Lecture 4: Two-Factor Studies with Equal Sample Sizes

STA 106: Analysis of Variance

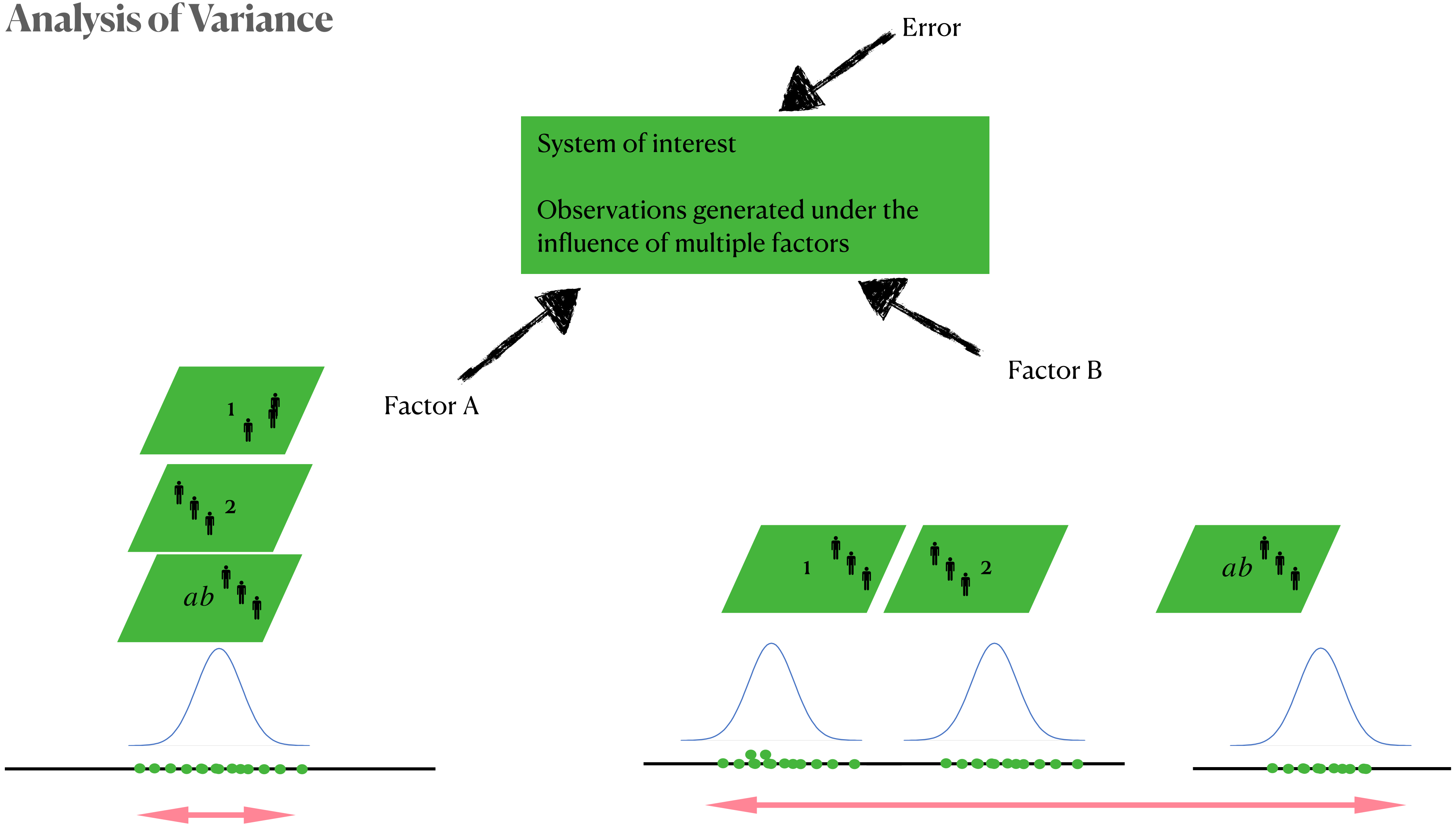
Suggested reading: ALSM Chapter 19

Xiner Zhou
Department of Statistics, University of California, Davis

Two-Factor Studies with Equal Sample Sizes

1. Two-Way ANOVA Model
2. Analysis of Variance
3. Tests for Interaction effects and main effects
4. Analysis of Factor A and B Main Effects (When Factors Do Not Interact)
5. Analysis of Treatment Means when Interactions Are Present

Analysis of Variance



Without factors A and B, the observations have some natural variation due to other extraneous factors, i.e. “error variance”

If some combinations of factor A and B indeed has some effects on the system, then we would expect more volatility .

Analysis of Variance

Partition of Total Sum of Squares

$$Y_{ijk} - \bar{Y}_{\dots} = \bar{Y}_{ij\cdot} - \bar{Y}_{\dots} + Y_{ijk} - \bar{Y}_{ij\cdot}$$

Total deviation

Deviation of estimated treatment mean around overall mean

Deviation around estimated treatment mean

$$\underbrace{\sum_i \sum_j \sum_k \left(Y_{ijk} - \bar{Y}_{\dots}\right)^2}_{\text{Total variation}} = \sum_i \sum_j \sum_k \left(\bar{Y}_{ij\cdot} - \bar{Y}_{\dots}\right)^2 + \sum_i \sum_j \sum_k \left(Y_{ijk} - \bar{Y}_{ij\cdot}\right)^2$$

$$= \underbrace{n \sum_i \sum_j \left(\bar{Y}_{ij\cdot} - \bar{Y}_{\dots}\right)^2}_{\text{Variation due to factor A and B}} + \underbrace{\sum_i \sum_j \sum_k \left(Y_{ijk} - \bar{Y}_{ij\cdot}\right)^2}_{\text{Variation due to extraneous factors}}$$

Let $SSTO = \sum_i \sum_j \sum_k \left(Y_{ijk} - \bar{Y}_{\dots}\right)^2$

$$SSTR = n \sum_i \sum_j \left(\bar{Y}_{ij} - \bar{Y}_{\dots}\right)^2$$

$$SSE = \sum_i \sum_j \sum_k \left(Y_{ijk} - \bar{Y}_{ij\cdot}\right)^2 = \sum_i \sum_j \sum_k e_{ijk}^2$$

$$\boxed{\text{SSTO} = \text{SSTR} + \text{SSE}}$$

Analysis of Variance

Partition of Treatment Sum of Squares.

$$\underbrace{\bar{Y}_{ij} - \bar{Y}_{\dots}}_{\substack{\text{Deviation of}\\ \text{estimated}\\ \text{treatment mean}\\ \text{around overall}\\ \text{mean}}} = \underbrace{\bar{Y}_{i..} - \bar{Y}_{...}}_{\text{A main effect}} + \underbrace{\bar{Y}_{.j.} - \bar{Y}_{...}}_{\text{B main effect}} + \underbrace{\bar{Y}_{ij} - \bar{Y}_{i..} - \bar{Y}_{j.} + \bar{Y}_{...}}_{\text{AB interaction effect}}$$

$$\underbrace{\sum_i \sum_j \sum_k \left(\bar{Y}_{ij} - \bar{Y}_{\dots}\right)^2}_{\substack{\text{SSTR : treatment sum of}\\ \text{squares}}} = bn \underbrace{\sum_i \left(\bar{Y}_{i..} - \bar{Y}_{...}\right)^2}_{\text{SSA: factor A sum of squares}} + an \underbrace{\sum_j \left(\bar{Y}_{.j.} - \bar{Y}_{...}\right)^2}_{\text{SSB: factor B sum of squares}} + \underbrace{\sum_i \sum_j \sum_k \left(\bar{Y}_{ij} - \bar{Y}_{i..} - \bar{Y}_{j.} + \bar{Y}_{...}\right)^2}_{\text{AB interaction sum of squares}}$$

$$\Longrightarrow \quad \boxed{\text{SSTR} \quad = \quad \text{SSA + SSB +SSAB}}$$

Analysis of Variance

Combined Partition of Total Sum of Squares.

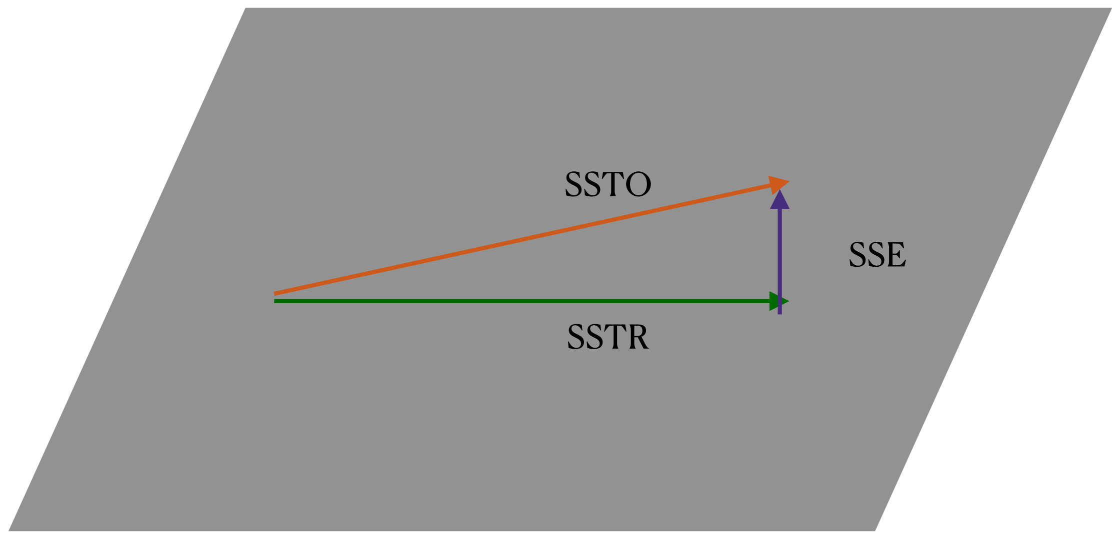
$$\underbrace{\sum_i \sum_j \sum_k \left(Y_{ijk} - \bar{Y}_{\ldots}\right)^2}_{\text{SSTO}} = \sum_i \sum_j \sum_k \left(\bar{Y}_{ij.} - \bar{Y}_{\ldots}\right)^2 + \sum_i \sum_j \sum_k \left(Y_{ijk} - \bar{Y}_{ij.}\right)^2$$

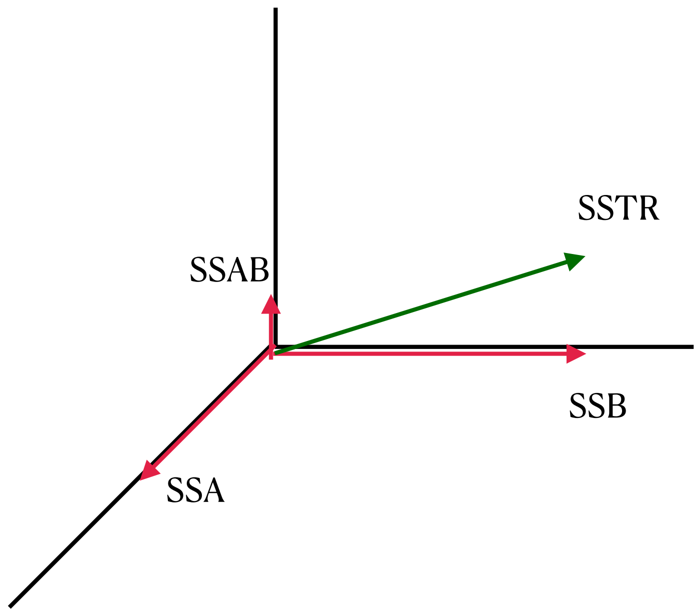
$$= \underbrace{n \sum_i \sum_j \left(\bar{Y}_{ij.} - \bar{Y}_{\ldots}\right)^2}_{\text{SSTR}} + \underbrace{\sum_i \sum_j \sum_k \left(Y_{ijk} - \bar{Y}_{ij.}\right)^2}_{\text{SSE}}$$

$$= \underbrace{bn \sum_i \left(\bar{Y}_{i..} - \bar{Y}_{\ldots}\right)^2}_{\text{SSA: factor A sum of squares}} + \underbrace{an \sum_j \left(\bar{Y}_{.j.} - \bar{Y}_{\ldots}\right)^2}_{\text{SSB: factor B sum of squares}} + \underbrace{\sum_i \sum_j \sum_k \left(\bar{Y}_{ij} - \bar{Y}_{i..} - \bar{Y}_{j.} + \bar{Y}_{\ldots}\right)^2}_{\text{SSAB: AB interaction sum of squares}} + \underbrace{\sum_i \sum_j \sum_k \left(Y_{ijk} - \bar{Y}_{ij.}\right)^2}_{\text{SSE}}$$

$$\Rightarrow \boxed{\text{SSTO} = \text{SSA} + \text{SSB} + \text{SSAB} + \text{SSE}}$$

Geometry of Decomposition of Variance:





Degrees of Freedom

Think of: dimensions of the space where an estimator lives in and allows to run free

$$\underbrace{\sum_i \sum_j \sum_k \left(Y_{ijk} - \bar{Y}_{\dots}\right)^2}_{\text{SSTO}} = \underbrace{bn \sum_i \left(\bar{Y}_{i..} - \bar{Y}_{...}\right)^2}_{\text{SSA: factor A sum of squares}} + \underbrace{an \sum_j \left(\bar{Y}_{.j.} - \bar{Y}_{...}\right)^2}_{\text{SSB: factor B sum of squares}} + \underbrace{\sum_i \sum_j \sum_k \left(\bar{Y}_{ij} - \bar{Y}_{i..} - \bar{Y}_{j.} + \bar{Y}_{...}\right)^2}_{\text{SSAB: AB interaction sum of squares}} + \underbrace{\sum_i \sum_j \sum_k \left(Y_{ijk} - \bar{Y}_{ij.}\right)^2}_{\text{SSE}}$$

How many independence pieces of information go into each quantity?

$Y_{ijk} - \overline{Y_{...}}$

abn pieces

But $\sum_i \sum_j \sum_k (Y_{ijk} - \overline{Y_{...}}) = 0$

$df(SSTO) = abn - 1$

$\overline{Y_{i..}} - \overline{Y_{...}}$

a pieces

But $\sum_i (\overline{Y_{i..}} - \overline{Y_{...}}) = 0$

$df(SSA) = a - 1$

$\overline{Y_{.j.}} - \overline{Y_{...}}$

b pieces

But $\sum_j (\overline{Y_{.j.}} - \overline{Y_{...}}) = 0$

$df(SSB) = b - 1$

$\bar{Y}_{ij} - \bar{Y}_{i..} - \bar{Y}_{j.} + \bar{Y}_{...} = \hat{\gamma}_{ij}$

ab pieces

But $\sum_i \hat{\gamma}_{ij} = 0 \; j = 1, \dots, b$

$\sum_j \hat{\gamma}_{ij} = 0 \; i = 1, \dots, a$

$Y_{ijk} - \overline{Y_{ij.}}$

abn pieces

But $\sum_k (Y_{ijk} - \overline{Y_{ij.}}) = 0$ for $i = 1 \dots a,\ j = 1 \dots b$

$df(SSE) = abn - ab = ab(n-1)$

$df(SSAB) = ab - (a + b - 1) = (a-1)(b-1)$

$$MSA = \frac{SSA}{a-1} \qquad MSB = \frac{SSB}{b-1} \qquad MSAB = \frac{SSAB}{(a-1)(b-1)} \qquad MSE = \frac{SSE}{ab(n-1)}$$

What's expected values of Mean Squares?

$$E\{MSE\} = \sigma^2$$

$$E\{MSA\} = \sigma^2 + nb\frac{\sum \alpha_i^2}{a-1} = \sigma^2 + nb\frac{\sum \left(\mu_{i.} - \mu_{..}\right)^2}{a-1}$$

$$E\{MSB\} = \sigma^2 + na\frac{\sum \beta_j^2}{b-1} = \sigma^2 + na\frac{\sum \left(\mu_{.j} - \mu_{..}\right)^2}{b-1}$$

$$E\{MSAB\} = \sigma^2 + n\frac{\sum\sum \gamma_{ij}^2}{(a-1)(b-1)}$$

$$= \sigma^2 + n\frac{\sum\sum \left(\mu_{ij} - \mu_{i.} - \mu_{.j} + \mu_{..}\right)^2}{(a-1)(b-1)}$$

ANOVA Table for Two-factor Studies (Two-Way ANOVA Table)

Source of Variation	SS	df	MS	E{MS}
Factor A	$SSA = nb\sum(\bar{Y}_{i..} - \bar{Y}_{...})^2$	$a-1$	$MSA = \frac{SSA}{a-1}$	$\sigma^2 + bn\frac{\sum(\mu_{i.} - \mu_{..})^2}{a-1}$
Factor B	$SSB = na\sum(\bar{Y}_{.j.} - \bar{Y}_{...})^2$	$b-1$	$MSB = \frac{SSB}{b-1}$	$\sigma^2 + an\frac{\sum(\mu_{.j} - \mu_{..})^2}{b-1}$
AB interactions	$SSAB = n\sum\sum(\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2$	$(a-1)(b-1)$	$MSAB = \frac{SSAB}{(a-1)(b-1)}$	$\sigma^2 + n\frac{\sum\sum(\mu_{ij} - \mu_{i.} - \mu_{.j} + \mu_{..})^2}{(a-1)(b-1)}$
Error	$SSE = \sum\sum\sum(Y_{ijk} - \bar{Y}_{ij.})^2$	$ab(n-1)$	$MSE = \frac{SSE}{ab(n-1)}$	σ^2
Total	$SSTO = \sum\sum\sum(Y_{ijk} - \bar{Y}_{...})^2$	$nab-1$		

Compare E(MSAB) vs E(MSE)

- If there are no interactions, i.e. the additive factor effects model holds
  - MSAB has same expiation as MSE (they tend to be similar in numeric values)
- Otherwise, when interaction exists
  - MSAB tend to be larger than MSE

Analysis of Variance

MSA has same expiation as MSE (they tend to be similar in numeric values)

Compare E(MSA) vs E(MSE)

- If there are no factor A main effects
- Otherwise

MSA tend to be larger than MSE

MSB has same expiation as MSE (they tend to be similar in numeric values)

Compare E(MSB) vs E(MSE)

- If there are no factor B main effects
- Otherwise

MSB tend to be larger than MSE

They suggest that ratios of Mean Squares provide evidence about the main effects and interactions, which will be the basis for F tests

Example

ANOVA Table

	SS	df	MS
factor A	0.441293435727881	1	0.441293435727881
factor B	3.20098397061084	2	1.60049198530542
interaction AB	0.11989261439729	2	0.0599463071986451
Error	5.46770181944887	54	0.101253737397201
Totoal	9.22987184018488	59	•