1.017/1.010 Class 21
Multifactor Analysis of Variance

Multifactor Models

We often wish to consider several factors contributing to variability ratherr than just one. Extend concepts of single factor ANOVA to multiple factors. Focus on the two-factor case.

Suppose there I treatments for Factor A and J treatments for factor B., giving IJ random variables described by CDFs $F_{\boldsymbol{x}ij}(x_{ij})$. The different $F_{\boldsymbol{x}ij}(x_{ij})$ are assumed **identical** (except for their means) and **normally distributed** (check this, as in single factor case).

A random sample $[\boldsymbol{x}_{ij1}, \boldsymbol{x}_{ij2}, ..., \boldsymbol{x}_{ijK}]$ of size K is obtained for treatment combination (i, j). Two-factor model describing $\boldsymbol{x}_{ijk}$:

$$\boldsymbol{x}_{ijk} = \mu_{ij} + \boldsymbol{e}_{ijk} = \mu + a_i + b_j + c_{ij} + \boldsymbol{e}_{ijk}$$

$\mu_{ij} = E[\boldsymbol{x}_{ijk}] = \mu + a_i + b_j + c_{ij}$ = unknown mean of $\boldsymbol{x}_{ijk}$ (for all k)

μ = unknown **grand mean** (average of μ_i's).

a_i = unknown **main effects** of **Factor *A***

b_j = unknown **main effects** of **Factor *B***

c_{ij} = unknown **interactions** between Factors A and B

$\boldsymbol{e}_{ijk}$ = **random residual** for treatment i, replicate j

$E[\boldsymbol{e}_{ijk}] = 0$, $Var[\boldsymbol{e}_{ijk}] = \sigma^2$, for all i, j, k

Note that c_{ij} can only be distinguished from $\boldsymbol{e}_{ijk}$ if number of replicates K>1. Constraints:

$$\sum_{i=1}^{I} a_i = \sum_{j=1}^{J} b_j = 0 \quad \sum_{i=1}^{I} c_{ij} = 0 \quad \forall j \quad\quad \sum_{j=1}^{J} c_{ij} = 0 \quad \forall i$$

Objective is to estimate/test values of a_i's, b_j's, and c_{ij}'s, which are distributional parameters for the $F_{\boldsymbol{x}ij}(x_{ij})$'s.

Formulating the Problem as a Hypothesis Test

Formulate three sum-of-squares hypotheses that insure that all a_i's, all b_i's, or all c_{ij}'s are zero:

$$\text{H0A}: \sum_{i=1}^{I} a_i^2 = 0$$

$$\text{H0B}: \sum_{j=1}^{J} b_i^2 = 0$$

$$\text{H0AB}: \sum_{j=1}^{J} \sum_{i=1}^{I} c_i^2 = 0$$

Derive test statistics based on sums-of-squares of data.

Sums-of-Squares Computations

Define treatment and grand sample means:

$$\boldsymbol{m_{xi}} = \frac{1}{JK} \sum_{j=1}^{J} \sum_{k=1}^{K} \boldsymbol{x}_{ijk} = \bar{\boldsymbol{x}}_{i\cdot\cdot} \qquad \boldsymbol{m_{xj}} = \frac{1}{IK} \sum_{i=1}^{I} \sum_{k=1}^{K} \boldsymbol{x}_{ijk} = \bar{\boldsymbol{x}}_{\cdot j\cdot}$$

$$\boldsymbol{m_{xij}} = \frac{1}{K} \sum_{j=1}^{K} \boldsymbol{x}_{ijk} = \bar{\boldsymbol{x}}_{ij\cdot}$$

$$\boldsymbol{m_{x}} = \frac{1}{IJK} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \boldsymbol{x}_{ijk} = \bar{\boldsymbol{x}}_{\cdot\cdot\cdot}$$

Test statistics are computed from sums-of-squares:

$$\boldsymbol{SSE} = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (\boldsymbol{x}_{ijk} - \boldsymbol{m_{xij}})^2$$

$$\boldsymbol{SSA} = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (\boldsymbol{m_{xi}} - \boldsymbol{m_{x}})^2 \qquad \boldsymbol{SSB} = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (\boldsymbol{m_{xj}} - \boldsymbol{m_{x}})^2$$

$$\boldsymbol{SSAB} = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (\boldsymbol{m}_{\boldsymbol{x}ij} - \boldsymbol{m}_{\boldsymbol{x}i} - \boldsymbol{m}_{\boldsymbol{x}j} + \boldsymbol{m}_{\boldsymbol{x}})^2$$

$$\boldsymbol{SST} = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (\boldsymbol{x}_{ijk} - \boldsymbol{m}_{\boldsymbol{x}})^2 = \boldsymbol{SSE} + \boldsymbol{SSA} + \boldsymbol{SSB} + \boldsymbol{SSAB}$$

Corresponding mean-sums-of-squares are:

$$\boldsymbol{MSE} = \frac{\boldsymbol{SSE}}{IJ(K-1)}$$

$$\boldsymbol{MSA} = \frac{\boldsymbol{SSA}}{I-1} \qquad \boldsymbol{MSB} = \frac{\boldsymbol{SSB}}{J-1}$$

$$\boldsymbol{MSAB} = \frac{\boldsymbol{SSAB}}{(I-1)(J-1)}$$

Expected values of these mean-sums-of-squares show depends on main effects and interactions:

$$E[\boldsymbol{MSE}] = \sigma^2$$

$$E[\boldsymbol{MSA}] = \sigma^2 + \frac{JK}{I-1}\sum_{i=1}^{I} a_i^2 \qquad E[\boldsymbol{MSB}] = \sigma^2 + \frac{IK}{I-1}\sum_{j=1}^{J} b_j^2$$

$$E[\boldsymbol{MSAB}] = \sigma^2 + \frac{K}{(I-1)(J-1)}\sum_{i=1}^{I}\sum_{j=1}^{J} c_{ij}^2$$

Test Statistic

Use ratios as test statistics for the three hypotheses:

$$\mathcal{F}_A(\boldsymbol{MSA}, \boldsymbol{MSE}) = \frac{\boldsymbol{MSA}}{\boldsymbol{MSE}}$$

$$\mathcal{F}_B(\boldsymbol{MSB}, \boldsymbol{MSE}) = \frac{\boldsymbol{MSB}}{\boldsymbol{MSE}}$$

$$\mathcal{F}_{AB}(\boldsymbol{MSAB}, \boldsymbol{MSE}) = \frac{\boldsymbol{MSAB}}{\boldsymbol{MSE}}$$

When H0 is true each statistic follows **F distribution** with degree of freedom parameters $\nu_A = I - 1$, $\nu_B = J - 1$, $\nu_{AB} = (I - 1)(J - 1)$, and $\nu_E = IJ(K-1)$.

One-sided rejection regions

$$R0A : \mathcal{F}(MSA, MSE) \geq F^{-1}_{\mathcal{F}, \nu_A, \nu_E}[\alpha]$$

$$R0B : \mathcal{F}(MSB, MSE) \geq F^{-1}_{\mathcal{F}, \nu_B, \nu_E}[\alpha] \quad :$$

$$R0AB : \mathcal{F}(MSAB, MSE) \geq F^{-1}_{\mathcal{F}, \nu_{AB}, \nu_E}[\alpha]$$

One-sided p-values:

$$p_A = 1 - F_{\mathcal{F}, \nu_A, \nu_E}[\mathcal{F}(MSA, MSE)]$$

$$p_B = 1 - F_{\mathcal{F}, \nu_A, \nu_E}[\mathcal{F}(MSB, MSE)]$$

$$p_{AB} = 1 - F_{\mathcal{F}, \nu_{AB}, \nu_E}[\mathcal{F}(MSAB, MSE)]$$

Unbalanced ANOVA problems with **different sample sizes for different treatments** can be handled by modifying formulas slightly.

Two-Factor ANOVA Tables

Source	SS	df	MS	$\mathcal{F}$	p
Factor A	SSA	$\nu_A = I - 1$	$MSA = SSA/\nu_A$	$\mathcal{F}_A = MSA/MSE$	$p = 1 - F_{\mathcal{F}, \nu A, \nu E}(\mathcal{F})$
Factor B	SSB	$\nu_B = J - 1$	$MSB = SSB/\nu_B$	$\mathcal{F}B = MSB/MSE$	$p = 1 - F_{\mathcal{F}, \nu B, \nu E}(\mathcal{F})$
Interaction AB	$SSAB$	$\nu_{AB} = (I-1)(J-1)$	$MSAB = SSAB/\nu_{AB}$	$\mathcal{F}_{AB} = MSAB/MSE$	$p = 1 - F_{\mathcal{F}, \nu AB, \nu E}(\mathcal{F})$
Error	SSE	$\nu_E = IJ(K-1)$	$MSE = SSE/\nu_E$		
Total	SST	$\nu_T = IJK - 1$			