# Within-subjects and mixed factorial designs

**Session 11**

MATH 80667A: Experimental Design and Statistical Methods
for Quantitative Research in Management
HEC Montréal

# Outline

## Hasse diagrams

## Two factors within-subjects designs

# Hasse diagrams

# To-do list

Need to determine

1. relation between factors (nested or crossed)
2. nature of factors (fixed or random effects)
3. which interactions should be in the model (maximal structure supported by experimental design)

The statistical model must match the experiment design.

Building tests require considering variability of effects (mean squared errors and degrees of freedom of denominators).

# What is a Hasse diagram?

A graph in which every effect is a node

- Node labels indicate fixed vs (random)
- superscript is number of instances
- subscript degrees of freedom
- lines joining variables show crossing/nesting relations and hierarchy

> Degrees of freedom for a term $U$ are found by starting with the superscript for $U$ and subtracting out the degrees of freedom for all terms above $U$.

This rule only applies for balanced designs.

# Examples of Hasse diagrams

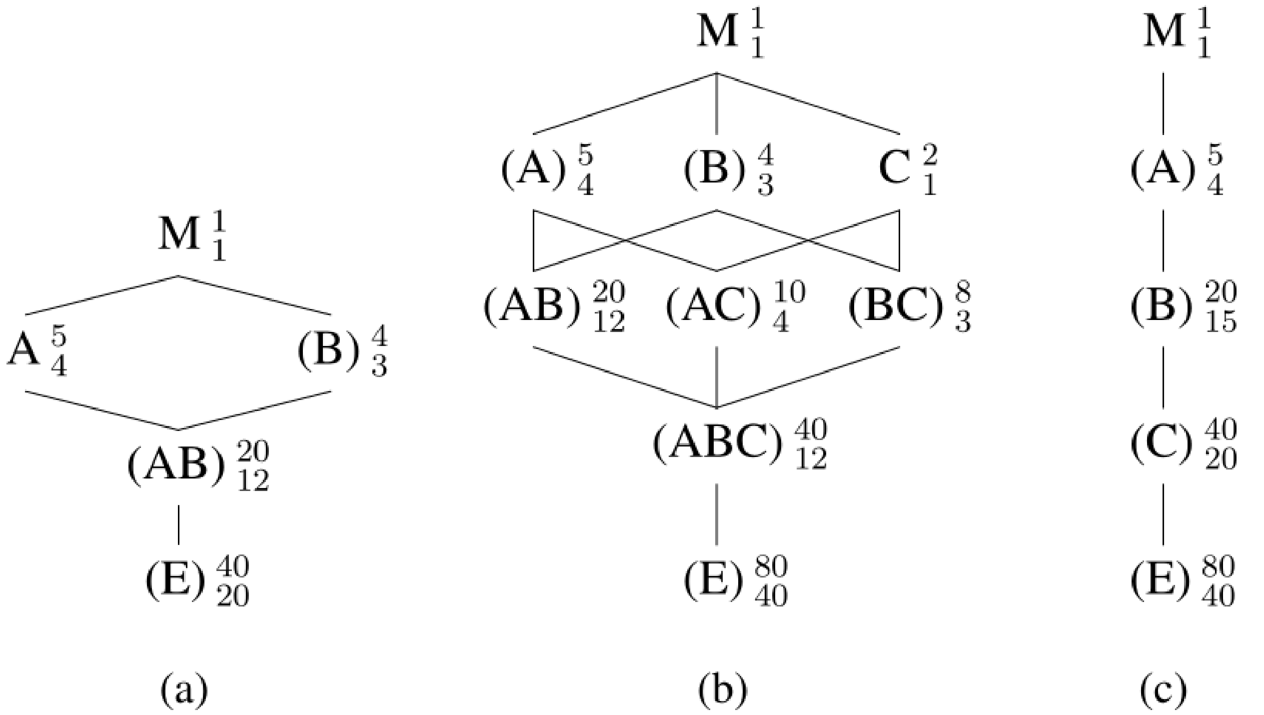


**Figure 12.1:** Hasse diagrams: (a) two-way factorial with A fixed and B random, A and B crossed; (b) three-way factorial with A and B random, C fixed, all factors crossed; (c) fully nested, with B fixed, A and C random. In all cases, A has 5 levels, B has 4 levels, and C has 2 levels.

# Rules for testing (denominators)

From Oehlert (2001), Display 12.3

1. The denominator for testing a term $U$ is the leading eligible random term below $U$ in the Hasse diagram.
2. An eligible random term $V$ below $U$ is leading if there is no eligible random term that is above $V$ and below $U$.
3. If there are two or more leading eligible random terms, then we must use an approximate test.
4. In the restricted model, all random terms below $U$ are eligible except those that contain a fixed factor not found in $U$.

# Example: chocolate rating

Example from L. Meier, adapted from Oehlert (2010)

> A group of 10 rural and 10 urban raters rated 4 different chocolate types. Every rater got to eat two samples from the same chocolate type in random order.

# Within-subjects and mixed designs

# Why use within-subject factors?

- (+) can decompose variance (more power)
- (+) fewer participants needed
- (-) crossover effects
- (-) more complicated
- (-) assumptions about correlation structure

# Two-factor within design

We observe $A \times B \times S$: each subject $S_i$ gets assigned to every treatment pair $a_j \times b_k$.

- $A$ and $B$ are crossed (interaction)
- Hasse diagram (board work)

# Testing in within-designs

Main-effect statistics are

- $A$: $\text{MS}_A/\text{MS}_{AS}$
- $B$: $\text{MS}_B/\text{MS}_{BS}$
- $AB$: $\text{MS}_{AB}/\text{MS}_{ABS}$

where $\text{MS}$ is the mean squared error.

The term $\text{MS}_{ABS}$ is equivalent to residuals if there are no replications.

# Analyzing contrasts in within-designs

From Table 18.5 of Keppel and Wickens (2004)

> Any effect involving a contrast on a within-subject factor is tested against an error term that includes the interaction of that contrast with subjects. To calculate this error term, create a contrast variable by applying the contrast to each subject's data, then analyse these computed data as if they come from a simpler design.
>
> Any within-subject effect involving a portion of the data (e.g., a pairwise comparison between two levels of one factor), is tested against an error term derived exclusively from those data. To calculate this error term, simply extract these data and analyze them as if they come from a simpler design.

# Two-factor mixed design

We observe $A \times (B \times S)$: each subject $S_i$ is nested in $A$ and we see each of the $b$ levels of $B$ per subject

- Hasse diagram (board work)

# Example: Curley et al. (2021+)

> Two variables were manipulated within participants: (a) evidence anchor (strong-first versus weak-first); (b) verdict system (two- versus three-verdict systems). Total pre-trial bias score was used as a covariate in the analysis (this score is based on the PJAQ and is explained further in the Materials section). Participants were also given two vignettes (Vignette 1 and Vignette 2); thus, the vignette variable was included in the data analysis [...]
>
> The dependent variable was the final belief of guilt score, which was measured on an accumulated scale from 0–14, with 0 representing no belief of guilt and 14 representing a total belief that the person is guilty

# Latin square

A latin square is a incomplete blocked design with one treatment with $T$ levels and two blocking factors with also $T$ levels each.

Experiment arranged so that experimental unit is assigned once to each row/column. Letters correspond to the different treatments:

| | **col 1** | **col 2** | **col 3** | **col 4** |
|---|---|---|---|---|
| **row 1** | $A$ | $B$ | $C$ | $D$ |
| **row 2** | $B$ | $C$ | $D$ | $A$ |
| **row 3** | $C$ | $D$ | $A$ | $B$ |
| **row 4** | $D$ | $A$ | $B$ | $C$ |

# Example of a $4 \times 4$ latin square

The model takes the form

$$\underset{\text{response}}{Y_{ijk}} = \mu + \underset{\text{treatment}}{\alpha_i} + \underset{\text{blocking (row)}}{\beta_j} + \underset{\text{blocking (col)}}{\gamma_k} + \underset{\text{error}}{\varepsilon_{ijk}}$$

There are $1 + 3 \cdot (T - 1)$ mean parameters to estimate, and **no interaction**.

This systematic assignment is called **counterbalancing**.

The order of the treatment to row/treatment is randomized.

**Only look at treatment effect $\alpha_i$.**