ANOVA for two or more factor experiments

Session 5

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

Outline

Factorial designs and interactions

Contrasts

Factorial designs and interactions

Motivating example

Consider a study on the retention of information by children aged 4 two hours after reading a story.

We expect the ending (happy/sad/neutral) and the complexity (easy/average/hard) to impact their retention.

complexity	happy	sad	neutral
complicated	😃	😟	😐
average	😃	😟	😐
simple	😃	😟	😐

Why factorial designs?

To study the impact of story complexity and ending, we could run a series of one-way ANOVA.

Factorial designs are more efficient: can study the impact of multiple variables simultaneously with **fewer overall observations**.

Estimates

Factorial design: study with multiple factors (subgroups)

- **simple effects**: difference between levels of one in a fixed combination of others (change in difficulty for happy ending)
- **main effects**: differences relative to average for each condition of a factor (happy vs neutral vs sad ending)
- **interaction effects**: when simple effects differ depending on levels of another factor

Vocabulary

- Simple effects are estimates from each ANOVA (either ending or difficulty of story)
- comparing cell averages within a given row or column
- Main effects are row/column averages
- Interactions effects are difference relative to the row or column average

Interaction

Interaction: when the effect of one factor depends on the levels of another factor.

Lack of interaction

If lines are parallel, there is no interaction

Assessing interactions

In practice, the subgroup averages are unknown!

- Plot sample averages with confidence intervals or ± 1 standard error.
- But difficult to judge based on graphs alone.

Better to proceed with hypothesis tests.

Model formulation

Formulation of the two-way ANOVA

Two factors: A (complexity) and B (ending) with a and b levels.

Write the average response Y_{ijk} of the kth measurement in group (a_i, b_j) as

$$Y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$$

where

- Y_{ijk} is the kth replicate for ith level of factor A and jth level of factor B
- ε_{ijk} are independent error terms with mean zero and variance σ^2.

One average for each subgroup

Hypothesis tests

Interaction between factors A and B

$\mathscr{H}_0$: no interaction between factors A and B vs $\mathscr{H}_a$: there is an interaction

Main effect of factor A

$\mathscr{H}_0$: $\mu_{1.} = \cdots = \mu_{a.}$ vs $\mathscr{H}_a$: at least two marginal means of A are different

Main effect of factor B

$\mathscr{H}_0$: $\mu_{.1} = \cdots = \mu_{.b}$ vs $\mathscr{H}_a$: at least two marginal means of B are different.

Reparametrization

- Mean of A_i (average of row i):

$$\mu_{i.} = \frac{\mu_{i1} + \cdots + \mu_{ib}}{b}$$

- Mean of B_j (average of column j):

$$\mu_{.j} = \frac{\mu_{1j} + \cdots + \mu_{aj}}{a}$$

- Overall average:

$$\mu = \frac{\sum_{i=1}^{a} \sum_{j=1}^{b} \mu_{ij}}{ab}$$

Formulation of the two-way ANOVA

Write the model for a response variable Y in terms of two factors A_i, B_j.

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$$

with the parameters in the sum-to-zero constraints

- $\alpha_i = \mu_{i.} - \mu$
- mean of level A_i minus overall mean.
- $\beta_j = \mu_{.j} - \mu$
- mean of level B_j minus overall mean.
- $(\alpha\beta)_{ij} = \mu_{ij} - \mu_{i.} - \mu_{.j} + \mu$
- the interaction term for A_i and B_j.

Sum-to-zero parametrization

The model in terms of α, β and $(\alpha\beta)$ is overparametrized.

For the sum-to-zero constraint, impose that

$$\sum_{i=1}^{a} \alpha_i = 0, \quad \sum_{j=1}^{b} \beta_j = 0, \quad \sum_{j=1}^{b} (\alpha\beta)_{ij} = 0, \quad \sum_{i=1}^{a} (\alpha\beta)_{ij} = 0.$$

which imposes $1 + a + b$ constraints.

Analysis of variance table

term	degrees of freedom	mean square	F
A	$a-1$	$\mathsf{MS}_A = \mathsf{SS}_A/(a-1)$	$\mathsf{MS}_A/\mathsf{MS}_{\text{res}}$
B	$b-1$	$\mathsf{MS}_B = \mathsf{SS}_B/(b-1)$	$\mathsf{MS}_B/\mathsf{MS}_{\text{res}}$
AB	$(a-1)(b-1)$	$\mathsf{MS}_{AB} = \mathsf{SS}_{AB}/\{(a-1)(b-1)\}$	$\mathsf{MS}_{AB}/\mathsf{MS}_{\text{res}}$
residuals	$n-ab$	$\mathsf{MS}_{\text{res}} = \mathsf{SS}_{\text{res}}/(n-ab)$	
total	$n-1$		

Intuition behind degrees of freedom

Example with $a = 3, b = 5$.

A \ B	b_1	b_2	b_3	b_4	b_5	**sum**
a_1	A_1B_1	A_1B_2	A_1B_3	A_1B_4	x	A_1
a_2	A_2B_1	A_2B_2	A_2B_3	A_2B_4	x	A_2
a_3	x	x	x	x	x	x
sum	B_1	B_2	B_3	B_4	x	total

Terms with x are fully determined by row/column/total averages

Contrasts for the main effects

In the interaction model, we cast the main effect in terms of parameters.

Suppose the order of the coefficients for factor A (drug, 3 levels) and factor B (deprivation, 2 levels).

test	μ_{11}	μ_{12}	μ_{21}	μ_{22}	μ_{31}	μ_{32}
main effect A (1 vs 2)	1	1	-1	-1	0	0
main effect A (1 vs 3)	1	1	0	0	-1	-1
main effect B (1 vs 2)	1	-1	1	-1	1	-1
interaction AB (1 vs 2, 1 vs 2)	1	-1	-1	1	0	0
interaction AB (1 vs 3, 1 vs 2)	1	-1	0	0	-1	1

Testing hypothesis of interest

We only tests hypothesis that are of interest

- If there is a significant interaction, the marginal means are **not** of interest
- Rather, compute the simple effects.

Controlling the FWER

- What is the number of hypothesis of interest? Often, this is pairwise comparisons within each level of the other factor
- much less than $\binom{ab}{2}$ pairwise comparisons
- Scheffé's method for all custom contrasts still applicable, but may be conservative
- Tukey's method also continues to hold.
- Omnibus procedures for controlling the FWER (Holm-Bonferroni) may be more powerful than either Scheffé or Tukey's methods.

Multifactorial designs

Beyond two factors

We can consider multiple factors A, B, C, $\ldots$ with respectively a, b, c, $\ldots$ levels and with r replications for each.

The total number of treatment combinations is

$$a \times b \times c \times \cdots$$

Curse of dimensionality

Full three-way ANOVA model

Each cell of the cube is allowed to have a different mean

$$\underset{\text{response}}{Y_{ijkr}} = \underset{\text{cell mean}}{\mu_{ijk}} + \underset{\text{error}}{\varepsilon_{ijkr}}$$

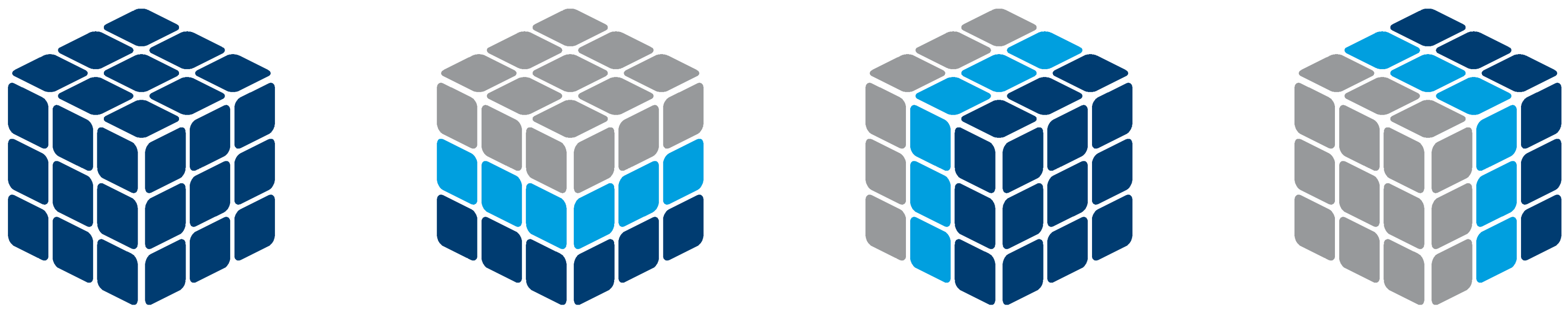
with ε_{ijkt} an independent $\mathsf{No}(0, \sigma^2)$ error term for

- row i
- column j
- depth k
- replication r

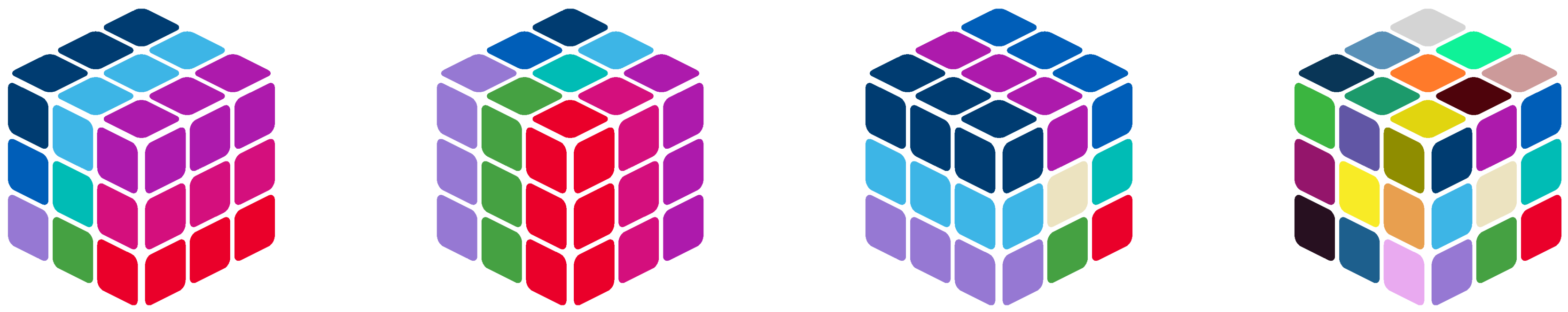
Parametrization of a three-way ANOVA model

With the **sum-to-zero** parametrization with factors A, B and C, write the response as

$$
\begin{aligned}
\underset{\text{theoretical average}}{\mathsf{E}(Y_{ijkr})} &= \underset{\text{global mean}}{\mu} \\
&\quad + \underbrace{\alpha_i + \beta_j + \gamma_k}_{\text{main effects}} \\
&\quad + \underbrace{(\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk}}_{\text{two-way interactions}} \\
&\quad + \underbrace{(\alpha\beta\gamma)_{ijk}}_{\text{three-way interaction}}
\end{aligned}
$$



global mean, row, column and depth main effects



row/col, row/depth and col/depth interactions and three-way interaction.

Example of three-way design

Petty, Cacioppo and Heesacker (1981). Effects of rhetorical questions on persuasion: A cognitive response analysis. Journal of Personality and Social Psychology.

A $2 \times 2 \times 2$ factorial design with 8 treatments groups and $n = 160$ undergraduates.

Setup: should a comprehensive exam be administered to bachelor students in their final year?

- **Response** Likert scale on -5 (do not agree at all) to 5 (completely agree)
- **Factors**
- A: strength of the argument (`strong` or `weak`)
- B: involvement of students `low` (far away, in a long time) or `high` (next year, at their university)
- C: style of argument, either `regular` form or `rhetorical` (Don't you think?, ...)

Interaction plot

Interaction plot for a $2 \times 2 \times 2$ factorial design from Petty, Cacioppo and Heesacker (1981)

The microwave popcorn experiment

What is the best brand of microwave popcorn?

- **Factors**
- brand (two national, one local)
- power: 500W and 600W
- time: 4, 4.5 and 5 minutes
- **Response**: ~~weight~~, ~~volume~~, ~~number~~, percentage of popped kernels.
- Pilot study showed average of 70% overall popped kernels (10% standard dev), timing values reasonable
- Power calculation suggested at least $r = 4$ replicates, but researchers proceeded with $r = 2$...

ANOVA | QQ-plot | R code | Interaction plot

```r
# Fit model with three-way interaction
model <- aov(percentage ~ brand*power*time,
data = popcorn)
# ANOVA table
anova_table <- anova(model)
# 'anova' is for balanced designs
# Check model assumptions
# plots + tests meaningless with r=2 replications per group...
# except for quantile-quantile plot
car::qqPlot(model, # points should be on straight line!
id = FALSE,
ylab = 'studentized residuals',
xlab = "Student-t quantiles")
```

ANOVA | QQ-plot | R code | Interaction plot

All points fall roughly on a straight line.

ANOVA QQ-plot **R code** Interaction plot

```
popcorn |>
   group_by(brand, time, power) |>
   summarize(mean_percentage = mean(percentage)) |>
ggplot(mapping = aes(x = power,
                     y = mean_percentage,
                     col = time,
                     group = time)) +
  geom_line() +
  facet_wrap(~brand) +
  labs(subtitle = "percentage of popped kernels",
       y = "",
       col = "time (min)",
       x = "power (W)") +
  theme_bw() +
```

ANOVA QQ-plot R code **Interaction plot**

No three-way interaction (hard to tell with $r = 2$ replications).

Analysis of variance table for balanced designs

terms	degrees of freedom
A	$a-1$
B	$b-1$
C	$c-1$
AB	$(a-1)(b-1)$
AC	$(a-1)(c-1)$
BC	$(b-1)(c-1)$
ABC	$(a-1)(b-1)(c-1)$
residual	$abc(r-1)$
total	$abcr-1$

Analysis of variance table for microwave-popcorn

	Degrees of freedom	Sum of squares	Mean square	F statistic	p-value
brand	2	331.10	165.55	1.89	0.180
power	1	455.11	455.11	5.19	0.035
time	2	1554.58	777.29	8.87	0.002
brand:power	2	196.04	98.02	1.12	0.349
brand:time	4	1433.86	358.46	4.09	0.016
power:time	2	47.71	23.85	0.27	0.765
brand:power:time	4	47.33	11.83	0.13	0.967
Residuals	18	1577.87	87.66		

Omitting terms in a factorial design

The more levels and factors, the more parameters to estimate (and replications needed)

- Costly to get enough observations / power
- The assumption of normality becomes more critical when $r = 2$!

It may be useful not to consider some interactions if they are known or (strongly) suspected not to be present

- If important interactions are omitted from the model, biased estimates/output!

Custom contrasts and marginal means

Guidelines for the interpretation of effects

Start with the most complicated term (top down)

- If the three-way interaction ABC is significative:
  - don't interpret main effects or two-way interactions!
  - comparison is done cell by cell within each level
- If the ABC term isn't significative:
  - can marginalize and interpret lower order terms

Analytical comparisons

- Preplanned
- *Post-hoc*: after seeing that the three-way interaction isn't significative, compare all pairwise differences within two-way.

What contrasts are of interest?

- Can view a three-way ANOVA as a series of one-way ANOVA or two-way ANOVAs…

Depending on the goal, could compare for variable A

- marginal contrast ψ_A (averaging over B and C)
- marginal conditional contrast for particular subgroup: ψ_A within c_1
- contrast involving two variables: ψ_{AB}
- contrast differences between treatment at $\psi_A \times B$, averaging over C.
- etc.

See helper code and chapter 22 of Keppel & Wickens (2004) for a detailed example.

Effects and contrasts for microwave-popcorn

Following preplanned comparisons

- Which combo (brand, power, time) gives highest popping rate? (pairwise comparisons of all combos)
- Best brand overall (marginal means marginalizing over power and time, assuming no interaction)
- Effect of time and power on percentage of popped kernels
- pairwise comparison of time $\times$ power
- main effect of power
- main effect of time

Preplanned comparisons using `emmeans`

Let A=brand, B=power, C=time

Compare difference between percentage of popped kernels for 4.5 versus 5 minutes, for brands 1 and 2

$$\mathscr{H}_0 : (\mu_{1.2} - \mu_{1.3}) - (\mu_{2.2} - \mu_{2.3}) = 0$$

```r
library(emmeans)
# marginal means
emm_popcorn_AC <- emmeans(model, specs = c("brand","time"))
contrast_list <- list(brand12with4.5vs5min = c(0, 0, 0, 1, -1, 0, -1, 1,0))
contrast(emm_popcorn_AC,  # marginal mean (no time)
 method = contrast_list) # list of contrasts
```

Preplanned comparisons

At level 99% with Tukey's method, compare all three times (4, 4.5 and 5 minutes)

- Careful! Potentially misleading because there is a `brand * time` interaction present.

```
# List of variables to keep go in `specs`: keep only time
emm_popcorn_C <- emmeans(model, specs = "time")
pairs(emm_popcorn_C, adjust = "tukey", level = 0.99, infer = TRUE)
```