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

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.

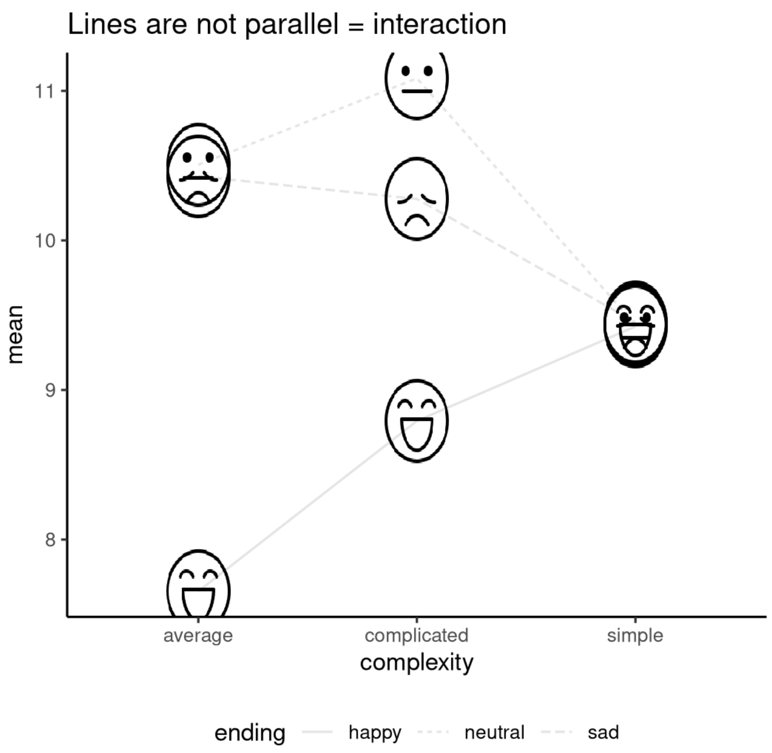
Factorial designs are more efficient: can study the impact of multiple variables simultaneously with **fewer overall observations**.

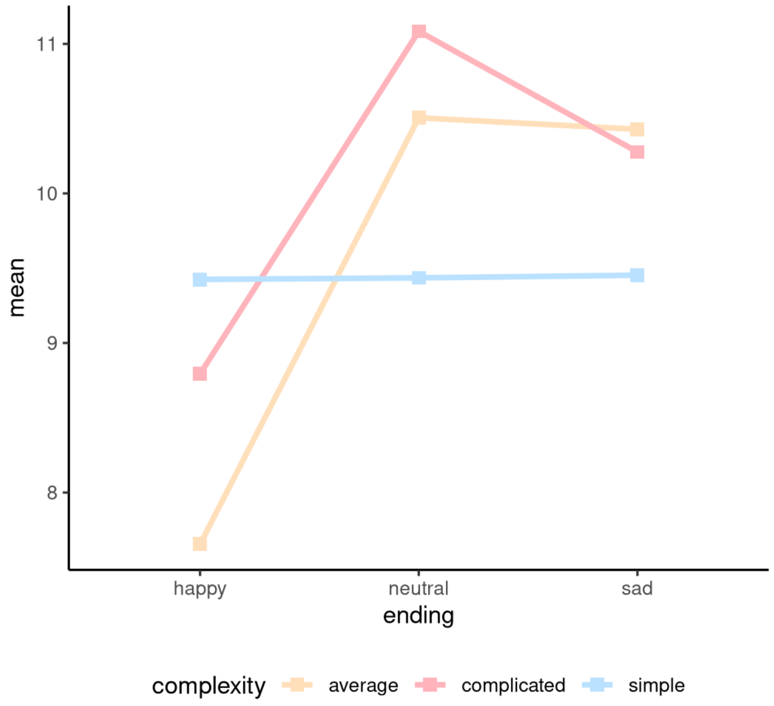
Estimates

- **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

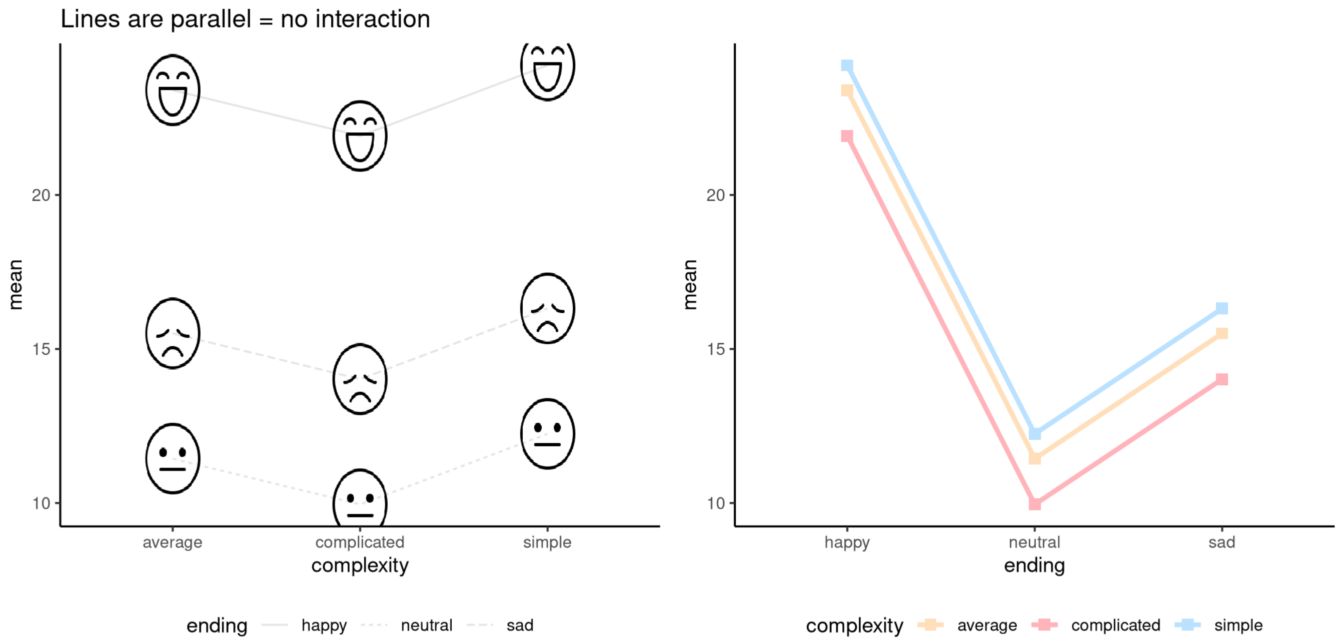
Interaction

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

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

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	$n-1$		

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

Effect size, contrasts and power

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.

Breaking down the variability

We can express the overall variability of the response around the global mean as

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

where $\sigma^2_A = a^{-1}\sum_{i=1}^a \alpha_i^2$, $\sigma^2_{AB} = (ab)^{-1}\sum_{i=1}^a\sum_{j=1}^b(\alpha\beta)^2_{ij}$, etc.

Partial Cohen's f

The **population** version of Cohen's partial f measures the proportion of variability that is explained by a main effect or an interaction, e.g.,

$$f_{\langle A \rangle} = \frac{\sigma^2_A}{\sigma^2_{\text{resid}}}, \qquad f_{\langle AB \rangle} = \frac{\sigma^2_{AB}}{\sigma^2_{\text{resid}}}.$$

These variance quantities are **unknown**, so need to be estimated somehow.

Partial effect size (variance)

Effect size are often reported in terms of variability via the ratio

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

- Both $\hat{\eta}^2_{\langle \text{effect} \rangle}$ (aka $\widehat{R}^2_{\langle \text{effect} \rangle}$) and $\widehat{\omega}^2_{\langle \text{effect} \rangle}$ are estimators of this quantity and obtained from the F statistic and degrees of freedom of the effect.
- $\widehat{\omega}^2_{\langle \text{effect} \rangle}$ is less biased than $\hat{\eta}^2_{\langle \text{effect} \rangle}$.

Estimation of partial ω^2

$$\widehat{\omega}^2_{\langle\text{effect}\rangle} = \frac{\text{df}_{\text{effect}}(F_{\text{effect}} - 1)}{\text{df}_{\text{effect}}(F_{\text{effect}} - 1) + n},$$

where n is the overall sample size.

In **R**, `effectsize::omega_squared` reports these estimates with one-sided confidence intervals.

Reference for confidence intervals: Steiger (2004), Psychological Methods

Converting ω^2 to Cohen's f

Given an estimation of $\eta^2_{\langle\text{effect}\rangle}$, convert it into an estimate of Cohen's partial $f^2_{\langle\text{effect}\rangle}$, e.g.,

$$\widehat{f}^2_{\langle\text{effect}\rangle} = \frac{\widehat{\omega}^2_{\langle\text{effect}\rangle}}{1 - \widehat{\omega}^2_{\langle\text{effect}\rangle}}.$$

Note that `effectsize::cohens_f` returns $\tilde{f}^2 = n^{-1} F_{\text{effect}} \text{df}_{\text{effect}}$, a transformation of $\hat{\eta}^2_{\langle\text{effect}\rangle}$.

Power for factorial experiments

- G*Power and **R** packages take Cohen's f (or f^2) as inputs.
- Calculation based on F distribution with
  - $\nu_1 = \mathrm{df}_{\mathrm{effect}}$ degrees of freedom
  - $\nu_2 = n - n_g$, where n_g is the number of mean parameters estimated.
  - noncentrality parameter $\phi = nf^2_{\langle\mathrm{effect}\rangle}$ where effect is either A, B or AB.

Example: if `a` and `b` denote the number of levels of each factor and `omega.sq` is $\omega^2_{\langle AB\rangle}$, then the sample size needed to detect the interaction AB with 80% power is

```
fhat <- sqrt(omega.sq/(1-omega.sq))
WebPower::wp.kanova(power = 0.8, f = fhat, ndf = (a-1)*(b-1), ng = ab)
```

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.