ANOVA for two factor experiments

Session 6

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

Outline

Factorial designs and interactions

Model formulation

Effect size, contrasts and power

Factorial designs and interactions

Example from the OSC psychology replication

> People can be influenced by the prior consideration of a numerical anchor when forming numerical judgments. [...] The anchor provides an initial starting point from which estimates are adjusted, and a large body of research demonstrates that adjustment is usually insufficient, leading estimates to be biased towards the initial anchor.

Replication of Study 4a of Janiszewski & Uy (2008, Psychological Science) by J. Chandler

Motivating example

Consider a study on the retention of information to children age 4 to which we read a story two hours after the reading.

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.

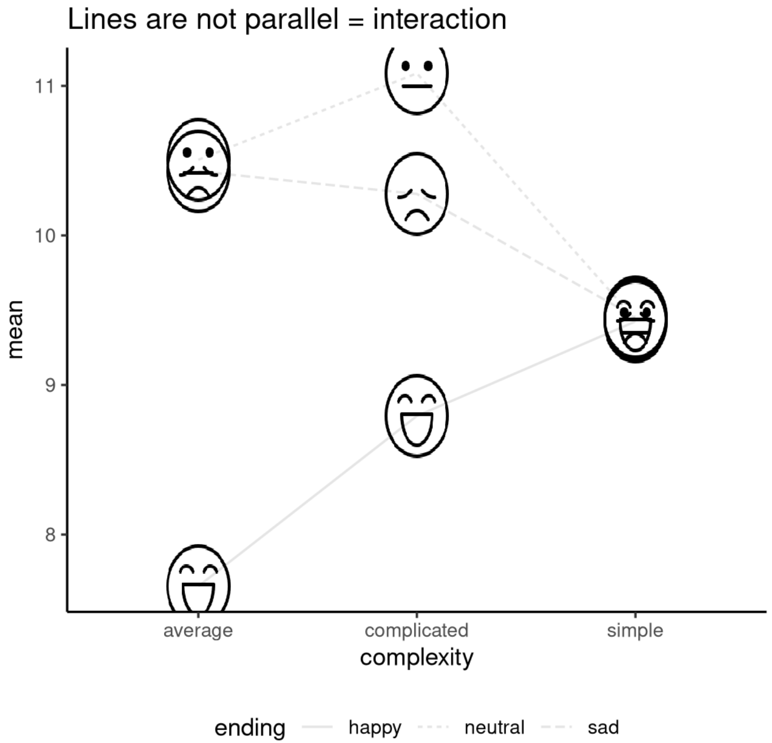
Factorial designs are more efficient: can study the impact of multiple variables simultaneously with **fewer overall observations**.

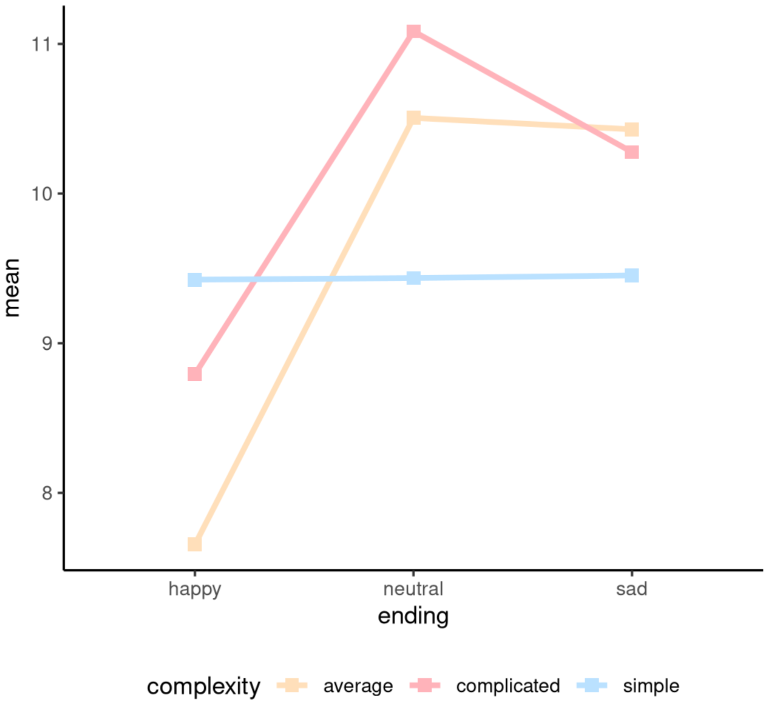
Estimates

- **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

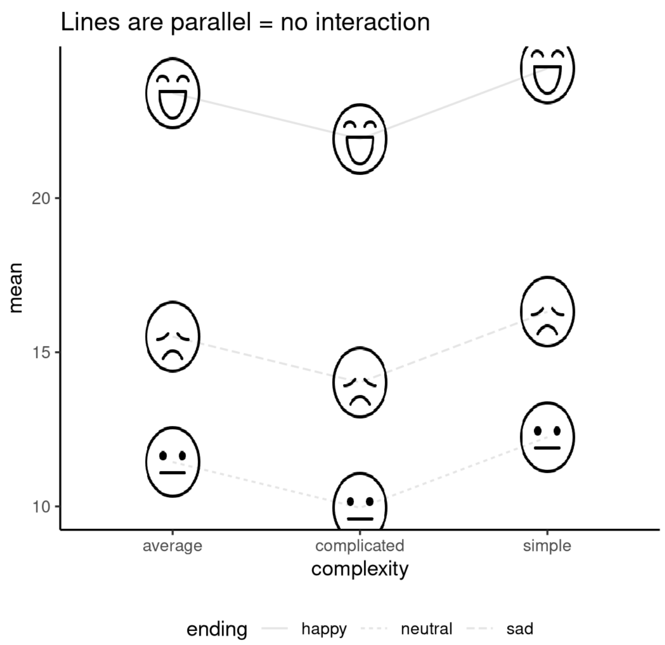
Interaction

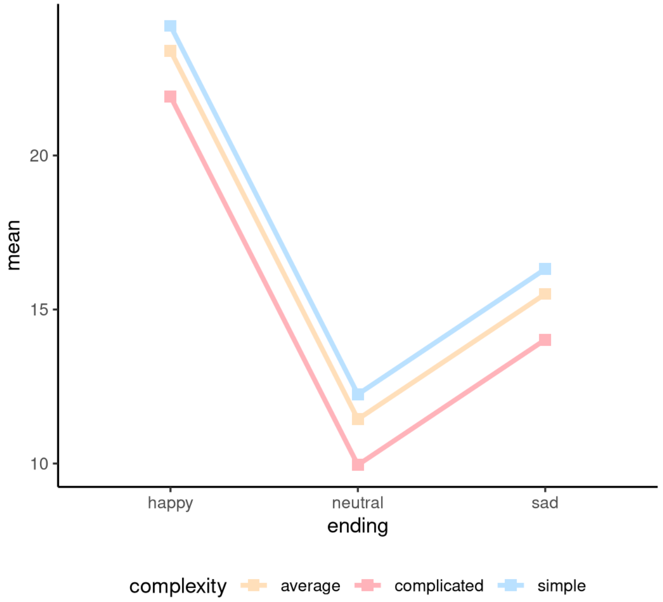
An interaction is present when the effect of one factor depends on the levels of another factor.





Lack of interaction





In practice, the sample average are uncertain!

- Plot averages with confidence intervals or ± 1 standard error.

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 the 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.

Two-way ANOVA model with interaction: 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

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

which imposes $1 + 1 + b + a$ constraints.

Breaking down the variability

For a **balanced** design, we can decompose the variability around sample means:

$$\begin{aligned}\mathsf{SS}_{\text{total}} &= \mathsf{SS}_{\text{model}} + \mathsf{SS}_{\mathsf{res}} \\ &= \mathsf{SS}_{A} + \mathsf{SS}_{B} + \mathsf{SS}_{AB} + \mathsf{SS}_{\mathsf{res}}\end{aligned}$$

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	$n-1$		

Factors are crossed, replicates are nested within AB groups

Example: Fiber strength

We consider a 5×3 factorial design (no interaction term).

Consider five levels of application of potash

- $T_1 = 36, T_2 = 54, T_3 = 72, T_4 = 108$ and $T_5 = 144$ lb K_2O per acre, applied to a cotton crop.

The response is a measure of single-fiber strength, an average of a number of tests on the cotton from each plot.

There were three blocks each containing five plots.

Data on fiber strength

Reordered data

block	T1	T2	T3	T4	T5
block 1	7.62	8.14	7.76	7.17	7.46
block 2	8.00	8.15	7.73	7.57	7.68
block 3	7.93	7.87	7.74	7.80	7.21

Residuals

block	T1	T2	T3	T4	T5
block 1	-0.14	0.18	0.11	-0.25	0.10
block 2	0.05	-0.01	-0.12	-0.05	0.13
block 3	0.09	-0.17	0.01	0.30	-0.23

Analysis of variance table

	sum of squares	df	F	p-value
treatment	0.732	4	4.192	0.040
block	0.097	2	1.112	0.375
Residuals	0.349	8		

Some pending questions

- Intuition behind degrees of freedom for the residuals?
- No interaction term (why?)

$A \setminus B$	b_1	b_2	b_3	b_4	b_5	**sum**
a_1	AB_{11}	AB_{12}	AB_{13}	AB_{14}	X	A_1
a_2	AB_{21}	AB_{22}	AB_{23}	AB_{24}	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

Example from Keppel and Wichern (table)

Consider errors by monkeys under three drug conditions (A) and two degrees of food deprivation (B)

Data for the 3×2 factorial design

A \ B	**1h deprivation**	**24h deprivation**
Control	1, 4, 0, 7	15, 6, 10, 13
Drug 1	13, 5, 7, 15	6, 18, 9, 15
Drug 2	9, 16, 18, 13	14, 7, 6, 13

R Demonstration

Effect size, contrasts and power

Noncentrality parameters

Consider a balanced design with $n = a \times b \times k$ observations

For the mean squared errors, the expected values are

$$\mathsf{E}(\mathsf{MS}_A) = \sigma^2 + \frac{bn}{a-1}\sum_{i=1}^{a}\alpha_i^2$$

$$\mathsf{E}(\mathsf{MS}_{AB}) = \sigma^2 + \frac{n}{(a-1)(b-1)}\sum_{j=1}^{b}\sum_{i=1}^{a}(\alpha\beta)_{ij}^2$$

Under the null hypothesis of no mean effect / interaction, these are thus unbiased estimators of the error variance.

Effect size

- We can report an estimate of the effect size for either of the main effects, for the interaction or overall.
- For a power calculation, do the calculations with each effect (whose size is of **scientific interest** and select the largest sample size per group.
- Given F statistics and degrees of freedom, we can find different measures: Cohen's f, (partial) η^2 and ω^2.
- Check the book for formulae and `effectsize` for estimates in **R**.
- Estimators of variability are noisy and biased

(Partial) ω^2

The proportion of variance explained by the effect T is

$$\omega^2_{\text{effect}} = \frac{\sigma^2_{\text{effect}}}{\sigma^2_{\text{total}}}.$$

where $\sigma^2_{\text{total}} = \sigma^2_A + \sigma^2_B + \sigma^2_{AB} + \sigma^2_{\text{error}}$.

The partial ω^2 for an effect is

$$\omega^2_{\langle\text{effect}\rangle} = \frac{\sigma^2_{\text{effect}}}{\sigma^2_{\text{effect}} + \sigma^2_{\text{resid}}}.$$

These get replaced by estimates based on F statistics and degrees of freedom (see Keppel & Wickens, p. 233).

The **R** package `effectsize` reports estimates with confidence intervals

One way for the two-way

We can cast the two-way model with an interaction as a one-way ANOVA with ab levels.

- Sometimes useful for using custom contrasts
- Used for some procedures that do not support two-way designs (unequal variance model) or Levene's test

Contrasts for the main effects

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

Say the order of the coefficients is drug (A, 3 levels) and deprivation (B, 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.
- 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 (or generalization thereof)
- Omnibus procedures for controlling the FWER (Holm-Bonferroni) may be more powerful.