Complete factorial designs

Session 5

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

Outline

Factorial designs and interactions

Model formulation

Multifactorial designs

Factorial designs and interactions

Complete factorial designs?

Factorial design

study with multiple factors (subgroups)

Complete

Gather observations for every subgroup

Motivating example

Response:
retention of information
two hours after reading a story

Population:
children aged four

experimental factor 1:
ending (happy or sad)

experimental factor 2:
complexity (easy, average or hard).

Setup of design

complexity	happy	sad
complicated		
average		
simple		

Factors are crossed

Efficiency of factorial design

Cast problem
as a series of one-way ANOVA
vs simultaneous estimation

Factorial designs requires
fewer overall observations

Can study interactions

Interaction

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

Effect together
$\neq$
sum of individual effects

Interaction or profile plot

Graphical display:
plot sample mean per category

with uncertainty measure
(1 std. error for mean
confidence interval, etc.)

Interaction: lines are not parallel

No interaction: parallel lines

Interaction plot for 2 by 2 design

Model formulation

Formulation of the two-way ANOVA

Two factors: A (complexity) and B (ending) with n_a and n_b levels.

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

$$\underbrace{Y_{ijr}}_{\text{response}} = \underbrace{\mu_{ij}}_{\text{subgroup mean}} + \underbrace{\varepsilon_{ijr}}_{\text{error term}}$$

where

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

One average for each subgroup

B **ending** A **complexity**	b_1 **(happy)**	b_2 **(sad)**	***row mean***
a_1 (complicated)	μ_{11}	μ_{12}	$\mu_{1.}$
a_2 (average)	μ_{21}	μ_{22}	$\mu_{2.}$
a_3 (simple)	μ_{31}	μ_{32}	$\mu_{3.}$
column mean	$\mu_{.1}$	$\mu_{.2}$	μ

Row, column and overall average

- Mean of A_i (average of row i):

$$\mu_{i.} = \frac{\mu_{i1} + \cdots + \mu_{in_b}}{n_b}$$

- Mean of B_j (average of column j):

$$\mu_{.j} = \frac{\mu_{1j} + \cdots + \mu_{n_aj}}{n_a}$$

- Overall average (overall all rows and columns):

$$\mu = \frac{\sum_{i=1}^{n_a} \sum_{j=1}^{n_b} \mu_{ij}}{n_a n_b}$$

Vocabulary of effects

Definitions

- **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 sad ending)
- **interaction effects**: when simple effects differ depending on levels of another factor

What it means relative to the table

- **simple effects** are comparisons between cell averages within a given row or column
- **main effects** are comparisons between row or column averages
- **interaction effects** are difference relative to the row or column average

Contrasts

Suppose the order of the coefficients is factor A (complexity, 3 levels, complicated/average/easy) and factor B (ending, 2 levels, happy/sad).

test	μ_{11}	μ_{12}	μ_{21}	μ_{22}	μ_{31}	μ_{32}
main effect A (complicated vs average)	1	1	-1	-1	0	0
main effect A (complicated vs simple)	1	1	0	0	-1	-1
main effect B (happy vs sad)	1	-1	1	-1	1	-1
interaction AB (comp. vs av, happy vs sad)	1	-1	-1	1	0	0
interaction AB (comp. vs easy, happy vs sad)	1	-1	0	0	-1	1

Hypothesis tests for main effects

Generally, need to compare multiple effects at once

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.

Equivalent formulation of the two-way ANOVA

Write the model for a response variable Y in terms of two factors a_i, b_j.

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

where

- $\alpha_i = \mu_{i.} - \mu$
  - 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.

One average for each subgroup

B ending / A complexity	b_1 **(happy)**	b_2 **(sad)**	***row mean***
a_1 (complicated)	$\mu + \alpha_1 + \beta_1 + (\alpha\beta)_{11}$	$\mu + \alpha_1 + \beta_2 + (\alpha\beta)_{12}$	$\mu + \alpha_1$
a_2 (average)	$\mu + \alpha_2 + \beta_1 + (\alpha\beta)_{21}$	$\mu + \alpha_2 + \beta_2 + (\alpha\beta)_{22}$	$\mu + \alpha_2$
a_3 (simple)	$\mu + \alpha_3 + \beta_1 + (\alpha\beta)_{31}$	$\mu + \alpha_3 + \beta_2 + (\alpha\beta)_{32}$	$\mu + \alpha_3$
column mean	$\mu + \beta_1$	$\mu + \beta_2$	μ

More parameters than data cells!

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

Sum-to-zero parametrization

Too many parameters!

Impose sum to zero constraints

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

which imposes $1 + n_a + n_b$ constraints.

Why use the sum to zero parametrization?

- Testing for main effect of A:

$$\mathscr{H}_0 : \alpha_1 = \cdots = \alpha_{n_a} = 0$$

- Testing for main effect of B:

$$\mathscr{H}_0 : \beta_1 = \cdots = \beta_{n_b} = 0$$

- Testing for interaction between A and B:

$$\mathscr{H}_0 : (\alpha\beta)_{11} = \cdots = (\alpha\beta)_{n_a n_b} = 0$$

In all cases, alternative is that at least two coefficients are different.

Seeking balance

Balanced sample
(equal nb of obs per group)

With n_r replications per subgroup, total sample size is $n = n_a n_b n_r$.

Why balanced design?

With equal variance, this is the optimal allocation of treatment unit.

maximize power

Estimated means for main and total effects correspond to marginal averages.

equiweighting

Unambiguous decomposition of effects of A, B and interaction.

orthogonality

Rewriting observations

$$
\begin{aligned}
\underbrace{(y_{ijr} - \widehat{\mu})}_{\text{obs vs grand mean}} &= \underbrace{(\widehat{\mu}_{i.} - \widehat{\mu})}_{\text{row mean vs grand mean}} \\
&+ \underbrace{(\widehat{\mu}_{.j} - \widehat{\mu})}_{\text{col mean vs grand mean}} \\
&+ \underbrace{(\widehat{\mu}_{ij} - \widehat{\mu}_{i.} - \widehat{\mu}_{.j} + \widehat{\mu})}_{\text{cell mean vs additive effect}} \\
&+ \underbrace{(y_{ijr} - \widehat{\mu}_{ij})}_{\text{obs vs cell mean}}
\end{aligned}
$$

Decomposing variability

Constructing statistics as before by decomposing variability into blocks.

We can square both sides and sum over all observations.

With balanced design, all cross terms cancel, leaving us with the **sum of square** decomposition

$$\mathsf{SS}_{\text{total}} = \mathsf{SS}_{A} + \mathsf{SS}_{B} + \mathsf{SS}_{AB} + \mathsf{SS}_{\text{resid}}.$$

Sum of square decomposition

The sum of square decomposition

$$\mathsf{SS}_{\text{total}} = \mathsf{SS}_A + \mathsf{SS}_B + \mathsf{SS}_{AB} + \mathsf{SS}_{\text{resid}}.$$

is an estimator of the population variance decomposition

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

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

Take ratio of variability (effect relative to residual) and standardize numerator and denominator to build an F statistic.

Analysis of variance table

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

Read the table backward (starting with the interaction).

- If there is a significant interaction, the main effects are **not** of interest and potentially misleading.

Intuition behind degrees of freedom

B **ending** / A **complexity**	b_1 (**happy**)	b_2 (**sad**)	***row mean***
a_1 (complicated)	μ_{11}	X	$\mu_{1.}$
a_2 (average)	μ_{21}	X	$\mu_{2.}$
a_3 (simple)	X	X	X
column mean	$\mu_{.1}$	X	μ

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

Multiplicity correction

With equal sample size and equal variance, usual recipes for ANOVA hold.

Correction depends on the effect: e.g., for factor A, the critical values are

- Bonferroni: $1-\alpha/(2m)$ quantile of $\mathsf{St}(n-n_an_b)$
- Tukey: Studentized range (`qtukey`)
  - level $1-\alpha/2$, n_a groups, $n-n_an_b$ degrees of freedom.
- Scheffé: critical value is $\{(a-1)\mathfrak{f}_{1-\alpha}\}^{1/2}$
  - $\mathfrak{f}_{1-\alpha}$ is $1-\alpha$ quantile of $\mathsf{F}(\nu_1=n_a-1, \nu_2=n-n_an_b)$.

Software implementations available in `emmeans` in **R**.

Numerical example

Multifactorial designs

Beyond two factors

We can consider multiple factors $A, B, C, \ldots$ with respectively $n_a, n_b, n_c, \ldots$ levels and with n_r replications for each.

The total number of treatment combinations is

$$n_a \times n_b \times n_c \times \cdots$$

Curse of dimensionality

Full three-way ANOVA model

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

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

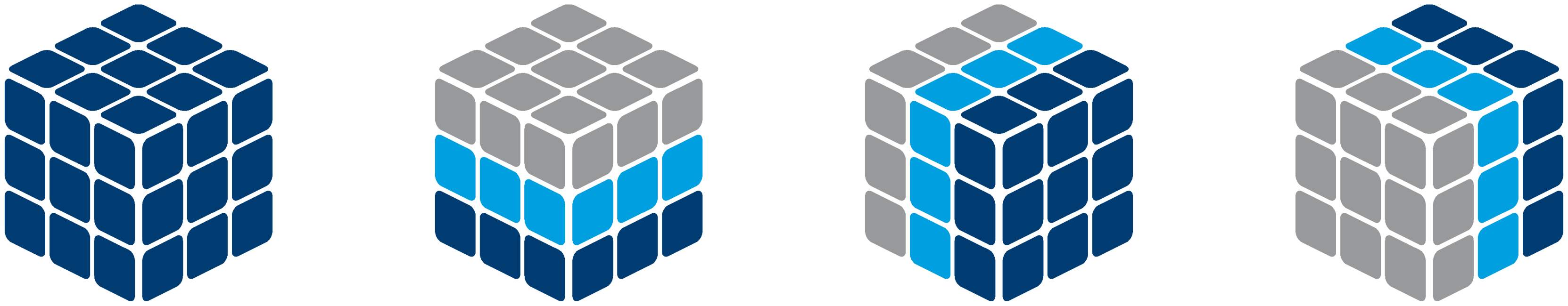
with ε_{ijkt} are independent error term for

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

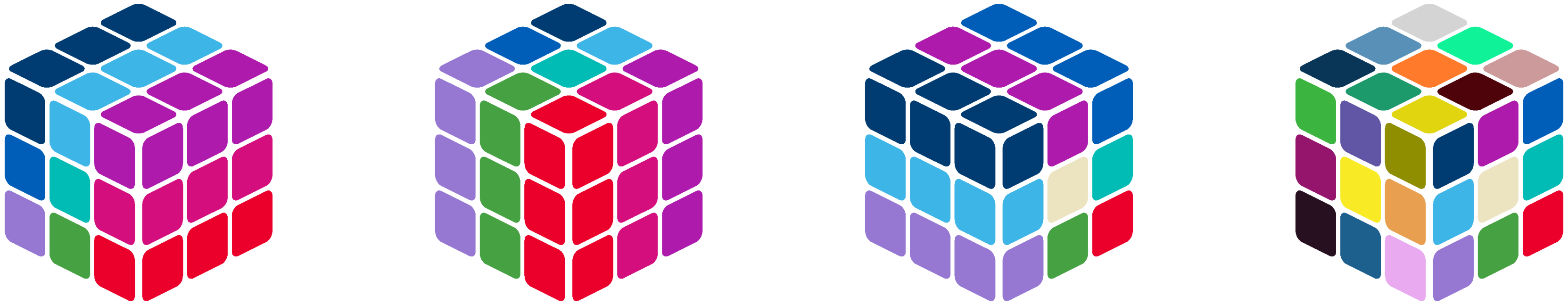
Parametrization of a three-way ANOVA model

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

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



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.

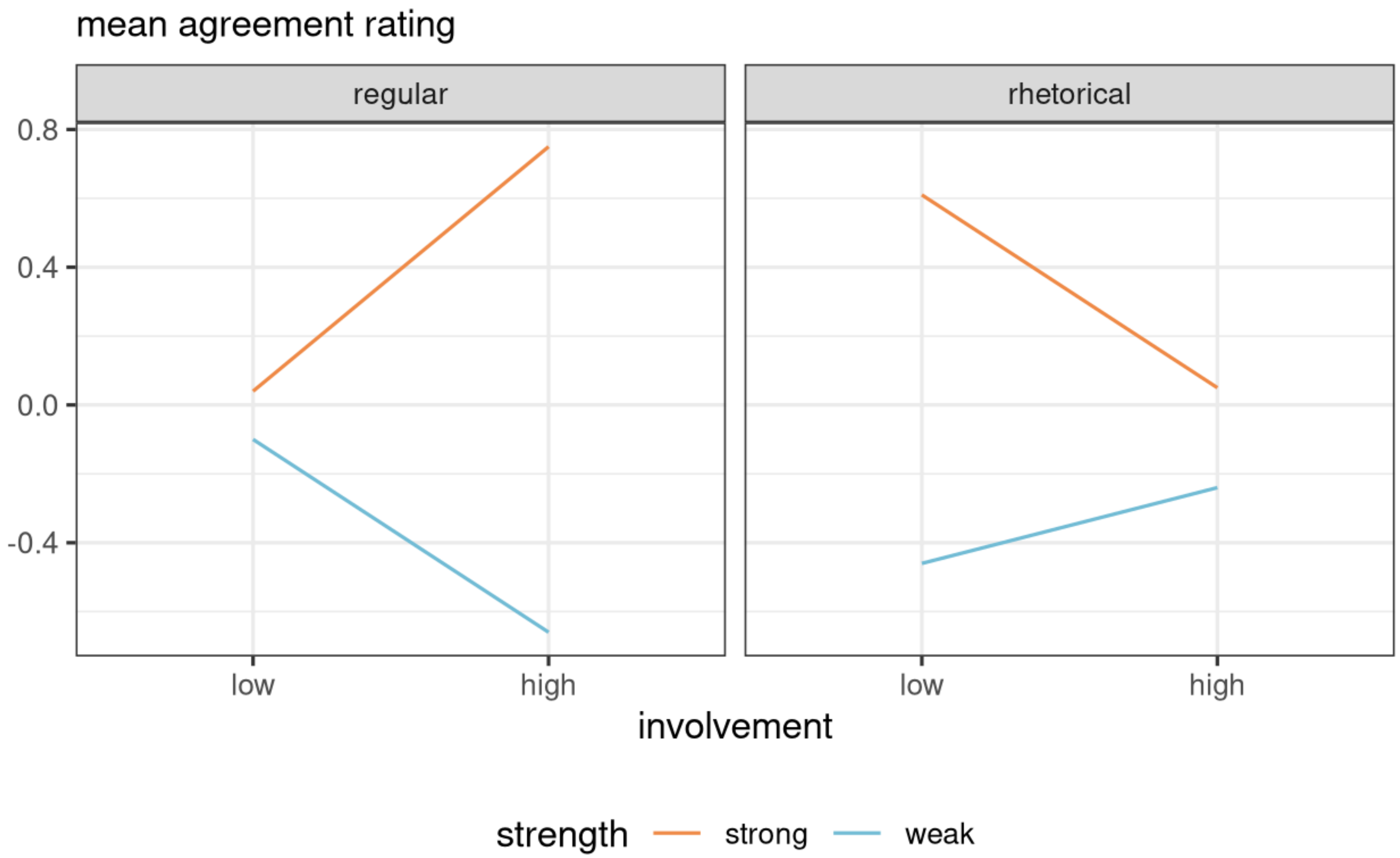
A $2 \times 2 \times 2$ factorial design with 8 treatments groups and $n = 160$ undergraduates.

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?

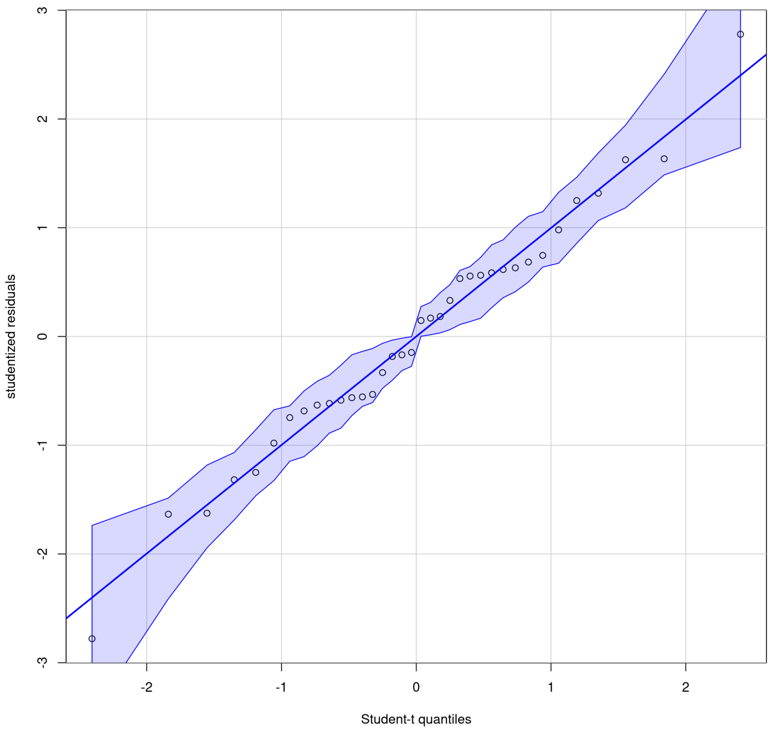
- **Factors**
- brand (two national, one local)
- power: 500W and 600W
- time: 4, 4.5 and 5 minutes
- **Response**: ~~weight~~, ~~volume~~, ~~number~~, percentage of popped kernels.
- Pilot study showed average of 70% overall popped kernels (10% standard dev), timing values reasonable
- Power calculation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

ANOVA | QQ-plot | R code | Interaction plot

```r
data(popcorn, package = 'hecedsm')
# Fit model with three-way interaction
model <- aov(percentage ~ brand*power*time,
             data = popcorn)
# ANOVA table - 'anova' is ONLY for balanced designs
anova_table <- anova(model)
# Quantile-quantile plot
car::qqPlot(model)
```

Model assumptions: plots and tests are meaningless with `(n_r=2)` replications per group...

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

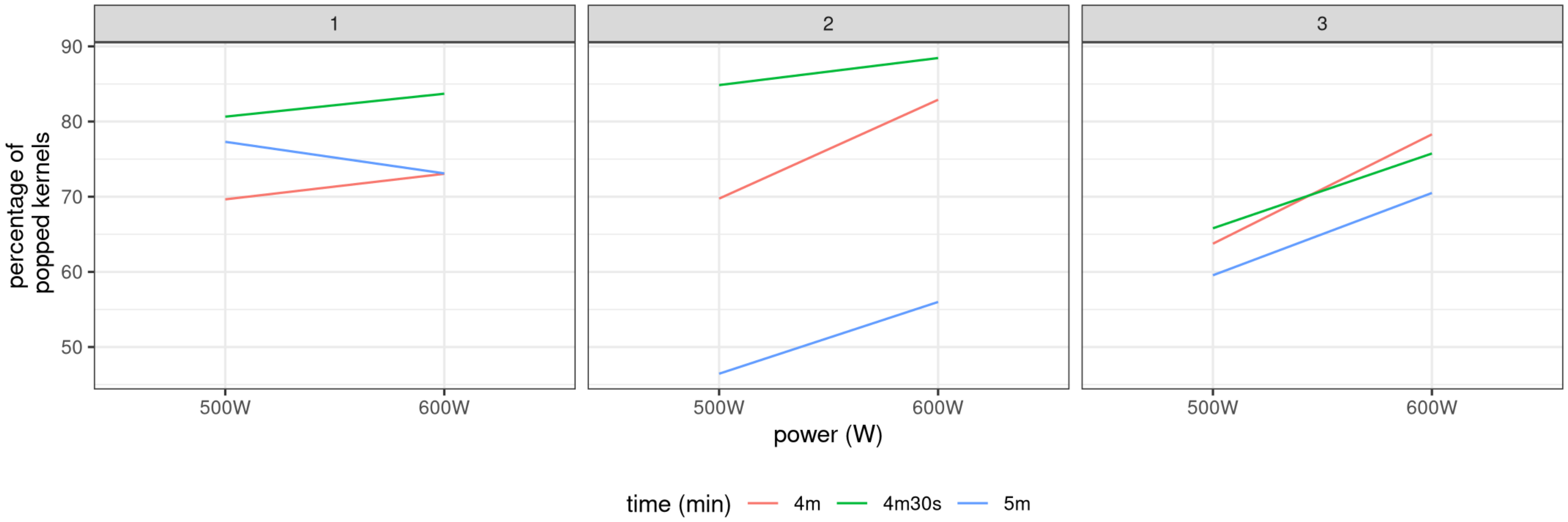


All points fall roughly on a straight line.

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

```
popcorn |>
   group_by(brand, time, power) |>
   summarize(meanp = mean(percentage)) |>
ggplot(mapping = aes(x = power,
                     y = meanp,
                     col = time,
                     group = time)) +
  geom_line() +
  facet_wrap(~brand)
```

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



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

Analysis of variance table for balanced designs

terms	degrees of freedom
A	$n_a - 1$
B	$n_b - 1$
C	$n_c - 1$
AB	$(n_a - 1)(n_b - 1)$
AC	$(n_a - 1)(n_c - 1)$
BC	$(n_b - 1)(n_c - 1)$
ABC	$(n_a - 1)(n_b - 1)(n_c - 1)$
residual	$n_a n_b n_c (R - 1)$
total	$n_a n_b n_c n_r - 1$

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!

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
  - back to a series of two-way ANOVAs

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

Compare all three times (4, 4.5 and 5 minutes)

At level 99% with Tukey's HSD method

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

```r
# 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)
```