes (d\mathbf{B}) &= cd(\mathbf{A} \otimes \mathbf{B}) \\
(\mathbf{A} \otimes \mathbf{B})^\intercal &= \mathbf{A}^\intercal \otimes \mathbf{B}^\intercal \\
(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) &= (\mathbf{AC}) \otimes (\mathbf{BD}) \quad \text{assuming the dimensions are compatible} \\
& \qquad\qquad\qquad\qquad\quad \text{such that } \mathbf{AC} \text{ and } \mathbf{BD} \text{ can be formed} \\
(\mathbf{A} \otimes \mathbf{B})^+ &= \mathbf{A}^+ \otimes \mathbf{B}^+ \\
\text{rank}(\mathbf{A} \otimes \mathbf{B}) &= \text{rank}(\mathbf{A})\text{rank}(\mathbf{B}) \\
\text{trace}(\mathbf{A} \otimes \mathbf{B}) &= \text{trace}(\mathbf{A})\text{trace}(\mathbf{B})
\end{aligned}$$

More properties can be found in e.g., "The Matrix Cookbook".

Assume $n_i = n$ for all $i = 1, \dots, a$ (balanced case). One can write

$$\mathbf{X}_1 = \mathbf{I}_a \otimes \mathbf{1}_n, \quad \mathbf{P_X} = \frac{1}{n} \mathbf{I}_a \otimes \mathbf{J}_n.$$

r.v.

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

Fixed. random.

$i = 1 \cdots a.$

① a is large.

② n_i vary. n_i large

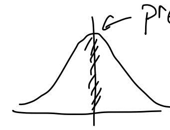
③ James Stein

$$\begin{bmatrix} \widehat{\mu+\alpha_1} \\ \vdots \\ \widehat{\mu+\alpha_a} \end{bmatrix}$$

2.3 One-way mixed ANOVA

So far, we assume that $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_a]^\intercal$ is nonrandom. What happens if it is random? For instance, if the groups are assumed to be drawn from a population. Say, y_{ij} is an attribute of student j sampled from school i, and school i is drawn from a population of schools. Then we can also model the school effect α_i as a random variable. We discuss this setting here. In this subsection, the distributional assumptions of the random quantities are specified as follows.

$$\boldsymbol{\alpha} \sim N_a(\mathbf{0}, \sigma_a^2\mathbf{I}), \quad \mathbf{e} \sim N_N(\mathbf{0}, \sigma^2\mathbf{I})$$



prefer 0.

n_i small shrink $\widehat{\mu+\alpha_i}$ to $\hat{\mu}$

where all these two quantities are mutually independent.

Our goal here is to investigate various sums of squares (similar to the fixed-$\boldsymbol{\alpha}$ setting), in order to provide estimation and inference tools for the variance component σ_a^2. In the mixed model, σ_a^2 describes the properties of group effects. When σ_a^2 is large, there is large variability for the group effects. As for the extreme case when $\sigma_a^2 = 0$, there is no difference between groups.

Fixed Test $\alpha_1 = \alpha_2 \cdots = \alpha_a$ μ $\sum_{i=1}^a \alpha_i = 0$

Random Test $\sigma_a^2 = 0.$ μ $E(\alpha) = 0$

$\alpha_1 \cdots \alpha_a \sim N(0, \sigma^2 I)$

2.3.1 ANOVA estimators

Write $SSA = \mathbf{y}^\intercal\mathbf{A}_1\mathbf{y}$. Since

$$SSA/\sigma^2 \mid \boldsymbol{\alpha} \sim \chi^2_{a-1}\left(\phi_a = \frac{1}{2}\sum_{i=1}^a n_i(\alpha_i - \bar{\alpha})^2/\sigma^2\right),$$

we have $\mathsf{E}(SSA/\sigma^2 \mid \boldsymbol{\alpha}) = a - 1 + \sum_{i=1}^a n_i(\alpha_i - \bar{\alpha})^2/\sigma^2$ due to Proposition 6.14. Using $\boldsymbol{\alpha} \sim N_a(\mathbf{0}, \sigma_a^2\mathbf{I})$ and $\bar{\alpha} \sim N(0, \sigma_a^2(\sum_i n_i^2)/N^2)$, we get

$$\begin{aligned}\frac{1}{a-1}\mathsf{E}(SSA) = \frac{1}{a-1}\mathsf{E}(\mathsf{E}(SSA \mid \boldsymbol{\alpha})) &= \frac{1}{a-1}\mathsf{E}\left((a-1)\sigma^2 + \sum_{i=1}^a n_i(\alpha_i - \bar{\alpha})^2\right) \\ &= \sigma^2 + \frac{1}{a-1}\mathsf{E}\left(\sum_{i=1}^a n_i\alpha_i^2 - N\bar{\alpha}^2\right) \\ &= \sigma^2 + \frac{\sigma_a^2}{a-1}\left(N - \sum_{i=1}^a n_i^2/N\right)\end{aligned}$$

① plug in

② solve σ_a^2

$$\frac{1}{a-1} \times SSA = \frac{SSE}{N-a} + \frac{\sigma_a^2}{a-1}\left(N - \sum \frac{n_i^2}{N}\right)$$

$$SSA = \frac{a-1}{N-a}SSE + \sigma_a^2\left(N - \sum\frac{n_i^2}{N}\right)$$

?

$$\frac{1}{N-a}\mathsf{E}(SSE) = \frac{1}{N-a}\mathsf{E}(\mathsf{E}(SSE|\boldsymbol{\alpha})) = \sigma^2 \quad \text{since } SSE/\sigma^2 \mid \boldsymbol{\alpha} \sim \chi^2_{N-a}$$

Therefore we can derive unbiased estimators for σ^2 and σ_a^2 under the mixed model as follows.

$$\hat{\sigma}^2 = SSE/(N-a), \qquad \hat{\sigma}_a^2 = \frac{SSA - \frac{a-1}{N-a}SSE}{N - \sum_{i=1}^a n_i^2/N}$$

get ③ $\rightarrow E(\hat{\sigma}_a^2) = \sigma_a^2$

which are known as the ANOVA estimators for variance components σ^2 and σ_a^2 respectively.

2.3.2 Hypothesis test

Recall that we utilize the distributions of the sums of squares to conduct hypothesis tests in Notes 7. However, the corresponding distributions of the sums of squares under a general one-way mixed ANOVA model are difficult to derived. Here we only focus on the balanced setting. With the above assumptions,

$$\mathbf{y} \sim N_N(\mu\mathbf{1}, \mathbf{V}),$$

α_i ϵ_{ij}

Balanced Setting. $n_1 = n_2 = \cdots = n_a = n$.

$$X = \begin{bmatrix} 1_n & 1_n & & \\ \vdots & & \ddots & \\ 1_n & & & 1_n \end{bmatrix}$$

$X_0 = 1_a \otimes 1_n$. $\quad X_1 = I \otimes 1_n$.

where

$$\sigma_a^2 I$$

$$\begin{aligned}
\mathbf{V} = \mathsf{Var}(\mathbf{y}) = \mathsf{Var}(\mathbf{1}_N\mu + \mathbf{X}_1\boldsymbol{\alpha} + \mathbf{e}) &= \mathbf{X}_1\mathsf{Var}(\boldsymbol{\alpha})\mathbf{X}_1^\intercal + \mathsf{Var}(\mathbf{e}) \\
&= \sigma_a^2\mathbf{X}_1\mathbf{X}_1^\intercal + \sigma^2\mathbf{I}_N \\
&= \sigma_a^2\mathbf{I}_a \otimes \mathbf{J}_n + \sigma^2\mathbf{I}_N \\
\text{Set } \rho = \frac{\sigma_a^2}{\sigma^2}. \quad &= \sigma^2(\rho\mathbf{I}_a \otimes \mathbf{J}_n + \mathbf{I}_N) = \sigma^2\mathbf{I}_a \otimes (\rho\mathbf{J}_n + \mathbf{I}_n),
\end{aligned}$$

$I_a \otimes I_n$

with

$$\mathbf{X}_1\mathbf{X}_1^\intercal = (\mathbf{I}_a \otimes \mathbf{1}_n)(\mathbf{I}_a \otimes \mathbf{1}_n)^\intercal = (\mathbf{I}_a \otimes \mathbf{1}_n)(\mathbf{I}_a \otimes \mathbf{1}_n^\intercal) = (\mathbf{I}_a\mathbf{I}_a) \otimes (\mathbf{1}_n\mathbf{1}_n^\intercal) = \mathbf{I}_a \otimes \mathbf{J}_n$$

and $\rho = \sigma_a^2/\sigma^2$. If you compare the distribution of $\mathbf{y}$ in the mixed model with the above fixed-$\boldsymbol{\alpha}$ model (fixed model), one immediately notice the difference in terms of the mean and variance of $\mathbf{y}$. Under the mixed model, $\mathbf{y}|\boldsymbol{\alpha} \sim N_N(\mathbf{Xb}, \sigma^2\mathbf{I})$ and so conditioning on $\boldsymbol{\alpha}$, we get the fixed model. However, the unconditioned case requires new developments, especially for the distributions of various sums of squares in the ANOVA table. In fact, whether these sums of squares are independent or not has to be re-investigated.

One relatively easier term is SSE. Conditioned on $\boldsymbol{\alpha}$, $SSE/\sigma^2 \sim \chi^2_{N-a}$. Since the distribution is unchanged with respect to $\boldsymbol{\alpha}$, SSE/σ^2 (hence SSE) is independent of $\boldsymbol{\alpha}$. As a result, $SSE/\sigma^2 \sim \chi^2_{N-a}$ under the mixed model (the unconditioned setting).

$A_1 = P_X - P_1 \qquad N = n_1 + \cdots + n_a = a \cdot n$.

Next, let us investigate $SSA = \mathbf{y}^\intercal\mathbf{A}_1\mathbf{y}$. Note that $\mathbf{A}_1 = n^{-1}\mathbf{I}_a \otimes \mathbf{J}_n - N^{-1}\mathbf{J}_N = (\mathbf{I}_a - a^{-1}\mathbf{J}_a) \otimes (n^{-1}\mathbf{J}_n)$.
Let $\tilde{\mathbf{A}}_1 = \sigma^{-2}(1+n\rho)^{-1}\mathbf{A}_1$. We have

$(an)^{-1} J_a \otimes J_n$.

$y^T A y \not\sim \chi^2$

$y^T \frac{A}{c} y \sim \chi^2$?

$A_1 V = c \cdot A_1$

$\left(\frac{A_1}{c}\right) V = A_1$, idempotent

$$\begin{aligned}
\tilde{\mathbf{A}}_1\mathbf{V} &= \frac{1}{1+n\rho}\left\{(\mathbf{I}_a - \tfrac{1}{a}\mathbf{J}_a) \otimes (\tfrac{1}{n}\mathbf{J}_n)\right\}\left\{\mathbf{I}_a \otimes (\mathbf{I}_n + \rho\mathbf{J}_n)\right\} \\
&= \frac{1}{1+n\rho}\left\{(\mathbf{I}_a - \tfrac{1}{a}\mathbf{J}_a)\mathbf{I}_a\right\} \otimes \left\{(\tfrac{1}{n}\mathbf{J}_n)(\mathbf{I}_n + \rho\mathbf{J}_n)\right\} \\
&= (\mathbf{I}_a - \tfrac{1}{a}\mathbf{J}_a) \otimes (\tfrac{1}{n}\mathbf{J}_n) = \mathbf{A}_1\,,
\end{aligned}$$

$\frac{1}{n}J_n + \rho J_n = \frac{1+n\rho}{n} J_n$.

which is an idempotent matrix. Following from Theorem 6.23,

$$\mathbf{y}^\intercal\tilde{\mathbf{A}}_1\mathbf{y} \sim \chi^2_{a-1}\left(\frac{1}{2}\boldsymbol{\mu}^\intercal\tilde{\mathbf{A}}_1\boldsymbol{\mu}\right), \tag{9.1}$$

with $\boldsymbol{\mu} = \mu\mathbf{1}_N = \mu(\mathbf{1}_a \otimes \mathbf{1}_n)$. We find that

$$\boldsymbol{\mu}^\intercal\mathbf{A}_1\boldsymbol{\mu} = \mu^2(\mathbf{1}_a^\intercal \otimes \mathbf{1}_m^\intercal)\left[(\mathbf{I}_a - \tfrac{1}{a}\mathbf{J}_a) \otimes (\tfrac{1}{m}\mathbf{J}_m)\right](\mathbf{1}_a \otimes \mathbf{1}_m) = 0\,.$$

Thus the RHS of (9.1) is a central χ^2 distribution and

$$\mathbf{y}^\intercal\tilde{\mathbf{A}}_1\mathbf{y} = \frac{1}{\sigma^2 + n\sigma_a^2}\,\mathbf{y}^\intercal\mathbf{A}_1\mathbf{y} \sim \chi^2_{a-1}\,.$$

Since $\tilde{\mathbf{A}}_1\mathbf{V}\mathbf{A}_2 = \mathbf{A}_1\mathbf{A}_2 = \mathbf{0}$, we conclude that SSE and SSA are independent due to Problem 4 of Homework 4. Together with $SSE/\sigma^2 \sim \chi^2_{N-a}$, we have

$$\frac{SSA/(a-1)}{SSE/(N-a)} \times \frac{\cancel{\sigma^2}\ 1}{\cancel{\sigma^2}\ 1 + n\cancel{\sigma_a^2}\ \rho} \sim F_{a-1,N-a}\,. \qquad \rho = \sigma_a^2/\sigma^2$$

Define the usual F test statistic

$$F = \frac{SSA/(a-1)}{SSE/(N-a)} \sim (1+n\rho)\cdot F_{a-1,N-a}, \quad \text{then} \quad F(\rho) = \frac{1}{1+n\rho}\,F \sim F_{a-1,N-a}\,. \tag{9.2}$$

Thus, $F(\rho_0)$ provides a test statistic for testing the null hypothesis $H_0 : \rho \le \rho_0$ against $H_0 : \rho > \rho_0$. In the special case $\rho_0 = 0$, where H_0 corresponds to no treatment effects ($\sigma_a^2 = 0$), the test statistic is again the usual F statistic.

$H_0: \sigma_a^2 = 0 \Rightarrow \rho = 0 \qquad F$ stat $\sim F_{a-1,N-a}$.

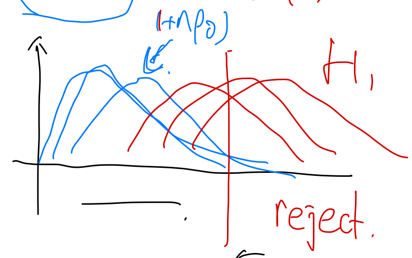
$H_1: \sigma_a^2 \neq 0 \Rightarrow \rho > 0 \qquad F$ stat $\sim (1+\rho) F_{a-1,N-a}$

Under H_0 true

$\text{Fstat} \sim (1+n\rho) F_{a-1,N-a}$

$\underset{(\rho \le \rho_0)}{P(\text{Fstat} > C_\alpha)} \le P_{(\rho=\rho_0)}(\text{Fstat} > C_\alpha) = \alpha$

H_0 $\rho \le \rho_0$ H_1 $\rho \geq \rho_1$. F



$(1+n\rho_0) F_{a-1,N-a,\alpha} = C_\alpha$

3 Two-way model

3.1 Two-way (fixed) ANOVA (with interaction)

In this subsection, we assume the balanced setting and consider the two-way model:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}, \quad i = 1, \dots, a, \quad j = 1, \dots, b, \quad k = 1, \dots, n,$$

where $e_{ijk} \overset{i.i.d.}{\sim} N(0, \sigma^2)$. Write $N = abn$, $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_a]^\intercal$, $\boldsymbol{\beta} = [\beta_1, \dots, \beta_b]^\intercal$ and $\boldsymbol{\gamma} = [\gamma_{11}, \gamma_{12}, \dots, \gamma_{1b}, \dots, \gamma_{a1}, \dots, \gamma_{ab}]^\intercal$. We can write the above model in matrix notations using Kronecker product:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}, \quad \mathbf{b} = \begin{bmatrix} \mu \\ \boldsymbol{\alpha} \\ \boldsymbol{\beta} \\ \boldsymbol{\gamma} \end{bmatrix}, \quad \mathbf{e} \sim N_N(\mathbf{0}, \sigma^2 \mathbf{I})$$

where

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_0 & \mathbf{X}_1 & \mathbf{X}_2 & \mathbf{X}_3 \end{bmatrix}, \quad \mathbf{X}_0 = \mathbf{1}_N, \quad \mathbf{X}_1 = \mathbf{I}_a \otimes \mathbf{1}_{nb}, \quad \mathbf{X}_2 = \mathbf{1}_a \otimes \mathbf{I}_b \otimes \mathbf{1}_n, \quad \mathbf{X}_3 = \mathbf{I}_{ab} \otimes \mathbf{1}_n.$$

Write

$$\mathbf{X}_{(i)} = \begin{bmatrix} \mathbf{X}_0 & \dots & \mathbf{X}_i \end{bmatrix}, \quad \mathbf{A}_0 = \mathbf{P}_{\mathbf{X}_0}, \quad \mathbf{A}_i = \mathbf{P}_{\mathbf{X}_{(i)}} - \mathbf{P}_{\mathbf{X}_{(i-1)}} \text{ for } i = 1, \dots, 3, \quad \mathbf{A}_4 = \mathbf{I} - \mathbf{P}_{\mathbf{X}}$$

We have

$$\begin{aligned}
&\operatorname{rank}(\mathbf{X}_{(0)}) = 1, && \mathbf{P}_{\mathbf{X}_{(0)}} = \frac{1}{N}\mathbf{J}_N \\
&\operatorname{rank}(\mathbf{X}_{(1)}) = a, && \mathbf{P}_{\mathbf{X}_{(1)}} = \frac{1}{bn}\mathbf{I}_a \otimes \mathbf{J}_{bn} \\
&\operatorname{rank}(\mathbf{X}_{(2)}) = a + b - 1, && \mathbf{P}_{\mathbf{X}_{(2)}} = \frac{1}{bn}\mathbf{I}_a \otimes \mathbf{J}_{bn} + \frac{1}{an}\mathbf{J}_a \otimes \mathbf{I}_b \otimes \mathbf{J}_n - \frac{1}{N}\mathbf{J}_N \\
&\operatorname{rank}(\mathbf{X}_{(3)}) = ab, && \mathbf{P}_{\mathbf{X}_{(3)}} = \frac{1}{n}\mathbf{I}_{ab} \otimes \mathbf{J}_n
\end{aligned}$$

The quadratic forms (sums of squares) $\mathbf{y}^\intercal \mathbf{A}_i \mathbf{y}$ are summarized in the following ANOVA table:

Source	df	Projection	Sum of squares
Mean	1	$\mathbf{A}_0$	$SSM = N\bar{y}_{...}^2$
Factor A, $\boldsymbol{\alpha}$	$a-1$	$\mathbf{A}_1$	$SSA = bn\sum_{i=1}^a (\bar{y}_{i..} - \bar{y}_{...})^2$
Factor B, $\boldsymbol{\beta}$	$b-1$	$\mathbf{A}_2$	$SSB = an\sum_{j=1}^b (\bar{y}_{.j.} - \bar{y}_{...})^2$
Interaction, $\boldsymbol{\gamma}$	$(a-1)(b-1)$	$\mathbf{A}_3$	$SSC = n\sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2$
Error	$ab(n-1)$	$\mathbf{A}_4$	$SSE = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (y_{ijk} - \bar{y}_{ij.})^2$

When $\mathbf{b}$ is non-random, by Cohran's theorem, we can show that $\mathbf{y}^\intercal \mathbf{A}_i \mathbf{y}/\sigma^2$, $i = 0, \dots, 4$ are independent and each distributes as (central or noncentral) chi-square distribution.

3.2 Two-way mixed ANOVA

Now, consider a similar model as in the last subsection. Instead of fixing $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, we assume that they are random. The detailed distributional assumptions are listed as follows.

$$\boldsymbol{\alpha} \sim N_a(\mathbf{0}, \sigma_a^2 \mathbf{I}), \quad \boldsymbol{\beta} \sim N_b(\mathbf{0}, \sigma_b^2 \mathbf{I}), \quad \boldsymbol{\gamma} \sim N_{ab}(\mathbf{0}, \sigma_c^2 \mathbf{I}), \quad \mathbf{e} \sim N_N(\mathbf{0}, \sigma^2 \mathbf{I}),$$

where all these quantities are mutually independent.

Then $\mathbf{y} \sim N_N(\mu \mathbf{1}_N, \mathbf{V})$, where

$$\begin{aligned}
\mathbf{V} = \mathsf{Var}(\mathbf{y}) &= \mathsf{Var}(\mathbf{X}_0\mu + \mathbf{X}_1\boldsymbol{\alpha} + \mathbf{X}_2\boldsymbol{\beta} + \mathbf{X}_3\boldsymbol{\gamma} + \mathbf{e}) \\
&= \mathsf{Var}(\mathbf{X}_1\boldsymbol{\alpha}) + \mathsf{Var}(\mathbf{X}_2\boldsymbol{\beta}) + \mathsf{Var}(\mathbf{X}_3\boldsymbol{\gamma}) + \mathsf{Var}(\mathbf{e}) \\
&= \mathbf{X}_1\mathsf{Var}(\boldsymbol{\alpha})\mathbf{X}_1^\intercal + \mathbf{X}_2\mathsf{Var}(\boldsymbol{\beta})\mathbf{X}_2^\intercal + \mathbf{X}_3\mathsf{Var}(\boldsymbol{\gamma})\mathbf{X}_3^\intercal + \sigma^2\mathbf{I}_N \\
&= \sigma_a^2\mathbf{X}_1\mathbf{X}_1^\intercal + \sigma_b^2\mathbf{X}_2\mathbf{X}_2^\intercal + \sigma_c^2\mathbf{X}_3\mathbf{X}_3^\intercal + \sigma^2\mathbf{I}_N \\
&= \sigma_a^2\mathbf{I}_a \otimes \mathbf{J}_{nb} + \sigma_b^2\mathbf{J}_a \otimes \mathbf{I}_b \otimes \mathbf{J}_n + \sigma_c^2\mathbf{I}_{ab} \otimes \mathbf{J}_n + \sigma^2\mathbf{I}_N \\
&= \sigma_a^2\mathbf{I}_a \otimes \mathbf{J}_b \otimes \mathbf{J}_n + \sigma_b^2\mathbf{J}_a \otimes \mathbf{I}_b \otimes \mathbf{J}_n + \sigma_c^2\mathbf{I}_a \otimes \mathbf{I}_b \otimes \mathbf{J}_n + \sigma^2\mathbf{I}_N,
\end{aligned}$$

since

$$\begin{aligned}
&\mathbf{X}_1\mathbf{X}_1^\intercal = \mathbf{I}_a \otimes \mathbf{J}_{nb} \\
&\mathbf{X}_2\mathbf{X}_2^\intercal = (\mathbf{1}_a \otimes \mathbf{I}_b \otimes \mathbf{1}_n)(\mathbf{1}_a^\intercal \otimes \mathbf{I}_b \otimes^\intercal \mathbf{1}_n^\intercal) = (\mathbf{1}_a\mathbf{1}_a^\intercal) \otimes ((\mathbf{I}_b \otimes \mathbf{1}_n)(\mathbf{I}_b \otimes \mathbf{1}_n^\intercal)) = \mathbf{J}_a \otimes \mathbf{I}_b \otimes \mathbf{J}_n \\
&\mathbf{X}_3\mathbf{X}_3^\intercal = \mathbf{X}_3\mathbf{X}_3^\intercal = \mathbf{I}_{ab} \otimes \mathbf{J}_n.
\end{aligned}$$

We omit the detailed computation and only list the results. The expected mean squares are:

$$\begin{array}{rcccccccc}
\mathsf{E}(SSA)/(a-1) & = & bn\sigma_a^2 & & & + & n\sigma_c^2 & + & \sigma^2 \\
\mathsf{E}(SSB)/(b-1) & = & & & an\sigma_b^2 & + & n\sigma_c^2 & + & \sigma^2 \\
\mathsf{E}(SSC)/[(a-1)(b-1)] & = & & & & & n\sigma_c^2 & + & \sigma^2 \\
\mathsf{E}(SSE)/[ab(n-1)] & = & & & & & & & \sigma^2
\end{array}$$

The sums of squares, SSA, SSB, SSC and SSE are independent, and

$$\begin{array}{lcl lcl}
\dfrac{1}{bn\sigma_a^2 + n\sigma_c^2 + \sigma^2}\, SSA & \sim & \chi^2_{a-1}\ , & \dfrac{1}{\sigma^2} SSE & \sim & \chi^2_{ab(n-1)} \\
\dfrac{1}{an\sigma_b^2 + n\sigma_c^2 + \sigma^2}\, SSB & \sim & \chi^2_{b-1}\ , & \dfrac{1}{n\sigma_c^2 + \sigma^2} SSC & \sim & \chi^2_{(a-1)(b-1)}
\end{array}$$

3.3 ANOVA estimators

The ANOVA estimators for variance components can be derived from the above expressions of expected mean squares as follows.

$$\begin{aligned}
&\hat{\sigma}^2 = \frac{SSE}{ab(n-1)}, \quad \hat{\sigma}_c^2 = \frac{1}{n}\left[\frac{SSC}{(a-1)(b-1)} - \hat{\sigma}^2\right] \\
&\hat{\sigma}_a^2 = \frac{1}{bn}\left[\frac{SSA}{a-1} - \frac{SSC}{(a-1)(b-1)}\right], \quad \hat{\sigma}_a^2 = \frac{1}{an}\left[\frac{SSB}{b-1} - \frac{SSC}{(a-1)(b-1)}\right]
\end{aligned}$$

3.4 Hypothesis test

Any ratio of two sums of squares has a scaled F distribution, this fact provides F tests of no effects. For example,

H_0	F-statistic	Distribution
$\sigma_c^2 = 0$	$\dfrac{SSC/[(a-1)(b-1)]}{SSE/[ab(n-1)]} \times \dfrac{\sigma^2}{n\sigma_c^2 + \sigma^2}$	$F_{(a-1)(b-1),ab(n-1)}$
$\sigma_a^2 = 0$	$\dfrac{SSA/(a-1)}{SSC/[(a-1)(b-1)]} \times \dfrac{n\sigma_c^2 + \sigma^2}{bn\sigma_a^2 + n\sigma_c^2 + \sigma^2}$	$F_{a-1,(a-1)(b-1)}$
$\sigma_b^2 = 0$	$\dfrac{SSB/(b-1)}{SSC/[(a-1)(b-1)]} \times \dfrac{n\sigma_c^2 + \sigma^2}{an\sigma_b^2 + n\sigma_c^2 + \sigma^2}$	$F_{b-1,(a-1)(b-1)}$

For the more general hypotheses of small effects, the null hypotheses are in terms of ratios of variance components. For example, the null hypothesis of small interaction effect is

$$H_0: \frac{\sigma_c^2}{\sigma^2} \leq \rho_0 \quad \text{vs.} \quad H_1: \frac{\sigma_c^2}{\sigma^2} > \rho_0$$

and the test statistic is $\frac{1}{1+n\rho_0} MSC/MSE$ where $MSC = SSC/[(a-1)(b-1)]$ and $MSE = SSE/[ab(n-1)]$. It distributes as $F_{(a-1)(b-1),ab(n-1)}$ under H_0.