

# General factorial designs

## **Session 7**

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

# Outline

**Blocking**

**More complex factorial designs**

**Custom contrasts and marginal means**

# Blocking

# Terminology for *nuisance*

## Block

Source of variation, but  
of no interest  
known and controllable

## Example

timing  
lab technician  
machine

## Covariates

Explanatory measured  
**before** the experiment  
Cannot be acted upon

## Example

socioeconomic  
variables  
environmental  
conditions

## Noise factor

Under which setting is  
response least  
affected?

## Example

temperature  
processing

# Why blocking?

Design experiment to reduce the effect of uncontrolled variations

In general, increases the power of the  $F$  test for treatment effects.

Group units in sets as alike as possible.

(Often) compare only treatments, so interactions are not included.

# Assignment to treatment

**Divide subjects within each block**

**Randomly allocate to treatment within block**

**(stratified sampling)**

# Block-treatment design

Without interaction,

$$\underset{\text{response}}{Y_{ij}} = \mu + \underset{\text{treatment}}{\alpha_i} + \underset{\text{blocking}}{\beta_j} + \underset{\text{error}}{\varepsilon_{ij}}$$

Compromise between

- reduced variability for residuals
- loss of degrees of freedom due to estimation of  $\beta$ 's.

# Example: Resting metabolic rate

From Dean, Voss and Draguljić (2017), Example 10.4.1 (p. 311)

experiment that was run to compare the effects of inpatient and outpatient protocols on the in-laboratory measurement of resting metabolic rate (RMR) in humans. A previous study had indicated measurements of RMR on elderly individuals to be 8% higher using an outpatient protocol than with an inpatient protocol. If the measurements depend on the protocol, then comparison of the results of studies conducted by different laboratories using different protocols would be difficult. The experimenters hoped to conclude that the effect on RMR of different protocols was negligible.



---

Clean

---

Fit

Plot

```
library(tidyverse)
url <- "https://edsm.rbind.io/data/resting_metabolic_rate.txt"
# Load data, transform integers to factors (categorical)
resting <- read.table(url, header = TRUE) %>%
  mutate(subject = factor(subject), #blocking factor
         protocol = factor(protocol),
         rate = rate/1000)
```

Clean

Fit

Plot

```
# Force sum-to-zero parametrization for unordered factors
options(contrasts = c("contr.sum", "contr.poly"))
# Fit model with blocking
model_block <- lm(rate ~ subject + protocol, data = resting)
# One-way ANOVA (no blocking)
model_raw <- lm(rate ~ protocol, data = resting)
# anova(model_block)
# anova(model_raw)
```

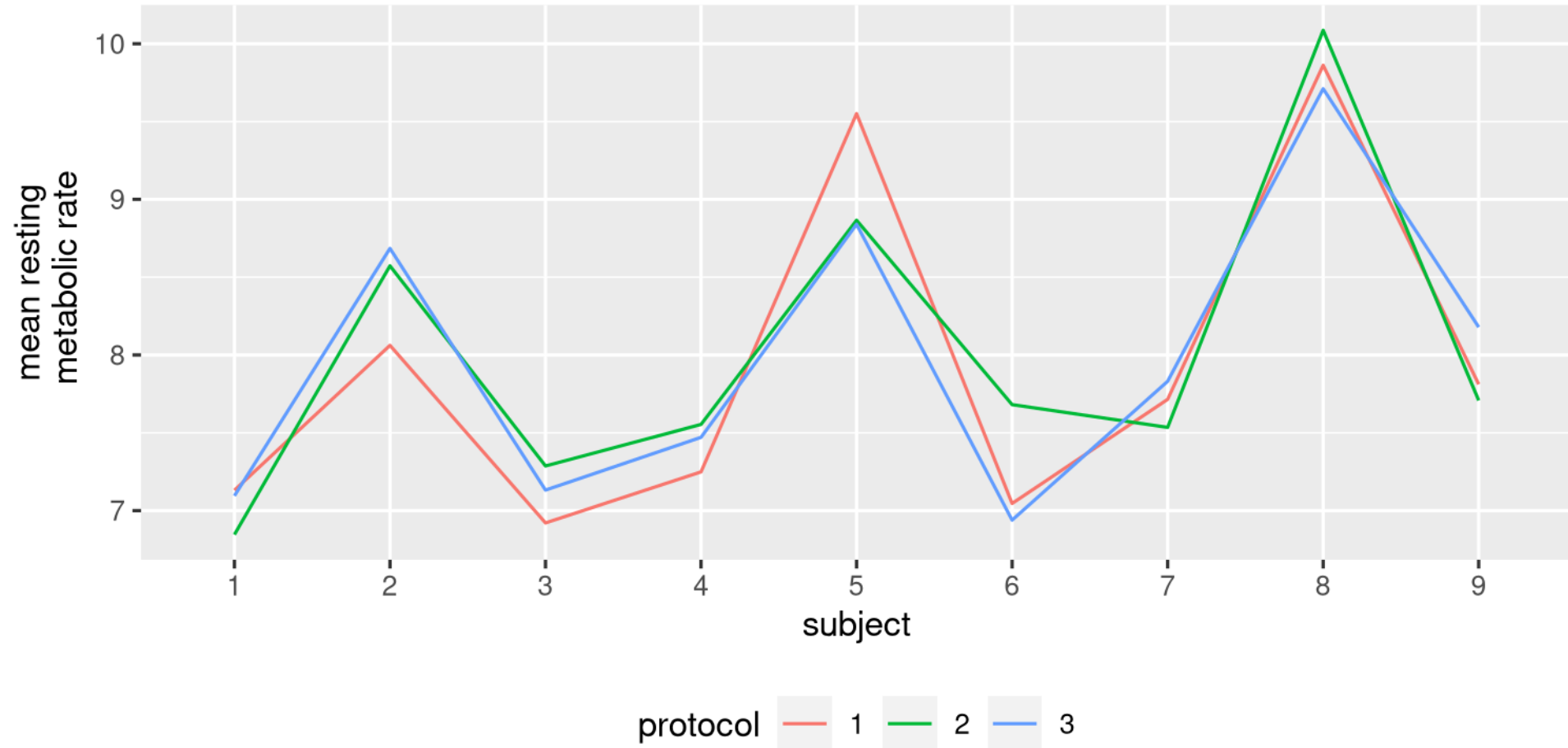
Clean

Fit

Plot

```
ggplot(data = resting,  
       aes(x = subject,  
           y = rate,  
           group = protocol,  
           color = protocol)) +  
  geom_line() +  
  labs(y = "mean resting\n metabolic rate") +  
  theme(legend.position = "bottom")
```

# Interaction plot



# Impact of blocking

ANOVA table (with blocking)

ANOVA table (without blocking)

Analysis of variance table - with blocking

	<b>Degrees of freedom</b>	<b>Sum of squares</b>	<b>Mean square</b>	<b>F statistic</b>	<b>p-value</b>
subject	8	23.12	2.89	37.42	0.000
protocol	2	0.04	0.02	0.23	0.795
Residuals	16	1.24	0.08		

# Impact of blocking

ANOVA table (with blocking)

ANOVA table (without blocking)

Analysis of variance table - without blocking

	<b>Degrees of freedom</b>	<b>Sum of squares</b>	<b>Mean square</b>	<b>F statistic</b>	<b>p-value</b>
protocol	2	0.04	0.02	0.02	0.982
Residuals	24	24.35	1.01		

# Multifactorial designs

# Beyond two factors

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

The total number of treatment combinations is

$$a \times b \times c \times \dots$$

**Curse of dimensionality**



# Full three-way ANOVA model

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

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

with  $\varepsilon_{ijk r}$  an independent  $\text{No}(0, \sigma^2)$  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

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



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.

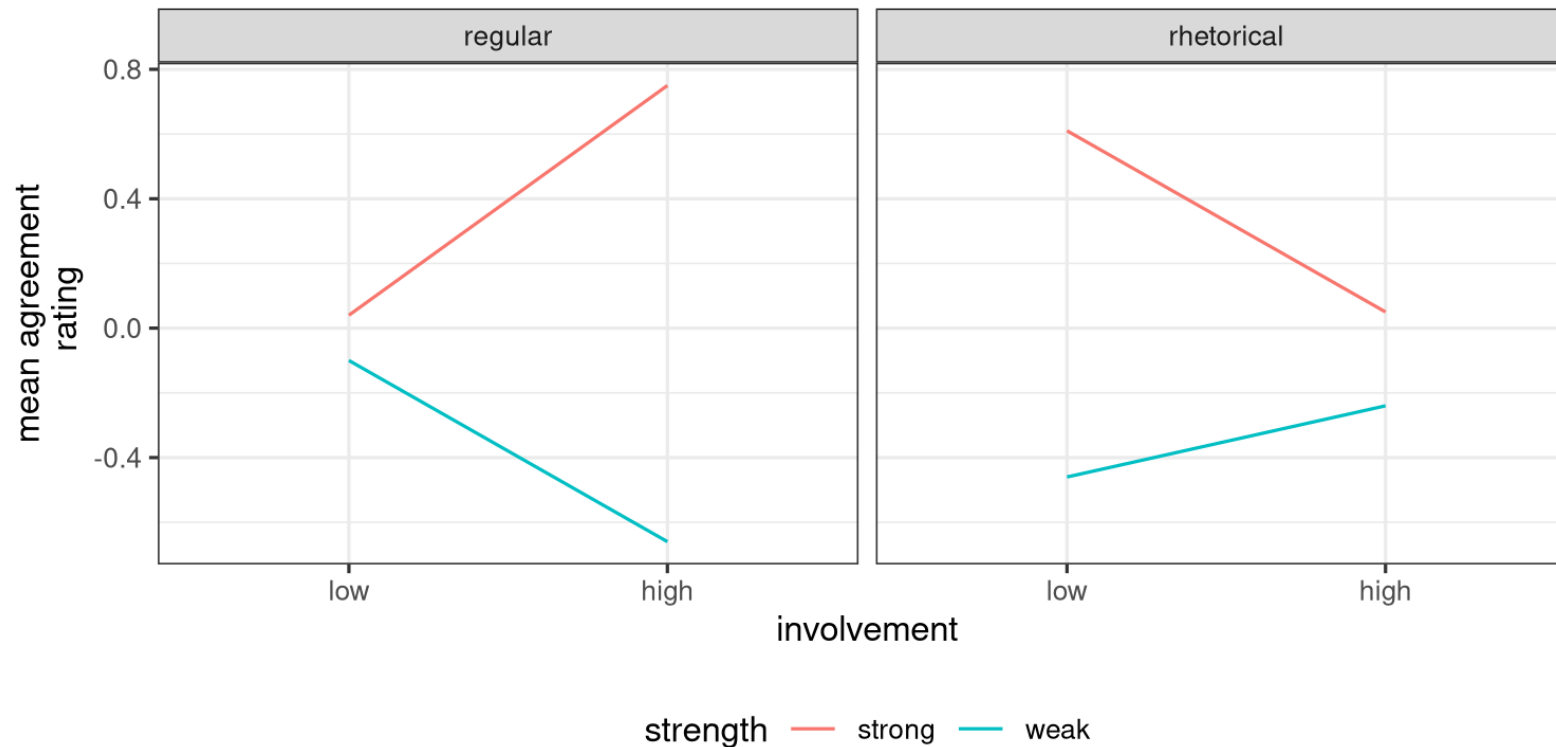
A  $2 \times 2 \times 2$  factorial design with 8 treatments groups and  $n = 160$  undergraduates.

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 ( $_{\text{strong}}$  or  $_{\text{weak}}$ )
  - $B$ : involvement of students  $_{\text{low}}$  (far away, in a long time) or  $_{\text{high}}$  (next year, at their university)
  - $C$ : style of argument, either  $_{\text{regular form}}$  or  $_{\text{rhetorical}}$  (Don't you think?, ...)

# Interaction plot

An 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  $r = 2...$

```
library(tidyverse) #load packages
# Sum-to-zero parametrization
options(contrasts = rep("contr.sum", 2))
url <- "https://edsm.rbind.io/data/popcorn.txt"
popcorn <- read.table(url,
                      header = TRUE,
                      stringsAsFactor = TRUE) %>%
  mutate(brand = factor(brand),
         power = factor(power, labels = c("500", "600")),
         time = factor(time, labels = c("4", "4.5", "5")))
```

Data cleaning

**ANOVA**

QQ-plot

R code

Interaction plot

```
# Fit model with three-way interaction
model <- lm(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')
```



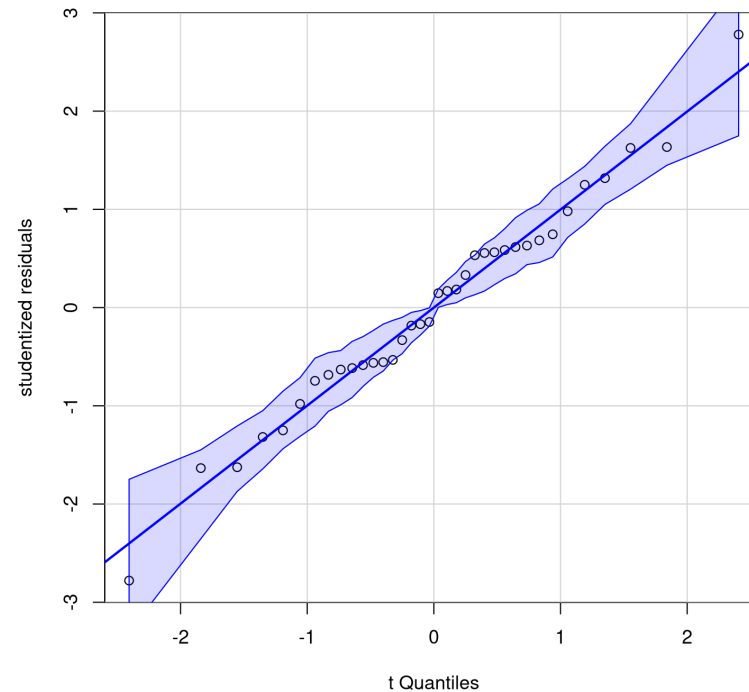
Data cleaning

ANOVA

QQ-plot

R code

Interaction plot



All points fall roughly on a straight line.

```
popcorn %>%  
  group_by(brand, time, power) %>%  
  summarize(mean_percentage = mean(percentage)) %>%  
  ggplot(aes(x = power,  
             y = mean_percentage,  
             col = time,  
             group = time)) +  
  geom_line() +  
  facet_wrap(~brand) +  
  labs(y = "percentage of\n popped kernels",  
       col = "time (min)",  
       x = "power (W)") +  
  theme_bw() +  
  theme(legend.position = "bottom")
```

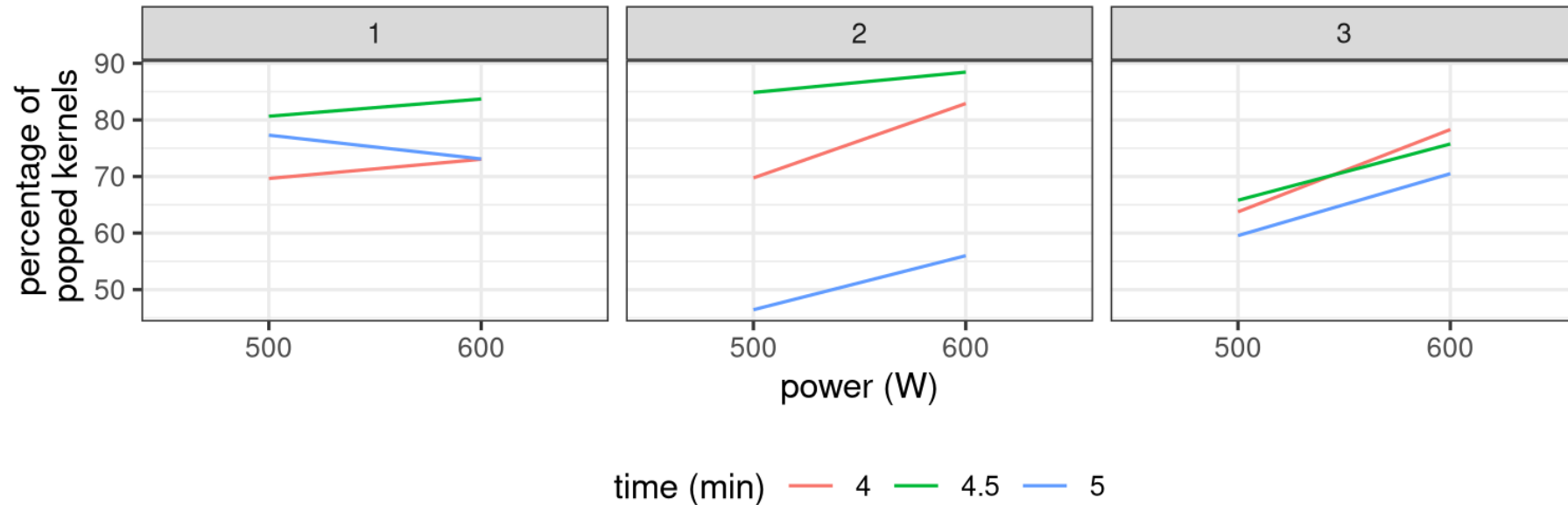
Data cleaning

ANOVA

QQ-plot

R code

Interaction plot



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

# Analysis of variance table

terms	degrees of freedom	sum of squares
$A$	$a - 1$	$[A] - [T]$
$B$	$b - 1$	$[B] - [T]$
$C$	$c - 1$	$[C] - [T]$
$AB$	$(a - 1)(b - 1)$	$[AB] - [A] - [B] + [T]$
$AC$	$(a - 1)(c - 1)$	$[AC] - [A] - [C] + [T]$
$BC$	$(b - 1)(c - 1)$	$[BC] - [B] - [C] + [T]$
$ABC$	$(a - 1)(b - 1)(c - 1)$	$[ABC] - [AB] - [AC] - [BC] + [A] + [B] + [C] - [T]$
residual	$abc(r - 1)$	$[Y] - [ABC]$
total	$abc r - 1$	$[Y] - [T]$

## Analysis of variance table for microwave-popcorn

	<b>Degrees of freedom</b>	<b>Sum of squares</b>	<b>Mean square</b>	<b>F statistic</b>	<b>p-value</b>
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!

# Latin square

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

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

	col 1	col 2	col 3	col 4
row 1	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>
row 2	<i>B</i>	<i>C</i>	<i>D</i>	<i>A</i>
row 3	<i>C</i>	<i>D</i>	<i>A</i>	<i>B</i>
row 4	<i>D</i>	<i>A</i>	<i>B</i>	<i>C</i>

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

The model takes the form

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

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

Only look at treatment effect  $\alpha_i$ .



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

## Analytical comparisons

- Preplanned
- *Post-hoc*: after seeing that the three-way interaction isn't significant, 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  $\psi_A$  (averaging over  $B$  and  $C$ )
- marginal conditional contrast for particular subgroup:  $\psi_A$  within  $c_1$
- contrast involving two variables:  $\psi_{AB}$
- 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

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

```
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)
```

# Semipartial effect sizes

Normally, partial effect sizes are of the form

- if only experimental factors,  $\sigma_{\text{effect}}^2 / (\sigma_{\text{effect}}^2 + \sigma_{\text{res}}^2)$  for the given effect
- if only blocking factors, use total effect size  $\sigma_{\text{effect}}^2 / \sigma_{\text{total}}^2$
- if a mix, include variance of all blocking factors and interactions (only with effect!) in denominator.

If  $A$  is effect of interest,  $B$  is a blocking factor and  $C$  is another experimental factor, use

$$\text{semipartial } \omega_A^2 = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_B^2 + \sigma_{AB}^2 + \sigma_{\text{res}}^2}.$$

# Power for complex studies

One can estimate (semi)partial effects and proceed as if you had a one-way ANOVA by specifying the missing ingredients.

- In more general settings, it may be easier to simulate from the fitted model directly and calculate the power as you change the number of observations per group.