# Statistical Methods for Economists

## Lecture (7 & 8)d

Four-Way Analysis of Variance (ANOVA)
— Græco-Latin Squares

[ Graeco / Greco ]

SILESIAN UNIVERSITY
SCHOOL OF BUSINESS ADMINISTRATION IN KARVINA

**David Bartl**
Statistical Methods for Economists
INM/BASTE

# Outline of the lecture

- Four-way ANOVA:  Introduction
- Græco-Latin squares
- Four-way ANOVA simplified by using Græco-Latin squares

# Four-factor ANOVA:  Motivation:  Example

We have:

- a set of distinct cars
- a set of distinct drivers
- several types of car-fuel (e.g. fuel with various additives)
- several types of tyres

We wish to test whether the mileage (the fuel consumption per 100 km) of the car depends also upon the driver who drives the car, the type of the fuel, and on the type of the tyres.

# Four-factor ANOVA: Motivation: Example

In particular, we have:

- $I$ distinct cars $(i = 1, 2, \ldots, I)$
- $J$ distinct drivers $(j = 1, 2, \ldots, J)$
- $K$ distinct types of fuel $(k = 1, 2, \ldots, K)$
- $L$ distinct types of tyres $(l = 1, 2, \ldots, L)$

# Four-factor ANOVA: Motivation: Example

There are four factors in this example:

- factor A = the car $(i = 1, 2, \dots, I)$
- factor B = the driver $(j = 1, 2, \dots, J)$
- factor C = the type of the fuel $(k = 1, 2, \dots, K)$
- factor D = the type of the tyres $(l = 1, 2, \dots, L)$

There are $IJKL = I \times J \times K \times L$ distinct combinations of the factors.

# Four-factor ANOVA: Motivation & Introduction

Considering the $IJKL = I \times J \times K \times L$ distinct combinations of the factors (the Cartesian product), we assume that each combination is tested $n_{ijkl}$-times.

We thus have a sample

$$y_{ijklm}$$

of the underlying random variables

$$Y_{ijklm}: \Omega \to \mathbb{R}$$

$$\text{for} \begin{cases} i = 1, 2, \dots, I \\ j = 1, 2, \dots, J \\ k = 1, 2, \dots, K \\ l = 1, 2, \dots, L \\ m = 1, 2, \dots, n_{ijkl} \end{cases}$$

These random variables are assumed to be normal $(Y_{ijkl} \sim \mathcal{N}(\mu_{ijk}, \sigma^2))$, independent (uncorrelated) and homoscedastic (with the same variance $\sigma^2$).

# Four-factor ANOVA: Introduction: Assumptions

We assume that the effect of the factors A, B, C, D is additive.

Moreover, it is possible to distinguish many combinations of interactions, e.g.:

- No interactions between / among the factors:

$$Y_{ijklm} \approx \mu + \alpha_i + \beta_j + \gamma_k + \delta_l$$

- Etc.
- All interactions between and among the factors:

$$Y_{ijklm} \approx \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + \\ + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{il}^{AD} + \lambda_{jk}^{BC} + \lambda_{jl}^{BD} + \lambda_{kl}^{CD} + \\ + \lambda_{ijk}^{ABC} + \lambda_{ijl}^{ABD} + \lambda_{ikl}^{ACD} + \lambda_{jkl}^{BCD}$$

$$\text{for } \begin{cases} i = 1, 2, \dots, I \\ j = 1, 2, \dots, J \\ k = 1, 2, \dots, K \\ l = 1, 2, \dots, L \\ m = 1, 2, \dots, n_{ijkl} \end{cases}$$

# Four-factor ANOVA: Introduction: Assumptions

We assume that the effect of the factors A, B, C, D is additive.

Moreover, it is possible to distinguish many combinations of interactions, e.g.:

- $Y_{ijklm} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + \varepsilon_{ijklm}$
- Etc.
- $Y_{ijklm} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l +$

$$+ \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{il}^{AD} + \lambda_{jk}^{BC} + \lambda_{jl}^{BD} + \lambda_{kl}^{CD} +$$

$$+ \lambda_{ijk}^{ABC} + \lambda_{ijl}^{ABD} + \lambda_{ikl}^{ACD} + \lambda_{jkl}^{BCD}$$

$$\text{for} \begin{cases} i = 1, 2, \dots, I \\ j = 1, 2, \dots, J \\ k = 1, 2, \dots, K \\ l = 1, 2, \dots, L \\ m = 1, 2, \dots, n_{ijkl} \end{cases}$$

# Four-factor ANOVA: Introduction: Assumptions

For simplicity, we shall consider the model with no interactions only:

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + \varepsilon_{ijkl}$$

where the parameters

$$\mu, \alpha_i, \beta_j, \gamma_k, \delta_l \in \mathbb{R}$$

are <u>unknown</u> and normalized so that:

$$\sum_{i=1}^{I} \alpha_i = 0 \qquad \sum_{j=1}^{J} \beta_j = 0 \qquad \sum_{k=1}^{K} \gamma_k = 0 \qquad \sum_{l=1}^{L} \delta_l = 0$$

# Four-factor ANOVA:  Simplification

It is possible to consider the general situation with a general number $n_{ijkl} \geq 1$ of observations for each combination of the factors and it is also possible to formulate and test various null hypotheses ($\alpha_i = 0$ / $\lambda_{ij}^{\mathbf{AB}} = 0$ / $\lambda_{ijk}^{\mathbf{ABC}} = 0$ / $\lambda_{ijkl}^{\mathbf{ABCD}} = 0$ / etc.),  but there are plenty of calculations and the resulting formulas are complicated.

This is why we shall study the following special case only:

(see the next slide)

# Four-factor ANOVA: Simplification

**We assume *for simplicity* that the number of the levels of all four factors is the same:**

$$I = J = K = L$$

In addition to that, we do not perform all the observations for each combination of the factors (because the number of the necessary experiments would soon become infeasible).

Instead, we perform either exactly one observation $(n_{ijkl} = 1)$ or no observation at all $(n_{ijkl} = 0)$ according to the scheme called **Græco-Latin square.**

# Græco-Latin squares in Four-way ANOVA

# Latin square

A **Latin square** of order $N$ is an arrangement of $N$ symbols, such as $\{1, 2, \ldots, N\}$, where each symbol is repeated $N$-times, into an $N \times N$ square table in such a way that

- in each column, each symbol occurs exactly once and
- in each row, each symbol occurs exactly once.

For example:

| | | |
|---|---|---|
| 1 | 2 | 3 |
| 2 | 3 | 1 |
| 3 | 1 | 2 |

| | | |
|---|---|---|
| 1 | 2 | 3 |
| 3 | 1 | 2 |
| 2 | 3 | 1 |

| | | | |
|---|---|---|---|
| 1 | 2 | 3 | 4 |
| 2 | 3 | 4 | 1 |
| 3 | 4 | 1 | 2 |
| 4 | 1 | 2 | 3 |

| | | | |
|---|---|---|---|
| 1 | 2 | 3 | 4 |
| 3 | 4 | 1 | 2 |
| 4 | 3 | 2 | 1 |
| 2 | 1 | 4 | 3 |

| | | | | |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |
| 2 | 3 | 4 | 5 | 1 |
| 3 | 4 | 5 | 1 | 2 |
| 4 | 5 | 1 | 2 | 3 |
| 5 | 1 | 2 | 3 | 4 |

# Orthogonal Latin squares

Let $A$ and $B$ be two Latin squares (i.e. matrices) of order $N$.

We say that the two Latin squares are **orthogonal** if and only if,

for each pair $(r, s)$ of symbols $(r, s \in \{1, 2, \dots, N\})$,

there exists (exactly one) pair $(i, j)$ of indices $(i, j \in \{1, 2, \dots, N\})$ so that

$$a_{ij} = r \qquad \text{and} \qquad b_{ij} = s$$

| | | |
|---|---|---|
| 1 | 2 | 3 |
| 2 | 3 | 1 |
| 3 | 1 | 2 |

&

| | | |
|---|---|---|
| 1 | 2 | 3 |
| 3 | 1 | 2 |
| 2 | 3 | 1 |

=

| | | |
|---|---|---|
| 1,□ 1 | 2,□ 2 | 3,□ 3 |
| 2,□ 3 | 3,□ 1 | 1,□ 2 |
| 3,□ 2 | 1,□ 3 | 2,□ 1 |

# Orthogonal Latin squares

Let $A$ and $B$ be two Latin squares (i.e. matrices) of order $N$.

We say that the two Latin squares are **orthogonal** if and only if,

for each pair $(r, s)$ of symbols $(r, s \in \{1, 2, \dots, N\})$,

there exists (exactly one) pair $(i, j)$ of indices $(i, j \in \{1, 2, \dots, N\})$ so that

$$a_{ij} = r \qquad \text{and} \qquad b_{ij} = s$$

| | | | |
|---|---|---|---|
| 1 | 2 | 3 | 4 |
| 2 | 1 | 4 | 3 |
| 3 | 4 | 1 | 2 |
| 4 | 3 | 2 | 1 |

&

| | | | |
|---|---|---|---|
| 1 | 2 | 3 | 4 |
| 3 | 4 | 1 | 2 |
| 4 | 3 | 2 | 1 |
| 2 | 1 | 4 | 3 |

=

| | | | |
|---|---|---|---|
| 1,□ 1 | 2,□ 2 | 3,□ 3 | 4,□ 4 |
| 2,□ 3 | 1,□ 4 | 4,□ 1 | 3,□ 2 |
| 3,□ 4 | 4,□ 3 | 1,□ 2 | 2,□ 1 |
| 4,□ | 3,□ | 2,□ | 1,□ |

# Orthogonal Latin squares

**The following statements are equivalent:**

- The Latin squares $A$ and $B$ of order $N$ are **orthogonal.**
- For each pair $(r, s)$ of symbols $(r, s \in \{1, 2, \dots, N\})$, there exists (exactly one) pair $(i, j)$ of indices $(i, j \in \{1, 2, \dots, N\})$ so that

$$a_{ij} = r \qquad \text{and} \qquad b_{ij} = s$$

- For each pair $(i, j)$ of indices $(i, j \in \{1, 2, \dots, N\})$, there exists (exactly one) pair $(r, s)$ of symbols $(r, s \in \{1, 2, \dots, N\})$ so that

$$a_{ij} = r \qquad \text{and} \qquad b_{ij} = s$$

- For every $i, j, k, l \in \{1, 2, \dots, N\}$

# Orthogonal Latin squares

Let $A_1, A_2, \dots, A_k$ be a collection of Latin squares (i.e. matrices) of order $N$.

We say that the Latin squares $A_1, A_2, \dots, A_k$ are **pairwise orthogonal** if and only if the squares $A_i$ and $A_j$ are orthogonal whenever $i, j \in \{1, 2, \dots, k\}$ and $i \neq j$.

Let $\mathcal{N}(N)$ denote the maximal number $k$ of (elements in a collection of) pairwise orthogonal Latin squares of order $N$.

It is easy to see:

$$1 \leq \mathcal{N}(N) \leq N - 1$$

# Orthogonal Latin squares

A collection $A_1, A_2, \dots, A_k$ of pairwise orthogonal Latin squares of order $N$ is called **complete** if and only if

$$k = N - 1$$

<u>The following is known:</u>

- If $N$ is a power of a prime number (that is $N = p^k$ for some prime number $p$ and a natural number $k$), then $\mathcal{N}(N) = N - 1.$
- We have $\mathcal{N}(6) = 1.$
- If $N \geq 3$ and $N \neq 6$, then $\mathcal{N}(N) \geq 2.$
- If $N \geq 4$ and $\mathcal{N}(N) \geq N - 3$, then $\mathcal{N}(N) = N - 1.$

# Orthogonal Latin squares:  The following is known:

| $N$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $\mathcal{N}(N)$ | — | 1 | 2 | 3 | 4 | 1 | 6 | 7 | 8 | ≥ 2 |

| $N$ | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|
| $\mathcal{N}(N)$ | 10 | ≥ 5 | 12 | ≥ 3 | ≥ 4 | 15 | 16 | ≥ 3 | 18 | ≥ 4 |

| $N$ | 21 | 22 | 23 | 24 | 25 | … |
|---|---|---|---|---|---|---|
| $\mathcal{N}(N)$ | ≥ 4 | ≥ 3 | 22 | ≥ 4 | 24 | … |

# Græco-Latin squares

A pair of orthogonal Latin squares is also called a **Græco-Latin square** or **Graeco-Latin square** or **Greco-Latin square**.

The name "*Græco-Latin square*" is inspired by the work of **Leonhard Euler** (1707–1783), a Swiss mathematician, physicist, astronomer, geographer, logician, and engineer who used the upper-case letters of the Latin alphabet and the lower-case letters of the Greek alphabet as the symbols in the respective square:

| | | | |
|---|---|---|---|
| A | B | C | D |
| B | A | D | C |
| C | D | A | B |
| D | C | B | A |

&

| | | | |
|---|---|---|---|
| α | β | γ | δ |
| γ | δ | α | β |
| δ | γ | β | α |
| β | α | δ | γ |

=

| | | | |
|---|---|---|---|
| Aα | Bβ | Cγ | Dδ |
| Bγ | Aδ | Dα | Cβ |
| Cδ | Dγ | Aβ | Bα |
| Dβ | Cα | Bδ | Aγ |

# Four-factor ANOVA & Græco-Latin squares

- Assume that each of the four factors A, B, C, D has the same number of levels

$$I = J = K = L = N$$

for some natural number $N \geq 2$. — ¡¡¡ Assume also that $N \neq 6$ !!!

- Arrange (or denote) the factors A, B, C, D so that

— we assume that factors A, B, C ("blocking factors") do have some effect on the observed values

— we ask (test the hypothesis) whether factor D ("treatment") has any effect on the observed values

# Four-factor ANOVA & Græco-Latin squares

- Arrange the $N \times N = N^2$ pairs $(k, l)$ into a Græco-Latin square of order $N$. (We consider $k, l \in \{1, 2, \dots, N\}$.)
- There are plenty of distinct Græco-Latin squares of order $N$.
- It is recommended: <u>The Latin square should be chosen randomly.</u>

Now, given the Græco-Latin square of type $N \times N$, define the numbers $n_{ijkl}$ as follows:

$$n_{ijkl} = \begin{cases} 1, & \text{if the pair } (k,l) \text{ is at the position } (i,j), \\ 0, & \text{otherwise.} \end{cases} \quad \text{for} \quad \begin{cases} i = 1, 2, \dots, N \\ j = 1, 2, \dots, N \\ k = 1, 2, \dots, N \\ l = 1, 2, \dots, N \end{cases}$$

# Four-factor ANOVA & Græco-Latin squares

Then,

- for each $i = 1, 2, \dots, N$ and for each $j = 1, 2, \dots, N$,
- find the unique pair $k, l \in \{1, 2, \dots, N\}$ such that $n_{ijkl} = 1$,
- set up Factor A to the level $i$ and set up Factor B to the level $j$,
- set up Factor C to the level $k$ and set up Factor D to the level $l$,
- carry out the experiment,
- observe the numerical outcome $y_{ijkl1}$ of the random variable $Y_{ijkl1}$

Recall that Factors A, B, C are assumed to have some effect ("blocks").

# Four-factor ANOVA: Further assumptions

We assume that the effect of Factors A, B, C, D is additive and that there are no interactions between / among the factors:

$$Y_{ijkl1} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + \varepsilon_{ijkl1} \qquad \text{for} \begin{cases} i = 1, 2, \dots, N \\ j = 1, 2, \dots, N \\ k = 1, 2, \dots, N \\ l = 1, 2, \dots, N \\ \& \; n_{ijkl} = 1 \end{cases}$$

where the meaning of the (unknown) parameters $\mu, \alpha_i, \beta_j, \gamma_k \delta_l \in \mathbb{R}$ is as follows:

- $\mu$ — the common mean value
- $\alpha_i$ — the effect of the level $i$ of Factor A (for $i = 1, 2, \dots, N$)
- $\beta_j$ — the effect of the level $j$ of Factor B (for $j = 1, 2, \dots, N$)
- $\gamma_k$ — the effect of the level $k$ of Factor C (for $k = 1, 2, \dots, N$)
- $\delta_l$ — the effect of the level $l$ of Factor D (for $l = 1, 2, \dots, N$)

# Four-factor ANOVA: Further assumptions

We assume that the effect of Factors A, B, C, D is additive and that there are no interactions between / among the factors:

$$Y_{ijkl1} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + \varepsilon_{ijkl1} \qquad \text{for} \quad \begin{cases} i = 1, 2, \dots, N \\ j = 1, 2, \dots, N \\ k = 1, 2, \dots, N \\ l = 1, 2, \dots, N \\ \& \; n_{ijkl} = 1 \end{cases}$$

Moreover, we assume that the (unknown) parameters $\alpha_i, \beta_j, \gamma_k \delta_l \in \mathbb{R}$ are normalized so that:

$$\sum_{i=1}^{N} \alpha_i = 0 \qquad \sum_{j=1}^{N} \beta_j = 0 \qquad \sum_{k=1}^{N} \gamma_k = 0 \qquad \sum_{l=1}^{N} \delta_l = 0$$

# Four-factor ANOVA: Further assumptions

We assume that the effect of Factors A, B, C, D is additive and that there are no interactions between / among the factors:

$$Y_{ijkl1} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + \varepsilon_{ijkl1} \qquad \text{for} \quad \begin{cases} i = 1, 2, \dots, N \\ j = 1, 2, \dots, N \\ k = 1, 2, \dots, N \\ l = 1, 2, \dots, N \\ \& \; n_{ijkl} = 1 \end{cases}$$

We assume

- Factor A has some effect (that is $\alpha_i \neq 0$ for some $i \in \{1, 2 \dots, N\}$)
- Factor B has some effect (that is $\beta_j \neq 0$ for some $j \in \{1, 2 \dots, N\}$)
- Factor C has some effect (that is $\gamma_k \neq 0$ for some $k \in \{1, 2 \dots, N\}$)

We ask – **test the null hypothesis** – whether Factor D has any effect:

$$H_{\mathrm{D}}: \quad \delta_1 = \delta_2 = \cdots = \delta_N = 0$$

# Four-Way ANOVA without interactions and simplified by using Græco-Latin squares

$$Y_{ijkl1}=\mu+$$
$$+\alpha_i+\beta_j+\gamma_k+\delta_l+\varepsilon_{ijk1}$$

# Four-way ANOVA with no interactions

Assume that we have a sample

$$y_{ijkl1}$$

of observations of the random variables

$$Y_{ijk1} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + \varepsilon_{ijkl1} \qquad \text{for} \quad \begin{cases} i = 1, 2, \dots, N \\ j = 1, 2, \dots, N \\ k = 1, 2, \dots, N \\ l = 1, 2, \dots, N \\ \& \, n_{ijkl} = 1 \end{cases}$$

where $\mu, \alpha_i, \beta_j, \gamma_k, \delta_l \in \mathbb{R}$ are fixed (but unknown)

parameters normalized

so that

$$\sum_{i=1}^{N} \alpha_i = 0 \qquad \sum_{j=1}^{N} \beta_j = 0 \qquad \sum_{k=1}^{N} \gamma_k = 0 \qquad \sum_{l=1}^{N} \delta_l = 0$$

# Four-way ANOVA with no interactions

Assume that we have a sample

$$y_{ijkl1}$$

of observations of the random variables

$$Y_{ijk1} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + \varepsilon_{ijkl1} \quad \text{for} \begin{cases} i = 1, 2, \dots, N \\ j = 1, 2, \dots, N \\ k = 1, 2, \dots, N \\ l = 1, 2, \dots, N \\ \& \, n_{ijkl} = 1 \end{cases}$$

…and

$$\varepsilon_{ijkl1} \sim \mathcal{N}(0, \sigma^2)$$

are mutually independent random variables

with the same variance $\sigma^2 \in \mathbb{R}^+$

(the variance $\sigma^2$ is also unknown).

# Four-way ANOVA with no interactions: Notation

Denote the index set:

$$\mathcal{S} = \{ (i, j, k, l, 1) : i, j, k, l \in \{1, 2, \dots, N\},\ n_{ijkl} = 1 \}$$

Notice that there are exactly $N^2$ elements in the index set $\mathcal{S}$.

Indeed, for each pair $(i, j)$ of the indices for $i = 1, 2, \dots, N$ and for $j = 1, 2, \dots, N$,

there exists exactly one pair $(k, l)$ of symbols $k, l \in \{1, 2, \dots, N\}$

such that $n_{ijkl} = 1$.

# Four-way ANOVA with no interactions: Notation

Stack the observations $y_{ijk1}$ into the $N^2$-dimensional vector

$$y = \left(y_{ijkl1}\right)_{(i,j,k,l.1)\in S} \quad \in \quad \mathbb{R}^S$$

and introduce **the sample mean:**

$$\bar{y} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{\substack{l=1 \\ n_{ijkl}=1}}^{N} y_{ijkl1}$$

This sample mean is an estimate of the parameter $\mu$ (the common mean value):

$$\bar{y} \approx \mu$$

# Four-way ANOVA with no interactions: Notation

Let $\mathbf{1} = (1)_{(i,j,k,l,1)\in\mathcal{S}} \in \mathbb{R}^{\mathcal{S}}$ be the vector of $N^2$ ones and introduce the line

$$L = \{\mathbf{1}\lambda : \lambda \in \mathbb{R}\} =$$

$$= \{\mathbf{z} \in \mathbb{R}^{\mathcal{S}} : z_{ijkl1} = \mu,\ \mu \in \mathbb{R}\}$$

which corresponds to the null hypothesis that

$$H_0: \quad Y_{ijkl1} = \mu + \varepsilon_{ijkl1} \qquad \text{for} \begin{cases} i = 1, 2, \dots, N \\ j = 1, 2, \dots, N \\ k = 1, 2, \dots, N \\ l = 1, 2, \dots, N \\ \& \, n_{ijkl} = 1 \end{cases}$$

that is

$$\alpha_1 = \cdots = \alpha_N = 0 \qquad \beta_1 = \cdots = \beta_N = 0 \qquad \gamma_1 = \cdots = \gamma_N = 0 \qquad \delta_1 = \cdots = \delta_N = 0$$

(cf. one-way ANOVA)

# Four-way ANOVA with no interactions: Notation

Moreover, introduce the subspace

$$H_{\mathrm{A}} = \left\{ \begin{array}{r} z \in \mathbb{R}^S : z_{ijkl1} = \mu + \beta_j + \gamma_k + \delta_l, \quad \mu, \beta_j, \gamma_k, \delta_l \in \mathbb{R}, \\ \sum_{j=1}^{N} \beta_j = \sum_{k=1}^{N} \gamma_k = \sum_{l=1}^{N} \delta_l = 0 \end{array} \right\}$$

which corresponds to the null hypothesis

$$H_{\mathrm{A}}: \quad Y_{ijkl1} = \mu + \beta_j + \gamma_k + \delta_l + \varepsilon_{ijkl1} \qquad \text{for} \quad \begin{cases} i = 1, 2, \dots, N \\ j = 1, 2, \dots, N \\ k = 1, 2, \dots, N \\ l = 1, 2, \dots, N \\ \& \; n_{ijkl} = 1 \end{cases}$$

that is

$$\alpha_1 = \cdots = \alpha_N = 0$$

Observe that the line

$$L \subset H_{\mathrm{A}}$$

# Four-way ANOVA with no interactions: Notation

Introduce also the subspace

$$H_{\mathrm{B}} = \left\{ \begin{array}{r} z \in \mathbb{R}^S : z_{ijkl1} = \mu + \alpha_i + \gamma_k + \delta_l, \quad \mu, \alpha_i, \gamma_k, \delta_l \in \mathbb{R}, \\ \sum_{i=1}^N \alpha_i = \sum_{k=1}^N \gamma_k = \sum_{l=1}^N \delta_l = 0 \end{array} \right\}$$

which corresponds to the null hypothesis

$$H_{\mathrm{B}}: \quad Y_{ijkl1} = \mu + \alpha_i + \gamma_k + \delta_l + \varepsilon_{ijkl1} \qquad \text{for} \left\{ \begin{array}{c} i = 1, 2, \dots, N \\ j = 1, 2, \dots, N \\ k = 1, 2, \dots, N \\ l = 1, 2, \dots, N \\ \& \, n_{ijkl} = 1 \end{array} \right.$$

that is

$$\beta_1 = \dots = \beta_N = 0$$

Observe that the line

$$L \subset H_{\mathrm{B}}$$

# Four-way ANOVA with no interactions: Notation

Introduce also the subspace

$$H_C = \left\{ \begin{aligned} z \in \mathbb{R}^S : z_{ijkl1} = \mu + \alpha_i + \beta_j + \delta_l, \quad \mu, \alpha_i, \beta_j, \delta_l \in \mathbb{R}, \\ \textstyle\sum_{i=1}^{N} \alpha_i = \sum_{j=1}^{N} \beta_j = \sum_{l=1}^{N} \delta_l = 0 \end{aligned} \right\}$$

which corresponds to the null hypothesis

$$H_C: \quad Y_{ijkl1} = \mu + \alpha_i + \beta_j + \delta_l + \varepsilon_{ijkl1} \qquad \text{for} \begin{cases} i = 1, 2, \dots, N \\ j = 1, 2, \dots, N \\ k = 1, 2, \dots, N \\ l = 1, 2, \dots, N \\ \& \, n_{ijkl} = 1 \end{cases}$$

that is

$$\gamma_1 = \cdots = \gamma_N = 0$$

Observe that the line

$$L \subset H_C$$

# Four-way ANOVA with no interactions: Notation

And introduce also the subspace

$$H_D = \left\{ \begin{array}{r} z \in \mathbb{R}^S : z_{ijkl1} = \mu + \alpha_i + \beta_j + \gamma_k, \quad \mu, \alpha_i, \beta_j, \gamma_k \in \mathbb{R}, \\ \sum_{i=1}^N \alpha_i = \sum_{j=1}^N \beta_j = \sum_{k=1}^N \gamma_k = 0 \end{array} \right\}$$

which corresponds to the null hypothesis

$$H_D: \quad Y_{ijkl1} = \mu + \alpha_i + \beta_j + \delta_l + \varepsilon_{ijkl1} \qquad \text{for} \left\{ \begin{array}{c} i = 1, 2, \dots, N \\ j = 1, 2, \dots, N \\ k = 1, 2, \dots, N \\ l = 1, 2, \dots, N \\ \& \, n_{ijkl} = 1 \end{array} \right.$$

that is

$$\delta_1 = \dots = \delta_N = 0$$

Observe that the line

$$L \subset H_D$$

# Four-way ANOVA with no interactions: Notation

Finally, introduce the subspace

$$M = \left\{ \begin{array}{r} z \in \mathbb{R}^S : z_{ijk1} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l,\ \ \mu, \alpha_i, \beta_j, \gamma_k, \delta_l \in \mathbb{R}, \\ \sum_{i=1}^N \alpha_i = \sum_{j=1}^N \beta_j = \sum_{k=1}^N \gamma_k = \sum_{l=1}^N \delta_l = 0 \end{array} \right\}$$

which corresponds to the model under consideration:

$$Y_{ijkl1} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + \varepsilon_{ijkl1} \quad \text{for} \quad \begin{cases} i = 1, 2, \dots, N \\ j = 1, 2, \dots, N \\ k = 1, 2, \dots, N \\ l = 1, 2, \dots, N \\ \& \, n_{ijkl} = 1 \end{cases}$$

# Four-way ANOVA with no interactions: Dimensions

Notice that the dimension of the line

$$L = \left\{ z \in \mathbb{R}^S : z_{ijk1} = \mu,\ \mu \in \mathbb{R} \right\}$$

is

$$\dim L = 1$$

# Four-way ANOVA with no interactions: Dimensions

Notice that the dimension of the subspace

$$H_A = \left\{ \begin{array}{r} \mathbf{z} \in \mathbb{R}^S : z_{ijkl1} = \mu + \beta_j + \gamma_k + \delta_l, \quad \mu, \beta_j, \gamma_k, \delta_l \in \mathbb{R}, \\ \sum_{j=1}^{N} \beta_j = \sum_{k=1}^{N} \gamma_k = \sum_{l=1}^{N} \delta_l = 0 \end{array} \right\}$$

is

$$\dim H_A = (1 + N + N + N) - 1 - 1 - 1 = \quad 3N - 2$$

# Four-way ANOVA with no interactions: Dimensions

Notice that the dimension of the subspace

$$H_{\mathrm{B}} = \left\{ \begin{array}{l} z \in \mathbb{R}^S : z_{ijkl1} = \mu + \alpha_i + \gamma_k + \delta_l, \quad \mu, \alpha_i, \gamma_k, \delta_l \in \mathbb{R}, \\ \qquad \sum_{i=1}^N \alpha_i = \sum_{k=1}^N \gamma_k = \sum_{l=1}^N \delta_l = 0 \end{array} \right\}$$

is

$$\dim H_{\mathrm{B}} = (1 + N + N + N) - 1 - 1 - 1 = \quad 3N - 2$$

# Four-way ANOVA with no interactions: Dimensions

Notice that the dimension of the subspace

$$H_C = \left\{ \begin{array}{l} z \in \mathbb{R}^S : z_{ijkl1} = \mu + \alpha_i + \beta_j + \delta_l, \quad \mu, \alpha_i, \beta_j, \delta_l \in \mathbb{R}, \\ \qquad \sum_{i=1}^{N} \alpha_i = \sum_{j=1}^{N} \beta_j = \sum_{l=1}^{N} \delta_l = 0 \end{array} \right\}$$

is

$$\dim H_C = (1 + N + N + N) - 1 - 1 - 1 = \quad 3N - 2$$

# Four-way ANOVA with no interactions: Dimensions

Notice that the dimension of the subspace

$$H_{\mathrm{D}} = \left\{ \begin{array}{r} \mathbf{z} \in \mathbb{R}^S : z_{ijkl1} = \mu + \alpha_i + \beta_j + \gamma_k, \quad \mu, \alpha_i, \beta_j, \gamma_k \in \mathbb{R}, \\ \sum_{i=1}^{N} \alpha_i = \sum_{j=1}^{N} \beta_j = \sum_{k=1}^{N} \gamma_k = 0 \end{array} \right\}$$

is

$$\dim H_{\mathrm{D}} = (1 + N + N + N) - 1 - 1 - 1 = \quad 3N - 2$$

# Four-way ANOVA with no interactions: Dimensions

And notice that **the dimension** of the subspace

$$M = \left\{ \begin{array}{r} \boldsymbol{z} \in \mathbb{R}^S : z_{ijk1} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l,\ \ \mu, \alpha_i, \beta_j, \gamma_k, \delta_l \in \mathbb{R}, \\ \sum_{i=1}^N \alpha_i = \sum_{j=1}^N \beta_j = \sum_{k=1}^N \gamma_k = \sum_{l=1}^N \delta_l = 0 \end{array} \right\}$$

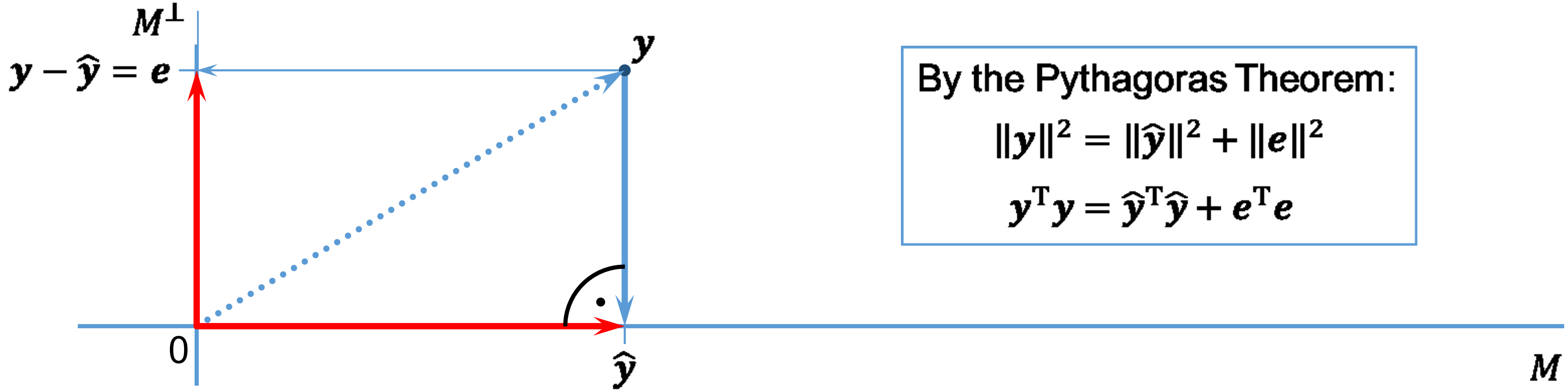
is

$$\dim M = (1 + N + N + N + N) - 1 - 1 - 1 - 1 = \quad 4N - 3$$

# Four-way ANOVA with no interactions

**Letting $\hat{y}_{ijkl1} = \mu - \alpha_i - \beta_j - \gamma_k - \delta_l$, solve the Least Squares Problem:**

$$\min_{\hat{y}\in M} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \sum_{\substack{l=1 \\ n_{ijkl}=1}}^{N} \left(y_{ijkl1} - \mu - \alpha_i - \beta_j - \gamma_k - \delta_l\right)^2 = \min_{\hat{y}\in M} \|y - \hat{y}\|^2 = \mathrm{RSS}$$




By the Pythagoras Theorem:

$$\|y\|^2 = \|\hat{y}\|^2 + \|e\|^2$$

$$y^{\mathrm{T}} y = \hat{y}^{\mathrm{T}} \hat{y} + e^{\mathrm{T}} e$$

$\dim M^{\perp} = N^2 - 4N + 3 = (N-1)(N-3)$

$\dim M^{\perp} + \dim M = N^2$

$\dim M = 4N - 3$

**Residual Sum of Squares:** $\mathrm{RSS} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} e_i^2 = e^{\mathrm{T}} e = \|e\|^2$

# Four-way ANOVA with no interactions

Letting

$$\hat{y}_{\mathrm{A}ijkl1} = \mu - \beta_j - \gamma_k - \delta_l \qquad \text{and} \qquad \hat{y}_{\mathrm{B}ijkl1} = \mu - \alpha_i - \gamma_k - \delta_l$$

solve also the problems:

$$\min_{\hat{\boldsymbol{y}}_{\mathrm{A}} \in H_{\mathrm{A}}} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \sum_{\substack{l=1 \\ n_{ijkl}=1}}^{N} \left(y_{ijkl1} - \mu - \beta_j - \gamma_k - \delta_l\right)^2 = \min_{\hat{\boldsymbol{y}}_{\mathrm{A}} \in H_{\mathrm{A}}} \|\boldsymbol{y} - \hat{\boldsymbol{y}}_{\mathrm{A}}\|^2$$

and

$$\min_{\hat{\boldsymbol{y}}_{\mathrm{B}} \in H_{\mathrm{B}}} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \sum_{\substack{l=1 \\ n_{ijkl}=1}}^{N} \left(y_{ijkl1} - \mu - \alpha_i - \gamma_k - \delta_l\right)^2 = \min_{\hat{\boldsymbol{y}}_{\mathrm{B}} \in H_{\mathrm{B}}} \|\boldsymbol{y} - \hat{\boldsymbol{y}}_{\mathrm{B}}\|^2$$

# Four-way ANOVA with no interactions

Letting

$$\hat{y}_{\mathrm{C}ijkl1} = \mu - \alpha_i - \beta_j - \delta_l \qquad \text{and} \qquad \hat{y}_{\mathrm{D}ijkl1} = \mu - \alpha_i - \beta_j - \gamma_k$$

solve also the problems:

$$\min_{\hat{\boldsymbol{y}}_{\mathrm{C}} \in H_{\mathrm{C}}} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \sum_{\substack{l=1 \\ n_{ijkl}=1}}^{N} \left(y_{ijkl1} - \mu - \alpha_i - \beta_j - \delta_l\right)^2 = \min_{\hat{\boldsymbol{y}}_{\mathrm{C}} \in H_{\mathrm{C}}} \|\boldsymbol{y} - \hat{\boldsymbol{y}}_{\mathrm{C}}\|^2$$

and

$$\min_{\hat{\boldsymbol{y}}_{\mathrm{D}} \in H_{\mathrm{D}}} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \sum_{\substack{l=1 \\ n_{ijkl}=1}}^{N} \left(y_{ijkl1} - \mu - \alpha_i - \beta_j - \gamma_k\right)^2 = \min_{\hat{\boldsymbol{y}}_{\mathrm{D}} \in H_{\mathrm{D}}} \|\boldsymbol{y} - \hat{\boldsymbol{y}}_{\mathrm{D}}\|^2$$

# Four-way ANOVA with no interactions

Finally, letting

$$\bar{y} = \mu$$

solve the Least Squares Problem:

$$\min_{1\bar{y}\in L} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \sum_{\substack{l=1 \\ n_{ijkl}=1}}^{N} \left(y_{ijkl1} - \mu\right)^2 = \min_{1\bar{y}\in L} \|\boldsymbol{y} - \mathbf{1}\bar{y}\|^2$$

# Four-way ANOVA with no interactions

We have and denote:

$$\underbrace{\|y - 1\bar{y}\|^2}_{\mathrm{SS_{TOTAL}}} = \underbrace{\|\hat{y} - \hat{y}_{\mathrm{A}}\|^2}_{\mathrm{SS_A}} + \underbrace{\|\hat{y} - \hat{y}_{\mathrm{B}}\|^2}_{\mathrm{SS_B}} + \underbrace{\|\hat{y} - \hat{y}_{\mathrm{C}}\|^2}_{\mathrm{SS_C}} + \underbrace{\|\hat{y} - \hat{y}_{\mathrm{D}}\|^2}_{\mathrm{SS_D}} + \underbrace{\|e\|^2}_{\mathrm{RSS}}$$

where, recall, we have:

$$\hat{y}_{ijkl1} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k$$

$$\hat{y}_{\mathrm{A}ijkl1} = \hat{\mu} + \hat{\beta}_j + \hat{\gamma}_k + \hat{\delta}_l \qquad \hat{y}_{\mathrm{B}ijkl1} = \hat{\mu} + \hat{\alpha}_i + \hat{\gamma}_k + \hat{\delta}_l$$

$$\hat{y}_{\mathrm{C}ijkl1} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\delta}_l \qquad \hat{y}_{\mathrm{D}ijkl1} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\delta}_l$$

$$\bar{y} = \mu$$

# Four-way ANOVA with no interactions

We thus have:

$$SS_{\mathrm{TOTAL}} = \|\boldsymbol{y} - \mathbf{1}\bar{y}\|^2 = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{\substack{l=1 \\ n_{ijkl}=1}}^{N} \left(y_{ijkl1} - \bar{y}\right)^2 =$$

$$= \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{\substack{l=1 \\ n_{ijkl}=1}}^{N} \left(y_{ijkl1} - \hat{\mu}\right)^2$$

# Four-way ANOVA with no interactions

**We thus have:**

$$\mathrm{RSS} = \|e\|^2 = \|y - \hat{y}\|^2 = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{\substack{l=1\\ n_{ijkl}=1}}^{N} \left(y_{ijkl1} - \hat{y}_{ijkl1}\right)^2 =$$

$$= \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{\substack{l=1\\ n_{ijkl}=1}}^{N} \left(y_{ijkl1} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j - \hat{\gamma}_k - \hat{\delta}_l\right)^2$$

# Four-way ANOVA with no interactions

**We thus have:**

$$SS_{\mathrm{A}} = \|\hat{y} - \hat{y}_{\mathrm{A}}\|^2 = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{\substack{l=1 \\ n_{ijkl}=1}}^{N} \left(\hat{y}_{ijkl1} - \hat{y}_{\mathrm{A}ijkl1}\right)^2 =$$

$$= \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{\substack{l=1 \\ n_{ijkl}=1}}^{N} \left(\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k + \hat{\delta}_l - \hat{\mu} - \hat{\beta}_j - \hat{\gamma}_k - \hat{\delta}_l\right)^2 =$$

$$= \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{\substack{l=1 \\ n_{ijkl}=1}}^{N} \hat{\alpha}_i^2 \quad = \quad N \sum_{i=1}^{N} \hat{\alpha}_i^2$$

# Four-way ANOVA with no interactions

**We thus have:**

$$SS_B = \|\hat{y} - \hat{y}_B\|^2 = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{\substack{l=1 \\ n_{ijkl}=1}}^{N} \left(\hat{y}_{ijkl1} - \hat{y}_{Bijkl1}\right)^2 =$$

$$= \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{\substack{l=1 \\ n_{ijkl}=1}}^{N} \left(\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k + \hat{\delta}_l - \hat{\mu} - \hat{\alpha}_i - \hat{\gamma}_k - \hat{\delta}_l\right)^2 =$$

$$= \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{\substack{l=1 \\ n_{ijkl}=1}}^{N} \hat{\beta}_j^2 \;=\; N\sum_{j=1}^{N}\hat{\beta}_j^2$$

# Four-way ANOVA with no interactions

We thus have:

$$SS_C = \|\hat{y} - \hat{y}_C\|^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \sum_{\substack{l=1 \\ n_{ijkl}=1}}^{N} \left(\hat{y}_{ijkl1} - \hat{y}_{Cijkl1}\right)^2 =$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \sum_{\substack{l=1 \\ n_{ijkl}=1}}^{N} \left(\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k + \hat{\delta}_l - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j - \hat{\delta}_l\right)^2 =$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \sum_{\substack{l=1 \\ n_{ijkl}=1}}^{N} \hat{\gamma}_k^2 = N \sum_{k=1}^{N} \hat{\gamma}_k^2$$

# Four-way ANOVA with no interactions

**We thus have:**

$$SS_D = \|\hat{y} - \hat{y}_D\|^2 = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{\substack{l=1 \\ n_{ijkl}=1}}^{N} \left(\hat{y}_{ijkl1} - \hat{y}_{Dijkl1}\right)^2 =$$

$$= \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{\substack{l=1 \\ n_{ijkl}=1}}^{N} \left(\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k + \hat{\delta}_l - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j - \hat{\gamma}_k\right)^2 =$$

$$= \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{\substack{l=1 \\ n_{ijkl}=1}}^{N} \hat{\delta}_l^2 = N\sum_{l=1}^{N}\hat{\delta}_l^2$$

# Four-way ANOVA with no interactions

By solving the above Least Squares Problems

$\left(\partial F/\partial\mu = \partial F/\partial\alpha_i = \partial F/\partial\beta_j = \partial F/\partial\gamma_k = \partial F/\partial\delta_l = 0\right.$ etc.; do it as an exercise),

we obtain:

$$\hat{\mu} \;=\; \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{\substack{l=1\\ n_{ijkl}=1}}^{N} y_{ijkl1} \;=\; \bar{y} \;=\; \bar{y}_{....}$$

# Four-way ANOVA with no interactions

By solving the above Least Squares Problems

($\partial F/\partial \mu = \partial F/\partial \alpha_i = \partial F/\partial \beta_j = \partial F/\partial \gamma_k = \partial F/\partial \delta_l = 0$ etc.; do it as an exercise),

we obtain:

$$\hat{\alpha}_i = \frac{1}{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{\substack{l=1 \\ n_{ijkl}=1}}^{N} y_{ijkl1} - \hat{\mu} = \bar{y}_{i\cdots} - \bar{y}$$

for $i = 1, 2, \ldots, N$

# Four-way ANOVA with no interactions

By solving the above Least Squares Problems

($\partial F/\partial\mu = \partial F/\partial\alpha_i = \partial F/\partial\beta_j = \partial F/\partial\gamma_k = \partial F/\partial\delta_l = 0$ etc.; <u>do it as an exercise</u>),

we obtain:

$$\hat{\beta}_j = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{N}\sum_{\substack{l=1 \\ n_{ijkl}=1}}^{N} y_{ijkl1} - \hat{\mu} = \bar{y}_{\cdot j \cdot\cdot} - \bar{y}$$

for $j = 1, 2, \ldots, N$

# Four-way ANOVA with no interactions

By solving the above Least Squares Problems

($\partial F/\partial\mu = \partial F/\partial\alpha_i = \partial F/\partial\beta_j = \partial F/\partial\gamma_k = \partial F/\partial\delta_l = 0$ etc.; do it as an exercise),

we obtain:

$$\hat{\gamma}_k = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{\substack{l=1 \\ n_{ijkl}=1}}^{N} y_{ijkl1} - \hat{\mu} = \bar{y}_{\cdot\cdot k\cdot} - \bar{y}$$

for $k = 1, 2, \ldots, N$

# Four-way ANOVA with no interactions

By solving the above Least Squares Problems

($\partial F/\partial \mu = \partial F/\partial \alpha_i = \partial F/\partial \beta_j = \partial F/\partial \gamma_k = \partial F/\partial \delta_l = 0$ etc.; do it as an exercise),

we obtain:

$$\hat{\delta}_l = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{\substack{k=1 \\ n_{ijkl}=1}}^{N} y_{ijkl1} - \hat{\mu} = \bar{y}_{\dots l} - \bar{y}$$

for $l = 1, 2, \dots, N$

# Four-way ANOVA with no interactions

Put together, we have:

$$SS_{TOTAL} = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{\substack{l=1\\ n_{ijkl}=1}}^{N}\left(y_{ijkl1}-\hat{\mu}\right)^2 = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{\substack{l=1\\ n_{ijkl}=1}}^{N}\left(y_{ijkl1}-\bar{y}\right)^2$$

$$RSS = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{\substack{l=1\\ n_{ijkl}=1}}^{N}\left(y_{ijkl1}-\hat{\mu}-\hat{\alpha}_i-\hat{\beta}_j-\hat{\gamma}_k-\hat{\delta}_l\right)^2 =$$

$$= \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{\substack{l=1\\ n_{ijkl}=1}}^{N}\left(y_{ijkl1}-\bar{y}_{i\ldots}-\bar{y}_{.j..}-\bar{y}_{..k.}-\bar{y}_{\ldots l}+3\bar{y}\right)^2$$

# Four-way ANOVA with no interactions

**Remark:** The quantity

$$s^2 = \frac{\text{RSS}}{\dim M^{\perp}} = \frac{\text{RSS}}{(N-1)(N-3)}$$

is an estimate of the unknown $\sigma^2$, that is $s^2 \approx \sigma^2$.

It holds

$$\mathrm{E}[s^2] = \sigma^2$$

Recall that $\dim M^{\perp} = N^2 - \dim M = (N-1)(N-3)$ — see the figure above!

# Four-way ANOVA with no interactions

Put together, we have:

$$SS_A = N\sum_{i=1}^{N}\hat{\alpha}_i^2 = N\sum_{i=1}^{N}(\bar{y}_{i\cdots} - \bar{y})^2$$

$$SS_B = N\sum_{j=1}^{N}\hat{\beta}_j^2 = N\sum_{j=1}^{N}(\bar{y}_{\cdot j\cdot\cdot} - \bar{y})^2$$

# Four-way ANOVA with no interactions

**Put together, we have:**

$$SS_C = N\sum_{k=1}^{N}\hat{\gamma}_k^2 = N\sum_{k=1}^{N}(\bar{y}_{..k.} - \bar{y})^2$$

$$SS_D = N\sum_{l=1}^{N}\hat{\delta}_l^2 = N\sum_{l=1}^{N}(\bar{y}_{...l} - \bar{y})^2$$

# Four-way ANOVA with no interactions

Recall that it holds:

$$SS_{TOTAL} = SS_A + SS_B + SS_C + SS_D + RSS$$

$$\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{\substack{l=1\\ n_{ijkl}=1}}^{N}\left(y_{ijkl1} - \bar{y}\right)^2 =$$

$$= N\sum_{i=1}^{N}(\bar{y}_{i\dots} - \bar{y})^2 + N\sum_{j=1}^{N}\left(\bar{y}_{.j..} - \bar{y}\right)^2 + N\sum_{k=1}^{N}(\bar{y}_{..k.} - \bar{y})^2 + N\sum_{l=1}^{N}(\bar{y}_{\dots l} - \bar{y})^2 +$$

$$+\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{\substack{l=1\\ n_{ijkl}=1}}^{N}\left(y_{ijkl1} - \bar{y}_{i\dots} - \bar{y}_{.j..} - \bar{y}_{..k.} - \bar{y}_{\dots l} + 3\bar{y}\right)^2$$

# Four-way ANOVA with no interactions: Test for $H_{\mathrm{D}}$

We use the theory of Linear Regression (Theorem 8) to test Hypothesis $H_{\mathrm{D}}$:

If the null hypothesis

$$H_{\mathrm{D}}: \quad \gamma_1 = \gamma_2 = \cdots = \gamma_N = 0$$

holds true, then

$$\frac{\mathrm{SS_D}}{\mathrm{RSS}} \Big/ \frac{\dim M - \dim H_{\mathrm{D}}}{N^2 - \dim M} = \frac{\mathrm{SS_D}}{\mathrm{RSS}} \Big/ \frac{N-1}{(N-1)(N-2)} \sim F_{N-1,(N-1)(N-2)}$$

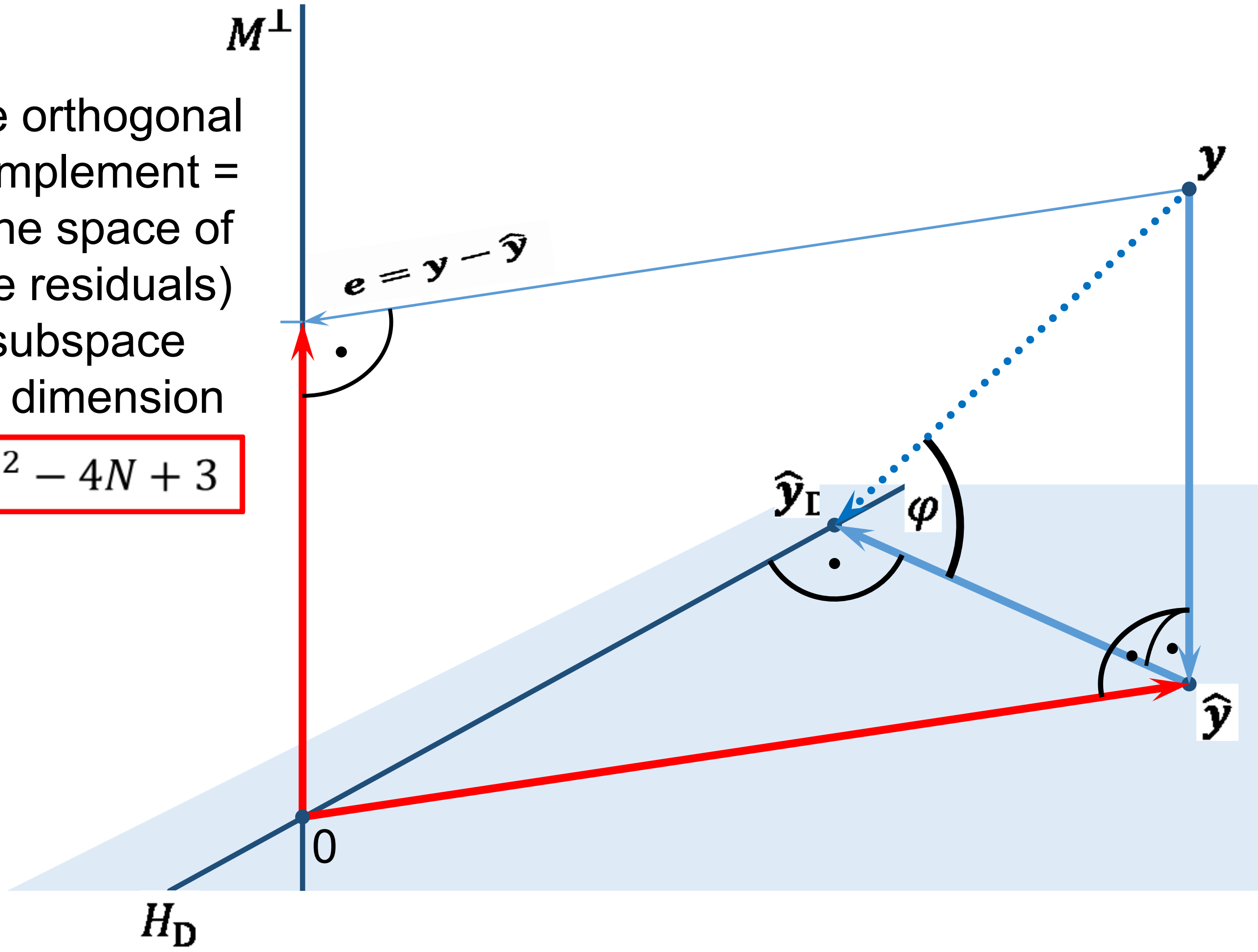
# Four-way ANOVA with no interactions: Test for $H_D$

It holds:

$$\text{cotan}^2\varphi = \frac{(\hat{\mathbf{y}} - \hat{\mathbf{y}}_D)^T(\hat{\mathbf{y}} - \hat{\mathbf{y}}_D)}{\text{RSS}}$$

$$(\text{cotan}\,\varphi)^2 / \frac{N-1}{N^2-4N+3} \sim F_{N-1,\,(N-1)(N-3)}$$

(the orthogonal complement = = the space of the residuals) subspace of dimension $N^2 - 4N + 3$




$M$ subspace of dimension $4N - 3$

$H_D$ subspace of dimension $3N - 2$

the dimension of its complement within the subspace of dimension $4N - 3$ is $N - 1$

# Four-way ANOVA with no interactions: Test for $H_D$

- Given the sample $y_{ijkl1}$ of the random variables

  $Y_{ijkl1} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + \varepsilon_{ijkl1}$

  where $\mu, \alpha_i, \beta_j, \gamma_k, \delta_l \in \mathbb{R}$ are (unknown) parameters such that

  $\sum_{i=1}^{N} \alpha_i = \sum_{j=1}^{N} \beta_j = \sum_{k=1}^{N} \gamma_k = \sum_{l=1}^{N} \delta_l = 0$ and $\varepsilon_{ijkl1} \sim \mathcal{N}(0, \sigma^2)$

  are mutually independent random variables for $i = 1, 2, \dots, N$, $j = 1, 2, \dots, N$, $k = 1, 2, \dots, N$, $l = 1, 2, \dots, N$ such that $n_{ijkl} = 1$, formulate the null hypothesis:

$$H_D: \quad \delta_1 = \delta_2 = \cdots = \delta_N = 0$$

# Four-way ANOVA with no interactions: Test for $H_D$

- Calculate the statistic

$$F = \frac{\mathrm{SS_D}}{\mathrm{RSS}} \Big/ \frac{\mathrm{DF_D}}{\mathrm{DF_{RSS}}} =$$

$$= \frac{N \sum_{l=1}^{N} (\bar{y}_{\dots l} - \bar{y})^2}{\sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \sum_{n_{ijkl}=1}^{N} {}_{l=1} \left( y_{ijkl1} - \bar{y}_{i\dots} - \bar{y}_{\cdot j \cdot\cdot} - \bar{y}_{\cdot\cdot k \cdot} - \bar{y}_{\dots l} + 3\bar{y} \right)^2} \Big/ \frac{N-1}{N^2 - 4N + 3}$$

- If the null hypothesis is true, then we have by the Theorem

$$F \sim F_{N-1,\,(N-1)(N-3)}$$

## Four-way ANOVA with no interactions: Test for $H_D$

- Choose **the level of significance**, a small number $\alpha > 0$, such as $\alpha = 5\,\%$

- Find the **critical value**

$$c = F_{N-1,(N-1)(N-3)}\,(1-\alpha)$$

so that $\int_c^{+\infty} f(x)\,\mathrm{d}x = \alpha$ where $f$ is the density of the *F*-distribution with $N-1$ and $(N-1)(N-3)$ degrees of freedom

- If $F \in [c, +\infty)$, **the critical region**, then **reject** the null hypothesis
- If $F \in [0, c)$, then **do not reject** (or fail to reject) the null hypothesis